Technical Physics, Vol. 45, No. 4, 2000, pp. 381–388. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 1–7.
Original Russian Text Copyright © 2000 by Shchukin, Grigor'ev.

GASES AND FLUIDS

Energy Analysis of Possible Channels for Decomposition of a Charged Drop into Two Parts

S. I. Shchukin and A. I. Grigor'ev
Yaroslavl State University, Sovetskaya ul. 14, Yaroslavl, 150000 Russia

Received October 14, 1998

Abstract—Based on numerical analysis of the mathematical expression for potential energy of an isolated charged spheroidal drop and two approximate spheroidal daughter drops, mechanisms for decomposition of a multiply charged drop in two nearly equal parts were studied taking into account the electrostatic interaction of the daughter drops. It was shown that, as the distance between the daughter drops increased, both the probability of spontaneous decomposition of a multiply charged drop in two daughter drops and the decomposition symmetry degree increase at the moment of breaking the connection between the daughter drops. *© 2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The study of mechanisms of instability of multiply charged drops and clusters is of interest because of numerous applications of the phenomenon in geophysics, liquid mass-spectrometry, liquid-metal epitaxy and lithography, in obtaining refractory metal powders, etc. [1–3]. The decomposition of a drop that is unstable with respect to its charge may proceed along various channels depending on the values of an initial deformation of the drop, its charge, size, viscosity, and electrical conductivity [1, 3–9].

The decomposition of a multiply charged drop of low-viscosity fluid with high conductivity has been studied in the greatest detail [1, 3, 10–13]. During this decomposition, a drop with a charge value equal to or larger than the limiting Rayleigh's charge decomposes, emitting about two hundreds daughter drops, which are two orders of magnitude smaller than the parent drop, and losing only 0.05 and 0.23 of its initial mass and charge, respectively. At the same time, the breakup of charged drops into a few nearly equal parts was observed in a number of experiments (see, for example, [4–7, 14, 15]). In experiments [6], the decomposition of a drop in nearly equal parts is the result of the influence of substantial viscosity of the medium, which damps the instability of high modes of the capillary oscillations of the unstable drop. In experiments [4], such a decomposition at subcritical charge is due to considerable mechanical deformations of an initially spherical drop; in [5], the same effect is due to deceleration in a highly inhomogeneous external electric field.

Theoretical studies of decomposition of a multiply charged drop or a cluster in two parts of nearly equal size have been carried out repeatedly [2, 8, 16–18]. In the latest edition of the book [19], this problem is presented as an example. However, based on comparison of the full potential energy of an initial drop with the potential energy of two daughter drops, all the investigations of that kind have used rather crude models to estimate the energy of the electrostatic interaction of the daughter drops. This has led to incorrect reasoning and wrong quantitative criteria. It should be noted that the idea of the energy approach, which is used in [2, 16–19], has been borrowed from the works of Frenkel' [20] and Bohr [21], which are devoted to the drop model of the nucleus. However, while the neglect of the electrostatic interaction of daughter drops in the nuclear fission problem could be justified, this is not the case with charged clusters and drops.

In what follows, the energy analysis of the possibility of decomposition of a multiply charged drop in two nearly equal parts is made taking rigorously into account the energy of the electrostatic interaction between the daughter drops and using the expressions for the factor of mutual electrical capacitance of two closely spaced spheres given in [22].

1. Considering drops that are formed at the final stage of the breakup as contacting spheres, let us determine the possibility of the appearance of various channels for the decomposition with different charge values of the initial drop using the analysis of the total potential energy of the system. By the potential energy of the system of two drops we mean the sum of the energy of their surface tension forces and their full electrostatic energy

$$U = U_s + U_e, \tag{1}$$

where the energy of the surface tension forces has the form

$$U_s = 4\pi\alpha(R_1^2 + R_2^2), \tag{2}$$

while the electrostatic energy of a two-conductor system is as follows:

$$U_e = \frac{1}{2}C_{11}V_1^2 + C_{12}V_1V_2 + \frac{1}{2}C_{22}V_2^2.$$

Here, C_{11} and C_{22} are the capacitance factors for the first and second conductors, respectively; C_{12} is the mutual capacitance factor; and V_1 and V_2 represent the conductor potentials related to their charge and capacitance by the formulas

$$V_1 = \frac{Q_2C_{12} - Q_1C_{22}}{C_{12}^2 - C_{11}C_{22}}, \quad V_2 = \frac{Q_1C_{12} - Q_2C_{11}}{C_{12}^2 - C_{11}C_{22}}.$$

Denoting by x the ratio of the drops' charges

$$Q_2 = xQ_1 \tag{3}$$

and using the law of conservation of charge, we arrive at

$$Q_1 = \frac{Q}{(1+x)}, \tag{4}$$

where Q is the total charge.

Then, the expressions for the potentials and energy take the form

$$V_1 = \frac{Q(C_{22} - xC_{12})}{(1+x)(C_{11}C_{22} - C_{12}^2)},$$

$$V_2 = \frac{Q(xC_{11} - C_{12})}{(1+x)(C_{11}C_{22} - C_{12}^2)},$$

$$U_e = \frac{(x^2C_{11} + C_{22} - 2xC_{12})Q^2}{(2C_{11}C_{22} - 2C_{12}^2)(1+x)^2}.$$

Assuming that the drops have equal potentials ($V_1 = V_2$), the charge ratio x is expressed in terms of the capacitance factors by the simple formula

$$x = \frac{C_{22} + C_{12}}{C_{12} + C_{11}}, \tag{5}$$

and the expression for the electrostatic energy becomes simpler; i.e.,

$$U_e = \frac{Q^2}{2(C_{11} + C_{22} + 2C_{12})}. \tag{6}$$

It should be noted that, according to the Thompson theorem, well known in electrostatics, the solution to the equation $\partial U_e/\partial x = 0$ for x is also represented by expressions (5) and (6); that is, the condition for equality of potentials of two charged spheres is, at the same time, the one for the electrostatic energy of the system to achieve its minimum, provided that the total charge is preserved.

The expressions for the capacitance factors for two conducting spheres have the form [22]

$$C_{11} = R_1R_2\sinh(t) \times \sum_{n=1}^{\infty}(R_2\sinh(nt) + R_1\sinh((n-1)t))^{-1},$$

$$C_{22} = R_1R_2\sinh(t) \times \sum_{n=1}^{\infty}(R_1\sinh(nt) + R_2\sinh((n-1)t))^{-1},$$

$$C_{12} = -\frac{R_1R_2\sinh(t)}{R_{12}}\sum_{n=1}^{\infty}(\sinh(nt))^{-1},$$

where t is to be determined from

$$\cosh(t) = \frac{R_{12}^2 - R_1^2 - R_2^2}{2R_1R_2},$$

R_1 and R_2 are radii of the drops, and R_{12} is the distance between their centers.

Let y be the ratio of the radii of the drops

$$R_2 = yR_1. \tag{7}$$

Using the law of conservation of mass, we have

$$R_1 = \frac{R}{(1+y^3)^{1/3}}. \tag{8}$$

Then, assuming that

$$R_{12} = (1 + y + z)R_1, \tag{9}$$

we rewrite expression (6) for the electrostatic energy in the form

$$U_e = Q^2\sqrt[3]{1+y^3}\left\{2Ry\sum_{n=1}^{\infty}(C_{1nt} + C_{2nt} + C_{12nt})\right\}^{-1}, \tag{10}$$

where

$$C_{1nt} \equiv \frac{\sinh(t)}{y\sinh(nt) + \sinh((n-1)t)},$$

$$C_{2nt} \equiv \frac{\sinh(t)}{\sinh(nt) + y\sinh((n-1)t)},$$

$$C_{12nt} \equiv -2\frac{\sinh(t)}{\sinh(nt)(1+y+z)},$$

$$t \equiv \operatorname{arccosh}\left(\frac{2y + 2z + 2yz + z^2}{2y}\right).$$

For contacting spheres, $z = 0$. In this case, (5) can be written as

$$x = \left\{ \sum_{n=1}^{\infty} (y\sinh(nt) + \sinh((n-1)t))^{-1} - \frac{1}{\sinh(nt)(1+y)} \right\} \left\{ \sum_{n=1}^{\infty} (\sinh(nt) + y\sinh((n-1)t))^{-1} - \frac{1}{\sinh(nt)(1+y)} \right\}^{-1}$$

and, if $t = 0$, takes the form

$$x = \frac{\Psi\left(\frac{1}{1+y}\right) + \gamma}{\Psi\left(\frac{y}{1+y}\right) + \gamma}. \qquad (11)$$

Then, (10) becomes

$$U_e = \frac{-Q^2\sqrt[3]{1+y^3}(1+y)}{2Ry\{\Psi((1+y)^{-1}) + \Psi(y(1+y)^{-1}) + 2\gamma\}}, \qquad (12)$$

where Ψ is the digamma function, and γ is Euler's constant

$$\gamma = \lim_{n\to\infty} \left\{ \sum_{i=1}^{\infty} (i)^{-1} - \ln(n) \right\} \approx 0.5772156649.$$

It should be noted that condition (11), which is obtained under the assumption that the potentials of contacting charged spheres are equal, substantially differs from the often-used relation

$$\frac{Q_1}{Q_2} = \frac{R_1}{R_2},$$

which is obtained from the condition of equality with respect to infinity of potentials of infinitely spaced isolated charged spheres.

It follows from the analysis of the function given by (12), which represents the electrostatic energy of a two-sphere system, that its minimum is attained at $y = 1$ and equals

$$U_{\min} = \frac{\sqrt[3]{2}}{2\ln(2)} \approx 0.90884.$$

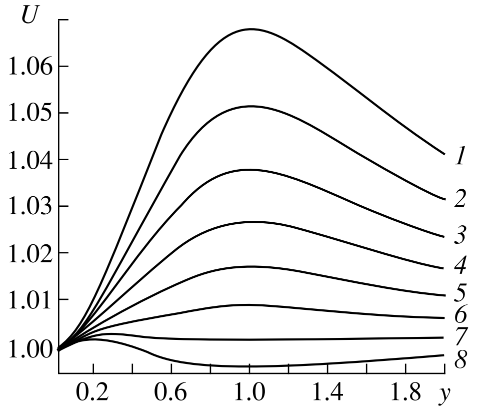


Fig. 1. The dimensionless total energy U of two charged contacting spherical drops as a function of the ratio of their radii for W = (*1*) 2.5, (*2*) 3.0, (*3*) 3.5, (*4*) 4.0, (*5*) 4.5, (*6*) 5.0, (*7*) 5.5, and (*8*) 6.0.

Using (1), (2), and (12), we can write the expression for the total potential energy of a system of two contacting charged spherical drops with identical potentials (this potential energy is rendered dimensionless by dividing it by the total potential energy of a spherical drop having the same mass and energy)

$$U = \left\{ \frac{1+y^2}{(1+y^3)^{2/3}} - \frac{W\sqrt[3]{1+y^3}(1+y)}{2y\left(\Psi((1+y)^{-1}) + \Psi\left(\frac{y}{1+y}\right) + 2\gamma\right)} \right\} \left(1 + \frac{W}{2}\right)^{-1}, \qquad (13)$$

where $W \equiv Q^2/(4\pi\alpha R^3)$ is Rayleigh's parameter.

The plots of the dimensionless full energy of two charged contacting spherical drops as a function of the ratio of their radii are shown in Fig. 1. They were obtained by calculating (13) for various values of Rayleigh's parameter W. It can be seen from Fig. 1 that, as the parameter W varies, the function undergoes not only quantitative but also qualitative changes. For small values of Rayleigh's parameter, the energy steadily increases from unity to some maximal value as the ratio of the radii of daughter drops increases from close to zero (one drop is very small and the other is close in volume to the parent drop) to unity (two daughter drops with equal volumes). As Rayleigh's parameter increases, a local minimum appears in the plots of $U = U(y)$, which then becomes global, changing its position from small values of y to $y = 1$.

The dependence of the dimensionless total energy on Rayleigh's parameter (Fig. 2) is a steadily decreasing function and does not exhibit any qualitative changes when varying the ratio of the radii. Setting the

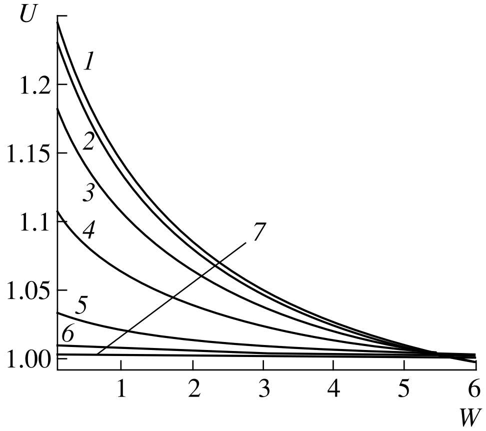


Fig. 2. The dependence of U on Rayleigh's parameter for y = (*1*) 1.0, (*2*) 0.8, (*3*) 0.6, (*4*) 0.4, (*5*) 0.2, (*6*) 0.1, and (*7*) 0.05.

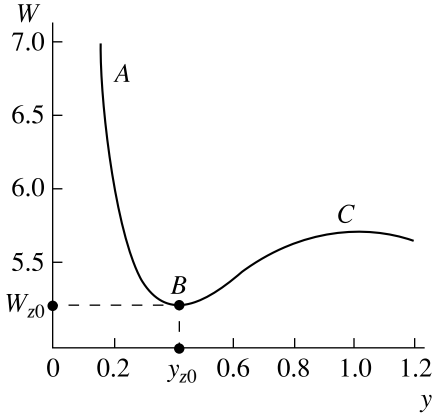


Fig. 3. Extrema of the function $U = U(y)$.

derivative $\partial U/\partial y$ equal to zero

$$\frac{\partial U}{\partial y} = \left\{\frac{2y}{(1+y^3)^{2/3}} - 2\frac{(y^2+1)y^2}{(1+y^3)^{5/3}} - \frac{W(y+1)y}{2(1+y^3)^{2/3}A}\right.$$

$$\left. + \frac{W\sqrt[3]{1+y^3}}{2yA}\left[\frac{1}{y} + \frac{\Psi\left(1, \dfrac{y}{y+1}\right)}{y+1} - \frac{\Psi(1, (y+1)^{-1})}{(y+1)A}\right]\right\}$$

$$\times\left(1+\frac{W}{2}\right)^{-1} = 0,$$

$$A = 2\gamma + \Psi\left(\frac{y}{y+1}\right) + \Psi\left(\frac{1}{y+1}\right) \qquad (14)$$

we find the function $W = W(y)$ that corresponds to extreme values of the total potential energy

$$W = \frac{4yA}{(1+y)^{2/3}}\left[1 - \frac{(y^2+1)y}{1+y^3}\right]\left\{\frac{(y+1)y}{(1+y^3)^{2/3}A} - \frac{\sqrt[3]{1+y^3}}{y}\right. \qquad (15)$$

$$\left.\times\left[\frac{1}{y} + \frac{\Psi\left(1, \dfrac{y}{y+1}\right)}{y+1} - \frac{\Psi(1, (y+1)^{-1})}{(y+1)A}\right]\right\}^{-1},$$

where $\Psi(1, x)$ is the trigamma function.

For $y = 1$ and $y = 0$, solution (15) becomes degenerate. Substitution of $y = 1$ into (14) yields identically zero; the same is true for the limit

$$\lim_{y \to 0}\frac{\partial U}{\partial y} \equiv 0.$$

Hence, these values always correspond to at least local extrema of the total potential energy of two contacting spherical drops as a function of the ratio of their radii. A plot of the function (15) shown in Fig. 3 means that, for Rayleigh's parameter smaller than certain W_{z0}, extrema different from $y = 1$ and $y = 0$ are not present. A numerical calculation yields, for coordinates of the minimum (point B in Fig. 3), values $W_{z0} \approx 5.279$ and $y_{z0} \approx 0.4124$.

To find out whether an extremum is maximum or minimum, we have to analyze the sign of the expression obtained by substituting solution (15) into the second derivative, which has the form

$$\frac{\partial^2 U}{\partial y^2} = \left\{-\frac{W\sqrt[3]{1+y^3}\left(\Psi\left(1, \dfrac{y}{y+1}\right)\right)^2}{(y+1)^3 yA^3} + \left[\frac{y(y+1)}{(1+y^3)A^2}\right.\right.$$

$$\left. + 2\frac{\Psi(1, (y+1)^{-1})}{(y+1)yA^3} - \frac{(y+1)}{y^2A^2} + \frac{2}{yA^2}\right]\frac{W\sqrt[3]{1+y^3}}{(y+1)^2}$$

$$\times\Psi\left(1, \frac{y}{y+1}\right) - \frac{W\sqrt[3]{1+y^3}(\Psi(1, (y+1)^{-1}))^2}{(y+1)^3 yA^3}$$

$$+\left(-\frac{y}{(1+y^3)^{2/3}A^2} + \frac{\sqrt[3]{1+y^3}}{y^2A^2}\right)\frac{W\Psi(1, (y+1)^{-1})}{(y+1)}$$

$$+\frac{2}{(1+y^3)^{2/3}} - 8\frac{y^3}{(1+y^3)^{5/3}} + 10\frac{(y^2+1)y^4}{(1+y^3)^{8/3}}$$

$$-4\frac{(y^2+1)y}{(1+y^3)^{5/3}} + \frac{W(y+1)y^3}{(1+y^3)^{5/3}A} - \frac{Wy}{(1+y^3)^{2/3}A}$$

$$-\frac{W\sqrt[3]{1+y^3}(y+1)}{y^3A}+\frac{W\sqrt[3]{1+y^3}}{y^2A}+\frac{W\sqrt[3]{1+y^3}}{2y(y+1)^3A^2}$$

$$\left.\times\left[\Psi\left(2,\frac{y}{y+1}\right)+\Psi(2,(y+1)^{-1})\right]\right\}\left(1+\frac{W}{2}\right)^{-1},$$

where $\Psi(2, x)$ is the tetragamma function.

An analysis of the resulting dependence, an analytical expression for which is not written here because of its cumbersome form, shows that point B in Fig. 3 corresponds to the transition from maxima of the function $U = U(y)$ to its minima; i.e., for $0 < y < y_{z0}$ (line AB in Fig. 3) the extremum is maximum, while, for $y_{z0} < y < 1$ (line BC in Fig. 3), it corresponds to a minimum. At points $y = 0$ and $y = 1$, the function under consideration vanishes. Analyzing the function $U = U(W, y)$, we can find out whether the total potential energy takes a maximum or a minimum value at each of these points.

Based on (13), a plot of the function $U = U(W, y)$ representing the dependence of the total energy of two contacting charged spherical drops on both the ratio of their radii and Rayleigh's parameter is shown in Fig. 4 as a three-dimensional surface. Lines *pq*, *cfk*, and *mn* correspond to extreme values of the energy when total volume and charge are preserved and the ratio of the radii of drops varies. Lines *ab*, *efg*, and *hj* correspond to a change of sign of the second derivative $\partial^2U/\partial y^2$: to the left of line *ab*, as well as between lines *efg* and *hj*, the second derivative is positive; between lines *efg* and *ab*, as well as to the right of line *hj*, it is negative. Line *di* corresponds to the solution to the equation $U = 1$, and the points of the surface below this line correspond to the states of the system at which the energy of two daughter drops is lower than that of the parent drop.

As seen in the figure, at the point $y = 0$, which corresponds to the case of a single drop, the potential energy attains its minimum. For W smaller than some W_1 $y = 1$, which corresponds to the case of two identical daughter drops, it is related to the maximum (line *nf*) of the total potential energy; whereas, for $W > W_1$ (line *fm*), it corresponds to the minimum. The quantity W_1 is determined by substituting $y = 1$ into (15)

$$W_1 = -\frac{32(\ln 2)^2}{7\zeta(3)-16\ln 2},$$

where $\zeta = \zeta(x)$ is the Riemann zeta function, and $W_1 \approx 5.7454$.

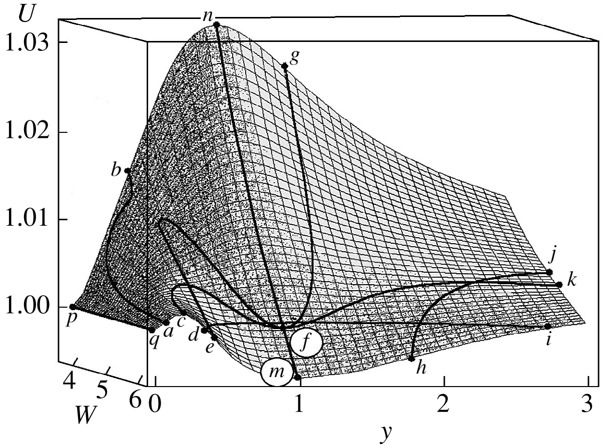


Fig. 4. The dimensionless total energy of two charged contacting spherical drops as a function of both Rayleigh's parameter and the ratio of the radii of the drops.

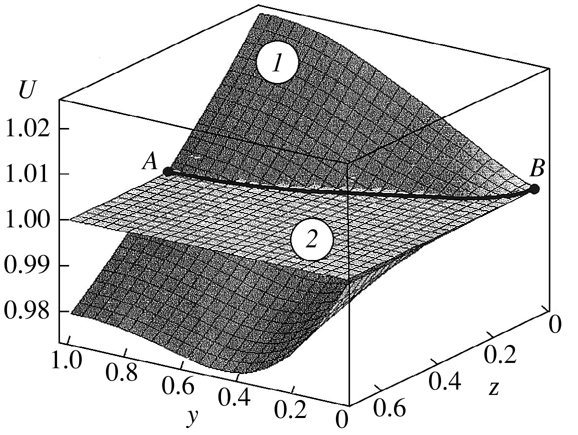


Fig. 5. The dimensionless total potential energy of two charged drops (*1*, *2*) connected by a bridge as a function of distance z between the drops and the ratio y of their radii for $W = 4$ is represented by surface *1*. Surface *2* defines the energy of the initial drop (i.e., $U = 1$).

The portion of line *ck* between points of intersection with lines *ef* and *hj* contains, along with point *f*, minima of the function $U = U(y)$; in the portion of line *ck* between points of intersection with line *di*, the minima are global and correspond to the states at which the energy of two drops is lower than that of a single drop. In the portion of line *ck* between points of intersection with lines *ef* and *di* and in the portion of line *dg* between points of intersection with lines *ef* and *di*, minima of the function $U = U(y)$ are local and correspond to the states at which the energy of two drops is higher than that of a single drop. A numerical calculation produces for the transition point $W_y = 5.702$, $y_1 = 0.8445$, and $y_2 = 1/0.8445$.

Thus, the analysis of the total energy as a function of the characteristic parameters of the system of two contacting spherical drops allows us to separate a number of intervals of Rayleigh's parameter values at which, under the proposed conditions, the following channels for decomposition of a drop exist:

(i) If $0 < W < 5.279$, the only minimum of the function $U = U(y)$ exists at the point $y = 0$; this is the case of a single drop.

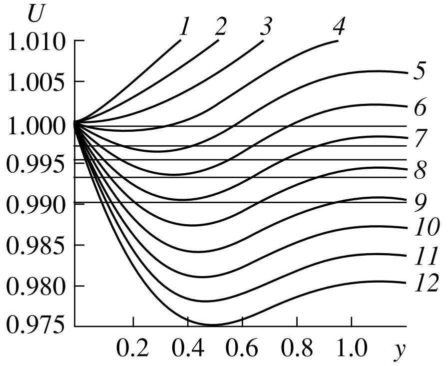


Fig. 6. The dimensionless total potential energy of two charged drops as a function of the ratio of their radii.

(ii) If $5.279 < W < 5.702$, two minima of the function $U = U(y)$ exist at the point $y = 0$ and in the range of $0.4124 < y < 0.8445$; the energy of a single drop is lower than that of two drops.

(iii) If $5.702 < W < 5.7454$, two minima of the function $U = U(y)$ exist at the point $y = 0$ and in the range of $0.8445 < y < 1$; the energy of two drops is lower than that of a single drop.

(iv) If $W > 5.7454$, two minima of the function $U = U(y)$ exist at the points $y = 0$ and $y = 1$; the energy of two identical drops is the lowest one.

Using the analysis made in this section, in particular, we can conclude that, since the instability of separate charged drops sets in at $W \approx 4$, the scheme of decomposition of a multiply charged drop leading to the formation of two contacting drops of nearly spherical form is nonrealizable in the context of the least action principle.

2. We assume that a disintegrating drop consists of two spheres connected by a thin bridge and consider the total energy of the two-drop system as a function of both the ratio of the radii and the distance between these drops. To make the interpretation of calculation results more convenient, we represent the distance between the centers of the spheres as

$$R_{12} = (1+y)R_1 + zR. \qquad (16)$$

As distinct from (9), in (16), z represents a measure of the distance between the spheres expressed in fractions of the radius of initial, rather than daughter, drop. Then, the electrostatic energy (6) can be written as

$$U_e = \frac{Q^2\sqrt[3]{1+y^3}}{2Ry}\left\{\sum_{n=1}^{\infty}\frac{\sinh(t)}{y\sinh(nt)+\sinh((n-1)t)} + \frac{\sinh(t)}{\sinh(nt)+y\sinh((n-1)t)} - \frac{2\sinh(t)}{\sinh(nt)[1+y+z(1+y^3)^{1/3}]}\right\}^{-1},$$

where

$$t = \operatorname{arccosh}\left(\frac{2y+2z(1+y^3)^{1/3}+2yz(1+y^3)^{1/3}+z^2(1+y^3)^{2/3}}{2y}\right).$$

Neglecting the volume and capacitance of the bridge, the total energy that corresponds to two electrostatically interacting drops and is divided by the total energy of the initial spherical drop can be written as

$$U = \left[\frac{1+y^2}{(1+y^3)^{2/3}} + \frac{W}{2}U_{e0}\right]\left(1+\frac{W}{2}\right)^{-1}, \qquad (17)$$

where

$$U_{e0} = \frac{2U_eR}{Q^2}.$$

Figure 5 shows a plot of the function (17) at $W = 4$ and the surface $U = 1$ that corresponds to the total energy of the spherical parent drop. The line AB is related to the states of the two-drop system at which the system energy is equal to the energy of the spherical drop. It can be seen that, for closely spaced drops, their total energy is higher than the energy of the spherical drop and the function $U = U(y)$ has a maximum for some $y > 0$. As the distance between the drops increases, the energy of the system decreases, and the function $U = U(y)$ exhibits a clearly pronounced minimum.

Figure 6 shows the plots of the total dimensionless potential energy $U = U(y)$ as a function of the ratio of the radii of the drops with $W = 4$ for the values of the dimensionless distance between the drops $z = 0.2$, 0.08, 0.14, 0.2, 0.26, 0.32, 0.38, 0.44, 0.5, 0.56, 0.62, and 0.68 (curves *1–12*, respectively). The horizontal lines indicate the values of the dimensionless total potential energy of a spheroidal drop with the same charge and volume and with ratio of its semiaxes $a/b = 1.5$, 2, 2.5, 3, and 4 (from top to bottom). A maximum can be seen in Fig. 6 for $y = 1$. A shift of position of the minimum to the larger values of y as the distance between the drops increases can be also observed. Thus, the study of the total energy of two drops connected by a thin bridge shows that an increase in the length of the bridge not only enhances the probability of decomposition of the drop, but also leads to a decrease in the asymmetry during the decomposition.

3. Assuming that the bridge that connects the drops formed as a result of decomposition has the form of a

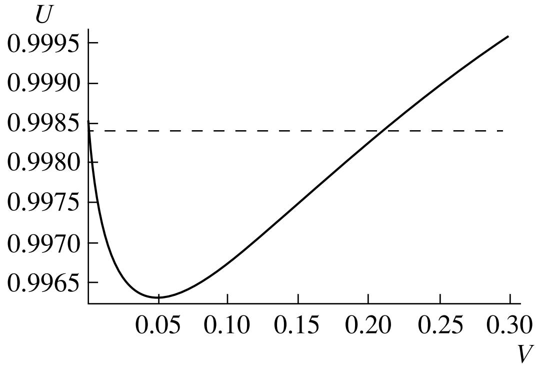


Fig. 7. The energy of two drops as a function of the ratio of their volumes for $W = 4$ and the limiting distance between them when the connecting bridge has the form of a catenoid.

catenoid, it is easy to obtain the maximum value of the dimensionless distance between the centers of the drops that precedes a break in the bridge

$$R_{12} \approx 1.2\frac{1+y}{(1+y^3)^{1/3}}.$$

Results of numerical calculation in this case, as obtained using the above formulas with the help of the Euler–Maclaurin method, are shown in Fig. 7 in the form of a plot of the total potential energy U of two spherical drops as a function of the ratio of their volumes $V = (R_1/R_2)^3$ with $W = 4$ (the solid line). The dashed line in Fig. 7 corresponds to the total potential energy of the parent drop that has the form of a prolate spheroid with ratio of the semiaxes $a/b = 1.8$. The presence of the minimum of the system energy for $V \approx 0.05$ means that, as a result of the spontaneous two-part decomposition of a charged drop, the volume of the smaller daughter drop is about 4.7% of the volume of the initial drop. The calculation according to (5) shows that the smaller daughter drop carries 18% of the charge of the initial drop. The values calculated are close to experimental results obtained in [12] (5% of mass and 23% of charge) but considerably differ from values measured in [13] (25% of mass and 30% of charge). The significant spread of the experimental data can be explained, first, by the fact that the minimum of the energy of the two-drop system under consideration differs from the energy of the initial drop by 0.4% only; second, it could occur because, as can be seen from Fig. 7, the energy of the resulting drops is lower than the energy of the parent drop in a wide range of ratios of masses of the resulting drops $V = 0.01$–0.22. Consequently, small fluctuations in the process or external influences can significantly affect the result of decomposition.

Thus, the assumption that the bridge has the form of a catenoid leads to the conclusion that the decomposition of a drop with Rayleigh's parameter $W = 4$ must proceed asymmetrically, and the most probable ratio of masses of the resulting drops is 1 : 20.

CONCLUSION

In conclusion, we note in the above consideration that both parent and daughter drops have a spherical form, while before the start of the decomposition process, the parent drop has a spheroidal form. Since the total potential energy of spheroidal drops with not very large values of the ellipticity (such as $a/b \leq 2$) differs from the energy of a spherical drop at most by 0.3%, consideration of the spheroidal shape should not bring about considerable corrections to the ratio of masses of resulting drops obtained in this paper; however, it may essentially affect the results of calculation of charge distribution (increasing the charge of a daughter drop). As follows from the above, a small daughter drop is characterized by Rayleigh's parameter $W_1 = 4(0.18)^2/0.047 \approx 2.75$. For the originating small drop to be unstable according to Rayleigh, it is sufficient for the given ratio of masses of resulting drops that the charge of the small drop equals 22% of the charge of the initial drop. In this case, the originating daughter drop is destroyed right after, or even during, the process of separation from a big drop, which was observed in some experiments.

REFERENCES

1. A. I. Grigor'ev and S. O. Shiryaeva, Zh. Tekh. Fiz. **61** (3), 19 (1991) [Sov. Phys.–Tech. Phys. **36**, 258 (1991)].
2. A. V. Eletskiĭ and B. M. Smirnov, Usp. Fiz. Nauk **159** (1), 45 (1989).
3. A. I. Grigor'ev and S. O. Shiryaeva, Izv. Akad. Nauk, Mekh. Zhidk. Gaza **3**, 3 (1994).
4. J. R. Adam, N. R. Lindblad, and C. D. Hendricks, J. Appl. Phys. **39**, 5173 (1968).
5. T. G. O. Berg, R. J. Trainor, Jr., and U. Vaughan, J. Atm. Sci. **27**, 1173 (1970).
6. S. A. Ryce and D. A. Patriarche, J. Phys. **43**, 2192 (1965).
7. V. I. Krasnitskiĭ, A. M. Apasov, and S. M. Kontush, Pis'ma Zh. Tekh. Fiz. **16** (18), 77 (1990) [Sov. Tech. Phys. Lett. **16**, 717 (1990)].
8. A. I. Grigor'ev, S. O. Shiryaeva, I. D. Grigor'eva, *et al.*, Zh. Tekh. Fiz. **61** (4), 25 (1991) [Sov. Phys.–Tech. Phys. **36**, 385 (1991)].
9. V. A. Koromyslov, A. I. Grigor'ev, and S. O. Shiryaeva, Zh. Tekh. Fiz. **68** (8), 31 (1998) [Tech. Phys. **43**, 904 (1998)].
10. A. I. Grigor'ev and S. O. Shiryaeva, J. Phys. D **23**, 1361 (1990).
11. A. Doyle, D. R. Moffett, and B. Vonnegut, J. Colloid Sci. **19**, 136 (1964).
12. J. M. Schweizer and D. N. Hanson, J. Colloid Interface Sci. **35**, 417 (1971).

13. M. A. Abbas and J. Latham, J. Fluid Mech. **30**, 663 (1997).
14. A. Hoareau, P. Melinov, D. Cabaud, *et al*., Chem. Phys. Sci. **143**, 602 (1988).
15. W. Schulze, B. Winter, J. Urban, *et al*., Z. Phys. D **4**, 373 (1987).
16. K. G. Emelius and A. C. Breslin, J. Electron. **54**, 195 (1983).
17. S. A. Ryce and R. R. Wyman, Can. J. Phys. **42**, 2185 (1964).
18. H. M. A. Elghazaly and G. S. P. Castle, IEEE Trans. Ind. Appl. **1A-23**, 108 (1987).
19. L. D. Landau and E. M. Lifshits, *Electrodynamics of Continuous Media* (Nauka, Moscow, 1982).
20. Ya. I. Frenkel’, Zh. Éksp. Teor. Fiz. **9**, 641 (1939).
21. N. Bohr and J. A. Wheeler, Phys. Rev. **56**, 426 (1939).
22. W. R. Smythe, *Static and Dynamic Electricity*, 2nd ed. (McGraw-Hill, New York, 1950; Moscow, 1954).

Translated by V. Gurskiĭ

Technical Physics, Vol. 45, No. 4, 2000, pp. 389–395. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 8–13.
Original Russian Text Copyright © 2000 by Grigor'ev, Zharov.

GASES AND FLUIDS

Stability of the Equilibrium States of a Charged Bubble in a Dielectric Liquid

A. I. Grigor'ev and A. N. Zharov

Yaroslavl State University, Sovetskaya ul. 14, Yaroslavl, 150000 Russia

Received December 23, 1998

Abstract—The stability of equilibrium states of a charged spherical bubble in dielectric fluid with respect to centrally symmetric variations of its volume is studied by analyzing a nonlinear equation describing radial oscillations of the bubble in the neighborhoods of its singularities. It is shown that only one of the two possible equilibrium states of the bubble is stable. The boundaries separating the domains in the physical parameter space are found that correspond to stable and unstable states. It is found that when the bubble carries an electric charge, the domains of physical parameters corresponding to equilibrium states of the bubble expand. *© 2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Stability analysis of charged bubbles in fluids is of interest because of numerous technical applications involving such physical objects. It relates to studies of electric breakdown of liquid dielectrics, in which an important role is played at the initial stage by cavitation complicated by the ejection of an uncompensated electric charge into the cavity from the electrode surface where the bubble appears [1–4]; electrical explosion in liquids, when a pulsating cavity filled by a long-lived discharge plasma appears in a liquid [5–7]; cavitation phenomena in dielectric liquids accompanied by generation of plasmas that emit electron, neutron, and x-ray radiation [8]; electric flotation, when bubble formation in a liquid is associated with electrolysis [9]; and cavitation [10]. However, the problem is still at an early stage of theoretical understanding, even though some preliminary results concerning the disintegration of charged bubbles have been obtained [11–14]. According to [11, 14], the spontaneous breakup of a bubble that is unstable with respect to its own charge (as well as that of a charged drop [15, 16]) into secondary ones can proceed by different scenarios depending on the surface conductivity of the bubble, which may substantially exceed the conductivity of the ambient dielectric liquid. Symmetry considerations [17] make it clear that the conductivities of both the bubble walls and the ambient dielectric liquid do not affect the stability of radial centrally symmetric oscillations of the bubble. In [18], a qualitative analysis of the stability of a charged bubble was performed in the neighborhood of one of the two equilibrium states of the bubble, characterized by the larger radius, and the state proved to be unstable.

In this paper, we present a qualitative analysis of the stability of a charged gas–vapor bubble in a liquid dielectric.

1. Consider a spherical bubble of radius R_0 in a space-filling liquid dielectric with permittivity ε. The bubble wall is subject to capillary forces with surface tension σ and to the electric-field pressure generated by the charge Q of the bubble. A steady spherical bubble in equilibrium must satisfy the following condition of pressure balance on its wall:

$$P_0 = P_V + P_g - \frac{2\sigma}{R_0} + \frac{Q^2}{8\pi\varepsilon R_0^4}, \qquad (1)$$

where P_0 is the initial liquid pressure in the neighborhood of the bubble determined by the atmospheric pressure on the free surface and the hydrostatic pressure of the liquid column above the bubble, P_V denotes the saturated-vapor pressure of the ambient liquid inside the bubble, and P_g is the gas pressure in the bubble. The third term on the right-hand side of (1) describes the capillary pressure under the spherical surface, while the fourth one represents the electric-field pressure generated by the bubble charge.

Unlike the ambient liquid, which is treated as incompressible, we allow for the gas and vapor in the bubble being essentially compressible. We consider the stability of the bubble with respect to radially symmetric variations of its volume; the variations may be associated with radially symmetric motion of the bubble wall. For small bubbles, such motion is characterized by small times that nevertheless exceed the gas-pressure relaxation time in the bubble, which is determined

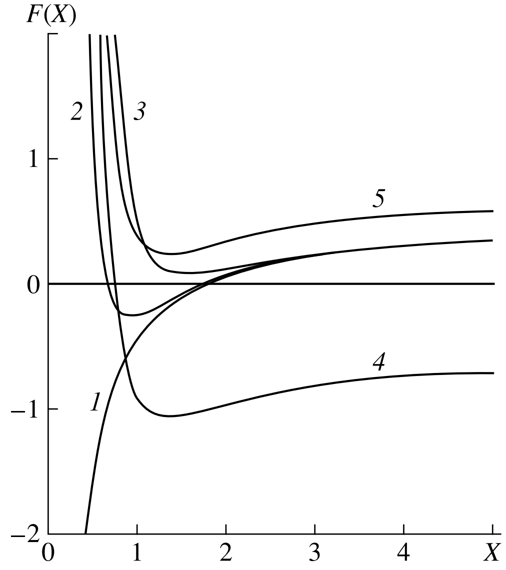


Fig. 1. Dependence of the resultant dimensionless pressure F acting on the bubble wall on the dimensionless bubble radius X for adiabatic variation of the state of a gas–vapor mixture with $n = 4/3$: $\beta_* = -0.55$ (*1–3*), –0.8 (*4*), 0.5 (*5*); $\beta = 0$ (*1–3*), 0.1 (*4*, *5*); $W = 0$ (*1*), 0.2 (*2*), 1 (*3*), 0.1 (*4*, *5*).

by the order of magnitude of the ratio of the bubble radius to the sonic velocity for the gas–vapor mixture filling the bubble. Under these conditions, we assume that the saturated-vapor pressure P_V is constant and the gas pressure P_g varies according to a polytropic law. Then, (1) can be rewritten as

$$P = P_V + \left(P_0 - P_V + \frac{2\sigma}{R_0} - \frac{Q^2}{8\pi\varepsilon R_0^4}\right)\left(\frac{R_0}{R}\right)^{3n} - \frac{2\sigma}{R} + \frac{Q^2}{8\pi\varepsilon R^4}, \qquad (2)$$

where n is the polytropic exponent and P denotes the liquid pressure in the neighborhood of the bubble. The second term on the right-hand side of (2) describes the gas pressure variation in the bubble associated with variation of the volume according to a polytropic law.

In dimensionless form, this relation can be written as

$$F(X) \equiv \frac{\beta}{X^{3n}} + \frac{W}{X^4} - \frac{1}{X} - \beta_* = 0,$$

$$X = \frac{R}{R_0}, \quad W = \frac{Q^2}{16\pi\varepsilon\sigma R_0^3}, \quad \beta_* = \frac{(P - P_V)R}{2\sigma}, \qquad (3)$$

$$\beta = \frac{R_0}{2\sigma}\left(P_0 - P_V + \frac{2\sigma}{R_0} - \frac{Q^2}{8\pi\varepsilon R_0^4}\right),$$

where X is the dimensionless radius of the bubble, W is a dimensionless parameter characterizing the bubble stability with respect to electric charge (introduced by analogy with Rayleigh’s parameter, which characterizes the stability of a charged drop with respect to thermal capillary oscillation [15, 16], and is also called Rayleigh’s parameter in what follows), β_* is a dimensionless parameter characterizing the liquid pressure, and β is the dimensionless initial gas pressure in the bubble.

The bubble volume increases when $F(X) > 0$ and decreases when $F(X) < 0$. When $F(X) = 0$, the bubble is in equilibrium. The qualitative behavior of $F = F(X)$ is illustrated in Fig. 1 for various typical situations.

Applying the Descartes sign rule to (3), one finds that equation (3) can have a different number of real positive roots (two, one, or none) corresponding to the static equilibrium of the bubble for various values of parameters β_*, W, and β (see Fig. 1). In the space of parameters β_*, W, and β, we single out three characteristic domains where solutions to equation (3) exist and denote them by A, B, and C.

In domain A, where the parameter β_*, which characterizes the pressure in the liquid, is positive ($\beta_* \geq 0$) and both Rayleigh’s parameter W and parameter β, which characterize the initial gas pressure, are arbitrary, there exists a unique solution to equation (3) describing an equilibrium state of a bubble with radius $X = X_*$.

In domain B, these exist two real positive solutions $X = X_{*1}$ and $X = X_{*2}$ to equation (3) corresponding to two possible stationary equilibrium radii of the bubble. In the space of parameters β_*, β, and W, domain B is defined by the following conditions: the parameter β_*, which characterizes the liquid pressure, is negative but lies above a certain critical value ($\beta_{*\mathrm{cr}} < \beta_* < 0$), whereas both Rayleigh’s parameter W and the gas pressure parameter β lie below their critical values ($W < W_{\mathrm{cr}}$ and $\beta < \beta_{\mathrm{cr}}$).

In domain C, where the parameter β_* is negative and lies below its critical value ($\beta_* < \beta_{*\mathrm{cr}}$) and both Rayleigh’s parameter W and the gas pressure parameter β lie above their critical values ($W > W_{\mathrm{cr}}$ and $\beta < \beta_{\mathrm{cr}}$), no solutions to equation (3) exist; i.e., equilibrium bubbles cannot exist in a liquid in this domain of parameters.

The critical values $\beta_{*\mathrm{cr}}$, β_{cr}, and W_{cr} are interpreted as the values that separate domains B and C in the space of β_*, β, and W. The boundary of domain A in the space of β_*, β, and W is defined by the plane $\beta_* = 0$.

Because the function $F = F(X)$ has a single minimum (Fig. 1), we can find a relation between the physical parameters β_*, β, and W that defines the boundary between domains *B* and *C* by requiring that the following conditions hold simultaneously:

$$F(X) = 0, \quad \frac{dF}{dX} = 0. \tag{4}$$

Under conditions (4), both roots of the equation $F(X) = 0$ merge into a multiple root at the point of minimum of the curve $F = F(X)$ where the curve is tangent to the abscissa axis. Thus, the equation $F(X) = 0$ has two identical solutions on the boundary surface.

The desired relation when $n = 4/3$ (variations of the gas volume in the bubble are adiabatic) has the following simple analytical form:

$$W + \beta + \frac{27}{256\beta_*^3} = 0.$$

Numerical calculations show that, if $\beta \neq 0$, then the critical values $\beta_{*\mathrm{cr}}$, W_{cr}, and β_{cr} decrease with the polytropic exponent *n*.

Physically, domain *C* corresponds to the case of a highly superheated liquid (or a liquid moving fast in a narrowing pipe or channel), when the saturated-vapor pressure in the bubble exceeds the liquid pressure. A charged gas–vapor bubble that accidentally forms in such a liquid grows indefinitely, because the saturated-vapor pressure inside it, which acts to expand the bubble, does not depend on the radius, whereas the Laplacian pressure decreases as the bubble radius increases. The growth rate of such a bubble is controlled by the rate of evaporation of liquid from the bubble wall. The energy required to increase the radius is derived from the thermal energy of the superheated liquid (or from the kinetic energy of the moving liquid in the case when the saturated-vapor pressure exceeds the liquid pressure because of the high flow velocity in a narrowing pipe).

Domain *B* also corresponds to a superheated liquid, but with a lower degree of superheating. The smaller equilibrium bubble radius is determined in the zeroth approximation by the balance between the sum of the Coulomb and gas pressures on the one hand and the Laplacian pressure on the other hand. The larger equilibrium bubble radius is determined in the zeroth approximation by the balance between the saturated-vapor and Laplacian pressures. A change in its radius can bring the bubble to the stable state with a smaller radius or make its radius indefinitely increase under the action of the saturated-vapor pressure. Indeed, when the radius decreases, the saturated vapor pressure does not change whereas the Laplacian pressure increases. The Coulomb and gas pressures, which depend more strongly on the radius and determine the equilibrium state of the bubble with the smaller radius, are weak at the larger radius corresponding to a nonequilibrium bubble and their influence on the other equilibrium radius can be neglected in zeroth approximation. When the radius increases, the saturated-vapor pressure does not change and the Laplacian pressure decreases. The role of an energy source for such variations of volume is played by the thermal energy of superheated liquid (or the kinetic energy of the moving liquid in the case when the saturated-vapor pressure exceeds the liquid pressure because of the high flow velocity in a narrowing pipe).

In a real physical situation, the motion of the bubble wall typical of domains *B* and *C* may be associated with the possibility of cavitation phenomena in various technical and technological devices. Such motions may also occur during electric discharges in liquids when the energy released in a pulse is high and results in unsteady nonuniform heating of the liquid in the neighborhood of an incipient bubble.

Domain *A* is not associated with any uncommon properties of the liquid and gas–vapor mixture in the bubble, and the physical meaning of the laws that govern the existence of equilibrium gas–vapor bubbles in this domain is obvious.

2. We now derive the differential equation describing radial centrally symmetric oscillations of a charged bubble in a viscous dielectric liquid by analogy with the way Lord Rayleigh did it for an uncharged bubble in an inviscid liquid [19]. The mathematical statement of the hydrodynamic problem of calculating the velocity field $\mathbf{V} = \mathbf{V}(\mathbf{r}, t)$ of the viscous flow in the neighborhood of a bubble of radius *R* has the form

$$\frac{\partial \mathbf{V}}{\partial t} + (\mathbf{V}\nabla)\mathbf{V} = -\frac{1}{\rho}\nabla P + \nu\Delta\mathbf{V}, \tag{5}$$

$$\operatorname{div}\mathbf{V} \equiv \Delta\psi = 0; \quad \mathbf{V} = \operatorname{grad}\psi, \tag{6}$$

$$r = R\colon V = \frac{dR}{dt} = \frac{\partial\psi}{\partial r}, \tag{7}$$

$$-P + 2\rho\nu\frac{\partial V}{\partial r} + \frac{2\sigma}{R} - \frac{Q^2}{8\pi\varepsilon R^4} = P_g, \tag{8}$$

$$r \longrightarrow \infty\colon V \longrightarrow 0, \tag{9}$$

$$r < R\colon p_g = P_g(V). \tag{10}$$

Here, ν and ρ are the kinematic viscosity and density of the liquid, respectively. The velocity field $\mathbf{V} = \mathbf{V}(\mathbf{r}, t)$ is centrally symmetric and, therefore, irrotational. Applying the operator curl to both sides of the Navier–Stokes equation (5), one finds that $\mathbf{V} = \operatorname{grad}\psi$ eliminates the viscous term. The influence of viscosity on the motion of the bubble wall and the liquid in its neighborhood is represented by the dynamical boundary condition (8). Relation (10) is the equation of state for the medium inside the bubble.

Let us assume that the gas–vapor mixture filling the bubble is both dynamically and thermally uniform at any moment of time and analyze the motion of the liquid outside the bubble only. We suppose that the physical parameters characterizing the gaseous medium inside the bubble are directly related to the corresponding quantities on the moving bubble wall, which is true when the speed of the bubble wall, being essentially dependent on the viscosity, is much lower than the sonic velocity in the gas. The motion of the bubble wall and the liquid in its neighborhood associated with variation of its volume is assumed to be spherically symmetric. Then, the liquid in the neighborhood of the bubble executes a potential motion, and the velocity potential $\psi = \psi(\mathbf{r}, t)$ is a harmonic function.

The solution to equation (6) satisfying boundary conditions (7) and (9) is readily written out:

$$\psi(\mathbf{r}, t) = -\frac{R^2}{r}\frac{dR}{dt}. \tag{11}$$

The velocity field in the neighborhood of the bubble is then expressed as

$$V(\mathbf{r}, t) = \frac{R^2}{r^2}\frac{dR}{dt}. \tag{12}$$

The first integral of equation (5) can easily be written in the well-known Cauchy–Lagrange form:

$$\frac{\partial\psi}{\partial t} + \frac{V^2}{2} - \frac{1}{\rho}(P - P_0) = f(t). \tag{13}$$

Consider this equation on the surface of the bubble (at $r = R$) and set $f(t) = 0$, which can be done by virtue of some freedom in defining the velocity potential $\psi(\mathbf{r}, t)$. Then, the pressure exerted by the liquid on the bubble wall can be determined from the dynamical boundary condition (8):

$$\begin{aligned} P = P_V &+ \left(P_0 - P_V - \frac{Q^2}{8\pi\varepsilon R_0^4} + \frac{2\sigma}{R_0}\right)\left(\frac{R_0}{R}\right)^n \\ &- \frac{2\sigma}{R} + \frac{Q^2}{8\pi\varepsilon R^4} - 2\rho\nu\frac{\partial V}{\partial r}. \end{aligned} \tag{14}$$

Substituting (14) into (13) and using (7), (11), and (12), we obtain the differential equation of motion for the bubble wall:

$$\begin{aligned} R\frac{d^2R}{dt^2} &+ \frac{3}{2}\left(\frac{dR}{dt}\right)^2 - \frac{1}{\rho}\left(P_0 - P_V + \frac{2\sigma}{R_0} - \frac{Q^2}{8\pi\varepsilon R_0^4}\right)\left(\frac{R_0}{R}\right)^{3n} \\ &+ \frac{4\nu}{R}\frac{dR}{dt} + \frac{2\sigma}{\rho R} - \frac{Q^2}{8\pi\varepsilon\rho R^4} - \frac{P_V - P_0}{\rho} = 0. \end{aligned} \tag{15}$$

Taking (2) into account and introducing dimensionless variables such that $\rho = 1$, $\sigma = 1$, and $R_0 = 1$, we rewrite equation (15) in the form

$$\begin{gathered} X\frac{d^2X}{d\tau^2} + \frac{3}{2}\left(\frac{dX}{d\tau}\right)^2 + \frac{\eta}{X}\frac{dX}{d\tau} - F(X) = 0, \\ \tau = \frac{t}{R_0^{3/2}}\left(\frac{2\sigma}{\rho}\right)^{1/2}, \quad \eta = 2\nu\left(\frac{2\rho}{\sigma R_0}\right)^{1/2}, \end{gathered} \tag{16}$$

where τ is the dimensionless time and η is the dimensionless kinematic viscosity of the liquid.

Depending on the values of the physical parameters W, β_*, and β, the last equation has one, two, or no singular points corresponding to the equilibrium states of the bubble. We perform a qualitative analysis of the differential equation (16) to examine the dynamics of the bubble wall near the singularities.

Equation (16) is equivalent to a system of two first-order differential equations:

$$\frac{dY}{d\tau} = -\frac{3}{2X}Y^2 - \frac{\eta}{X^2}Y + \frac{F(X)}{X},$$

$$\frac{dX}{d\tau} = Y.$$

The singular points $X = X_*$ and $Y = Y_*$ of such a system in the plane (X, Y) are defined by the following simultaneous conditions:

$$\frac{dY}{d\tau} = 0, \quad \frac{dX}{d\tau} = 0, \tag{17}$$

which yield the relations

$$Y = Y_* = 0, \quad F(X_*) = 0.$$

The coordinate X of a singular point is the radius of an equilibrium bubble, whereas the coordinate Y is zero, which implies zero velocity of the bubble wall. It is clear that the number of singular points is completely determined by the number of solutions to equation (3), which may be one, two, or zero. Thus, to describe the dynamics of the bubble wall deviating from equilibrium, it is sufficient to determine the type of singular point of system (17) and use the qualitative theory of differential equations.

3. Linearizing the system of differential equations (17) in the neighborhood of a singular point, we have

$$\begin{gathered} \frac{d\xi}{d\tau} = -\frac{\eta}{X_*^2}\xi + \frac{D}{X*}\zeta, \quad \frac{d\zeta}{d\tau} = \xi, \\ D = \left(\frac{dF}{dX}\right)_{X = X_*}, \end{gathered} \tag{18}$$

where D is the derivative of the resultant pressure with respect to the bubble radius at the singular point, and $\zeta \equiv X - X_*$ and $\xi \equiv Y - Y_*$ are first-order quantities describing an instantaneous deviation of the point on the phase plane from the singular point.

System (18) is a homogeneous system of two first-order linear differential equations, and the corresponding secular equation is

$$S^2 + \frac{\eta}{X_*^2}S - \frac{D}{X_*} = 0. \qquad (19)$$

The roots of secular equation (19) have the form

$$S_{1,\,2} = -\frac{\eta}{2X_*^2} \mp \left(\frac{\eta^2}{4X_*^4} + \frac{D}{X_*}\right)^{1/2}. \qquad (20)$$

According to (20), the roots of the secular equation can be negative, positive, or complex, depending on the value and sign of the derivative $D \equiv [dF/dX]_{X = X_*}$.

4. In domain A, which contains a single equilibrium state of the bubble with dimensionless radius $X = X_*$, the derivative D is negative. If $D \geq -\eta^2/(4X_*^3)$, which is possible when the viscosity is high, roots (20) of secular equation (19) are real and negative and the singular point represents a stable node. In this case, a change in the volume of the bubble leads to an aperiodic motion of the bubble wall toward equilibrium. This situation is illustrated by Fig. 2. The arrows in Fig. 2b indicate the direction of trajectories in the phase plane.

If $D < -\eta^2/(4X_*^3)$, the roots of secular equation (19) are complex with negative real parts and the singular point is a stable focus. In this case, the equilibrium state of the bubble is also stable and any variation of the bubble volume results in a decaying periodic motion of the bubble wall toward the equilibrium value (Fig. 3) with the frequency

$$\omega \equiv \operatorname{Im} S = \left(\frac{3n\beta}{X_*^{3n+2}} + \frac{4W}{X_*^6} - \frac{1}{X_*^3} - \frac{\eta^2}{4X_*^4}\right)^{1/2}. \qquad (21)$$

The dependence of the frequency ω of capillary oscillations of the bubble in the neighborhood of equilibrium on the dimensionless viscosity η is clear from (21). The dependence of ω on the dimensionless parameters W, β, and β_*, which is determined both explicitly by (21) and implicitly by the dependence $X_* = X_*(W, \beta, \beta_*)$ of the singular point on the parameters, is shown in Fig. 4. It is clear that the frequency increases with the liquid pressure β_* and decreases as Rayleigh's parameter W, which characterizes the bubble charge, and the initial pressure β increase. In addition, numerical calculations show that the frequency ω increases with the polytropic exponent n.

In domain B, where two equilibrium states, $X = X_{*1}$ and $X = X_{*2}$, exist, the derivative D is negative for the smaller equilibrium radius, $X = X_{*1}$ and positive for the larger one, $X = X_{*2}$. It follows from (20) that, for the smaller value of the equilibrium radius, the roots of the

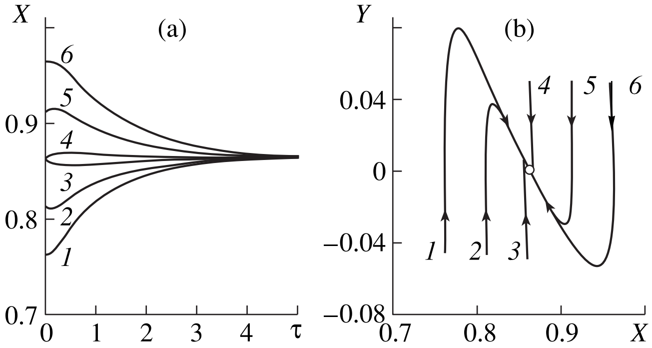


Fig. 2. Motion of the bubble wall in domain A initiated by deviation from equilibrium for adiabatic variation of the state of the gas–vapor mixture ($n = 4/3$) when the singular point is a stable node with $W = 0.2$, $\beta = 0.5$, $\beta_* = 0.1$, and $\eta = 5$: (a) time dependence of the bubble radius; (b) phase diagram illustrating the relation between the dimensionless velocity of the bubble wall and its dimensionless radius. The circle represents the singular point. (*1*) $X_0 = X_* - 0.1$, $Y_0 = -0.05$; (*2*) $X_0 = X_* - 0.05$, $Y_0 = -0.05$; (*3*) $X_0 = X_*$, $Y_0 = -0.05$; (*4*) $X_0 = X_*$, $Y_0 = -0.05$; (*5*) $X_0 = X_* + 0.05$, $Y_0 = 0.05$; (*6*) $X_0 = X_* + 0.1$, $Y_0 = 0.05$.

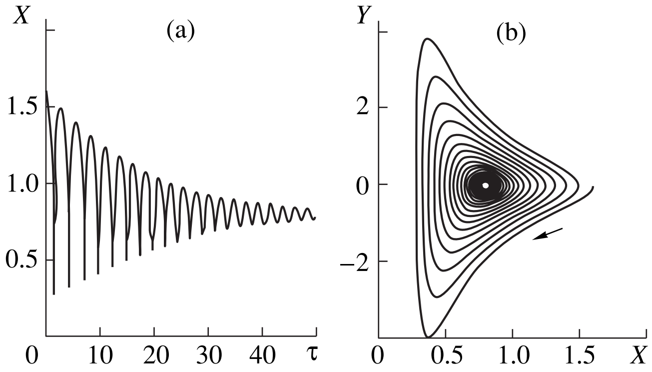


Fig. 3. Same as Fig. 2 when the singular point is a stable focus: (a) time dependence of the bubble radius for $W = 0.3$, $\beta = 0.5$, $\beta_* = 0.5$, $\eta = 8 \times 10^{-2}$, $Y_0 = 0$, and $X_0 = X_* + 0.8$; (b) phase diagram illustrating the relation between the dimensionless velocity of the bubble wall and its dimensionless radius. The arrow indicates the direction of trajectories.

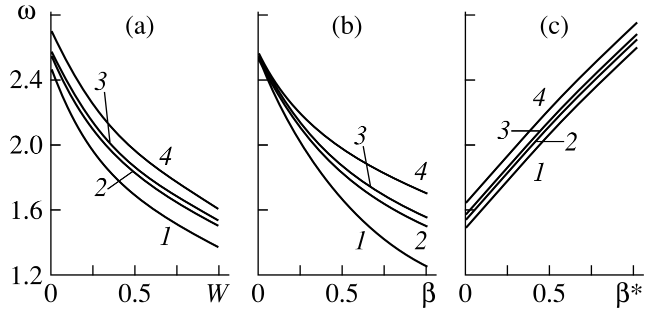


Fig. 4. Dependence of the frequency of capillary oscillations of the bubble in the neighborhood of equilibrium on W, β, and β_* for $\eta = 8 \times 10^{-2}$, $\beta = 0.5$ (a, c), $\beta_* = 0.05$ (a–c), $W = 0.5$ (b, c), and $n = 1$ (*1*), 4/3 (*2*), 7/5 (*3*), and 5/3 (*4*).

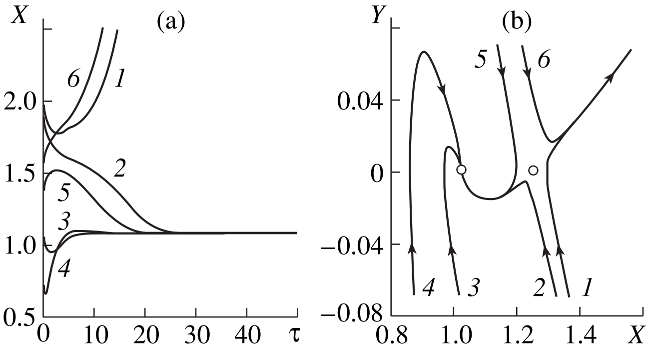


Fig. 5. Motion of the bubble wall in domain B initiated by deviation from equilibrium for adiabatic variation of the state of the gas–vapor mixture ($n = 4/3$) when the singular point with smaller radius is a stable node and the one with larger radius is a saddle with $W = 0.21$, $\beta = 0.15$, $\beta_* = -0.65$, and $\eta = 1$: (a) time dependence of the bubble radius; (b) phase diagram illustrating the relation between the dimensionless velocity of the bubble wall and its dimensionless radius. (*1*) $X_0 = X_{*2} + 0.12$, $Y_0 = -0.07$; (*2*) $X_0 = X_{*2} + 0.08$, $Y_0 = -0.07$; (*3*) $X_0 = X_{*1}$, $Y_0 = -0.07$; (*4*) $X_0 = X_{*1} - 0.15$, $Y_0 = -0.07$; (*5*) $X_0 = X_{*2} - 0.12$, $Y_0 = 0.07$; (*6*) $X_0 = X_{*2} - 0.04$, $Y_0 = 0.07$.

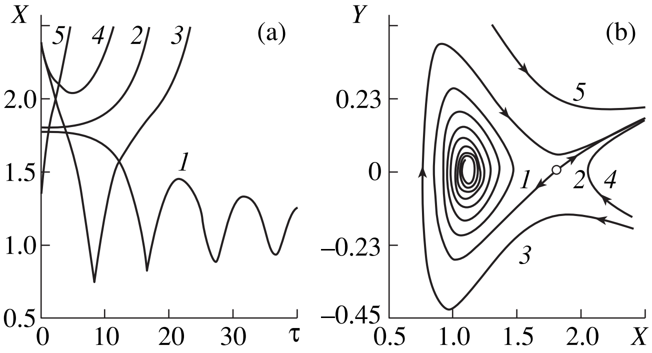


Fig. 6. Same as Fig. 5 when the singular point with smaller radius is a stable focus and the one with larger radius is a saddle for $W = 0.2$, $\beta = 0.4$, $\beta_* = -0.5$, and $\eta = 8 \times 10^{-2}$: (a) time dependence of the bubble radius; (b) phase diagram illustrating the relation between the dimensionless velocity of the bubble wall and its dimensionless radius. (*1*) $X_0 = X_{*2} - 0.01$, $Y_0 = 0$; (*2*) $X_0 = X_{*2} + 0.01$, $Y_0 = 0$; (*3*) $X_0 = X_{*2} + 0.6$, $Y_0 = -0.18$; (*4*) $X_0 = X_{*2} + 0.6$, $Y_0 = -0.12$; (*5*) $X_0 = X_{*2} - 0.5$, $Y_0 = 0.45$.

secular equation are real and negative if $D \geq -\eta^2/(4X_*^3)$ and complex with a negative real part if $D < -\eta^2/(4X_*^3)$. The singular point corresponding to this state can be either a stable node or a stable focus. For the larger equilibrium radius of the bubble, the roots of secular equation (20) are real and have opposite signs; therefore, the singular point is a saddle. Any change in the bubble volume in the second equilibrium state drives it away from this state, aperiodically decreasing or increasing its radius. These motions are illustrated by Figs. 5 and 6. It is clear that the bubble in the second equilibrium (with the larger radius) can expand or contract and shift toward the first equilibrium (with the smaller radius). It is also clear that the bubble can reach the stable equilibrium only from a bounded domain in the phase plane in the neighborhood of the first equilibrium. A bubble with a radius different from an equilibrium value and with a sufficiently large positive or negative wall velocity may pass both equilibrium states and start to expand indefinitely (within the limits of validity of the present model).

CONCLUSIONS

When a bubble is electrically charged, its stability increases with respect to axially symmetric radial variations of its volume in the sense that the domains of gas–vapor pressure in the bubble and ambient liquid pressure where equilibrium states of the bubble exist become larger.

A charged bubble in a dielectric fluid may have one, two, or no equilibrium states depending on its electric charge, initial gas pressure in it, ambient liquid pressure, and saturated-vapor pressure of the liquid. The equilibrium states of the bubble can be stable or unstable. Depending on viscosity, electric charge on the bubble wall, and other physical parameters, a change in the bubble volume in a stable equilibrium leads to periodic or aperiodic motion of the bubble wall toward the equilibrium. In the case of periodic motion, the frequency of small-amplitude oscillations of the bubble increases with the polytropic exponent and liquid pressure and decreases as the charge on the bubble wall and initial pressure of the gas in the bubble increase. A change in the charged-bubble volume in an unstable equilibrium results in transition to the stable equilibrium or indefinite expansion, which drives the system out of the scope of the proposed idealized model if its energy is sufficiently high.

REFERENCES

1. C. G. Garton and Z. Krasucki, Trans. Faraday Soc. **60**, 211 (1964).
2. V. F. Klimkin, Zh. Tekh. Fiz. **60**, 161 (1990).
3. I. V. Pylaeva, O. A. Sinkevich, and P. V. Smirnov, Teplofiz. Vys. Temp. **30**, 367 (1992).
4. S. M. Korobeĭnikov, Doctoral Dissertation in Mathematical Physics (Tomsk, 1998).
5. V. V. Ivanov, I. S. Shvets, and A. V. Ivanov, *Underwater Spark Discharges* [in Russian] (Naukova Dumka, Kiev, 1982).
6. P. I. Golubnichiĭ, V. M. Gromenko, and V. M. Krutov, Zh. Tekh. Fiz. **60**, 183 (1990).
7. V. G. Kovalev, Zh. Tekh. Fiz. **66**, 24 (1996).
8. A. I. Koldamasov, Zh. Tekh. Fiz. **61**, 188 (1991).
9. Ya. E. Geguzin, *Bubbles* [in Russian] (Nauka, Moscow, 1985).

10. N. I. Zharov, Zh. Tekh. Fiz. **60**, 22 (1990).
11. A. I. Grigor'ev, V. A. Koromyslov, and A. N. Zharov, Pis'ma Zh. Tekh. Fiz. **23**, 60 (1997) [Tech. Phys. Lett. **23**, 760 (1997)].
12. A. I. Grigor'ev, A. N. Zharov, V. A. Koromyslov, *et al.*, Izv. Akad. Nauk, Mekh. Zhidk. Gaza, No. 5, 205 (1998).
13. A. N. Zharov, Pis'ma Zh. Tekh. Fiz. **24**, 49 (1998) [Tech. Phys. Lett. **24**, 849 (1998)].
14. S. O. Shiryaeva and A. N. Zharov, Pis'ma Zh. Tekh. Fiz. **24**, 30 (1998) [Tech. Phys. Lett. **24**, 959 (1998)].
15. A. I. Grigor'ev and S. O. Shiryaeva, Zh. Tekh. Fiz. **61**, 19 (1991).
16. A. I. Grigor'ev and S. O. Shiryaeva, Izv. Akad. Nauk, Mekh. Zhidk. Gaza, No. 3, 3 (1994).
17. A. I. Grigor'ev and S. O. Shiryaeva, Izv. Akad. Nauk, Mekh. Zhidk. Gaza, No. 5, 108 (1997).
18. A. I. Grigor'ev, Pis'ma Zh. Tekh. Fiz. **25**, 25 (1999) [Tech. Phys. Lett. **25**, 10 (1999)].
19. Lord Rayleigh (J. W. Strutt), Philos. Mag. **34**, 93 (1917).

Translated by V.V. Gurskiĭ

Technical Physics, Vol. 45, No. 4, 2000, pp. 396–399. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 14–17.
Original Russian Text Copyright © 2000 by Aleksandrov, Rukhadze.

GAS DISCHARGES, PLASMA

Kinetic Theory of Surface Waves in a Semibounded Plasma Flow

A. F. Aleksandrov and A. A. Rukhadze

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

Received September 28, 1998

Abstract—A study is made of the spectrum of surface waves in a semibounded plasma flow. The frequency spectra and damping rates of the waves propagating along the flow are analyzed both in the high-frequency range (in which the spatial dispersion is weak and the wave damping is governed primarily by electron collisions) and the low-frequency range (in which the spatial-dispersion effects dominate), with focus on the effect of the flow velocity on the propagation of ion-acoustic waves. Special attention is paid to the penetration of a static field into a plasma flowing at a supersonic velocity. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION AND FORMULATION OF THE PROBLEM

Surface waves in a plasma have been studied for a long time [1–3], primarily because of the features of their frequency spectrum $\omega(\mathbf{k})$. Since surface waves exist in the frequency range $\omega < \omega_{Le}$, they can be used to produce an overdense plasma, such as in plasma sources known as surfatrons [4].

Here, we study a surface wave in a plasma flow as a possible mechanism for ionizing a supersonic gas stream with the purpose of creating a supersonic plasma source capable of producing high-density plasmas [5, 6]. We consider the simplest case of a semibounded flow of an isotropic (unmagnetized) plasma, assuming that the surface wave propagates along the plasma flow. We examine both the high-frequency range, in which the surface wave is a purely electron wave and the spatial dispersion in the plasma permittivity is negligible, and the low-frequency range, in which the ion motion and spatial dispersion effects play the dominant role. In particular, we investigate an ion-acoustic surface wave in a plasma flow and determine whether it is possible for a static field to penetrate into a plasma flowing at a supersonic velocity.

First, we formulate the problem for the case of a collisionless plasma on the basis of the Vlasov equation. Then, we generalize the dispersion relation obtained to the case of a collisional but weakly ionized plasma described by the model Bhatnagar–Gross–Krook (BGK) collision integral [3]. Following [3], we write the Vlasov equation for small perturbations of the equilibrium distribution functions and the boundary condition corresponding to the specular reflection of electrons and ions as follows:

$$-i(\omega - k_z v_z)\delta f + v_x \frac{\partial \delta f}{\partial x} + \frac{e}{m}\left\{\mathbf{E} + \frac{1}{c}[\mathbf{vB}]\right\}\frac{\partial f_0(\mathbf{v}-\mathbf{u})}{\partial \mathbf{v}} = 0, \tag{1}$$

$$\delta f(x = 0, v_x > 0) = \delta f(x = 0, v_x < 0).$$

Here, the perturbation is specified as $\delta f \approx \delta f(x) c^{-i\omega t + ik_z}$; $f_0(\mathbf{v} - \mathbf{u})$ is the Maxwellian distribution function in which the directed flow velocity $\mathbf{u}$ is parallel to the plasma surface, $\mathbf{u} \parallel OZ$; and, for brevity, the subscripts characterizing the particle species are omitted. We supplement the Vlasov equation (1) with the Maxwell equations

$$\operatorname{div}\mathbf{E} = 4\pi\rho, \quad \operatorname{curl}\mathbf{B} = +\frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c}\mathbf{J}, \qquad \operatorname{curl}\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}, \quad \operatorname{div}\mathbf{B} = 0, \tag{2}$$

where the charge density ρ and the current density $\mathbf{j}$ are defined by

$$\rho = \sum e\int d\mathbf{p}\,\delta f, \quad \mathbf{j} = \sum e\int d\mathbf{p}\mathbf{v}\,\delta f \tag{3}$$

with the summation over all species of the charged particles.

The boundary conditions for the fields $\mathbf{E}$ and $\mathbf{B}$ can be derived by directly integrating equations (2) near the boundary $x = 0$ between the plasma ($x < 0$) and vacuum

1063-7842/00/4504-0396$20.00 © 2000 MAIK "Nauka/Interperiodica"

($x > 0$) regions:

$$\{E_t\}_{x=0} = 0, \quad \{B_n\}_{x=0} = 0,$$
$$\{E_n\}_{x=0} = 4\pi\rho_n, \quad \{B_t\}_{x=0} = \frac{4\pi}{c}\mathbf{j}_n, \tag{4}$$

where the surface densities of the induced charges and currents, ρ_n and $\mathbf{j}_n$, can be determined by integrating the first and second equations in (2).

In order to complete the formulation and to make the method used in [3] applicable to solving the problem at hand, we represent the fields in (2) as $\mathbf{E}$, $\mathbf{B} \approx \mathbf{E}$, and $\mathbf{B}\exp(-i\omega t + ik_z z)$ and pass over to the rest frame of the plasma flow. The latter is equivalent to the replacement $\mathbf{v} - \mathbf{u} \longrightarrow \mathbf{v}$, in which case $f_0(\mathbf{v} - \mathbf{u})$ becomes an isotropic Maxwellian distribution function. As a result, we obtain

$$-i(\omega - k_z v_z)\delta f + v_x\frac{\partial \delta f}{\partial x} = -\frac{e}{m}\left\{\frac{\omega'}{\omega}\mathbf{E} + \frac{\mathbf{k}(\mathbf{uE})}{\omega}\right\}\frac{\partial f_0(v)}{\partial \mathbf{v}},$$
$$\delta f(x = 0, v_x > 0) = \delta f(x = 0, v_x < 0) \tag{5}$$

$$\rho = \rho' = \sum e\int \delta f d\mathbf{p},$$
$$\mathbf{j} = \sum e\int d\mathbf{p}(\mathbf{v} - \mathbf{u})\delta f \tag{6}$$
$$= e\int d\mathbf{p}\mathbf{v}\delta f + \mathbf{u}\rho = \mathbf{j}' + \mathbf{u}\rho',$$

where $\omega' = \omega - \mathbf{ku} = \omega - k_z u$.

Here, we do not write out the Maxwell equations, because they coincide with (2) to within the replacement $\rho \longrightarrow \rho'$ and $\mathbf{j} \longrightarrow \mathbf{j}'$ [3].

2. SOLUTION OF THE PROBLEM AND THE DISPERSION RELATION FOR SURFACE WAVES

The solution of the problem as formulated is completely analogous to that in [3]. Specifically, from the equations for the Fourier-transformed field components,

$$\left\{k^2\delta_{ij} - k_i k_j - \frac{\omega^2}{c^2}\varepsilon_{ij}(\omega_i\mathbf{k})\right\}E_j = \frac{4\pi i\omega}{c^2}j_{0i},$$
$$\frac{4\pi}{c}j_{0i} = 2B_y(0)\delta_{iz}. \tag{7}$$

we determine $E_z(z)$ as a function of $B_y(0)$ in the regions $x < 0$ and $x > 0$. Then, we can derive the desired dispersion relation for surface waves from the condition that the field components E_z and B_y be continuous at the plasma–vacuum boundary $x = 0$. Our approach differs from that used in [3] only with regard to the specific form of $\varepsilon_{ij}(\omega, \mathbf{k})$:

$$\varepsilon_{ij}(\omega, \mathbf{k}) = \delta_{ij} + \frac{\omega'^2}{\omega^2}\left[\left(\delta_{ij} - \frac{k_i k_j}{k^2}\right)(\varepsilon^{\mathrm{tr}}(\omega', k) - 1)\right.$$
$$\left. + \left(\frac{k_i k_j}{k^2} + \frac{u_i k_j + k_i u_j}{\omega} + \frac{u_i u_j k^2}{\omega'^2}(\varepsilon^{\mathrm{l}}(\omega', k) - 1)\right)\right]. \tag{8}$$

This expression is obtained by integrating the kinetic equation (1) with allowance for relationships (6); in this case, the $\varepsilon^{\mathrm{tr},\,\mathrm{l}}(\omega', k)$ are described by the familiar expressions that apply to a Maxwellian plasma and in which ω should be replaced by ω'.

Omitting simple but fairly laborious manipulations similar to those performed in [3], we write out the final dispersion relation for a surface wave in a semibounded plasma flow:

$$\sqrt{\frac{k_z^2 c^2}{\omega^2} - 1} + \frac{2}{\pi}\frac{c}{\omega}\int_0^{+\infty} dk_x\left[-\frac{(\omega'^2)k_x^2}{c^2 k^2 A} + \frac{k_z^2}{k^2}\frac{1}{\varepsilon^{\mathrm{l}}}\right] = 0. \tag{9}$$

Here, we introduced the function

$$A = k^2 - \frac{\omega'^2}{c^2} - \frac{\omega'^2}{c^2}(\varepsilon^{\mathrm{tr}}(\omega', k) - 1), \tag{10}$$

whose zeros correspond to the roots of the dispersion relation for transverse waves in a nonrelativistic ($u \ll c$) plasma flow.

We are going to investigate equation (9) in the following limiting cases, which will be analyzed in more detail in the next section. First, we examine the case of a cold plasma, when the thermal motion of the particles can be ignored and the ions can be assumed to be immobile. If we also neglect collisions (which have not been taken into account up to this point), then we obtain

$$\varepsilon^{\mathrm{l}}(\omega', \mathbf{k}) = \varepsilon^{\mathrm{tr}}(\omega', \mathbf{k}) = \varepsilon(\omega') = 1 - \frac{\omega_{Le}^2}{\omega'^2}. \tag{11}$$

Substituting (11) into (9) and (10), we can readily obtain the following dispersion relation, which can also be derived in the hydrodynamic model of a cold plasma:

$$\sqrt{k_z^2 - \frac{\omega^2}{c^2}\left[1 - \frac{\omega'^2}{\omega^2}(\varepsilon(\omega') - 1)\right]}$$
$$+ \varepsilon(\omega')\sqrt{k_z^2 - \frac{\omega^2}{c^2}} = 0 \tag{12}$$

or, in explicit form,

$$\sqrt{k_z^2 - \frac{\omega^2}{c^2}\left(1 - \frac{\omega_{Le}^2}{\omega^2}\right)} + \left(1 - \frac{\omega_{Le}^2}{(\omega - k_z u)^2}\right)\sqrt{k_z^2 - \frac{\omega^2}{c^2}} = 0. \tag{13}$$

Of course, in the limit $u = 0$, this equation goes over to the familiar equation for a surface wave in a "cold" collisionless semibounded plasma [1, 3]. However, for $u \neq 0$, the surface waves in a flowing plasma exhibit a new feature—they are nonreciprocal; i.e., $\omega(k_z) \neq \omega(-k_z)$.

Another new feature is that the surface wave decays into a fast wave and a slow wave, both satisfying the condition

$$(\omega - k_z u)^2 < \omega_{Le}^2. \tag{14}$$

The decay is especially pronounced in a short-wavelength range, in which $|k_z c| \gg \omega_{Le}$ and the field can be assumed to be potential, $\mathbf{E} = -\nabla\Phi$. In this range, equation (12) reduces to the equation

$$2 - \frac{\omega_{Le}^2}{(\omega - k_z u)^2} = 0, \tag{15}$$

which has the following two solutions describing a fast wave and a slow wave:

$$\omega = k_z u \pm \frac{\omega_{Le}}{\sqrt{2}}. \tag{16}$$

We are primarily interested in the potential approximation, because the low flow velocity $u \ll c$ implies that the phase velocities of the waves under discussion are also low, $\omega/k_z \ll c$. In this limit, equation (9) becomes

$$1 + \frac{2|k_z|}{\pi}\int_0^{+\infty} \frac{dk_x}{k^2 \varepsilon^1(\omega - k_z u, k)} = 0. \tag{17}$$

Before proceeding to a thorough investigation of this equation (see the next section), note that, for the above analysis to be valid for collisional (and the more so, weakly ionized) plasmas, all of the formulas in this section, in particular, equation (17), as well as the general dispersion relation (9) and expression (8) for $\varepsilon_{ij}(\omega, \mathbf{k})$, should be taken with $\varepsilon^1(\omega', \mathbf{k})$ and $\varepsilon^{tr}(\omega', \mathbf{k})$ derived with allowance for particle collisions [3].

3. SPECTRA OF LOW-FREQUENCY SURFACE WAVES IN A PLASMA FLOW

To analyze equation (17), we first consider the ion-acoustic frequency range, assuming that the conditions

$$k v_{Te} \gg |\omega - k_z u| \gg k v_{Ti} \tag{18}$$

are satisfied and neglecting particle collisions. Under these conditions, $\varepsilon^1(\omega', k)$ can be described by [3]

$$\varepsilon^1(\omega', k) = 1 - \frac{\omega_{Li}^2}{\omega'^2} + \frac{\omega_{Le}^2}{k^2 v_{Te}^2}\left(1 + i\sqrt{\frac{\pi}{2}}\frac{\omega'}{k v_{Te}}\right). \tag{19}$$

In the limit $k v_{Te} \gg \nu_e$, which corresponds to a collisionless plasma, the dissipative term in (19) stems from the Cherenkov absorption of the waves by plasma electrons. In the opposite case of a collision-dominated plasma, we must examine the following two limits. The condition $\omega \nu_e \gg k^2 v_{Te}^2$, under which expression (19) is meaningless, refers to a cold plasma described by the dispersion relation (13); in the potential approximation, which is of interest to us, (13) passes over to

$$2 - \frac{\omega_{Le}^2}{(\omega - k_z u)(\omega - k_z u + i\nu_e)} = 0. \tag{20}$$

This dispersion relation, which differs from (15) in that it incorporates collisions, provides no fundamentally new information, except that it reflects a weak absorption of surface waves. Under the condition $\omega' \nu_e \ll k^2 v_{Te}^2$, the dissipation is purely collisional and the dissipative term in parentheses in (19) should be replaced with $i(\omega' \nu_e / k^2 v_{Te}^2)$. This case of a purely collisional dissipation will be examined below, along with expression (19).

We substitute (19) into (17) and use the smallness of the dissipative term to immediately obtain the condition

$$1 - \frac{\omega_{Li}^2}{(\omega - k_z u)^2} < -\frac{\omega_{Li}^2}{k^2 v_s^2}, \tag{21}$$

where $v_s = \sqrt{T_e/M}$ is the ion-acoustic speed in a nonisothermal plasma.

Near the resonance, condition (21) is equivalent to the inequality $(\omega - k_z u) < k v_s$. It is also an easy matter to find the oscillation spectra of slow and fast ion-acoustic surface waves and the rate at which they are damped ($\omega \longrightarrow \omega + i\delta$):

$$\omega = k_z u \pm \begin{cases} |k_z| v_s & \text{for} \quad |k_z| v_s \ll \omega_{Li} \\ \dfrac{\omega_{Li}}{\sqrt{2}} & \text{for} \quad |k_z| v_s \gg \omega_{Li}, \end{cases} \tag{22}$$

$$\delta = \begin{cases} -\sqrt{\dfrac{\pi}{8}\dfrac{m}{M}}|\omega - k_z u| & \text{for} \quad k_z v_s \ll \omega_{Li} \\ -\dfrac{1}{6}\sqrt{\dfrac{m}{M\pi}}\dfrac{\omega_{Li}^2}{|k_z|^3 v_s^3}|\omega - k_z u| & \text{for} \quad k_z v_s \gg \omega_{Li}. \end{cases} \qquad (23)$$

The damping rate (23) can be readily generalized to the case of a collision-dominated plasma under the condition $\omega\nu_e \ll k^2 v_{Te}^2$:

$$\delta = \begin{cases} -\dfrac{m}{M}\nu_e \dfrac{|k_z|^3 v_s^3}{\omega_{Li}^3} & \text{for} \quad |k_z| v_s \ll \omega_{Li} \\ -\dfrac{3}{8}\dfrac{\omega_{Li}^4}{k_z^4 v_s^4}\nu_e & \text{for} \quad |k_z| v_s \gg \omega_{Li}. \end{cases} \qquad (24)$$

The formulas derived for the spectra of ion-acoustic surface waves in a plasma flow are analogous to the relevant formulas for a semibounded plasma at rest. The only difference is that, as can be seen from (23), they incorporate the Doppler frequency shift. However, it is this circumstance that makes it possible to describe the essential features of the propagation of surface waves along the plasma boundary. In the static approximation, equation (17), when applied to the plasma at rest, admits no solutions with the real wave vector k_z, because the static field of a particle is screened at a distance of about the Debye radius from the particle. The situation with a flowing plasma is radically different: under conditions (18), equation (17) taken with (19) and at $\omega = 0$ yields

$$1 - \frac{\omega_{Li}^2}{k_z^2 u^2} + \left(1 + \frac{\omega_{Li}^2}{v_s^2 k_z^2}\frac{1}{1 - \dfrac{\omega_{Li}^2}{k_z^2 u^2}}\right)^{-1/2} = 0. \qquad (25)$$

(For simplicity, the dissipative term is omitted.) Equation (25) admits real solutions in the range $k_z^2 u^2 < \omega_{Li}^2$ (in the plasma region, we have $k_z^2 u^2 \ll \omega_{Li}^2$) only under the condition

$$v_{Ti}^2 < u^2 < v_s^2. \qquad (26)$$

This condition, in which the left-hand inequality gives the basis for the neglect of ion thermal motion, requires that the plasma be nonisothermal, so that $T_e \gg T_i$. The ion-acoustic waves under discussion can only exist is such a plasma. In this case, equation (26) has the solution

$$k_{z0}^2 = \frac{\omega_{Li}^2}{u^2}\left(1 - \frac{u^2}{v_s^2}\right)^{1/2}, \qquad (27)$$

which provides evidence that the quasistatic field changes periodically with a period k_{z0}^{-1} along the plasma surface. As $u \longrightarrow v_s$, the period k_{z0}^{-1} increases with increasing ratio u/v_s. At $u = v_s$, solution (27) becomes physically meaningless.

For a gas discharge sustained by such an alternating electric field, this effect should manifest itself as a longitudinal modulation of the plasma temperature and plasma density.

REFERENCES

1. A. N. Kondratenko, *Surface and Internal Waves in a Bounded Plasma* (Atomizdat, Moscow, 1987).
2. A. N. Kondratenko, *Penetration of a Field into a Plasma* (Atomizdat, Moscow, 1974).
3. A. F. Aleksandrov, L. S. Bogdankevich, and A. A. Rukhadze, *Principles of Plasma Electrodynamics* (Vysshaya Shkola, Moscow, 1978; Springer-Verlag, Berlin, 1984).
4. P. S. Bulkin, A. P. Ershov, G. S. Solntsev, *et al.*, Moscow Univ. Phys. Bull. **47** (1), 50 (1992).
5. A. F. Aleksandrov, A. P. Ershov, I. Imad, *et al.*, Teplofiz. Vys. Temp. **31**, 850 (1993).
6. A. P. Ershov, S. N. Chuvashev, V. M Shibkov, and I. B. Timofeev, in *Proceedings of 9th International Conference on Space Planes and Hypersonic Systems and Technologies, Norfolk, 1999,* paper A5.

Translated by O. E. Khadin

Technical Physics, Vol. 45, No. 4, 2000, pp. 400–405. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 18–23.
Original Russian Text Copyright © 2000 by Nikulin, Kuleshov.

Generation of Homogeneous Plasma in a Low-Pressure Glow Discharge

S. P. Nikulin and S. V. Kuleshov
Institute of Electrophysics, Ural Division, Russian Academy of Sciences,
Komsomol'skaya ul. 34, Yekaterinburg, 620049 Russia

Received December 1, 1998

Abstract—The possibility of producing homogeneous plasma in a low-pressure discharge with the use of a hollow anode or hollow cathode is analyzed. It is shown that, in contrast to the high-pressure discharge, where uniform ionization is needed to produce homogeneous plasma, in the low-pressure discharge, nearly uniform radial distribution of the plasma parameters can be achieved under nonuniform ionization conditions by increasing the ionization probability at the system periphery and reducing it near the system axis. It is shown that the magnetic field can facilitate generation of the homogeneous plasma instead of interfering with it. *© 2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Interest in the study of low-pressure discharges is mainly associated with their application in ion sources. In the last few years, in this area, more and more attention has been given to the problem of producing beams of large cross section with a uniform current density distribution over the beam cross section. To solve this problem, it is necessary to devise a plasma emitter with a uniform or close to uniform distribution of the emission current density.

Generally speaking, homogeneous beams can be produced even with a nonhomogeneous plasma using an emitter electrode of nonuniform transparency that is more transparent in areas where the plasma density is low and less transparent where it is high. However, this method leads to a lowering of the ion extraction efficiency and, in addition, its application is complicated by the possibility of the current density distribution being altered when the working gas, discharge current, or some other factors are changed. Therefore, fabrication of an optical system suitable for a wide range of discharge conditions seems problematic. Usually, electrodes of uniform transparency are used and beam uniformity is achieved by creating uniform plasma density near the emission boundary.

In contrast to narrow-beam sources—in which magnetic fields are often applied to produce a highly nonuniform plasma with the maximum density near the emission exit—in the sources of large-cross-section beams, a discharge in a hollow cathode without a magnetic field is used as a rule [1]. In such a discharge, oscillating high-velocity electrons produce a highly uniform ionization. However, such a uniform ionization would lead to the emergence of a homogeneous plasma of density $n = \sqrt{G/\beta_r}$, where G is the number of ionization events in a unit volume per unit time and β_r is the recombination rate in the high-pressure discharge, where the charged particles annihilate as a result of recombination. As to the low-pressure discharge where the particles escape to the walls, here the situation is essentially different and the uniform ionization does not guarantee that the generated plasma will be uniform. To explore the possibility of producing the homogeneous plasma at low pressures, it is necessary to carry out a combined analysis of the conditions for generating charged particles and their transport. Such an analysis is the objective of this paper. The most widely used cylindrical geometry of the system will be considered in the case where one of the electrodes, either the cathode or the anode, is cylindrical.

1. THE DISCHARGE WITH A HOLLOW CATHODE

Let us assume that the density distribution of plasma electrons is described by the Boltzmann law

$$n_e = n_0 \exp\left(\frac{e\varphi}{kT_e}\right), \tag{1}$$

where the potential φ is assumed to be zero, n_0 is the concentration at the system center, e is the electron charge, k is the Boltzmann constant, and T_e is the electron temperature.

As for the ions, at low pressures, they escape into the walls in a collisionless regime. Neglecting the pressure tensor [2], let us write the continuity equation and the equations of ion motion in the following simplified form:

$$\frac{n_i v_i}{r} + \frac{d(n_i v_i)}{dr} = G(r), \tag{2}$$

$$\frac{n_i v_i^2}{r} + \frac{d(n_i v_i^2)}{dr} = -\frac{e n_i}{M}\frac{d\varphi}{dr}, \tag{3}$$

where n_i, v_i, and M are the concentration, the average radial velocity, and the mass of the ions, respectively, and r is the radial coordinate.

In a discharge with a hollow cathode, apart from slow plasma electrons, there is usually some quantity of so-called high-velocity particles formed at the cathode as a result of γ-processes which acquire high energy as they move through the cathode layer; however, one may neglect the contribution of these particles to the total electronic charge and write the equation of quasi-neutrality in the form

$$n_i = n_e = n. \tag{4}$$

It should be noted that attainment of the uniform plasma density over the range of radial coordinate r values from 0 to R, where R is the cathode radius, is impossible, because the plasma density drastically decreases as the cathode potential drop region is approached. Neglecting the length of the near-cathode layer in comparison with R, let us assume that $dn/dr \longrightarrow -\infty$ if $r \longrightarrow R$. It is clear that this condition is incompatible with the condition $n = \text{const}$ from which it follows that $dn/dr = 0$. Therefore, we shall be seeking a distribution $G(r)$, such that $n = n_0$ within a certain region $r \le r_0$. The ionization at the system periphery G_p (at $r_0 < r < R$) will be assumed to be a constant meeting the following condition:

$$G_p \pi L(R^2 - r_0^2) = \frac{I_d}{e} - \int_0^{r_0} G(r) 2\pi r L dr, \tag{5}$$

where L is the cathode length and I_d is the discharge current practically equal to the ion current at the cathode.

The sought-for solution of the set of equations (1)–(4) satisfying the condition $v_i(0) = 0$ is a trivial one. We obtain a homogeneous plasma of density n_0 in the region $r < r_0$, where there is no ionization and the radial ion velocity is zero. This solution apparently has no physical meaning, because it is not clear how the ion emission in the axial direction can be accounted for if there is no ionization and no radial flow of ions towards the system center, while the elucidation of conditions under which this sort of emission will occur uniformly is the aim of this study.

Keeping within the framework of the one-dimensional problem, let us roughly account for the escape of particles in the axial direction, assuming [3] that the axial escape of the ions is governed by the Bohm velocity $v_B = \sqrt{kT_e/M}$. Then equations (2) and (3) take the following form:

$$\frac{n_i v_i}{r} + \frac{d(n_i v_i)}{dr} = G(r) - \frac{n_i v_B}{L}, \tag{2'}$$

$$\frac{n_i v_i^2}{r} + \frac{d(n_i v_i^2)}{dr} = -\frac{e n_i}{M}\frac{d\varphi}{dr} - \frac{n_i v_i v_B}{L}. \tag{3'}$$

Escape in the axial direction of the electrons may be neglected on the assumption that the gas-discharge chamber ends have the same potential as the cathode. Analysis of the set of equations (1), (2'), (3'), and (4') shows that generation of homogeneous plasma with the density n_0 in the central region is possible at two different ionization levels in this region:

$$G = n_0 v_B / L \tag{6}$$

and

$$G = n_0 v_B / 3L. \tag{7}$$

In the first case, all ions produced in the central region leave it in the axial direction and the radial flow of ions in this region is zero; and, in the second case, the emission from the central region is due both to ionization in this region and to the inflow of ions from the system periphery. To realize the second situation, the concentration and potential profiles at the periphery should be nonmonotonic and have maxima.

In neither of the two cases did we succeed in deriving an analytical solution at the system periphery. The equations were written in a nondimensional form using the relations

$$\tilde{n} = \frac{e n v_B 2\pi R L}{I_d}, \quad g = \frac{e G \pi R^2 L}{I_d}, \qquad \tilde{v} = \frac{v}{v_B}, \quad \eta = \frac{e\varphi}{kT_e}, \quad \tilde{r} = \frac{r}{R} \tag{8}$$

and solved numerically by the ranging parameter method, with the quantity $\tilde{n}_0$ as the ranging parameter. In the calculations, the discharge current was assumed to be known and condition (5) had to be fulfilled; at point $r = R$, fulfillment of the condition $v = v_B$, or $\tilde{v} = 1$ was checked, which is equivalent to $dn/dr = -\infty$ for this system of equations. Let us demonstrate that. Multiplying equation (2') by dr/dn we obtain at point R, where $dr/dn = 0$, the following relationship:

$$n\frac{dv}{dn} = -v. \tag{9}$$

Applying the same procedure to equation (3'), we get

$$v^2 + v_B^2 + 2nv\frac{dv}{dn} = 0. \tag{10}$$

Substitution of (9) into (10) gives

$$v^2 + v_B^2 - 2v^2 = 0; \tag{11}$$

hence, $v = v_B$, which was to be proved.

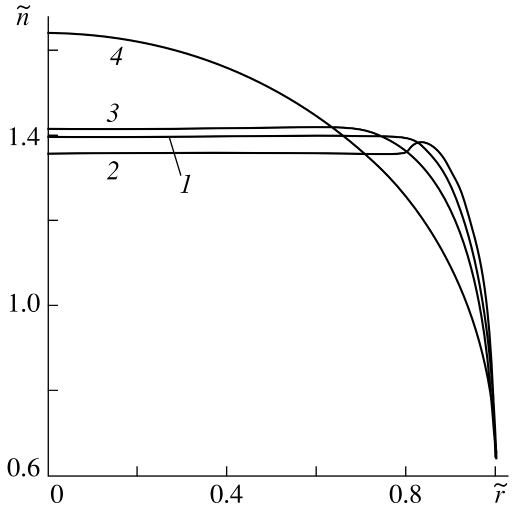


Fig. 1. Radial variation of the plasma density in the hollow cathode under different ionization conditions: (*1*) $G = n_0 v_B/L$ at $r < 0.8R$, (*2*) $G = n_0 v_B/3L$ at $r < 0.8R$, (*3*) $G \sim r^2$, (*4*) $G \sim$ const.

The distributions $\tilde{n}(\tilde{r})$ are shown in Fig. 1 for the first (curve *1*) and the second (curve *2*) cases. In spite of the large difference in the ionization levels in the central region, the curves obtained differ little and both solutions yield ionization values at the system periphery much in excess of those in the central region. If the ionization is increased gradually from the level corresponding to the first case to that of the second one, a set of distributions is obtained, which is only slightly different from the uniform distributions within the central region. Thus, we can obtain nearly uniform distributions with widely varying forms of the function G.

Distributions which are quite uniform within fairly extensive regions are possible not only with functions of the form

$$G(r) = \begin{cases} G_1 = \text{const}, & r < r_0 \\ G_2 = \text{const}, & r > r_0, \end{cases} \tag{12}$$

but also with many other functions, their common property being higher ionization probability at the system periphery and lower in the vicinity of the axis. For example, curve *3* in Fig. 1 represents a density distribution obtained with $G(r) \sim r^2$.

As mentioned above, in the self-sustained discharge with a hollow cathode at low, close to critical pressures, at which burning of the discharge ceases, the oscillating high-velocity particles produce uniform ionization throughout the volume of the cathode hollow and, in this case (G = const for all values of r from 0 to R), the density decreases fairly quickly with r (Fig. 1, curve *4*). The ionization probability at the system periphery can be made higher by increasing the pressure to such values at which the relaxation length of the high-velocity electrons becomes less than the hollow cathode radius. In this case, close to uniform radial distributions can be obtained, as confirmed by experimental results [4]. As the pressure increases further and the ionization processes concentrate more and more in the near-cathode region, it becomes possible to obtain distributions sagging in the center. Some authors [4] link such distributions to the volume recombination, indicating, however, that its probability is negligible. In our opinion, the droop around the system axis and the maximum at the periphery may well be explained by the highly nonuniform ionization and taking account of the ions leaving the cathode hollow in the axial direction. For example, curves *2* and *3* in Fig. 1 have just such a qualitative character. However, it should be noted at this point that at higher pressures, the ion motion regime is no longer collisionless, as assumed in writing equation (3), though, taking account of the collisions will scarcely lead to a qualitative change in the situation.

Thus, the close to uniform radial plasma density distribution in a self-sustained discharge with the hollow cathode can be produced by increasing pressure to a certain optimal level at which the probability of ionization processes is higher (but not excessively) at the system periphery and accordingly lower in the near-axis region. However, such an approach to generation of the uniform plasma can hardly be applied to sources of charged particles, where, as a rule, the discharge pressure is kept just above the critical value so as not to weaken the electrical strength of the accelerating gap. At these pressures, to intensify ionization at the system periphery to a certain optimal level, an external ionization source is needed, that is, the transition to the sustained discharge.

As for the self-sustained discharge, intensification of the ionization at the system periphery at low pressures may be achieved by applying a magnetic field along the system axis to the movement of high-velocity electrons toward the system center. As the magnetic field strength is gradually increased, the spatial distribution of the high-velocity ionizing particles and, consequently, the spatial distribution of the ionization processes will gradually transform from the uniform one, at the magnetic induction $B = 0$, to an increasingly more nonuniform distribution with higher ionization probability near the cathode. Proceeding from the analysis conducted, we expect that with increasing magnetic field at some optimal value of B, a nearly uniform distribution will be attained, at least within a certain region, and a further increase in B will make possible the formation of a distribution with a sag in the center. Results of the experiments carried out with the use of a gas-discharge system consisting of a hollow cathode 50 cm in both length and diameter, a bar anode 35 cm in length and 1 cm in diameter, and a solenoid fully confirm these conclusions (we note that this system may be looked upon as a system with the hollow cathode in an applied magnetic field or as a system of the inverted magnetron type). Figure 2 displays radial variation of the ion current I_p in the probe. The curves were recorded using a system of probes, whose potential was equal to that of the cathode, at the fixed discharge current and various values of the magnetic field induction. It is seen that for a zero magnetic field, the density of the emission current in the radial direction decreases rather quickly (Fig. 2, curve *1*). After application of the magnetic fields, the distribution changes; and at a certain B value, we obtain a distribution closest to the uniform one within a considerable part of the gap (Fig. 2, curve *3*). As the magnetic field is further increased, transformation of the distribution continues and it begins to sag in the system center.

Thus, both the calculation results and the experimental data show that in order to obtain uniform distribution of the plasma density in the low-pressure discharge with a hollow cathode, it is necessary, instead of maintaining uniform ionization as in a high-pressure discharge, to enhance ionization processes at the periphery of the gas-discharge system and, correspondingly, suppress them somewhat in the area near the system axis.

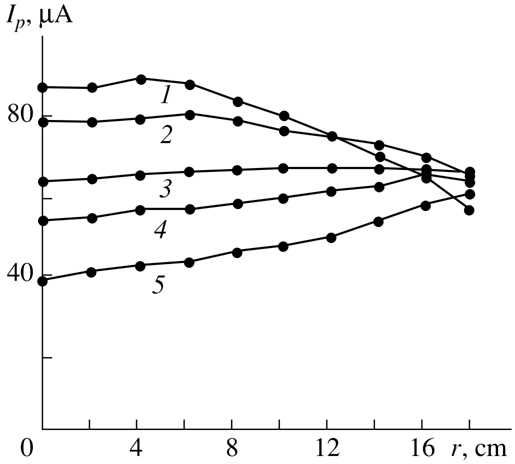


Fig. 2. Radial variation of the ion current in the probe in a discharge with the hollow cathode in applied magnetic field, B: (*1*) 0, (*2*) 0.2, (*3*) 0.3, (*4*) 0.4, and (*5*) 0.6 mT.

2. THE DISCHARGE WITH A HOLLOW ANODE

A disadvantage of gas-discharge systems with a hollow cathode is rather low ion extraction efficiency in the axial direction, which does not exceed, as a rule, a few percent. Most of the ions escape in the radial direction to the cathode. A considerable increase in the extraction efficiency may be expected if the cylindrical electrode is made the anode, in which the ion current is either low or zero. However, a decrease in the radial ion flow will not be brought about just by changing the polarity of the cylindrical electrode. In a low-pressure discharge with large-size anode, the plasma potential, as a rule, is higher than the anode potential and the ions are as free to travel to the anode as to the cathode. To reduce the positive charge and the corresponding radial electric field in a gas-discharge system, a magnetic field is usually applied, which primarily affects electrons as having considerably smaller mass, and so impedes the escape of electrons from the discharge. However, the magnetic field should not exceed a certain critical value B beyond which an electron layer forms near the anode. The electron layer is undesirable inasmuch as it causes a considerable increase in the discharge voltage or intensifies instabilities.

Thus, let us consider a discharge with the hollow anode in an axially applied magnetic field where the end electrodes operate as a cathode, in other words, a discharge in the Penning-type system [3]. If the Larmor radius of electrons is significantly less than the radius of the gas-discharge chamber, then the electron motion at a normal to the magnetic field can be described in the diffusion-drift approximation; however, the Boltzmann distribution (1) in this approximation is invalid. Let us write the continuity equation and the expression for the electron current density in the following form:

$$\frac{n_e v_e}{r} + \frac{d(n_e v_e)}{dr} = G(r), \tag{13}$$

$$n_e v_e = -D_e\frac{dn_e}{dr} + \mu_e n_e\frac{d\varphi}{dr}, \tag{14}$$

where v_e, D_e, and μ_e are the average velocity, the diffusion coefficient, and the electron mobility across the magnetic field, respectively.

The balance and motion equations for the ions will be used in the above form. Analysis of the set of equations (2'), (3'), (4), (13), and (14) shows that, similar to the discharge with the hollow cathode, the homogeneous plasma of density n_0 within a certain region $r < r_0$ is obtained in two cases when the ionization function in this region is defined by the following relation:

$$G = \frac{n_0 v_B}{L}\frac{2-\alpha \pm \sqrt{(2-\alpha)^2-3}}{3}, \tag{15}$$

where

$$\alpha = \frac{eL}{M\mu_e v_B}. \tag{16}$$

As the electron mobility in a strong magnetic field is inversely proportional to the square of the magnetic induction, then $\alpha \sim B^2$. With a decrease in the magnetic field, $\alpha \longrightarrow 0$ and (15) gives two solutions identical to those obtained for the discharge with the hollow cathode [equations (6) and (7)]. As α increases, that is, as the magnetic field strength rises, the two solutions described by expression (15) gradually become closer and, at $\alpha = 2 - \sqrt{3}$, they merge into a unique solution.

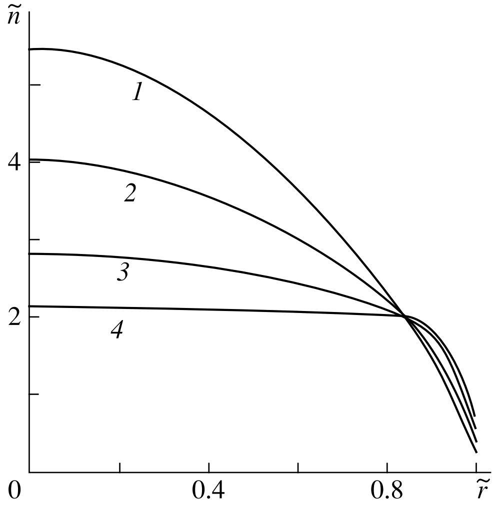


Fig. 3. Radial variation of the plasma density in the hollow anode under various ionization conditions: α = 20; $r_0 = 0.8R$; G_2/G_1: (*1*) 1, (*2*) 3, (*3*) 10, and (*4*) 100.

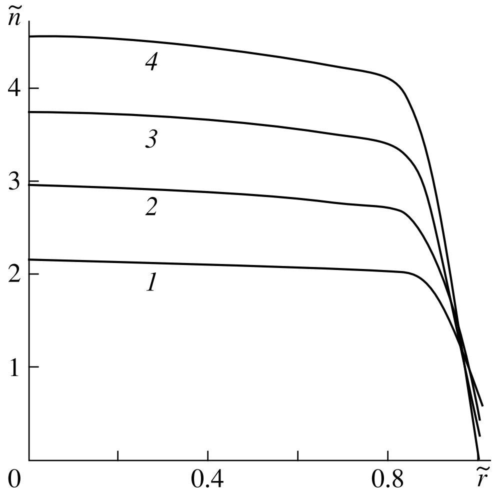


Fig. 4. Same as in Fig. 3 for different magnetic field strengths. $G_2/G_1 = 100$; $r_0 = 0.8R$; α: (*1*) 20, (*2*) 40, (*3*) 60, and (*4*) 80.

At α values in excess of $2 - \sqrt{3}$, highly constant plasma density is impossible at any G. Nevertheless, the calculation results show that even at α values considerably exceeding $2 - \sqrt{3}$, it is quite possible to obtain distributions only slightly differing from uniform ones. The way to obtain such distributions remains the same as before: it consists in decreasing the ionization intensity near the system center and increasing it at the system periphery, as illustrated in Fig.3. In calculating the distributions shown in Fig. 3, it was assumed that the ionization function is described by expression (12) and the ratio between G_1 and G_2 was varied. It is seen that as the ratio G_2/G_1 increases, the density distribution transforms from rather steeply sloping at $G_2/G_1 = 1$, that is, when $G \sim$ const. at all r, to fairly uniform at $G_2/G_1 = 100$. Unfortunately, this transformation is accompanied by a decrease in the plasma density; however, this drawback can be easily eliminated by further increasing the magnetic field strength (Fig. 4). In the case where the ionization is mainly concentrated at the system periphery, the magnetic field may be increased to rather large values without causing formation of the near-anode electron layer, because the lifetime of the electrons in the discharge is short. At the same time, the potential distribution in the plasma is nonmonotonic, and most of the ions produced at the periphery, instead of traveling to the anode, move toward the system center, increasing the plasma density in the central region, and then leave the discharge in the axial direction.

Thus, to produce a uniform plasma density distribution in the discharge with the hollow anode under an applied magnetic field, the same as in the discharge with the hollow cathode, it is necessary to reduce the ionization intensity near the system center and to increase it at the periphery. However, the radial distribution of the concentration of high-velocity ionizing electrons in the original Penning cell is quite the opposite. Indeed, these high-velocity particles oscillating between the opposite cathodes and gradually diffusing transversely the magnetic field will tend to form a typical diffusion profile, which has a maximum at the system center and decays to zero approaching the anode absorbing these particles. Corresponding qualitative features are found in the distributions of the plasma density and the emission current. Figure 5 (curve *1*) shows experimental values of the ion current in the probe, for probes in positions specified by probe number n_p. The results were obtained using a Penning cell of 15 cm both in length and diameter.

To produce a uniform plasma, it is necessary to intensify ionization at the system periphery, which was achieved in the experiment using an auxiliary discharge with the hollow ring cathode emitting electrons into the peripheral section of the main Penning discharge, where they produced additional ionization. This modification of the gas-discharge system allowed the radial distribution of the ion emission current to be made nearly uniform over a certain region (Fig. 5, curve *2*). Naturally, such a modification is not the only possible one. Another possibility is to use a nonequipotential cathode, which can be made, for example, by isolating the central section of at least one of the cathodes and applying to it a negative potential of lower absolute value than to other parts of the cathode. In this case, the high-velocity electrons emitted by that section of the cathode will acquire less energy in the cathode layer, the ionization intensity in the central region will decrease, and the plasma profile will become more uniform. Certainly, other solutions are possible.

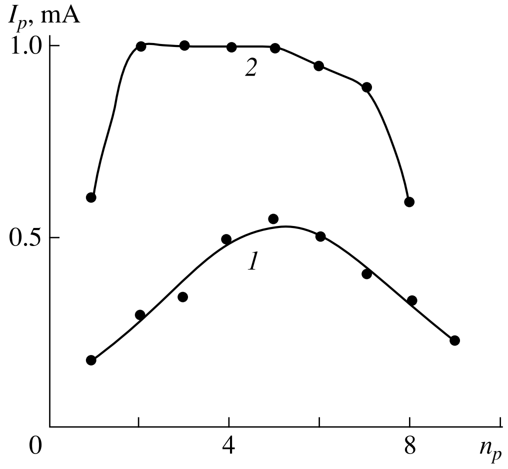


Fig. 5. Radial variation of the ion current in the probe in (*1*) Penning discharge and (*2*) Penning discharge with a hollow ring cathode.

It should be noted that use of these two modifications in the discharge with the hollow cathode and zero magnetic field instead of the Penning cell will hardly improve the situation, because the electrons, even though they have reached the hollow cathode periphery, after a single effective collision become isotropic and uniformly distributed throughout the cathode volume. In a similar way, in the system with a nonequipotential cathode, the electrons that acquired different energies will soon mix and no nonuniform ionization will result. Thus, contrary to widespread opinion, the magnetic field in the cases considered, as well as in the case of the discharge with hollow cathode analyzed in section 1, is a factor contributing to production of homogeneous plasma.

CONCLUSIONS

For producing fairly uniform radial density distributions of the plasma and the ion emission current in a low-pressure discharge with the hollow cathode or hollow anode, it is necessary to produce nonuniform ionization in the discharge gap with higher probability at the system periphery. In the discharge with the hollow cathode, it can be done by the application of a magnetic field to the discharge gap to impede the flow of high-velocity ionizing particles toward the system center. In the Penning system, modifications should be made with a view to reducing ionization at the center and/or increasing it at the periphery of the system. In particular, it has been experimentally shown that the homogeneous plasma can be produced in some region by supplying additional ionizing electrons to the periphery of the Penning system from an auxiliary discharge with a hollow ring cathode.

REFERENCES

1. S. P. Bugaev, Yu. E. Kreindel', and P. M. Shchanin, *Electron Beams of the Large Cross Section* (Énergoatomizdat, Moscow, 1984).
2. L. Bromberg and L. D. J. Smullin, J. Appl. Phys. **50**, 2634 (1979).
3. M. D. Gabovich, *Physics and Technology of Plasma Sources of Ions* (Atomizdat, Moscow, 1972).
4. V. I. Kirichenko, V. M. Tkachenko, and V. B. Tyutyunnik, Zh. Tekh. Fiz. **46**, 1857 (1976) [Sov. Phys.–Tech. Phys. **21**, 1080 (1976)].

Translated by N.P. Mende

Technical Physics, Vol. 45, No. 4, 2000, pp. 406–410. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 24–28.

On Distribution over the Cathode Surface of the Current Due to Positive Ions in Discharge with Oscillating Electrons

R. P. Babertsyan, É. S. Badalyan, G. A. Égiazaryan,
É. I. Ter-Gevorkyan, and V. N. Ogannisyan
Yerevan State University, ul. A. Manukyana 1, Yerevan, 375049 Armenia
Received March 17, 1999

Abstract—The mechanisms of processes determining distribution of positive ions on the cathode surface in a discharge with oscillating electrons are studied. At low pressures $P \leq 5 \times 10^{-5}$ torr over the entire range of anode length variation l_a = 0.5–11 cm, the ion current distribution over the cathode radius $J_i(r)$ features a single maximum in the center and drops steeply with distance from the axis. At pressures $P = 1–4 \times 10^{-4}$ torr, the distribution $J_i(r)$ for short anodes (l_a < 6 cm) is similar to the previous one but, for long anodes (l_a = 6–10 cm), new maxima at higher J_i values have been detected. A physical explanation for the obtained results is proposed. *© 2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Distribution of the charged particles over the cathode surface is influenced by the structure of the space charge in the tube and location of the intensive ionization areas. The importance of the study of this subject is evident from the diversity of applications of this type of discharge: vacuum pumps, sources of charged particles, growth of thin films, etc.

It has been shown [1–5] that the maximum ion current density at the cathode occurs in its center; at $P = 2 \times 10^{-5}$ torr the ion current density drops by half just 0.5 mm from the discharge axis. In [6, 7] the ion and electron currents flowing through the central area of the cathode have been separated. For the first discharge regime, a formula was derived [8] describing the distribution of the ion current density along the cathode radius. In a study by the present authors [9], distribution of the ion current density over the cathode surface as a function of anode length was measured for d_a = 20 mm. This study formed a natural sequel to earlier studies for d_a = 31 mm [9, 10].

EXPERIMENTAL

The measurements were carried out in the setup with an electrode system shown in Fig. 1a. The cylindrical anode *1* has inside it a fixed cathode *2*, in which there is one central hole and eight radial holes 2 mm in diameter; flat molybdenum disks *4* of identical diameter (2 mm) are mounted behind each hole for picking up currents emerging out of individual holes. Adjustable cathode *3* serves to smoothly vary the discharge gap l_a from 0.5 to 11 cm without breaking vacuum.

Figures 1c–1e show plots of the ion current density $J_i(r)$ as a function of the distance from the cathode center at a residual gas pressure of $P = 5 \times 10^{-5}$ torr. The magnetic induction was varied in a range B = 300–600 G; the anode voltage was U_a = 2–3 kV. The anode cylinder length l_a was a parameter. Simultaneously, variation of the discharge current $I_p(l_a)$ was recorded (Fig. 1b) and its variable component monitored with the use of a spectrum analyzer and oscilloscope. It is seen that in the entire range of anode lengths available, $I_p(l_a)$ has a single maximum. Corresponding densities $J_i(r)$ are also the maximum values. The shape of $J_i(r)$ curves is always the same: J_i has a maximum in the central part of the cathode and falls down steeply away from the center. As the anode voltage U_a is raised, the currents I_p and J_i increase monotonically. At B = 300 G, raising U_a from 2 to 3 kV makes $J_i(0)$ nearly twice as high (Figs. 1c, 1d). Conversely, with an increase in the magnetic induction from 300 to 600 G, J_i drops by half (Figs. 1d, 1e). High frequency oscillations have not been observed.

Analysis of the above electrical characteristics of the discharge shows that the obtained results correspond to a uniform low-density space charge of electrons. The effect of ions on the space charge is ignored, because under conditions of low pressure and a strong longitudinal electric field (the first regime), the ions experience no collisions and quickly escape to the cathodes. In this regime, the distribution of the electrical potential along radius r and axis z of the tube can be with good accuracy approximated with a quadratic function. The potential U_0 at the center of the discharge chamber is high enough, and ionization is produced by electrons moving along both r and z [10]. The ampli-

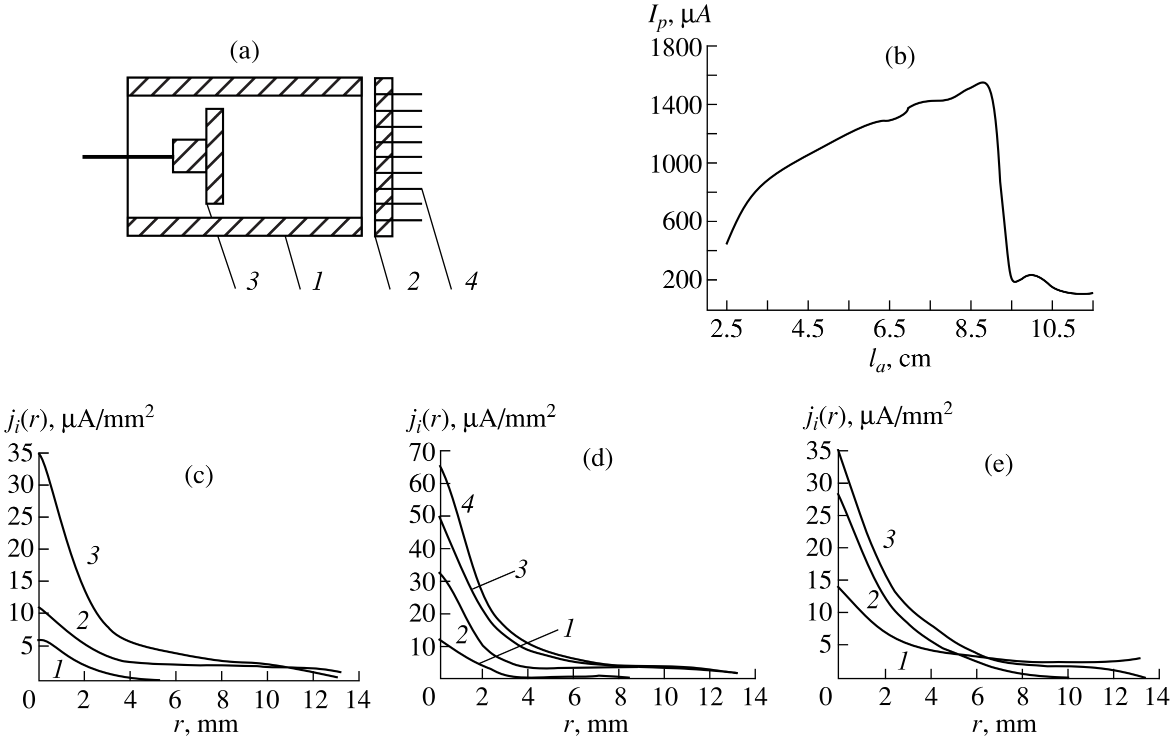


Fig. 1. (a) Electronic system of the cell; (b) dependence of the discharge current on anode length $I_p(l_a)$; (c–e) variation of the ion current with distance from cathode center. (b) $U_a = 3$ kV, $B = 300$ G; (c) $U_a = 2$ kV, $B = 300$ G, $l_a = 12$ (*1*), 3 (*2*), 8 cm (*3*); (d) $U_a = 3$ kV, $B = 300$ G, $l_a = 12$ (*1*), 3 (*2*), 6 (3), 8 cm (*4*); (e) $U_a = 3$ kV, $B = 600$ G, $l_a = 4$ (*1*), 10 (*2*), 6 cm (*3*). Pressure $P = 5 \times 10^{-5}$ torr.

tude of radial oscillations of an ion does not exceed the radius of the point r_0 where the ion was born, and obviously the ions can get to those points on the cathode surface whose separation from the center r does not exceed r_0 $(r \le r_0)$. The central part of the cathode can be hit by ions generated at any distance from the axis, but the actual point of incidence is determined not only by the coordinate of the point of generation but also by the ion trajectory. Calculations of the ion trajectories [8] show that the ions hit the cathodes at various angles of incidence with different energies and in different quantities. In the first discharge regime, the ion current density at the central part of the cathode is given by expression

$$J_i(r \le \Delta r_k) = \frac{\nu_i n_e e}{\pi \Delta r_k^2}\Omega'_0, \qquad (1)$$

where ν_i is the ionization rate; n_e is the electron concentration; e the electron charge; Δr_k is the radius of a circle at the center of cathode within which axial ions are scattered because of the scatter in their initial velocities; and Ω'_0 is the total volume of all regions from where ions reach the cathode center.

Examination of the size and structure of Ω'_0 indicates that the majority of ions hitting the central part of the cathode comprise ions generated inside a cylinder of radius Δr_k and length $l_a/2$ located along the cell axis z, as well as the ions coming from volumes (r, z) located closer to the cathodes (the nearer the volume to the cathode, the greater its size). Quantities ν_i and n_e under these conditions grow faster with increasing P and B (the dependence on U_a is weaker); therefore, it is possible [1] that the ion current density J_i also rises with pressure and magnetic induction. The maximum of the axial ion current is intimately related to the intensive ionization taking place close to the axis, wherefrom all ions arrive directly at the central part of the cathode. Some contribution also comes from an increase in the ion concentration n_i due to the greater “lifetime” of the axial ions because of radial oscillations.

With increasing distance to the cathode, the width of the region from which ions can get to their central part diminishes. Therefore, peripheral parts of the cathode receive some fraction of the ions produced in a small intensive ionization region near the anode. Here, the ion withdrawal rate is greater and the concentration n_i lower. The corresponding ion current density is given by formula [8]:

$$J_i(r > \Delta r_k) = \frac{4e\nu_i n_e l_a}{\pi^2}(r_a^2/r^2 - 1)^{1/2}\Omega'_0. \qquad (2)$$

Curves in Fig. 1 are in agreement with equation (2). Indeed, simultaneous reduction of n_i along the tube

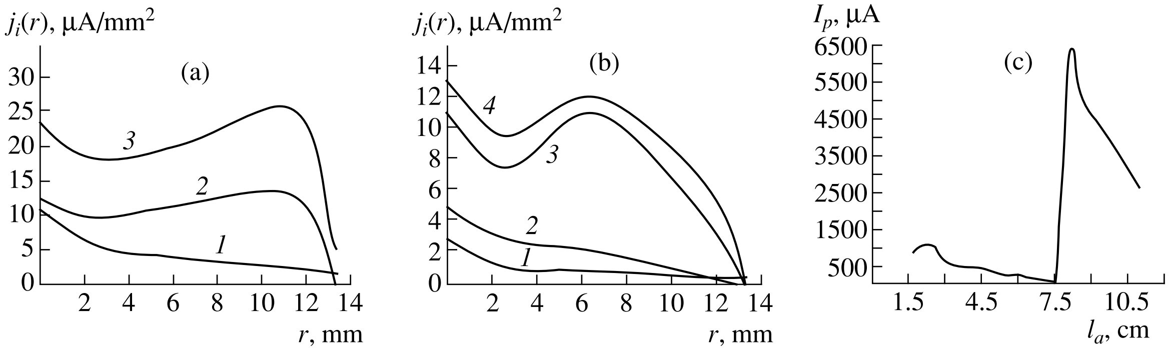


Fig. 2. Same as in Fig. 1, at $U_a = 1$ kV, $P = 2 \times 10^{-4}$ torr. (a) $B = 300$ G, $l_a = 4$ (*1*), 10 (*2*), 8 cm (3); (b) $B = 1200$ G, $l_a = 3$ (*1*), 5 (*2*), 10.5 (*3*), 9.5 cm (*4*); (c) $B = 300$ G.

radius and of the size of corresponding regions, whence ions come to point r at the cathode surface results in appreciable lowering of the ion current density at the periphery. So, for a Penning tube 31 mm in diameter at a pressure $P = 5 \times 10^{-5}$ torr with anode voltage and magnetic induction corresponding to the first discharge regime, changing the anode cylinder length l_a will affect both the discharge current value and the ion current density. Yet the shape of $J_i(r)$ remains the same: in the entire range of anode lengths $l_a = 0.5$–11 cm, the current density has a maximum at the cathode center. In other regimes, the distribution of $J_i(r)$ is drastically changed.

The above mechanism of ion production and distribution suggests that once the localization of separate ion beams at the cathode surface is established, one can tentatively outline the regions where the ions have been produced, that is, reconstruct intensive ionization regions inside the anode. Our experiments show that this correspondence is more pronounced under conditions of higher pressure and longer anodes. In Figs. 2a and 2b, distributions of the current density $J_i(r)$ along the cathode radius are given for $P = 2 \times 10^{-4}$ torr. It is seen that the shape of the $J_i(r)$ curve is quite close to that for short anodes at $P = 5 \times 10^{-5}$ torr ($l_a = 4$ cm, curve *1* in Fig. 2a). But for tubes with long anodes the $J_i(r)$ distribution is significantly changed with respect to both the shape and the absolute values (curves *2*, *3*). The density $J_i(r)$ is much higher over nearly the entire cathode surface and peaks not only in the central part but also against the third disk. Some reduction of $J_i(0)$ is also noticeable, and practically no ions arrive at the extreme periphery of the cathode: these areas remain unevaporated.

Plots of the discharge current versus the tube length (Fig. 2c) demonstrate that curve *1* in Fig. 2a corresponds to discharge conditions where l_a values are in the range of short anodes ($l_a \le 4$ cm) prior to formation of the second maximum of the function $I_p(l_a)$. That is, with short anodes, the shape of the $J_i(r)$ curve is the same at both $P = 5 \times 10^{-5}$ torr and $P = 2 \times 10^{-4}$ torr. Conversely, curves *2* and *3* in Fig. 2a with a different $J_i(r)$ distribution correspond to conditions favorable for the appearance of a second maximum in $I_p(l_a)$, i.e., to the long anode conditions.

As the gas pressure and the rate of electron collisions with neutrals increase, the ionization rate ν_1 will increase as well. Because of the increasing rate of withdrawal of electrons to the anode, the ratio n_e/n_i is decreasing. Thereby, the influence exerted by ions on the space charge is enhanced and the potential distribution $U(r, z)$ in the tube changes correspondingly. In particular, the radial drop of potential $U_a - U_0$ decreases and the longitudinal drop of potential $U_0 - U_k$ increases. The latter provides an increase in the number of secondary electrons owing to the higher energy of the ions bombarding the cathode surface. Also rising is the energy acquired by the close-to-axis electrons as they oscillate between the cathodes. On the other hand, a longer anode also causes an enhancement of $U(z)$, and makes the role of regions close to the cathode [10] greater. The axial and radial ionizations, by enhancing each other, proceed with much higher intensity. As a result, in the case of high P, for a long anode, a steep rise of the discharge current and the current density occurs at the cathode as the second maximum of $J_i(r)$ emerges. The radial ionization is nonuniform along the tube radius, and so a radial zone of intensive ionization exists in the tube. The longitudinal ionization, on the contrary, is less sensitive to the distance to the tube axis; therefore large quantities of ions impinge on the peripheral areas of the cathode as well (Fig. 2). As noted above, a certain correlation exists between the ion coordinate at the moment of birth and the point it will hit on the cathode; therefore, it can be argued that it is the ions produced in the intensive ionization zone that arrive at corresponding points on the cathode surface and form the second maximum of $J_i(r)$. The formation mechanism of the central ion current is similar to that presented above for conditions of Fig. 1. Curves obtained for different values of magnetic induction (Fig. 2) demonstrate that the above relationships persist: there is observed only one maximum of $J_i(r)$ for

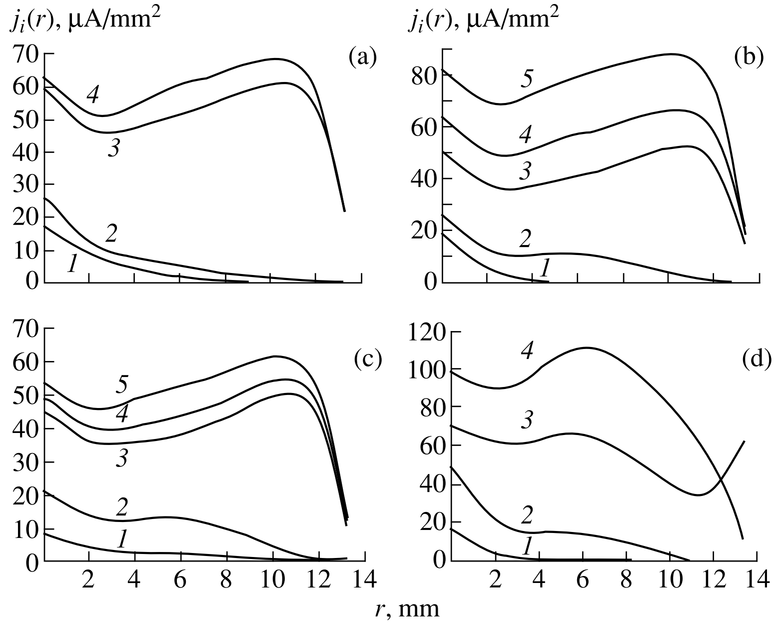


Fig. 3. Same as in Fig. 1, for higher anode voltages. (a) U_a = 1.5 kV, B = 450 G, l_a = 6.5 (*1*), 4 (*2*), 9.5 (*3*), 8 cm (*4*); (b) U_a = 1.5 kV, B = 750 G, l_a = 2 (*1*), 4 (*2*), 10 (*3*), 8 (*4*), 7 cm (*5*); (c) U_a = 1.5 kV, B = 1050 G, l_a = 2 (*1*), 5 (*2*), 10 (*3*), 9 (*4*), 8 cm (*5*); (d) U_a = 2 kV, B = 600 G, l_a = 1 (*1*), 4 (*2*), 10 (*3*), 7 cm (*4*).

short tubes and two maxima for long ones. The critical anode length at which the transition between the two shapes of $J_i(r)$ takes place depends on the magnetic induction B. Under conditions of containment of a large quantity of electrons (large B), the axial potential drops and all currents decrease considerably (Fig. 2b). Therefore, to restore sufficient longitudinal ionization it is necessary to increase l_a (with longer anodes the role of longitudinal ionization is greater). However, all this becomes manifest only under certain conditions of long anodes and higher pressures, when, due to significant enhancement of ionization along the tube, the ionization intensity and ion concentration in the radial zones and on the cathode surface rise considerably, and the ions from these zones produce on the cathode surface a corresponding ion current density profile.

With increasing anode voltage (U_a = 1.5; 2 kV), the character of the $J_i(r)$ distribution as a whole remains the same, while the absolute values of the currents I_a and J_i become larger (Fig. 3). With increasing magnetic induction, the central maximum of the ion current density $J_i(0)$ for long anodes (l_a > 7 cm) gradually lowers, equaling the second maximum and ultimately ending up much lower; i.e., the ion current density at the cathode periphery becomes higher than in the center (curves *3* and *4* in Fig. 3a). The height of the second peak of $J_i(r)$ is less affected by the magnitude of magnetic induction, in agreement with its formation mechanism described above. In the case where l_a = 10 cm and B = 600 G, a third peak appears in the $J_i(r)$ distribution (Fig. 3a); i.e., three belts of intensive ionization are found inside the anode. The presence of the radial intensive ionization zones is evidenced by circular evaporation patterns on the cathode. The width of these rings and their radii are determined by the relationships between U_a, B, and l_a. By adjusting the physical and geometrical parameters, a fairly uniform distribution of the ion current density over the cathode surface can be achieved (Figs. 3b, 3c); this may be used advantageously in producing uniform films of different materials. The highest ion current density, 110 μA/mm^2, is achieved in the second disk at U_a = 2 kV, B = 600 G, and l_a = 7 cm (Fig. 3d).

Thus, with a long anode ($6 \le l_a \le 10$ cm) at $P = 2 \times 10^{-4}$ torr, the distribution of the electrical potential is such that the ratio of radial and longitudinal ionizations is at an optimum and at all points maximum values are reached. In this case, the ion current density distribution over the cathode surface $J_i(r)$ is essentially different, featuring subordinate peaks. For still longer anodes (l_a > 10 cm), the optimum ratio of the ionizations is violated and the currents drop significantly.

CONCLUSIONS

In the low-pressure Penning discharge, $P \le 5 \times 10^{-5}$ torr, in the entire range of anode lengths l_a = 0.5–11 cm and for anode voltages and magnetic induction values used in the experiment, the distribution of ion current density $J_i(r)$ along the cathode radius is bell-shaped, peaking in the center and steeply sloping off in the radial direction away from the axis.

At a pressure $P = 1–4 \times 10^{-4}$ torr, the distribution of ion current density along the cathode radius is similar to the previous one in the case of short anodes ($l_a <$ 6 cm) and changes abruptly for long anodes ($l_a = 6$–10 cm): new intensive current peaks appear.

The abrupt change of the discharge characteristics is explained as being caused by attainment of the optimum ratio of radial and longitudinal ionizations. This optimum ratio is attainable in long anodes of particular length at specific values of U_a and B and only at high pressures.

The emergence of subsidiary peaks in the $J_i(r)$ distribution is related to the formation inside the tube of radial belts of considerably higher ionization. At pressures $P < 10^{-4}$ torr, the subsidiary peaks are weak or undetectable.

By adjusting l_a, U_a, B, and P values, either a narrow collimated ion beam or a uniform distribution of the ion current density over the entire cathode surface can be produced.

REFERENCES

1. Yu. E. Kreĭndel, Zh. Tekh. Fiz. **33**, 883 (1963).
2. Yu. E. Kreĭndel and A. S. Ionov, Zh. Tekh. Fiz. **34**, 1199 (1964).
3. Yu. E. Kreĭndel, Zh. Tekh. Fiz. **35**, 315 (1965).
4. O. K. Kurbatov, Zh. Tekh. Fiz. **36**, 1665 (1966).
5. E. M. Rudnitskiĭ, Zh. Tekh. Fiz. **37**, 927 (1967).
6. E. T. Kucherenko and V. A. Saenko, Zh. Tekh. Fiz. **37**, 112 (1967).
7. É. M. Reĭkhrudel’, G. V. Smirnitskaya, and A. N. Mavlyanov, Vestn. Mosk. Univ., Ser. 3: Fiz., Astron., No. 7, 49 (1969).
8. G. V. Smirnitskaya and Nguen Khyu Ti, Zh. Tekh. Fiz. **39**, 1625 (1969).
9. R. P. Babertsyan, G. A. Égiazaryan, V. Kh. Garibyan, and A. K. Chobanyan, Izv. Nats. Acad. Nauk Armenii, Fiz. **27**,103 (1992).
10. R. P. Babertsyan, É. S. Badalyan, G. A. Égiazaryan, and É. I. Ter-Gevorkyan, Zh. Tekh. Fiz. **68**, 29 (1998) [Tech. Phys. **43**, 1035 (1998)].

Translated by B.N. Kalinin

Technical Physics, Vol. 45, No. 4, 2000, pp. 411–413. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 29–32.
Original Russian Text Copyright © 2000 by Egorov, Belyaev, Lobachev.

Pressure Shaping of Titanium Nickelide upon Heating under Shear Stress

S. A. Egorov, S. P. Belyaev, and I. N. Lobachev

Research Institute of Mechanics and Mathematics, St. Petersburg State University, St. Petersburg, 198904 Russia

Received August 19, 1998; in final form, October 1, 1999

Abstract—A unidirectional variation of the pressure applied to Ti–Ni alloy was found to cause strains of different sign when the material temperature was changed in 10 K steps. The pressure dependence of twin yield stress was obtained. It is concluded that pressure may initiate twinning in crystals. *© 2000 MAIK “Nauka/Interperiodica”.*

As is known, pressure applied to alloys with martensite inelasticity may cause a number of unusual deformation processes at temperatures within the range of direct and reverse martensite transformations [1–4]. For example, isothermal variation of pressure can produce the shape memory effect, the reversible shape memory effect, plasticity of transformation, etc. The mechanism of pressure shaping is a shift of the temperature of thermodynamic interphase equilibrium in accordance with the Clausius–Clapeyron equation.

Until now, the influence of pressure on martensite inelasticity has been investigated only for processes that exhibit a unidirectional change of strain with monotonically varying temperature. In memory effect media, however, reversible strains are also feasible. Deformation of these alloys upon heating under constant shear stress is an example. In this case, at the first stage of heating, the strain builds up with applied load. With a further increase in temperature, deformation recovery occurs. Generally, such behavior is adequately explained by martensite transformations [5]. In titanium nickelide (TiNi), however, the twin yield stress decreases and the strain builds up by V19' martensite twinning as the temperature approaches the reverse phase transition point [6]. Thus, it is of interest to trace the pressure effect on the behavior of TiNi alloys upon heating under constant shear load at various stages of deformation. This is the purpose of our study.

The Ti–49.5 at. % Ni alloy used in experiments had the following characteristic temperatures of martensite transformation: M_I = 305 K, M_F = 345 K, A_I = 340 K, and A_F = 385 K. These are the initial and the final temperatures of transformation plasticity upon cooling and the shape memory effect upon heating, respectively, under the constant stress τ = 40 MPa and atmospheric pressure.

Experiments were performed in the setup described in [3]. Samples in the form of a coil spring were prepared from 0.8-mm-dia. wire. The number of turns was six, and their diameter was 5.6 mm. Samples were annealed for 40 min at 720 K before and after experiments. The strain was calculated from the spring shrinkage in an elastic approximation by the expression $\gamma = \Delta hd/(4\pi R^2 n)$, where d is the wire diameter, Δh is a change in the spring height, n is the number of turns, and $2R$ is the spring diameter. The change Δh was measured with an accuracy of ±10 μm, so that the strain γ was accurate to 0.002%.

A sample was placed in the working chamber and subjected to a shear stress of 40 MPa at a temperature of 292 K. Under these conditions, the material was strained 0.205%. Then, the sample was taken for experiments.

First, we studied pressure shaping under given conditions. To do this, after the sample had been placed in the chamber, the pressure P was increased to a certain level and remained constant during subsequent heating. Experimental results obtained for P_1 = 0.1 MPa and P_2 = 200 MPa are presented in Fig. 1. It is seen that, for either pressure, the temperature dependence of the

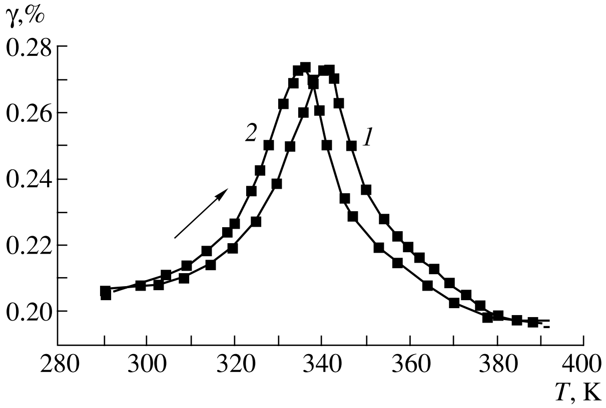


Fig. 1. Strain vs. temperature during heating under a constant shear stress of 40 MPa for applied pressures of (*1*) 0.1 and (*2*) 200 MPa.

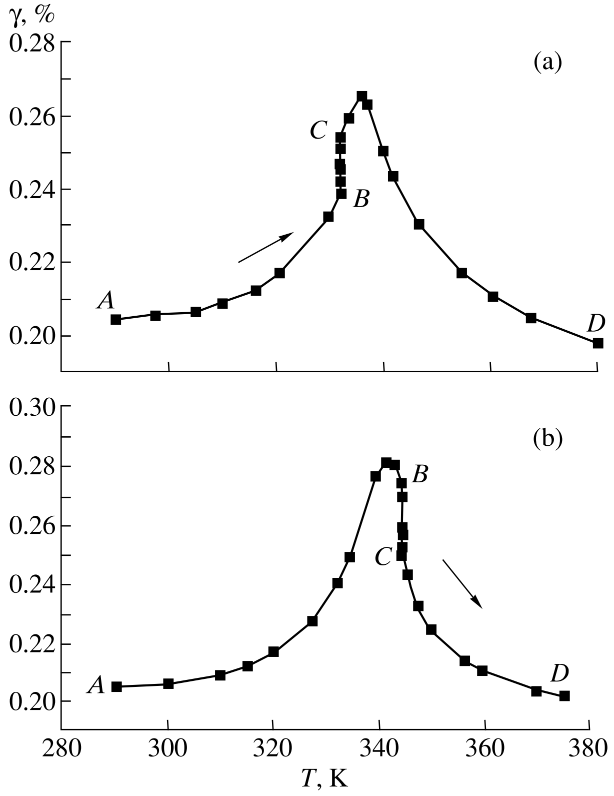


Fig. 2. Strain vs. temperature during heating and isothermal increase in pressure at (a) 332 and (b) 344 K. *AB*, atmospheric pressure; *BC*, pressure rises from 0.1 to 200 MPa; and *CD*, pressure equals 200 MPa.

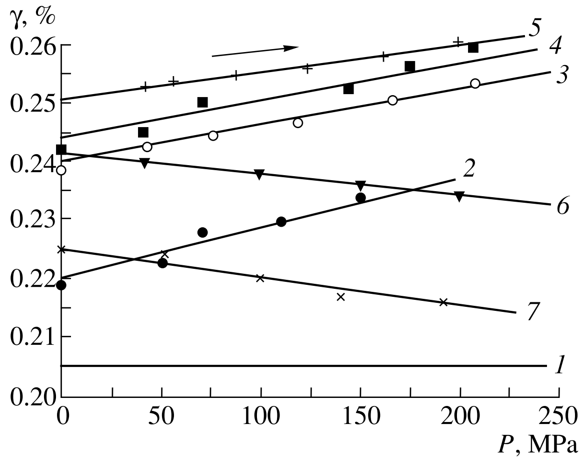


Fig. 3. Strain vs. pressure for temperatures of (*1*) 292, (*2*) 322.5, (*3*) 328, (*4*) 332, (*5*) 333, (*6*) 345, and (*7*) 356 K.

strain γ has a specific bell-like shape described in [5, 6]. An important point is that the curve for $P = 0.1$ MPa displays a maximum at $T = 340$ K, which is the initial temperature of reverse phase transition A_H. In other words, at temperatures below 340 K, the alloy was single-phase, the strain built up through twinning [6], and deformation recovery was due to martensite transformation [5].

Now turn to Fig. 1 again. Under a pressure of 200 MPa, the plot is shifted to lower temperatures by about 5 K. This means that the pressure applied to the alloy does more than lower the temperature of martensite transformation, which is well known. The range where the twin yield stress strongly depends on temperature also shifts to lower temperatures. In other words, in the temperature range from 292 to 340 K, the twin yield stress drops with an increase in pressure. Therefore, one can assume that an isothermal pressure rise during strain buildup should decrease the twin yield stress and hence contribute to the strain. At the stage of deformation recovery, the pressure rise causes the transformation plasticity effect upon heating, which encourages recovery.

In order to verify the aforesaid, another series of experiments was carried out. The sample was placed in a high-pressure chamber and heated at atmospheric pressure. At the stage of inelastic deformation, the heater was switched off and the pressure was isothermally raised to 200 MPa. As the pressure grew, the strain kept its sign. After the pressure was raised to 200 MPa, the heater was switched on again. Associated experimental results are shown in Fig. 2. Here, *AB*, *BC*, and *CD* denote successive experimental stages. One can see that, in the temperature ranges of 322.5–338 K and 346.5–356 K, isothermal growth of pressure (portion *BC*) strains the material by 0.014–0.02% both for the rising (Fig. 2a) and the falling (Fig. 2b) branches of the γ–T curve. It is significant that an increase in the pressure contributes, as expected, to the strain in the rising branch of the curve γ–T and favors deformation recovery in its falling branch. Moreover, pressure-induced strain variations may change the sign even at close temperatures. In Figs. 2a and 2b, the temperatures at which the pressure was increased differ by 12 K only.

As one could expect from Fig. 1, the sample was not strained when the pressure was isothermally increased at 292 K.

The pressure dependences of the strain are shown in Fig. 3. The curves are nearly linear whatever the temperature of isothermal pressure rise. At both stages of deformation (buildup and recovery), the magnitude of the ratio $d\gamma/dP$ roughly equals 8×10^{-5} MPa^{-1} in the temperature range of 322.5–356 K.

The obtained results suggest that, during reversible shaping in the TiNi alloy, an increase in the pressure initiates mass transfer mechanisms responsible for the rising and falling branches of the γ–T curve. One more series of experiments was aimed at elucidating the influence of isothermal pressure decrease on the alloy.

The sample was placed in the working chamber and heated under a pressure of 200 MPa. In the course of inelastic deformation, the heater was switched off and the pressure was isothermally decreased to the atmospheric value; then, heating recommenced. Some of the

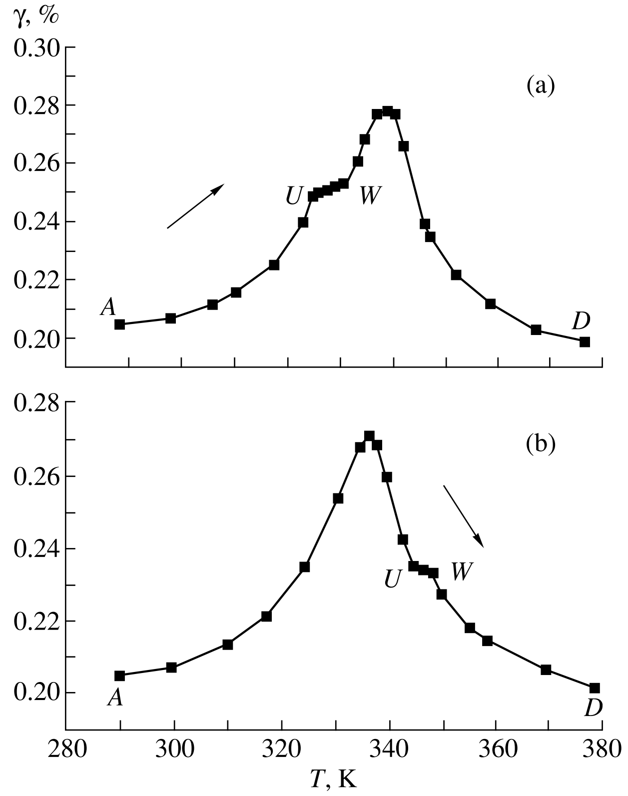


Fig. 4. Strain vs. temperature during heating and isothermal decrease in pressure at (a) 327.5 and (b) 344 K. *AU*, pressure equals 200 MPa; *U*, pressure drops from 200 MPa to the atmospheric value; and *UWD*, atmospheric pressure.

experimental results are presented in Fig. 4. After the pressure has fallen at the point *U*, further heating does not entail significant shaping; deformation proceeds only above a certain temperature (the point *W* in Fig. 4).

The discovered effects take place at the stages of both buildup and recovery (Figs. 4a, 4b). Note that, if the pressure drops from 200 MPa in the rising branch of the γ–T curve, a temperature delay of deformation ΔT is 6–6.5 K (Fig. 4a). For the falling branch of the curve, $\Delta T \approx 3.5$ K (Fig. 4b).

Our results provide vast potentialities for controlling titanium nickelide mechanical behavior by means of pressure. It was found that, during heating of the alloy, a strain of either sign can be initiated merely by an increase in the pressure (without cooling). The strain sign depends on the temperature at which the pressure was changed. It follows that temperature and pressure conditions can be selected such that a unidirectional variation of pressure (here, an increase) alone ensures strain reversibility. That is, a variation of pressure to some threshold initiates direct shaping and a further variation of pressure (in the same direction) stimulates reverse shaping. This is of special importance for creating TiNi-based devices with reversible operation due to increasing pressure.

The experimental results indicate that pressure-induced reversible deformation in Ti–49.5 at. % Ni is qualitatively similar to other memory effects [2–4]. In all these cases, an increase in the pressure during heating favors the same deformation process that proceeded before a change in pressure conditions. A decrease in the pressure always produces a temperature delay of deformation.

Also, we revealed the baromechanical effect in Ti–49.5 at. % Ni. Its essence is a low-temperature shift of the range where the twin yield stress is strongly temperature-dependent when pressure is applied. This feature is also typical of the characteristic temperatures of martensite transformation.

Finally, we should note the following. The deformation mechanism represented by the rising branch of the γ–T curve in Fig. 1 is twinning [6]. At the same time, the pressure applied to an isotropic solid may change its shape only by initiating processes responsible for shaping. Therefore, the results presented in Figs. 2a and 3 (curves *2*–*5*) suggest that, by pressure, one can induce twinning (reorientation) in crystals.

REFERENCES

1. V. A. Likhachev, V. G. Malinin, and S. Ya. Ovcharenko, *XXIV All-Union Workshop on Topical Problems of Strength, Rubezhnoe, 1990*, p. 183.
2. S. P. Belyaev, S. A. Egorov, and V. A. Likhachev, Zh. Tekh. Fiz. **66** (11), 36 (1996) [Tech. Phys. **41**, 1102 (1996)].
3. V. A. Likhachev, S. P. Belyaev, and S. A. Egorov, in *Proceedings of XXXII Workshop on Topical Problems of Strength, St. Petersburg, Russia, 1997*, p. 171.
4. V. A. Likhachev, S. P. Belyaev, and S. A. Egorov, in *XXXII Workshop on Topical Problems of Strength, St. Petersburg, Russia, 1996*, p. 44.
5. V. A. Likhachev and V. G. Malinin, *Structural Analytical Theory of Strength* (Nauka, St. Petersburg, 1993).
6. V. N. Hachin, V. G. Pushin, and V. V. Kondrat'ev, *Titanium Nickelide: Structure and Properties* (Nauka, Moscow, 1992).

Translated by A. A. Sidorova

Technical Physics, Vol. 45, No. 4, 2000, pp. 414–418. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 33–37.
Original Russian Text Copyright © 2000 by Cherenkov, Samoĭlenko, Pashchenko, Prokhorov, Gusakova.

SOLIDS

Structural Transformations in Y–Ba–Cu–Ti–O Metal Oxides Accompanying Substitution of Titanium for Y and Cu

O. P. Cherenkov, Z. A. Samoĭlenko, V. P. Pashchenko,
A. D. Prokhorov, and L. G. Gusakova

Donetsk Physicotechnical Institute, National Academy of Sciences of Ukraine, Donetsk, 340114 Ukraine

Received October 19, 1998

Abstract—Taking into account changes in the Cu charge states, general characteristics and distinctive features of structural transformations typical of high-temperature superconducting and ferroelectric phases in Y–Ba–Cu–Ti–O ceramic samples are studied. It is shown that with substitution of titanium for Y and Cu the concentration of Jahn–Teller ions Cu^{2+} detected by EPR technique does not rise higher than10%. According to X-ray spectroscopy data, the Cu charge state lowers from $Cu^{2.2+}$ to $Cu^{0.5+}$ with increasing titanium content. It is found by X-ray diffraction studies that with increasing Ti content, the fraction of the layered structure containing (00*l*) planes of orthorhombic structure decreases and the growth of planes of the (11*l*) type pertaining to tetragonal structure is stimulated. The structural phase transition from the orthorhombic to tetragonal crystal structure mainly occurs in the range $x = 0.3$–0.5. *© 2000 MAIK “Nauka/Interperiodica”.*

INTRODUCTION

In this paper, a high-temperature superconductor (HTSC)-ferroelectric (FE) phase transition was produced in the metal-oxide perovskite systems Y–Ba–Cu-Ti–O using an ion substitution method similar to that of [1]. The ion substitution (Y, Cu) $\longrightarrow$ Ti caused a gradual transition from the well-known superconducting $YBa_2Cu_3O_{7-\sigma}$ ceramics to the composition of the equally well-known ferroelectric $BaTiO_3$. For the $YBa_2Cu_3O_{7-\sigma}$ phase, the typical crystal lattice structure is orthorhombic; and for the FE $BaTiO_3$ phase, it is tetragonal. The aim of the work was to study the course of atomic rearrangements taking place when the material composition is gradually altered by way of concentrational changes in the chemical composition of the investigated solid solutions, as well as to estimate changes in the charge state of Cu^{2+} cation principally responsible, along with oxygen, for the superconducting properties of HTSC ceramics [2–8].

SAMPLES AND TECHNIQUES

Samples for the investigations were prepared by sintering powder mixtures. Content of the alkaline-earth element barium in the samples was kept constant as per the chemical formula (Ba_2); and the content of the 3*d* transition metal (Ti) in the cation sublattice, or the relation $x = Ti/(Y + Ba + Cu)$, increased from 0 to 1 to obtain the compositions $YBa_2Cu_3O_{7-\sigma}$ ($x = 0$), $Y_{0.7}Ba_2Cu_{2.1}Ti_{1.2}O_n$ ($x = 0.25$), $Y_{0.5}Ba_2Cu_{1.5}Ti_2O_n$ ($x = 0.5$), $Y_{0.2}Ba_2Cu_{0.6}Ti_2O_n$ ($x = 0.7$), and $BaTiO_3$ ($x = 1$). The initial mixtures were calcined for 5 h at 930°C. The subsequent grinding into powder was done in a portable mill; the pellets were pressed at a pressure of 0.1 GPa and then sintered for 5 h at different temperatures (from 950 to 1200°C).

The structure and properties of samples of the metal-oxide system were studied by X-ray spectroscopy, EPR (electron paramagnetic resonance), and optical methods. The X-ray diffraction analysis was carried out by a photographic method sensitive to diffusely scattered X-radiation and with the use of CrK_α radiation. The valence state of copper ions was determined from the energy shift of the CuK_{β_1} X-ray spectral line registered with a DRS-2M spectrometer with $(13\bar{4}0)$ quartz as a dispersive medium, as well as from EPR spectra. X-ray spectra of the investigated and reference samples (ZnK_{α_1} line) were excited simultaneously by a wide-angle electron beam at a low accelerating voltage (10 keV), so that with anode water-cooling the sample temperature did not rise above 50°C. The spectra of the reference compounds of copper oxides with known valence values were recorded under similar conditions.

Variations of the free carrier concentration were monitored by measuring the optical reflection coefficient R. A universal FOU-type photometer was used.

EPR spectra were recorded with a superheterodyne spectrometer operating in the 3-cm range, which could measure both the rate of change of absorption signal using modulation frequencies from 20 to 100 Hz and the absorption signal. Powdered $CuSO_4 \cdot 5H_2O$ samples were used as a standard.

RESULTS AND DISCUSSION

X-ray diffraction analysis of lattice parameters of the orthorhombic YBaCuTiO revealed that substitution of small amounts of titanium for Y and Cu ions ($0 \le x \le 0.2$) caused a drastic decrease in parameter c (by 9.4%) with practically unchanged parameters a and b (Fig. 1). As follows from the calculations, in this composition range, the decrease in the effective cation size in the unit cell compared to the starting compound is also considerable and amounts to $\Delta r/r = 0.18$ (Fig. 2, curve *2*).

As the Ti content is further increased in the range $0.22 \le x \le 0.7$, the "shrinking" of the cation size changes slightly from 0.2 to 0.3; also small is the change in parameter c (by 2%). But the increase in parameters a and b, 15 and 21%, respectively, is noticeable.

So, gradual substitution of titanium for Y and Cu in $YBa_2Cu_3O_{7-\sigma}$ first (up to $x \le 0.2$) causes a drastic contraction (>9%) of the crystal lattice of the initial structure along the crystallographic c-axis and then (at $0.22 \le x \le 0.7$) a lattice strain (>15%) in the (ab) plane in the diagonal direction. It follows from these data that the distortion accumulated in the initial crystalline structure is so great that it is ready for a transition into a different crystallographic order. The orthorhombic structure of the yttrium ceramics in the range of titanium contents from 0.7 to 1 transforms into the tetragonal structure of barium titanate. In this range, the disparity in size between the cations of the initial and final compositions increases drastically again from 0.3 to 0.82. Throughout the range of compositions, gradually accumulating distortions of the crystal structure lead to transformations typical of type II phase transitions of $YBa_2Cu_3O_{7-\sigma}$.

X-ray emission spectra studies revealed that substitution of titanium for yttrium and copper ions in Y–Ba–Cu–Ti–O caused electron density redistribution in the structure of the solid solution, such that the charge of copper ions decreased from $Cu^{2.2+}$ to $Cu^{1.6+}$ at $x = 0.2$ and down to $Cu^{0.5+}$ at $x = 0.75$ (inset in Fig. 2). In this case, the presence of Jahn–Teller Cu^{2+} ions in essentially lower concentrations than expected was detected by the EPR technique.

Lowering of the charge state of copper ions implies, first of all, a decrease in the concentration of free charge carriers taking part in the charge transfer and, second, weakening of the long-range bonding forces between copper cations and oxygen ions in the ceramic layered structure. The first factor causes degradation of the conducting and superconducting properties of the ceramics in this metal-oxide class. The second factor accounts for the destruction of crystallographic ordering in the solid solution in the form of chemical stratification with preferential development of layers of $Cu^{2+}O$ planes responsible for superconductivity. Weakening of the long-range bonding forces between atoms in a multicomponent solid solution causes weakening of the atomic order anisotropy and rearrangement of the crystal structure; this observation has been confirmed by X-ray diffraction spectra analysis (Fig. 3). Histograms of the corresponding experimental X-ray pictures are shown in Fig. 3. The line intensities are numbered starting with the brightest line, which is number 1.

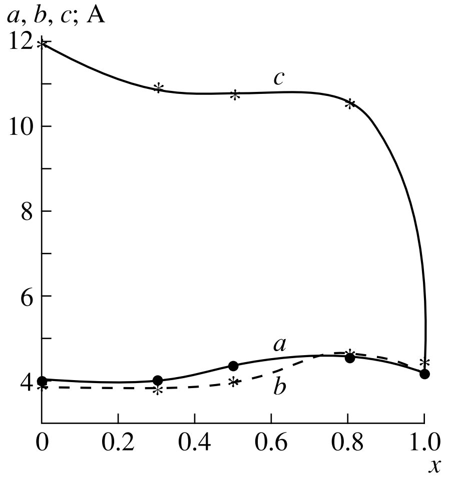


Fig. 1. Variation of the crystal lattice parameters a, b, and c in YBaCuTiO with increasing titanium content.

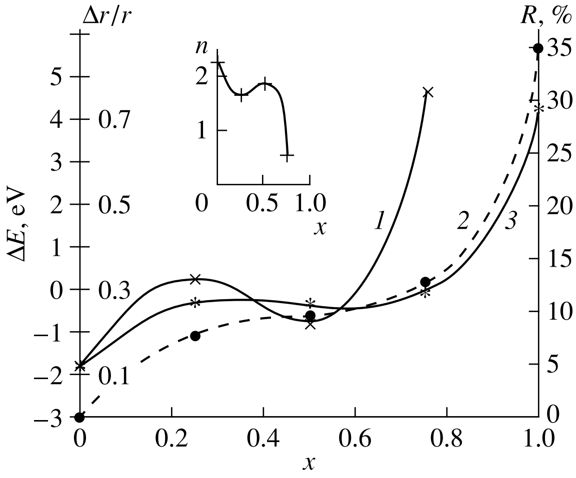


Fig. 2. Variation of the physical parameters of samples with increasing titanium content (x): (*1*) energy shift of X-ray spectral line $CuK\beta_1$; (*2*) relative variation of the cation radius ($\Delta r/r$); (*3*) optical reflection coefficient R. Dependence of the charge state (n) of copper cations on x is shown in the inset.

It has been noticed that there are usually only a few lines in the diffraction patterns from undoped $YBa_2Cu_3O_{7-\sigma}$ ceramics, with type (00l) reflections dominating, which is typical for structures of this class exhibiting superconductor properties (the critical temperature of the sample was $T_c = 94$ K). We note the following typical features of evolution of the diffraction pattern:

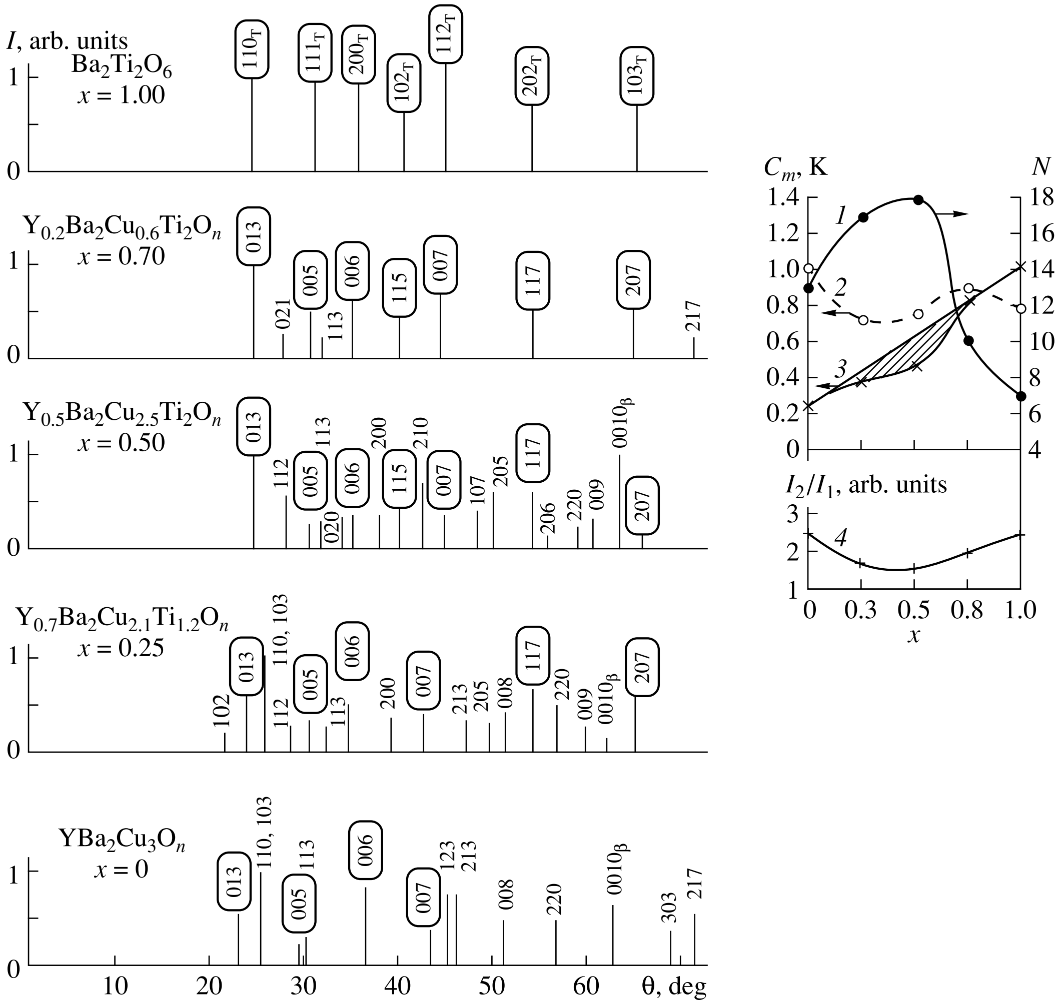


Fig. 3. X-ray histograms. Parameters obtained by analysis of the diffraction patterns are shown in inset: (*1*) number of diffraction lines N; (*2*) concentration C_m of structural components—(00l) planes; (*3*) structural coherence coefficient K; and (*4*) fraction of coherent scattering I_2/I_1.

(1) When titanium is substituted for Y and Cu ions, the number of diffraction lines N changes nonmonotonically: at first, it rises sharply (from 13 to 18) at $0 \le x \le 0.5$ (Fig. 3 and curve *1* in inset); then it drops to 7 (at $0.7 \le x \le 1$). The maximum corresponds to the most disordered state of the multicomponent solid solution Y–Ba–Cu–Ti–O based on the matrix structure of phase 123. In this state, the distortions accumulated in the crystal lattice as a result of substitution of titanium for Y and Cu destroy the initial regular arrangement of the cations in the solid solution, and there usually arise atomic arrangements in the form of layers in (00l)-type planes consisting of cations of only one of the metals and oxygen anions. With order destroyed, formation in the solid solution of every possible plane of the matrix-type crystallographic structure becomes equiprobable, which is seen as an increase in the number of diffraction lines (Fig. 3, $x \ne 0, 1$).

(2) The change in the ensemble of crystallographic planes corresponding to the atomic layers seen in the diffraction patterns is noteworthy as well. As shown earlier [2], among HTSC ceramics, superconducting properties are displayed by the structures with the prevailing atomic order represented by (00l) planes. Therefore, as an estimate of the volume concentration C_m of such structural elements in the ensemble of the crystallographic planes seen in the X-ray picture of coherent scattering from a particular sample, we get a ratio of the integrated line intensities $C_m = I_{00l}/I_1$. Equating to unity the maximum value of C_m, we obtain a nonlinear dependence of C_m on titanium content x (curve *2* in inset of Fig. 3) with a minimum in the range $x = 0.2$–0.5. This minimum corresponds to the maximum disturbance of the initial regular arrangement of the atoms of different elements in the matrix of the solid solution structure.

(3) Intensities of the major lines of the $YBa_2Cu_3O_{7-\sigma}$ phase (phase 123) vary nonmonotonically with increasing degree of substitution. In the beginning ($x = 0, 0.25$)

we can see a noticeable decrease in intensities of the reflections from planes (006), (007), (008), (009), and (0010), while reflections from (013), (110), (103) planes are practically unchanged. This fact suggests that the system still tends to form atomic arrangements dominated by (00*l*) planes. Titanium ions occupy cation positions mostly in these structural elements. The substantial drop in intensity is caused by the lower scattering ability of Ti ($Z = 22$) compared with Y ($Z = 39$) and Cu ($Z = 29$), because it is known that $I \sim Z^4$.

With a further increase in titanium content in the phase 123 solid solution ($x = 0.5, 0.75$), reflections from planes (013), (005), (006), (115), (007), (117), and (207) [having crystallographic parameters approaching those of the new structure of $BaTiO_3$ ($x = 1$)], become stronger. At the same time, the reflections from the initial matrix phase 123 disappear gradually as the titanium content is increased. The first to disappear is the brightest line, from (110) plane; next to disappear are lines of moderate intensities. So, in the diffraction pattern of a sample with $x = 0.25$, there are only three lines from the initial structure. The other seven lines show the readiness of the atomic ordering for replacement of the initial structure (orthorhombic $YBa_2Cu_3O_{7-\sigma}$ phase) by the tetragonal $BaTiO_3$ phase.

The ratio of the measured integrated intensity of reflections from the orthorhombic phase that survived among the tetragonal phase (marked by ovals in Fig. 3) to the total intensity gives a coherence coefficient of the two structures in the experimental samples $K = I_{0 \to T}/I_1$ (curve *3*, inset in Fig. 3). The range of compositions where the dependence $K = f(x)$ is nonmonotonic is shaded. From the plot one can see that the greatest mismatch in the structures of both phases occurs in the solid solution of composition $x = 0.5$. To the left of this composition the atomic order of the orthorhombic phase of $YBa_2Cu_3O_{7-\sigma}$ predominates; and to the right, the tetragonal phase of $BaTiO_3$. This conclusion is in good agreement with the curves (curves *1* and *2* in the inset in Fig. 3) obtained from analysis of diffraction patterns, as well as with the value of the minimum ratio of the intensity of coherently scattered X-radiation to the overall scattered intensity (curve *4*). The minimum in the dependence I_2/I_1 corresponds to the highest disorder in the atomic structure of the investigated solid solution and is evidence of a structural phase transition [4]. In Y–Ba–Cu–Ti–O this transition occurs at the extremum point $x = 0.5$ corresponding to the above-mentioned minimum.

Fine structure of the X-ray diffraction peaks changes as well. As Y and Cu ions are replaced by titanium, the atomic rearrangement of the crystallographic structure in the investigated series of metal oxides results in substantial changes in the shape of diffraction lines, especially distinct at large scattering angles (Fig. 4). As seen in the microphotometer curves, as the content of titanium is increased from 0 to 0.5, the sharp (220) peak decreases in intensity and becomes broader, assuming the form of a diffusion reflection; at $x = 0.75$ it disappears. At the same time, the weak diffuse (117) peak of the orthorhombic phase increases in intensity and broadens, "swelling" at $x = 0.5$ and, at $x = 0.75$, turning into a narrow sharp line of reflection from the (022) plane of the new tetragonal phase $BaTiO_3$. Continuous changes in angular position and shape are also observed for the reflection from the (207) plane of the orthorhombic phase; after "swelling" at $x = 0.75$, the peak transforms into the (013) peak of the new phase. Apart from above-mentioned, all other orthorhombic structure reflections detected in the initial or close to initial compositions and typical of the specified range of scattering angles disappear with increasing x (Figs. 3, 4). Described evolution of the fine structure of diffraction peaks indicates a gradual enhancement of the crystallographic components of the initial atomic arrangement "inherited" by the new phase. They transform from weak lines into intensive bands (halo) and then to bright lines: (107) $\longrightarrow$ (022), (207) $\longrightarrow$ (013). At the same time, the "inconvertible" crystallographic components are destroyed, even though their initial atomic ordering was perfect; this is exemplified by the disappearance of the line corresponding to the (220) planes when x is increased from 0 to 0.5.

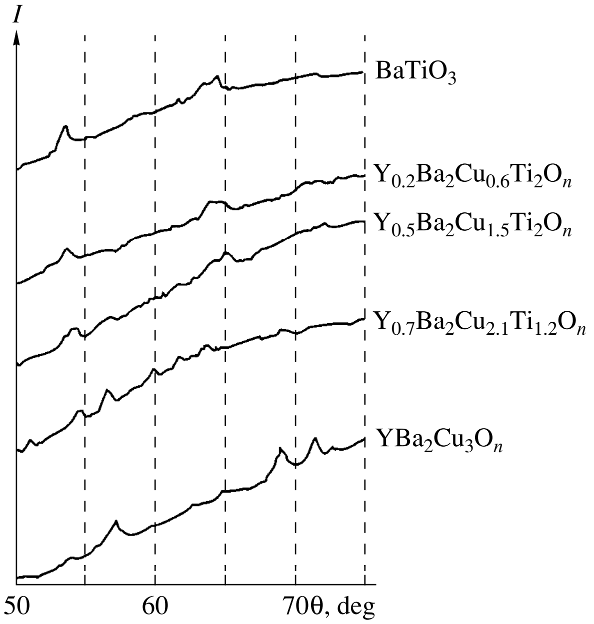


Fig. 4. Fine structure evolution of the diffraction patterns in the range of scattering angles θ from 52° to 67°.

CONCLUSIONS

The studies of YBaCuTiO ceramics in the range of compositions from the HTSC phase of $YBa_2Cu_3O_{7-\sigma}$ to the ferroelectric phase $BaTiO_3$ have revealed that titanium behaves as a chemically active component when it is substituted for Y and Cu. It has been determined by EPR and X-ray emission spectroscopy that the driving force of structural reconstruction occurring in the investigated multicomponent system is the redistribution of spatial electron density in the vicinity of copper cations caused by the introduction of titanium and leading to a decrease in the effective charge of Cu cations. These developments destroy superconductivity and the layered structure of the solid solution and cause the concentrational phase transition of type II, by which the orthorhombic structure of HTSC is converted to the tetragonal one of FE, with maximum distortions of the electronic and atomic subsystems occurring in the range a compositions x = Ti/(Y + Ba + Cu) = 0.3–0.5.

REFERENCES

1. V. P. Pashchenko, A. M. Nesterov, V. I. Arkharov, *et al.*, Dokl. Akad. Nauk SSSR **318**, 371 (1991).
2. Z. A. Samoĭlenko, V. D. Okunev, E. I. Pushenko, *et al.*, Neorg. Mater. **32**, 359 (1996).
3. V. D. Shveĭkin, V. A. Gubanov, A. A. Fotiev, *et al.*, *Electronic Structure and High Temperature Superconductors Physicochemical Properties* [in Russian] (Nauka, Moscow, 1990).
4. M. A. Krivoglaz, in *Electron Properties of Metals and Alloys* [in Russian] (Naukova Dumka, Kiev, 1988), pp. 3–39.
5. V. D. Okunev, Z. A. Samoĭlenko, N. N. Pafomov, *et al.*, Neorg. Mater. **33**, 737 (1997).
6. V. V. Shpeizman, T. S. Orlova, B. I. Smirnov, *et al.*, Cryst. Res. Technol. **25**, 827 (1990).
7. V. A. Antonov, P. A. Arsen’ev, and O. I. Koval’, Izv. Akad. Nauk SSSR, Neorg. Mater. **26**, 2625 (1990).
8. I. S. Lyubutin, E. M. Smirnovskaya, E. M. Terziev, *et al*., Pis’ma Zh. Éksp. Teor. Fiz. **47**, 196 (1988).

Translated by M.S. Astrov

Technical Physics, Vol. 45, No. 4, 2000, pp. 419–422. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 38–41.
Original Russian Text Copyright © 2000 by Zon, Ledovskiĭ, Likholet.

Lattice Diffusion of Impurity Enhanced by a Surface Heterogeneous Reaction

B. A. Zon, S. B. Ledovskiĭ, and A. N. Likholet

Voronezh State University, Universitetskaya pl. 1, Voronezh, 394693 Russia

Received November 10, 1998

Abstract—It is shown that a heterogeneous surface reaction affects impurity distribution in the bulk of a solid body. An analytic expression for the concentration of crystal defects as a function of coordinate and reaction time and a diffusion equation in view of this expression were derived. Numerical simulation shows that the impurity concentration varies as $\sim\exp(-(x/\sigma)^{\beta})$. Analytic results were compared with experimental data for phosphorus diffusion upon forming nickel silicide. *© 2000 MAIK “Nauka/Interperiodica”.*

INTRODUCTION

Radiation, electron-beam, and ion-plasma processing are known to enhance impurity diffusion in solids and, in particular, in semiconductors [1–3]. It has been noted that some solid-phase reactions proceeding on the surface of a sample may substantially affect impurity distribution in its bulk [4–6]. Wittmer and Tu [4, 5] suggested a model for the diffusion of an impurity atom into a semiconductor. In this model, silicide formation in the bulk of silicon produces a flux of point lattice defects that enhances (activates) impurity diffusion. Upon silicide formation, the diffusion coefficient of an impurity in silicon was shown to rise by several orders of magnitude in comparison with conventional diffusion [6].

In this work, we elaborated a mathematical model of this process. To be definite, we consider inward impurity diffusion from a film of a doped metal on the silicon surface due to the silicide formation reaction.

DERIVATION OF THE BASIC EQUATION

Consider a piece of undoped silicon occupying the half-space $x > 0$. It is covered by a metal film doped to a concentration far exceeding the dopant solubility limit in silicon at a given temperature. This can be a nickel film containing 5–7 wt% of phosphorus.

Let the temperature rise at $t = 0$. We assume for simplicity that the temperature rises instantaneously. A silicide layer begins to form on the silicon surface. Its thickness is a function of time: $l = l(t)$, $l(0) = 0$. The impurity diffuses into the semiconductor, and it is necessary to find the impurity distribution $C(x, t)$ in the silicon at any time instant. Distributions of impurity atoms in the metal and the silicide are of no practical value and will not be considered. Let

$$C(l(t), t) = C_0 = \text{const.} \tag{1}$$

Condition (1) suggests that the metal film is heavily doped and that the impurity concentration at the silicide / silicon interface is constant in the course of the reaction [6]. It is assumed that the solubility limit and the diffusion coefficient of the impurity in the silicide are such that they ensure an impurity flux from the metal film sufficient for (1) to be met. The problem thus stated is described by the diffusion equation

$$\frac{\partial C(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(D(x,t)\frac{\partial C(x,t)}{\partial x}\right), \quad l(t) \le x < \infty \tag{2}$$

where (1) is the edge condition and

$$C(x, 0) = 0 \tag{3}$$

is the initial condition. Here, $D(\cdot)$ is the time- and coordinate-dependent diffusion coefficient of the impurity. The form of this dependence is related to crystal defect generation due to silicide formation. As was noted, defect generation during silicide formation can raise the diffusion coefficient many times. Specifically, for the formation of nickel monosilicide, the diffusion coefficient of phosphorus in silicon at 600°C is of the order of 10^{-12} cm^2/s [6], while the coefficient of ordinary (interstitial) phosphorus diffusion is 7.4×10^{-21} cm^2/s at the same temperature [7]. We will therefore consider only defect diffusion. The dependence of the diffusion coefficient on the impurity concentration will also be ignored, since the contribution of this effect is comparable to that from interstitial diffusion (for this effect for phosphorus, see [8]).

The diffusion coefficient of an impurity can be represented as [9]

$$D = \frac{1}{6} f d^2 \Gamma,$$

where f is a correlation factor ($f = 1/2$ for diamond-like lattices), d is the lattice constant, and Γ is the jump rate. From [10],

$$\Gamma = n\nu \exp\left(\frac{G}{kT}\right).$$

Here, n is the number of paths whereby an impurity atom can pass from one position to another, ν is the eigenfrequency of vibration of an impurity atom, and G is the isothermal work needed for an impurity atom to change its position. For lattice defect diffusion, n can be written in the form [9]

$$n = \frac{Z}{S}\sum_i U_i \exp\left(-\frac{G_p + G_{qi}}{kT}\right),$$

where Z is the correlation factor, S is the number of lattice atoms, U_i is the concentration of defects of the ith type, G_p is the strain energy due to the overlap of closed shells of impurity and lattice atoms, and G_{qi} is the energy of Coulomb interaction between an impurity ion and a defect with a charge q_i.

Assume, for simplicity, that silicide formation generates defects of only one type. Then,

$$n = \frac{ZU}{S}\exp\left(-\frac{G_p + G_q}{kT}\right),$$

and the diffusion coefficient is proportional to the defect concentration:

$$D(x, t) = KU(x, t), \tag{4}$$

where

$$K = \frac{d^2 \nu Z}{12S}\exp\left(-\frac{G + G_p + G_q}{kT}\right). \tag{5}$$

It is assumed that the concentration S of lattice atoms does not depend on the defect concentration $U(x, t)$. This is valid if $U(x, t)$ is much smaller than S. The defect concentration is found from the solution to the diffusion problem

$$\frac{\partial U(x, t)}{\partial t} = \frac{\partial}{\partial x}\left(D_v \frac{\partial U(x, t)}{\partial x}\right), \quad l(t) \le x < \infty \tag{6}$$

with the initial condition $U(x, 0) = 0$, since most defects arise upon silicide formation. Here, D_v is the diffusion coefficient of crystal defects in silicon. We assume that D is independent of the impurity and defect concentrations and is constant.

In formulating a boundary condition to (6), it would be natural to suggest that the silicon / silicide interface is an efficient defect producer, since silicon is involved in the formation of silicide. Assuming that the defect concentration in silicon has some limit,

$$U(l(t), t) = U_0 = \text{const}. \tag{7}$$

We also assume that silicide growth follows a parabolic law $l(t) = \alpha\sqrt{t}$, which is usually the case [11]. Then, (6) and (7) have the regularly varying solution [12]

$$U(x, t) = U_0\left(1 - \operatorname{erf}\left(\frac{x}{\sqrt{4D_v t}}\right)\right) \Big/ \left(1 - \operatorname{erf}\left(\frac{\alpha}{\sqrt{4D_v}}\right)\right), \tag{8}$$

which is consistent with experimental data [13]. Here,

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z \exp(-y^2)\,dy$$

is an error function integral. Substituting (4) and (8) into (2), we obtain the equation that, together with initial condition (3) and boundary condition (1), describes impurity distribution during the formation of silicide:

$$\frac{\partial C(x, t)}{\partial t} = \frac{\partial}{\partial x} \times \left[KU_0\left(1 - \operatorname{erf}\left(\frac{x}{\sqrt{4D_v t}}\right)\right) \Big/ \left(1 - \operatorname{erf}\left(\frac{\alpha}{\sqrt{4D_v}}\right)\right) \times \frac{\partial C(x, t)}{\partial x}\right], \tag{9}$$

where K is found from (5).

Equation (9) has α, D_v, and KU_0 as parameters. The last, composite parameter is the diffusion coefficient of impurity at the silicon / silicide interface. If $D_v \gg KU_0$ and α^2, this problem has an analytic solution. For $x \sim \sqrt{KU_0 t}$, the probability integrals are neglected, since they are much smaller than unity. We then arrive at the equation

$$\frac{\partial C(x, t)}{\partial t} = \frac{\partial}{\partial x}\left(KU_0 \frac{\partial C(x, t)}{\partial x}\right),$$

which describes impurity diffusion for a constant concentration of defects. With regard for the initial and boundary conditions, one obtains

$$C(x, t) = C_0\left(1 - \operatorname{erf}\left(\frac{x}{\sqrt{4KU_0 t}}\right)\right) \Big/ \left(1 - \operatorname{erf}\left(\frac{\alpha}{\sqrt{4KU_0}}\right)\right). \tag{10}$$

If $KU_0 \gg D_v$ and α^2, equation (9) is impossible to simplify; however, an analytic solution can be obtained as follows. When $KU_0 \gg D_v$, an impurity occupies defects much faster than they migrate into silicon and it can be considered that all defects are occupied by impurity. Then, the impurity concentration $C(x, t)$ coincides with the defect concentration $U(x, t)$ [see (8)]. For other relationships between the parameters, (9) can be solved with numerical techniques.

COMPUTATION TECHNIQUE AND NUMERICAL EXAMPLE

In view of the edge and initial conditions, the solution to (9) can be found in the form [14]

$$C(\sigma(t), \beta(t)) = C_0 \exp\left(-\left(\frac{x}{\sigma(t)}\right)^{\beta(t)}\right). \tag{11}$$

The quantity $\sigma(t)$ has the meaning of effective penetration depth of impurity. Numerically, $\sigma(t)$ and $\beta(t)$ are found by an optimization technique. To do this, a time domain of interest is divided into small intervals $[t_i, t_{i+1}]$ within which σ and β are assumed to be constant:

$$\sigma(t) = \sigma_i = \sigma(t_i), \quad \beta(t) = \beta_i = \beta(t_i).$$

Within each interval $[t_i, t_{i+1}]$, σ_i and β_i are determined with the golden section technique from the minimality condition for the quantity

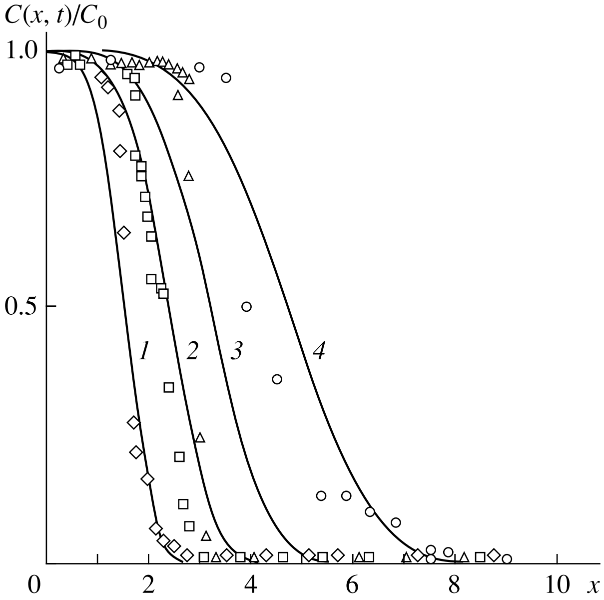
$$S_i \equiv \int\limits_{l(t_i)}^{\infty} \left\{ \frac{\partial C}{\partial t} - \frac{\partial}{\partial x} \right.$$
$$\times \left[K U_0 \left(1 - \operatorname{erf}\left(\frac{x}{\sqrt{4 D_V t}}\right)\right) \Big/ \left(1 - \operatorname{erf}\left(\frac{\alpha}{\sqrt{4 D_V}}\right)\right) \right. \tag{12}$$
$$\left. \left. \times \frac{\partial C(\sigma_i, \beta_i)}{\partial x} \right] \right\}^2 dx,$$

where the time derivative is calculated by the formula

$$\frac{\partial C}{\partial t} = \frac{C(\sigma_i, \beta_i) - C(\sigma_{i-1}, \beta_{i-1})}{t_i - t_{i-1}}$$

and the coordinate derivatives are taken analytically. With σ and β known, the impurity concentration is found from (11).

The method of computation is illustrated by numerically solving (9) for t varying from 0 to 1.5 h, $KU_0 = 1.98 \times 10^{-11}$ cm^2/s (this value best fits experimental data in [6]), $\alpha = 10^{-7}$ cm/s$^{1/2}$ (calculated from data in [15]), and $D_V = 10^{-12}$ cm^2/s [13]. With such values of the parameters, $\beta \approx 4$ and S_i, defined by (12), depends on β much more weakly than σ. Therefore, the only dependence to be optimized is $\sigma(t)$.



Analytic (this work) and experimental [6] phosphorus distribution in silicon after isothermal annealing at 600°C for (*1*, ◇) 2.6, (*2*, □) 21, (*3*, △) 42, and (*4*, ○) 90 min.

The figure shows analytic and experimental [6] coordinate dependences of the impurity concentration for different reaction (silicide formation) times ($x = 0$ at the silicide/silicon interface). The analytic curves fall more steeply at the interface and have a less pronounced plateau. The reason for these discrepancies may be that data in [6] were obtained for electrically active phosphorus, while our calculation was made for the entire concentration of phosphorus.

ACKNOWLEDGMENTS

The authors are indebted to V.S. Pitanov for valuable discussions.

REFERENCES

1. V. E. Vavilov, A. E. Kiv, and O. R. Niyazova, *Formation and Migration of Defects in Semiconductors* (Nauka, Moscow, 1981).
2. G. V. Dudko, M. A. Kolegaev, and V. A. Panteleev, Fiz. Tverd. Tela (Leningrad) **11**, 1356 (1969) [Sov. Phys. Solid State **11**, 1097 (1969)].
3. V. A. Labunov and V. E. Borisenko, Fiz. Tekh. Poluprovodn. (Leningrad) **13**, 604 (1979) [Sov. Phys. Semicond. **13**, 355 (1979)].
4. M. Wittmer, C. Y. Ting, I. Ohdomari, *et al.*, Appl. Phys. **53**, 6781 (1982).
5. M. Wittmer and K. N. Tu, Phys. Rev. B: Condens. Matter **29**, 2010 (1984).
6. V. S. Pitanov, I. N. Kornfel'd, and T. N. Il'ina, Izv. Akad. Nauk SSSR, Neorg. Mater. **20**, 1582 (1984).

7. B. I. Boltaks, *Diffusion and Point Defects in Semiconductors* (Nauka, Leningrad, 1972).
8. A. O. Konstantinov, Fiz. Tekh. Poluprovodn. (S.-Peterburg) **26**, 339 (1992) [Sov. Phys. Semicond. **26**, 191 (1992)].
9. *Atomic Diffusion in Semiconductors*, Ed. by D. Show (Plenum, New York, 1971; Mir, Moscow, 1975).
10. *Imperfections in Nearly Perfect Crystals*, Ed. by C. Zener and W. Shockley (Wiley, New York, 1952).
11. S. P. Murarka, *Silicides for VLSI Applications* (Academic, New York, 1983; Mir, Moscow, 1986).
12. A. N. Tikhonov and A. A. Samarskiĭ, *Equations of Mathematical Physics* (Nauka, Moscow, 1972).
13. E. Rosencher, S. Delage, E. Arnaud, *et al.*, J. Vac. Sci. Technol. B **3**, 762 (1985).
14. B. A. Zon, S. B. Ledovskiĭ, and A. N. Likholet, Zh. Tekh. Fiz. **68** (4), 75 (1998) [Tech. Phys. **43**, 416 (1998)].
15. V. A. Uskov, A. B. Fedotov, A. I. Rodionov, *et al.*, Izv. Akad. Nauk SSSR, Neorg. Mater. **20**, 1148 (1984).

Translated by V. Isaakyan

Technical Physics, Vol. 45, No. 4, 2000, pp. 423–426. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 42–45.
Original Russian Text Copyright © 2000 by I. Khaneft, A. Khaneft.

SOLIDS

Effect of the Duration of the Front Edge of the Voltage Pulse on the Electric Breakdown of Ammonium Perchlorate

I. G. Khaneft and A. V. Khaneft

Kemerovo State University, ul. Krasnaya 6, Kemerovo, 650043 Russia

Received December 1, 1998

Abstract—The results of a study of the effect of the front edge duration of a voltage pulse (τ) on the electric breakdown of ammonium perchlorate single crystals are submitted. Experiments have shown that at $\tau < 1.5$ μs, the breakdown threshold field increases with reduction in τ and does not depend on τ at $\tau > 1.5$ μs. It is shown that the electric breakdown of ammonium perchlorate results from an impact generation of electrons. The impact generation rate of electrons versus electric field and the diameter of the through channel being formed during electric breakdown in ammonium perchlorate are estimated. *© 2000 MAIK “Nauka/Interperiodica”.*

Ammonium perchlorate (AP) belongs to the compounds with a large surplus of oxygen. It is used as an oxidizing component of mixed solid rocket fuels [1, 2]. That is why it is necessary to study comprehensively the effect of various factors, including the electric field [3], on its thermal stability. The papers [4, 5] are devoted to studies of the electric strength of single crystals and polycrystalline pressed AP samples as a function of the distance between electrodes L and density ρ at a duration of the front edge of the voltage pulse τ of 1.5 μs. These studies revealed that the electric strength of AP single crystals at $L < 0.04$ cm was considerably higher than that of the pressed AP tablets. This is natural, because the electric breakdown of pressed tablets originates in air pores [6, 7]. It was shown in [4] that the electric breakdown of ammonium perchlorate is not of thermal nature. This paper presents experimental results describing the effect of the front edge duration of the voltage pulse on the electric strength of ammonium perchlorate crystals. The samples grown out of a supersaturated aqueous solution of ammonium perchlorate were used to study the electric strength of the AP single crystals. These crystals were colorless transparent plates, and the electric field was applied to the crystal along the (210) axis. Holes of diameter d = 0.7 mm were preliminarily drilled under the microscope on opposite sides of the single crystal. The depth of these holes was determined by the desired distance between electrodes. The saturated solution of AP (electrolyte) located in the holes was used as an electrode. Polarization microscopy revealed the absence of residual mechanical stresses in the space of the crystal between the electrodes [4]. The dependence of the breakdown voltage of the dielectric U* was determined at the front edge of a pulse of amplitude U_0 = 10 kV. Duration of the front edge τ of the voltage pulse was determined in the open-circuit generator regime. Depending on the spread of the AP breakdown voltage, the measurements of each experimental point were averaged over 11–20 samples. During the breakdown, an electric spark was observed in the inner space of the sample. The study of AP crystals under a microscope revealed that the electric breakdown of [thin] samples ($L < 0.02$ cm) causes their burning out accompanied by the formation of a through channel. The diameter of the channel increases from 3 to 5 μm as L decreases from 0.02 to 0.005 cm. The through channel was not observed for [thick] samples ($L > 0.02$ cm). Probably this was due to the insufficient resolution of the microscope used in the experiment (the least division was 5 μm). However, a yellow spot was observed on the surface of the crystals at the breakdown point. This is evidence of chlorine arising in the breakdown. Also, the probability of self-recovery of the channel as a result of condensation of decay products in it is higher for [thick] samples.

EXPERIMENTAL RESULTS AND THEIR DISCUSSION

The results of measurements of the electric strength of ammonium perchlorate crystals versus front edge duration of the voltage pulse at L = 0.01 (curve *1*) and 0.03 cm (curve *2*) are shown at Fig. 1. The dependence of the breakdown electric field E^* versus τ is equivalent to that of the electric strength versus the inverse rate of the increase in the voltage across the samples $(U_0/\tau)^{-1}$. It follows from Fig. 1 that the electric strength of AP increases with a decrease in the duration of the front edge of the electric pulse at $\tau < 1.5$ μs. E^* becomes constant at $\tau > 1.5$ μs. Note that the time τ at which the electric strength becomes constant is almost independent of the interelectrode distance. According to [8] the effect of a nonmonotonic dependence of the dielectric breakdown voltage versus the inverse rate of the voltage increase is conditioned by polarization relaxation of the

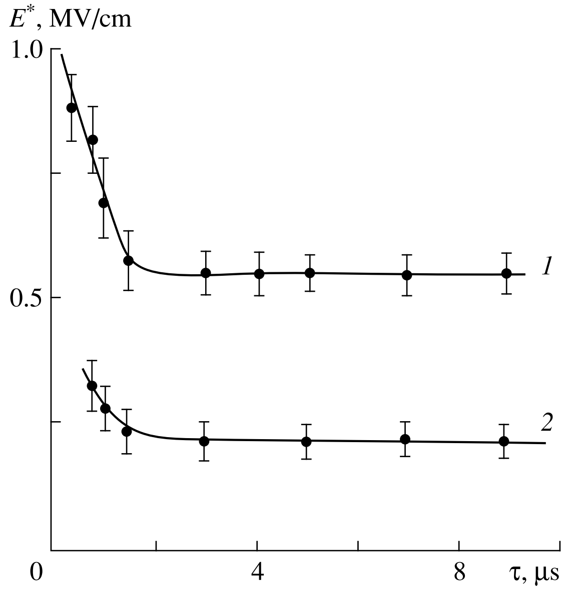


Fig. 1. The dependence of the electric strength of ammonium perchlorate as a function of duration of the front edge of the voltage pulse at $L = 0.01$ (*1*) and 0.03 cm (*2*).

dielectric. According to [8], the dependence of the breakdown voltage $U^* = f(\tau/U_0)$ should have a minimum, which shifts to the left as the thickness of the dielectric increases. Consequently, in our case, polarizable relaxation either does not affect or insignificantly affects the AP breakdown voltage. According to [9], an increase in the breakdown voltage of the dielectric with a decrease in the voltage action time is due to the delay of the electric discharge formation. The time of electric discharge delay involves the breakdown formation time t^* and statistical delay time t_s. The independence of the breakdown voltage of τ at large τ, according to [9], could be a consequence of the formation of a volume charge, its electric field being opposite to the external one.

Let us derive the qualitative dependence of the dielectric electric strength on the rate of increase in the voltage across the sample. Suppose also that in AP, $t^* \gg t_s$ similar to alkali-halide crystals [9]. In the first approximation, assume that the voltage across the sample increases linearly, $U(t) = ht$, where $h = U_0/\tau$ is the rate of the voltage change across the sample. Impact ionization determines the increase in the electron density in the dielectric. Then, the breakdown formation time t^* is estimated by [10]

$$qv(t^*)n_0S\exp\left(\int_0^{t^*}\alpha v\,dt\right) = 0.1ht^*/r, \tag{1}$$

where n_0 is the starting concentration of electrons before impact ionization, v is the electron drift velocity, r is the external electric circuit resistance, t^* is the time for the current to reach 10% of its maximum value, and S is the area of the electrodes.

Taking into account that $v(t^*) = \mu ht^*/L$, transform equation (1) into

$$\exp\left(\mu\int_0^{t^*}\alpha E\,dt\right) = 0.1R/r, \tag{2}$$

where μ is the electron mobility and $R_0 = L/(q\mu n_0 S)$ is the resistance of the sample at $t = 0$.

Let us assume that the impact electron multiplication rate versus electric field follows a power law $\alpha = AE^m$ [11]. Integrating the left side of equation (2), one derives after transformations

$$t^* = B(\tau L/U_0)^{(m+1)/(m+2)}, \tag{3}$$

where

$$B = \{[(m+2)/\mu A]\ln(0.1R_0/r)\}^{1/(m+2)}.$$

The constant B weakly depends on L. Considering that $E^* = t^*U_0/\tau L$, one gets from equation (3)

$$E^* = B(U_0/\tau L)^{1/(m+2)}. \tag{4}$$

Hence, it follows from equation (4) that the smaller the value of τ, the larger the electric strength E^* of the dielectric and the smaller the value of L, the steeper the dependence of E^* on τ. These conclusions are in qualitative agreement with the experimental results in Fig. 1 for $\tau < 1.5$ μs.

Earlier, the experimental data on electric strengthening of ammonium perchlorate crystals along with the shortening of the interelectrode distance were analyzed [4]. Using the electric breakdown condition [11]

$$\alpha L = 1 \tag{5}$$

the authors estimated the impact multiplication rate of electrons α. They applied the following empirical dependence on the electric field E for it [12, 13]:

$$\alpha = A\exp[-(b/E)^{\gamma}] \tag{6}$$

at $A = 1.82 \times 10^3$ cm^{-1}, $b = 8.4 \times 10^6$ V/cm, and $\gamma = 0.39$ [4]. The value of $\gamma < 1$ for α was predicted theoretically in [14] for electron scattering by polar optical vibrations of the lattice in ionic crystals. It should be noted that typical values of the impact multiplication rate of electrons in solids under electric breakdown are of the order of 10^3–10^4 cm^{-1} [13], i.e., a factor of ten higher than follows from criterion (5) applied to AP crystals. Let us consider, for example, rock salt to investigate the validity of determining the parameters A, b, and γ for the impact multiplication rate using the electric breakdown condition for dielectrics (5). The data are known for the dependence of the electric strength of NaCl on the distance between electrodes. The impact multiplication rate for this material is $\alpha = 2 \times 10^5\exp(-1.1 \times$

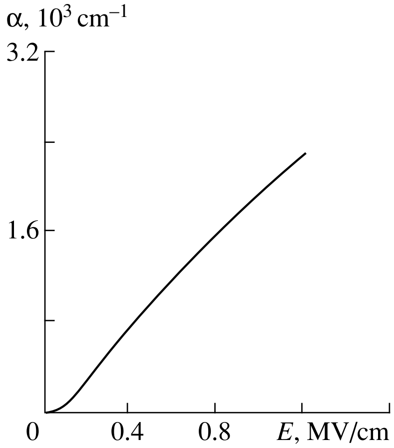


Fig. 2. The impact multiplication rate of electrons for ammonium perchlorate versus electric field.

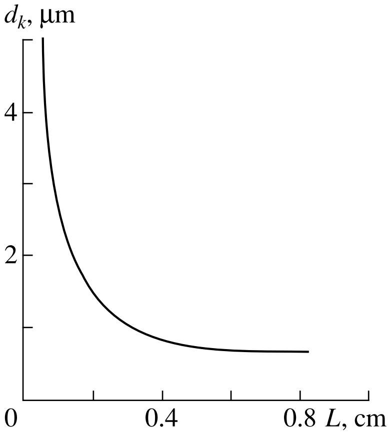


Fig. 3. The diameter of the through channel formed during electric breakdown in ammonium perchlorate versus distance between the electrodes (calculated).

$10^7/E$) cm^{-1} [15]. We processed these data and found out that α and *L* for NaCl are related by the formula

$$\alpha L = K, \tag{7}$$

where $K \approx 10.9$ in the interval $10^{-2} \geq L \geq 5 \times 10^{-4}$ cm.

The electric breakdown criterion (7) is qualitatively similar to that of avalanche-streamer (spark) discharge in gases at atmospheric pressure $\alpha L \approx 20$ [16]. Thus, from a comparison of the breakdown conditions (5) and (7), it appears that the breakdown criterion (5) allows the functional dependence of α on the electric field (parameters *b* and γ) to be determined correctly. However, it lowers by a factor of *K* the factor *A* of the exponential of the impact multiplication of electrons. Consequently, the true value of the factor *A* is approximately ten times higher than that derived from the condition for electric breakdown (5); i.e.,

$$\alpha \approx 2 \times 10^4 \exp[-(8.4 \times 10^6/E)^{0.39}] \text{ cm}^{-1}. \tag{8}$$

Consider the possibility of applying expression (4) to the quantitative estimations of the dependence of E^* on τ for ammonium perchlorate. Figure 2 demonstrates the dependence of α on E^* for AP calculated according to expression (8). It follows from this figure that, at $E >$ 200 kV/cm, the dependence of the impact multiplication rate of electrons on the electric field can be approximated well to a linear function $\alpha \sim E$, i.e., $m = 1$. Thus, according to (4), the following relation between two points on the curve $E^* = f(\tau)$ for $\tau < 1.5$ μs and $L = \text{const}$ can be established:

$$E_1^*/E_2^* = (\tau_2/\tau_1)^{1/3}. \tag{9}$$

In accordance with Fig. 1, the curve *1* ($L = 0.01$ cm) at $\tau_1 = 0.4$ μs has $E_1^* = 905$ kV/cm and at $\tau_2 = 1.5$ μs, $E_2^* = 570$ kV/cm. The values of $E_1^*/E_2^* \approx 1.59$ and $(\tau_2/\tau_1)^{1/3} \approx 1.55$ are close to each other and in good agreement with equality (9). Curve *2* ($L = 0.03$ cm) satisfies equality (9) to a lower degree. Thus, at $\tau_1 = 0.8$ μs and $E_1^* = 327$ kV/cm, $\tau_2 = 1.55$ μs and $E_2^* = 233$, we have $E_1^*/E_2^* \approx 1.44$ and $(\tau_2/\tau_1)^{1/3} \approx 1.25$. This might be due to the fact that, at $\tau_1 = 0.8$ μs and $L = 0.03$ cm, the AP breakdown voltage is close to the pulse amplitude; i.e., the voltage rise across the sample cannot be well approximated by a linear function.

Let us estimate the diameter of a through channel d_k forming during AP breakdown. Assume that at the moment the AP sample is burnt out, the energy supply from the generator to the breakdown channel is significant and the energy released in the channel is determined mostly by that which is stored in the geometric capacitance of the sample. In this case [17],

$$CU^{*2}/2 = \pi d_k^2 \rho L Q/4, \tag{10}$$

where *C* is the geometric capacity of the sample, *Q* is the amount of heat necessary to heat up and sublimate a unit mass of the substance, and ρ is the crystal density.

The heat *Q* is equal to 3394 kJ/kg [2] in the case of dissociative sublimation of AP in the breakdown channel. The effect of the residual voltage was neglected in equation (10). Considering that the capacitance of the sample is $C = \varepsilon\varepsilon_0\pi d^2/4L$, we derive the expression for d_k from (10):

$$d_k = dE^*(\varepsilon\varepsilon_0/2\rho Q)^{1/2}. \tag{11}$$

The dependence of the through channel diameter on the breakdown field $E^*(L)$ is calculated following formula (11) at $\tau = 1.5$ μs and $\rho = 1.95 \times 10^3$ kg/m^3. Figure 3 shows the dependence of d_k on the interelectrode distance. The diameter of the through channel evidently diminishes as the distance between the electrodes increases. This fact is in qualitative agreement with the experimental results. The theoretically calculated value of d_k is close to the experimental data for thin samples. Consequently, the supposition that the energy released in the sample during breakdown is equal to that stored in its geometric capacitance is valid.

Hence, the qualitative and quantitative explanation of the dependencies of the electric breakdown of ammonium perchlorate single crystals on the distance between the electrodes and duration of the pulse front edge of the voltage across the sample allows us to conclude that the breakdown of AP single crystals is determined by multiplication of electrons due to impact ionization.

REFERENCES

1. E. F. Khaĭretdinov, T. V. Mulina, and V. V. Boldyrev, *Thermal Mechanism of Decomposition of Ammonium Perchlorate* [in Russian] (IKhF AN SSSR, Chernogolovka, 1981), pp. 101–103.
2. L. K. Gusachenko, V. E. Zarko, V. Ya. Zyryanov, *et al.*, *Modeling of Processes of Solid Fuel Burning* (Nauka, Novosibirsk, 1985).
3. B. K. Laptenkov, A. V. Raevskiĭ, G. B. Manelis, *et al.*, Dokl. Akad. Nauk SSSR **250**, 1185 (1980).
4. I. G. Khaneft and A.V. Khaneft, Zh. Tekh. Fiz. **65**, 200 (1995) [Tech. Phys. **40** (6), 631 (1995)].
5. Yu. N. Sukhushin and I. G. Khaneft, Prikl. Mekh. Tekh. Fiz., No. 3, 12 (1990).
6. G. I. Skanavi, *Physics of Dielectrics (Strong Field Range)* [in Russian] (Fizmatgiz, Moscow, 1958).
7. V. V. Stengach, Prikl. Mekh. Tekh. Fiz., No. 1, 128 (1972).
8. Yu. N. Vershinin and A. S. Pleshanov, Élektrichestvo, No. 12, 54 (1988).
9. A. A. Vorob’ev and G. A. Vorob’ev, *Electric Breakdown and Destruction of Solids* [in Russian] (Vysshaya Shkola, Moscow, 1966).
10. V. G. Geĭman, S. A. Genkin, and K. A. Klimenko, Zh. Tekh. Fiz. **55**, 2347 (1985) [Sov. Phys.–Tech. Phys. **30** (12), 1394 (1985)].
11. W. Franz, *Dielektrischer Durchschlag* (Springer-Verlag, Berlin, 1956; Inostrannaya Literatura, Moscow, 1961).
12. I. V. Grekhov and Yu. N. Sereghkin, *Avalanche Breakdown of p–n-Function in Semiconductors* (Énergiya, Leningrad, 1980).
13. F. Capasso, *Semiconductors and Semimetals: A Treatise*, Ed. by W. T. Tsang (AT&T Bell Laboratories, New Jersey, 1985; Mir, Moscow 1988), Vol. 22, Part D.
14. V. A. Chuenkov, Fiz. Tekh. Poluprovodn. **9**, 48 (1967) [Sov. Phys.–Solid State **9** (1) (1967)].
15. G. A. Vorob’ev, Izv. Vyssh. Uchebn. Zaved., No. 10, 158 (1972).
16. Yu. P. Raĭzer, *Physics of Gas Discharge* [in Russian] (Nauka, Moscow, 1987).
17. G. A. Vorob’ev and V. A. Mukhachev, *Breakdown of Thin Dielectric Films* [in Russian] (Sov. Radio, Moscow, 1977).

Translated by C.V. Egorov

Technical Physics, Vol. 45, No. 4, 2000, pp. 427–431. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 46–51.
Original Russian Text Copyright © 2000 by Baranov, Drozdov, Chudinov, Bayankin.

Atomic Mechanisms of Microcrack Propagation in Pure and Hydrogen-Containing FCC and BCC Metals

M. A. Baranov, A. Yu. Drozdov, V. G. Chudinov, and V. Ya. Bayankin

Institute of Physics and Technology, Ural Division, Russian Academy of Sciences,
ul. Kirova 132, Izhevsk, 426001 Russia

Received January 11, 1999

Abstract—A molecular dynamics method was used to simulate crack propagation in pure and hydrogen-containing aluminum and α-iron for loads far exceeding the critical values. Pairwise interaction potentials calculated within the Heine–Abarenkov–Animalu pseudopotential approximation were applied. It was shown that cracks do not propagate in the pure metals. Their tips become blunt, mouths broaden, and internal stresses are released owing to arising dislocations and necking. This means that mechanisms of viscous fracture come into play. In the presence of hydrogen impurity, the situation is quite different. In aluminum, hydrogen desorbs and the material retains its ductility. In α-iron, hydrogen forms Cottrell clouds around dislocations, thus suppressing their movement and generation. In addition, an increase in the hydrogen concentration in iron near the crack mouth makes the material more prone to $\alpha \longrightarrow \gamma$ phase transition. As a result, crack propagation is observed; i.e., the material embrittles. *© 2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The decisive role of microcracks or cracklike defects in the fracture of high-strength materials was revealed as early as the 1920s [1]. Further studies of fracture gave rise to a new field of research—fracture mechanics. In spite of significant (especially experimental) advances in this field, a unified mechanism for crack propagation has not been worked out. For example, in the framework of the continual approximation [2], the threshold stress of crack propagation equals or even exceeds the ultimate strength predicted by the theory (about 0.1μ, where μ is the shear modulus). The actual values are, however, lower by several orders of magnitude. Most burning is the problem of brittle fracture in metals and alloys, where stress relaxation near the crack mouth may be due to the formation of different defects. Most researchers believe that impurities are of great concern in this case, but their behavior and the actual effect are treated ambiguously. It is also suggested that brittleness is related to definite lattice types, for example, body-centered cubic (bcc) crystals. To either validate or disprove one or another model, it is necessary to consider the problem at the atomic level.

The advent of powerful computers has given the chance to attack the problem with the molecular dynamics method (MDM), although only at a qualitative level [3]. Unfortunately, earlier studies had substantial drawbacks; here, we list the most important of them. First, only two-dimensional [3] or quasi-three-dimensional lattices [4] were simulated. However, it is well known that the properties of two-dimensional materials differ essentially from those of three-dimensional materials, at least because of the fundamental impossibility of two-dimensional defects originating in the latter. Second, the simplest pairwise interaction potentials (PIPs), such as the Lennard–Jones and Morse potentials, were applied. Such an approach means that the elastic sphere approximation is used and that only close-packed structures, that is, face-centered cubic (fcc) and hexagonal close-packed (hcp), can be stable upon simulation. The simulation of other structures can lead to artifacts. Third, on the atomic scale, structure transformations are slow processes. We believe that they were not completed in earlier simulations because of the limited capabilities of old computers and nonoptimal (in terms of machine time) MDM algorithms. Also, there exists the problem of adequate choice of metal atom–nonmetal impurity PIPs [4].

There are many models devoted to the effect of impurities on the embrittlement of metals and alloys. We will list only those of hydrogen-induced embrittlement [5]. They include (1) gaseous products forming inside a metal in the course of reactions between diffusing hydrogen and impurities or dopants, (2) molecular hydrogen in discontinuities, (3) brittle products due to metal–hydrogen interaction, (4) dissolved hydrogen, (5) decomposition products from solutions unstable relative to hydrogen that are under stress for a long time, (6) dissolved hydrogen capable of diffusing, (7) hydrogen from the environment, (8) water on the sample surface, and (9) pinning of dislocations by Cottrell clouds.

In this paper, we applied the MDM to simulate uniaxial tensile fracture in pure three-dimensional specimens of fcc aluminum and bcc α-iron, as well as to trace their behavior in the presence of hydrogen. Our goal was to reveal the role of the crystal lattice and to elucidate mechanisms of hydrogen-induced embrittlement. This kind of embrittlement is known to be pronounced in α-iron but virtually absent in aluminum [6]. Here, we do not consider atomic mechanisms of crack nucleation. Any atomic mechanism is possible: dislocation pileup with a small plastic shear, intersection of twin lamellas, formation of polygonization cracks, merging of parallel slip lines, etc.

COMPUTATIONAL METHODS

Computer simulation was performed with the MDM-based MMDYN program. Atom motion was governed by a set of ordinary Newton equations, which was solved numerically by the second-order Euler method. Energy dissipation was modeled by introducing viscous forces. The algorithm was optimized in terms of computer time (for details, see [7]). Initially, crystallites formed a three-dimensional (001) film 24 atomic layers thick for α-iron and 18 atomic layers thick for aluminum. The (001) surface was assumed to be free, and the (100) and (010) surfaces were subjected to repeated boundary conditions. The crystallite dimensions along the [100] and [010] directions were equal to 20 atomic layers. Thus, a system of about 3000 atoms was involved in each computer experiment. A crack was modeled by removing several atoms from two neighboring (010) planes. In all cases, the crack depth was equal to nine atomic layers. To study the evolution of the system, the film was subjected to uniaxial tensile stresses in the [100] direction that far exceeded the critical values. The dissipated energy was taken such that the kinetic energy at all applied loads and ~300 K was kept fixed. Unfortunately, we failed to simulate hydrogen diffusion to the crack mouth, because it was too long in terms of computer time. Only the trend of this process was observed, and we managed to simulate its final stage. A sufficiently large amount of hydrogen (about 5–7%) entered into octahedral pores. Therefore, the agreement with known models and experiments is qualitative rather than quantitative. Note that there is no consensus on the valent state of hydrogen in metals. It has been proposed to be equal to 1, 0, or –1. Our calculations show that reasonable metal–hydrogen PIPs correspond only to a valence of $+1e^+$ (Fig. 1). The corresponding valences of Fe and Al are $+3e$.

PURE AND HYDROGEN-CONTAINING ALUMINUM

It is well known that aluminum is a ductile material, and hydrogen is hard to dissolve in it. If hydrogen is introduced by implantation or electrolysis, it forms bubbles mainly at grain boundaries [8]. It was also hypothesized that the oxide film could play a signifi-

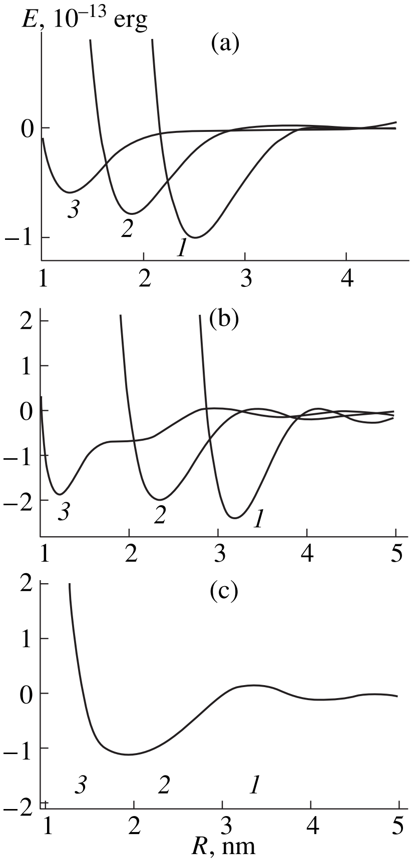


Fig. 1. PIPs in (a) fcc aluminum with hydrogen (*1,* Al–Al; *2,* Al–H; and *3,* H–H), (b) bcc iron with hydrogen (*1,* Fe–Fe; *2,* Fe–H; and *3,* H–H), and (c) pure bcc iron (Fe–Fe).

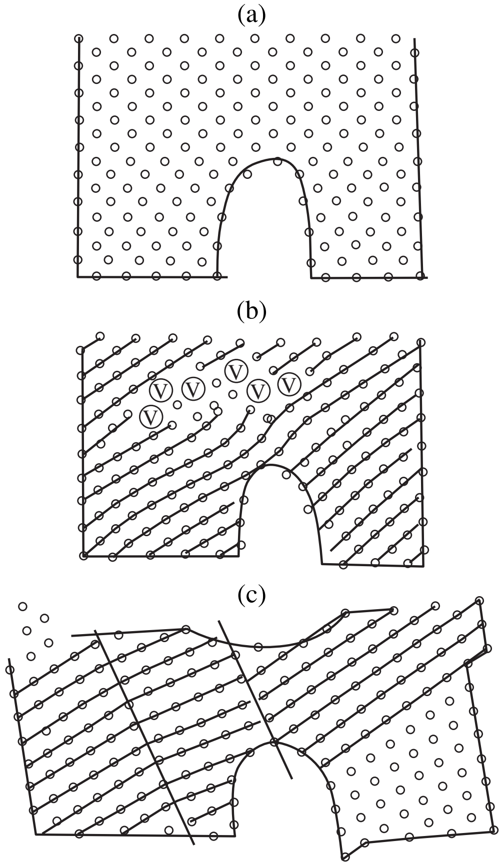


Fig. 2. Evolution of a central (001) plane in a pure aluminum crystallite: (a) initial state, no strain; (b) 10% strain; and (c) 30% strain.

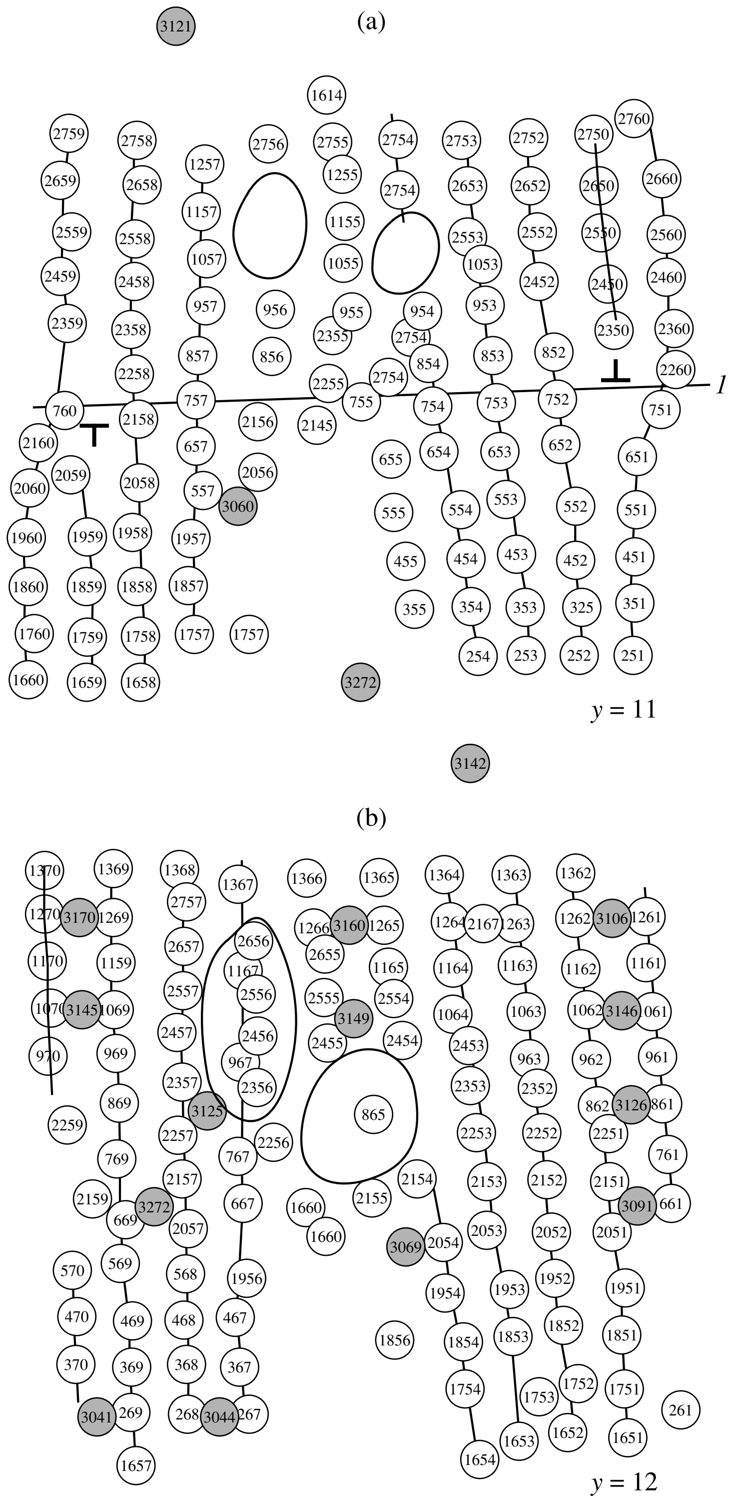


Fig. 3. Intermediate atomic structure of three neighboring planes near the crack mouth in α-Fe : H. *1*, slip line; (a)–(d), neighboring central (100) planes.

cant role. Physically reasonable Al–H PIPs were obtained only for the hydrogen valence +1*e*. They are illustrated in Fig. 1. It turned out that neither substitutional positions nor tetrahedral and octahedral pores are energetically favorable positions for a hydrogen ion.

The evolution ($\tau_{max} = 10^{-11}$ s) of a central (100) plane in Al on stretching the film to about 30% is shown in Fig. 2. At the initial stage, the relaxation is due to the generation of dislocations with various Burgers vectors. One can clearly see two partial Frank dislocations

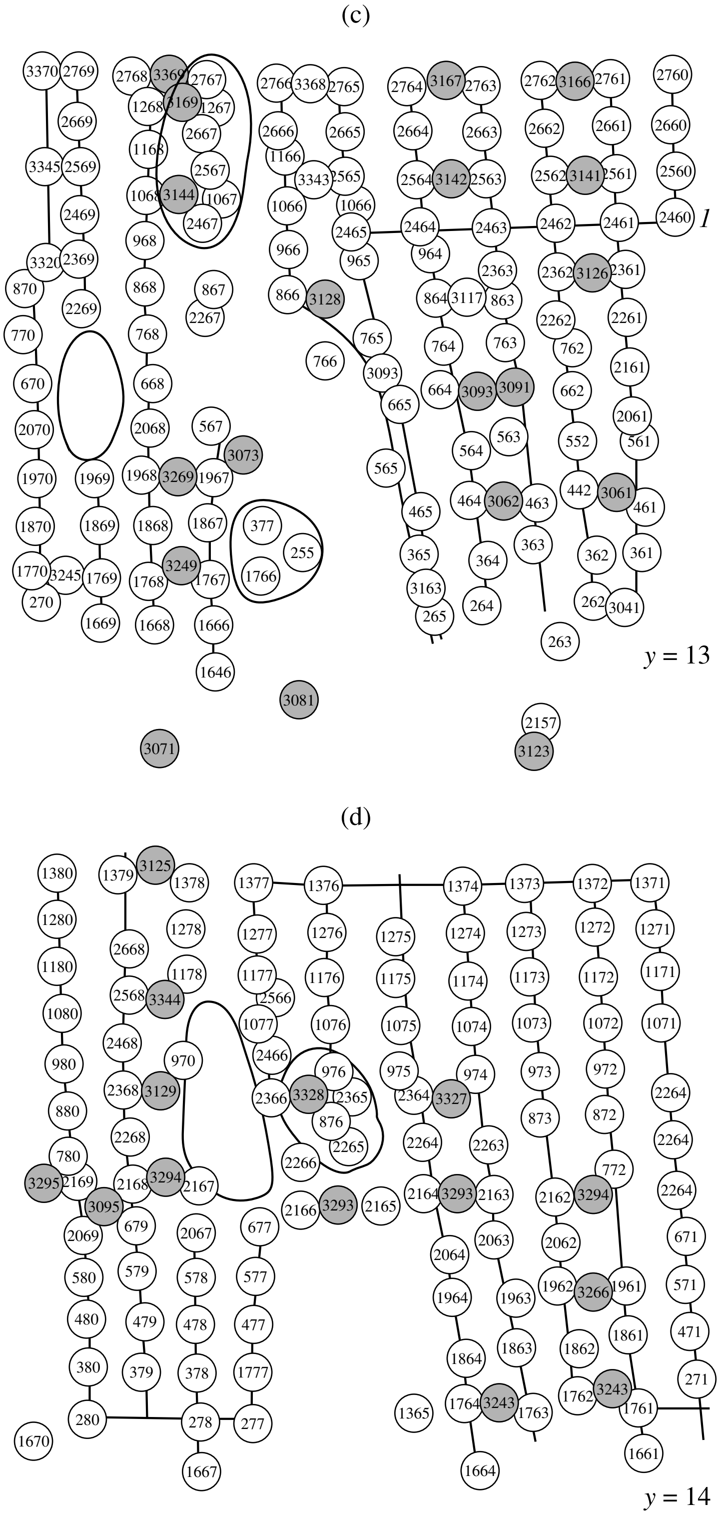


Fig. 3. (Contd.).

with a Burgers vector of *a*/3 ⟨111⟩, which form a dislocation dipole; four partial Shockley dislocations with a Burgers vector of *a*/6 ⟨112⟩, which form two dislocation dipoles; and two perfect *a*/2 ⟨110⟩ dislocations, which break on the crack surface. It is known that perfect dislocations are energetically unfavorable; here, they may arise due to the presence of a crack as indicated by their positions. The presence of these dislocations was verified by considering other planes parallel to the given one. As a result, diffusion-induced structure rearrangement took place. The (100) planes turned by an angle of about 10–15°, twins formed (Fig. 2c), and a neck appeared. In other words, this is a typical case of viscous fracture with plastic deformation at the crack mouth according to the dislocation mechanism. The crack, being blunted, broadens but does not propagate.

When hydrogen impurities are added to aluminum, the situation remains essentially the same, because hydrogen immediately leaves the metal through free boundaries. Even Cottrell clouds at dislocations are absent. It is clear that floccules actually form either at the boundaries or in the bulk of the host material. However, our technical possibilities do not allow us to simulate these processes before they are complete.

PURE AND HYDROGEN-CONTAINING IRON

In pure α-Fe, cracks also do not extend into the film. As in aluminum, the crack mouth becomes more rounded and the crack length decreases. This should naturally lead to stress relaxation [9]. The shift of some atoms from the (100) plane parallel to a crack to the neighboring plane where the crack was initially situated is shown in Fig. 3. Surface atoms in the vicinity of the crack tip are seen to move inside the crystallite. The movement is fragmentary: not entire atomic rows but their fragments move. For example, in Fig. 3, a fragment of four atoms (in view of the repeated boundary conditions) is displaced, which eventually leads to the formation of a vacancy-enriched region at the crack site. In essence, “self-healing” of the crack occurs. This disagrees with conclusions based on similar experiments [3] that brittleness is inherent in a bcc lattice, at least in two-dimensional bcc structures. The reason for this discrepancy seems to be that, first, a two-dimensional lattice was simulated in [3] and, second, that the PIPs used did not assure the stability of the lattice. The simulated behavior of pure α-Fe is, in general, the same as that of pure Al. However, hydrogen impurities change the situation drastically.

Figure 3 illustrates results of numerical simulation for a Fe + H system. It is known that hydrogen in concentrations as low as hundredths of a percent results in hydrogen embrittlement of iron. However, this process is too long compared to the simulation times (100 h or more). Hydrogen diffusion toward the crack mouth prevails in this case [8]. Therefore, we could simulate only the final stage of the process at a sufficiently high hydrogen concentration (about 7%). In this case, the crack remained sharp and was easily elongated, the neck did not form, and the crystal lattice did not rotate. However, the observed atomic mechanism of this process was not predicted earlier. It turned out that hydrogen changes the free electron density, and hence, the effect of screening iron atoms. This gives rise to the bcc-to-fcc phase transition. This transition is the most intense in the vicinity of the crack mouth. Radial distribution functions of atoms are shown in Fig. 4. One can see that, at the final stage, the system consists of three phases: residual bcc, fcc, and amorphous. Plastic deformation due to dislocations was not observed.

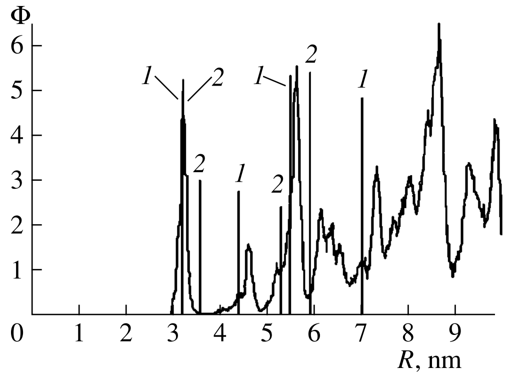


Fig. 4. Radial density distribution functions for α-Fe. Final structure at $\tau = 0.8 \times 10^{-11}$ s; (*1*) fcc and (*2*) bcc.

CONCLUSION

The numerical simulation of three-dimensional fcc and bcc systems demonstrated that interatomic forces play a decisive role in embrittlement. Pure aluminum and iron are not embrittled. In the presence of gaseous impurities, different situations may occur. If the impurities are concentrated at the crack mouth, a local phase transition accompanied by crack propagation, that is, embrittlement, is possible. After degasification, the material regains ductility.

ACKNOWLEDGMENTS

This work was supported by the program “State Support of Integration between Higher School and Fundamental Science for the Period 1997–2000,” project no. KO 422.

REFERENCES

1. O. D. Smiyan, Zh. Fiz. Khim. **54**, 2913 (1980).
2. V. I. Malkin, B. D. Kal’ner, L. G. Orlov, *et al.*, *Collection of Articles from the Central Research Institute of Ferrous Metals: Adaptive Metallurgy and Physics of Metals* (Metallurgiya, Moscow, 1980), No. 6.
3. D. De Celis, A. S. Argon, and S. Yip, J. Appl. Phys. **54**, 4864 (1983).
4. V. V. Kirsanov, T. É. Turkebaev, G. T. Temiralieva, and Yu. S. Pyatiletov, *Computer Simulation of Radiation-Induced and Other Effects in Crystals* (Ioffe Inst. Phys. Technol., Leningrad, 1989).
5. I. K. Pokhodnya, V. I. Shvachko, and V. N. Upyr’, Secondary Ion and Ion–Photon Emission, No. 2, 30 (1988).
6. B. A. Kolachev, *Hydrogen Embrittlement of Metals* (Metallurgiya, Moscow, 1985), p. 51.
7. V. M. Dyadin, V. G. Chudinov, O. G. Gondyreva, *et al.*, Available from VINITI (Moscow, 1991), No. 1537-V91.
8. A. B. Vaĭnman, R. K. Melekhov, and O. D. Smiyan, *Hydrogen Embrittlement of High-Pressure Boilers* (Naukova Dumka, Kiev, 1990).
9. A. P. Gulyaev, *Metal Science* (Metallurgiya, Moscow, 1986).

Translated by K. I. Kugel

Technical Physics, Vol. 45, No. 4, 2000, pp. 432–435. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 52–55.

SOLID-STATE ELECTRONICS

Kinetics of Breakdown Electroluminescence in Silicon Carbide *p–n* Structures

A. M. Genkin, V. K. Genkina, and L. P. Germash

Ukrainian National Technical University, Kiev, 252056 Ukraine

Received November 20, 1998; in final form, October 4, 1999

Abstract—Using a single photon statistics method providing time resolution of 0.25 ns, the effect on the breakdown electroluminescence kinetics in light-emitting diodes based on silicon carbide of particular features of the *p–n* junction structure was studied. Conditions were identified under which the relaxation times did not exceed a few tenths of a nanosecond and were outside the resolution range of the registration system. © *2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Prototype samples of light-emitting diodes (LED) in silicon carbide operating in the electrical breakdown regime, owing to their high speed of response and stability, have long been used as nano- and sub-nanosecond light-pulse generators applied in simulation of scintillator flashes and a controlling temporal parameters of photomultipliers [1]. Electroluminescence in such LEDs occurs in a wide spectrum similar to that of a black body at a temperature of 6000–8000 K. It is mainly due to radiative transitions of high-energy carriers, which acquire their energy in the strong electric field of the reverse-biased *p–n* junction. Time-of-flight of these carriers in the strong field region is below 10^{12}–10^{-13} s, and the thermalization time on the order of 10^{-10} s [2]. It can therefore be assumed that the physical limit for the LED speed of response is a quantity not exceeding 10^{-10} s. A theoretical analysis [3] has shown that the speed of response in real LEDs can be limited by the rise and relaxation times of the avalanche current, which depend on the barrier capacity and other *p–n* junction characteristics. For reliable operation of the LEDs, further experimental studies of their temporal characteristics are needed. This kind of research is apparently hampered by problems of methodological character, such as attainment of a subnanosecond resolution in combination with high enough sensitivity for registration of radiation relaxation and the formation of subnanosecond steps of the current that is passed through the structure under study.

In this study, the effect on the breakdown electroluminescence kinetics in alloyed silicon carbide *p–n* structures of distinctive structural features of the *p–n* junction is studied.

EXPERIMENTAL

LED emission relaxation measurements were performed by Ya. Yanson at the Institute of Solid-State Physics (Latvian University, Riga) with the use of techniques and equipment developed by the staff of the Institute. LED emission is registered with FÉU-165 microchannel photomultiplier. A setup for the statistical count of photons in the time domain [14] is utilized to determine the distribution in time of the emergence from the photomultiplier of one-electron pulses corresponding to the incidence of single photons on the photocathode. The event time is counted from the initiating pulse, which is locked to the LED driving pulse. In the course of every measurement cycle, the emergence time of the first three one-electron pulses is registered. The number of cycles (10^6–10^8) is chosen such as to insure low enough statistical error. Time resolution of the setup (time increments between adjacent time channels) is 0.25 ns. LEDs are driven with the use of a switch based on MIS (Metal–Insulator–Semiconductor) power transistor KP-907A (capable of switching currents of amplitude up to 2 A in less than 1 ns) controlled by a stage comprising an avalanche transistor GT 338B and an accumulating line [5]. The pulse former operates in external triggering mode, supplying to the 51-Ω load voltage pulses of nearly rectangular shape and of an amplitude of a few volts less than the final stage supply voltage. The pulse duration depends on the accumulating line length and equals 1–4 ns. The LED being measured is connected in parallel to the 51-Ω load resistor at the pulse former output. To estimate the resolution of the setup and check it for dynamic interference, the response to an extremely short emission pulse from a silicon carbide LED operating in the breakdown mode is recorded. The LED design is optimized for short pulses: it has minimum reaction and is powered by a stage incorporating an avalanche transistor and an accumulating line. The obtained normalized distribution in time of the number of one-electron pulses $N(t)$ (in the following, the measured emission pulse) is shown in Fig. 1, curve *10.* It is seen that its width does not exceed 0.5 ns at half-height and 0.7 ns at the 0.1 height level. The pulse decayed exponentially with a characteristic time of 0.15 ns over at least three orders of magnitude. The dynamic range for the emission intensity spans four orders of magnitude and is limited by photomultiplier noise. Therefore, longer processes observed in experiments on breakdown electroluminescence relaxation ought to be ascribed to the emission pulse shape rather than to the response time of the registration system.

In studies of fast relaxation processes in different spectral intervals, the possibility of spectral variation of the registration system characteristics should be considered. In this study, experimental verification of this possibility was accomplished by comparing the shape and position in time of pulses, corresponding to the same pulsed emission source, that were recorded with the use of light filters transparent in different spectral bands (other measurement conditions being maintained identical). Usually, a short delay (a few tenths of a nanosecond) was observed, with no appreciable difference in the shape of the pulse corresponding to the ultraviolet emission component compared with the red component. Possibly, the delay is due to the dependence on wavelength of the time-of-flight in the FÉU-165 multiplication system. A similar dependence with a slope of about 1 ps/nm is found in many types of photomultiplier tubes [6].

CONCLUSIONS

Alloyed *p*–*n* structures were studied. The breakdown in these structures was usually localized in microplasmas; their distribution and characteristics varied with doping level, surface treatment of the initial crystals, time and temperature regime of alloying, weight and composition of the emitter alloy, and other factors. With carefully adjusted technological regimes, structures were produced that fall into the following three groups.

Samples in group I were fabricated by alloying silumin disks with silicon carbide crystals of 6*H* and 15*R* polytypes grown by the Lely technique. Diameter of the structures was 150–300 μm. Alloying was carried out on natural crystal faces perpendicular to the crystallographic C axis. The concentration of uncompensated donors in the initial crystals was 10^{18}–10^{19} cm^{-3}. If defects were intentionally introduced in the near-surface layer of starting wafers of α-SiC, microplasmas in the structures obtained were dense and uniform over the whole structure area or localized in a ring occupying a considerable area of the *p*–*n* junction. In the optimal case, the microplasma turn-on voltages differed insignificantly. In the current–voltage characteristics, the emergence of microplasmas is seen as a kink preceding the region of steep current rise. Such samples have high long-term stability of emission power in dc mode when operated in the region of “abrupt” breakdown fairly high above the kink in the current–voltage characteristic [7]. Some characteristics of analogous samples are presented in [8]. If not influenced by defects, the obtained structures have microplasmas localized in a ring close to the *p*–*n* junction periphery.

Group II comprised samples fabricated from α-SiC crystals having fairly high doping levels (higher than 10^{19} cm^{-3}), in which defects had no influence on localization of microplasmas. The *p*–*n* junctions were alloyed on a carbon crystal face. Microplasmas in these samples were uniformly distributed over the *p*–*n* junction area, and their number increased considerably with rising current. The turn-on voltage for the microplasmas was a factor of 1.5–2 higher than the operating voltage of the samples in group I fabricated from crystals doped to the same level. At forward currents in excess of 5–10 mA, an injection electroluminescence, uniform over the structure area, was observed in group II samples. This is unusual for *p*–*n* structures in group I with uniformly distributed microplasmas, in which nonuniform injection electroluminescence occurred only when current pulses in excess of 0.5–1 A were passed. Microplasmas in group II samples seem to be caused by statistical fluctuations of the concentration of uncompensated impurity atoms in the base region of the structures.

Samples of these two groups were packaged in cases intended for low-power rf transistors. Contact to the *p*-region was accomplished by thermocompression welding of aluminum wire to a drop of emitter alloy. The emission was coupled out through the thin *p*-region. The capacity of such LEDs at zero bias was 200–400 pF, and differential resistance at currents above 0.1 A was 5–10 Ω. The LED operating voltage depended on the concentration of uncompensated donors in the starting α-SiC crystals and was equal to 15–27 V for direct current of 0.1 A.

Group III consisted of samples having small-area *p*–*n* junctions 40–50 μm in diameter [9]. The samples were packaged in coaxial cases 5 mm in diameter. The emission was coupled out through the base crystal of α-SiC. These LEDs showed high long-term stability under maximum operating currents [10]. The LED capacity at zero bias did not exceed 20 pF, and differential resistance at 0.1 A current was 10–30 Ω. The operating voltage was 18–30 V for direct current of 0.1 A. Typical pulse shapes for group I and III samples are shown in Fig. 1.

It was found that if the duration of the voltage pulse formed at the 51-Ω load was around 1 ns, a considerable share of the pulsed current in group I samples with uniform distribution of microplasmas was lost in recharging the LED capacity. This capacity is dominated by the barrier capacity of the *p*–*n* junction, which is a function of the reverse voltage (possibly of the rate of voltage variation as well). For this reason, the number of photons in the emission pulse of these LEDs is found to be considerably less than for samples of group III with small a *p*–*n* junction area. It is seen that for the above duration of the turn-on pulse and the maximum admissible operating supply voltage at the final stage of

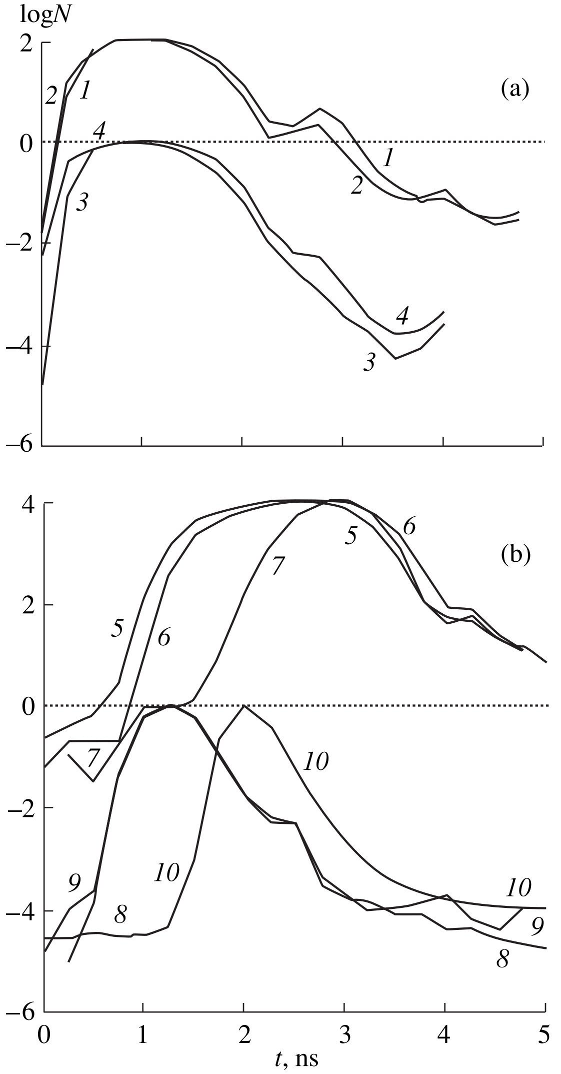


Fig. 1. Typical measured emission pulses for group I and III samples. Diameters of *p–n* junctions (a) 50 and (b) 200 μm (except curve *10*). Duration of the excitation pulse 1 (curves *1–4*, *8*, *9*) and 2 ns (curves *5–7*). Supply voltages at the pulse shaper final stage 49.5 (curves *1*, *2*, *5*, *6*, *8*, *9*), 20 (curve *3*), 30 (curve *4*), and 29.5 V (curve *7*). Light filter types: KS-14 (bandwidth > 640 nm, curves *2* and *5*); UFS-2 (bandwidth 290–380 nm, curves *6*, *9*). Curve *10*—measured pulse of minimum duration.

the pulse former, 49.5 V, typical measured pulse durations in group I samples with uniform distribution of microplasmas were in the subnanosecond range (curves *8*, *9*). The shape of these pulses varied insignificantly between the samples. Longer pulses close in shape to the driving voltage pulses and little affected by their amplitude (determined by supply voltage of the pulse shaper final stage) were observed in group III samples (curves *3*, *4*). Spectral composition of the emission had practically no effect on the pulse shapes (curves *1*, *2*, *5*, *6*, *8*, *9*).

For observation of the dependence of the emission pulse shape on the excitation level, it was necessary to increase the driving pulse duration to about 2 ns. As seen in Fig. 1b (curves *5*, *7*), at lower excitation levels, the duration of the measured pulses is shorter, which can be explained by the greater role of the capacitive current, which erodes the front of the current pulse produced at electrical breakdown of the *p–n* junction. In group III samples, this effect is practically undetectable because of considerably lower barrier capacitance of the *p–n* junction.

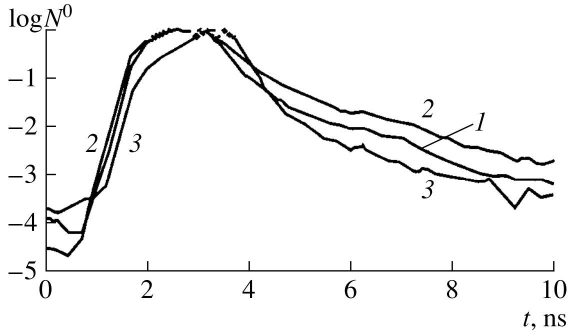


Fig. 2. Emission pulses measured on a group III sample. Curve *1*, integrated emission (without light filter); *2*, light filter KS-14; and *3*, light filter UFS-2.

In group I samples, based on α-SiC, in which microplasmas were localized in a ring at the *p–n* junction periphery, pulses similar to those in samples of group I with a uniform distribution of microplasmas were measured.

In a group II sample (curve *7*) fabricated on SiC-15*R* with the concentration of uncompensated donors of the order 10^{19} cm^{-3}, the breakdown electroluminescence relaxation time depended on the spectral position of the measured emission. In this sample, the *p–n* junction had a diameter about 300 μm. Figure 2 shows pulses for this sample (curve *7*) measured with a driving voltage pulse of 2 ns and a voltage of 49.5 V at the final stage of the pulse former. It is seen that the measured pulse, corresponding to the red spectral region (curve *2*), rises exponentially with a time constant of about 0.12 ns; has a bell-shaped upper part; decays exponentially, first with a time constant of about 0.5 ns (decreasing in magnitude by one order); and ends with a tail having a time constant of about 1.5 ns. The pulse measured in the ultraviolet region of the spectrum (curve *3*) features a peaked top corresponding to the end of the driving voltage pulse (apparently the emission process does not reach saturation). The pulse front displays a slow component with the time constant 0.7 ns; the trailing edge over about two orders of magnitude is dominated by a fast component with the time constant of 0.25 ns. The measured pulse of integrated emission (curve *1*; sensitivity of FÉU-165 is a maximum in the range 540–390 nm at half-width level) is a combination of the above components. To explain the observed highly complicated emission relaxation, further studies are needed.

It has thus been established that the alloyed *p–n* structures based on α-SiC, which belong to groups I and III, can be used in forming subnanosecond light pulses. There is no evidence to suggest that the breakdown electroluminescence relaxation times in these structures are longer than a few tenths of a nanosecond. However, high barrier capacitance of the structures 150–300 μm in diameter makes bleak the prospect of their use for generation of light pulses of duration less than 2 ns.

Relaxation times longer than 1 ns with a variable dependence on the emission spectrum range have been measured on structures belonging to group II. It appears that these structures cannot be used for generation of short light pulses.

REFERENCES

1. I. V. Ryzhikov, V. I. Rykalin, V. G. Lapshin, *et al.*, Poluprovodn. Tekh. Mikroelektron., No. 20, 87 (1975).
2. L. A. Kosyachenko, Uchebnye Zapiski Tartusskogo Universiteta, No. 655, 12 (1983).
3. N. V. Mulyukin and B. V. Pronin, Fiz. Tekh. Poluprovodn. **9**, 1065 (1975).
4. Ya. Tibergs, Yu. Veinberg, Ya. Yansons, *et al.*, *Methods and Measuring Equipment for Physical Research* [in Russian] (Riga, 1989), p. 55.
5. V. P. D’yakonov, *Circuit Engineering of Devices Based on Power Field-Effect Transistors*: *Reference Book* [in Russian] (Radio i Svyaz’, Moscow, 1994), p. 279.
6. S. S. Vetokhin, I. P. Gulakov, and A. N. Pertsev, *Single Electron Photodetectors* [in Russian] (Énergoatomizdat, Moscow, 1986), p. 161.
7. Yu. M. Altaĭskiĭ, A. M. Genkin, and V. K. Genkina, Élektron. Tekh. **190**, 76 (1987).
8. Yu. M. Altaĭskiĭ and A. M. Genkin, Zh. Tekh. Fiz. **52**, 543 (1982) [Sov. Phys. Tech. Phys. **27**, 350 (1982)].
9. Yu. M. Altaĭskiĭ, S. F. Avraamenko, and A. M. Genkin, Prib. Tekh. Éksp., No. 2, 245 (1986).
10. M. V. Belous, A. M. Genkin, and V. K. Genkina, Zh. Tekh. Fiz. **67**, 130 (1997) [Tech. Phys. **42**, 115 (1997)].

Translated by B. Kalinin

Technical Physics, Vol. 45, No. 4, 2000, pp. 436–442. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 56–62.
Original Russian Text Copyright © 2000 by Darznek, Rukin, Tsiranov.

Effect of Structure Doping Profile on the Current Switching-Off Process in Power Semiconductor Opening Switches

S. A. Darznek, S. N. Rukin, and S. N. Tsiranov
Institute of Electrophysics, Ural Division, Russian Academy of Sciences,
Komsomol'skaya ul. 34, Yekaterinburg, 620049 Russia

Received December 17, 1998

Abstract—Using a physicomathematical model, the process of current breaking in power semiconductor opening switches was investigated in p^+–p–n–n^+ structures with different doping profiles. The model takes account of the actual doping profile of a structure, diffusion and drift of current carriers in a strong electric field, recombination via deep impurities and Auger recombination, and impact ionization in a dense plasma. The calculation of the electrical circuit of an opening switch is based on solution of Kirchhoff's equations. It has been shown that in the nanosecond regime of breaking superhigh current densities with densities of the interrupted currents from a few to tens of kA/cm^2, the dominant factor in the current breaking process is the width of the p-region in the initial doping profile of a structure. An increase in the p-region width from 100 to 200 μm makes the velocity of the excess plasma front propagating in the p-region in the reverse pumping stage higher by a factor of 5–7. Higher propagation velocity of the plasma front makes the current breaking process more intensive, which is manifested in the shorter current breaking time and higher overvoltage across the opening switch. © *2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The phenomenon of nanosecond interruption of superhigh current densities or SOS-effect (semiconductor open switch) was first observed in ordinary high-voltage semiconductor diodes such as those used in alternating current rectifiers, in which the current density and pumping duration are adjusted in appropriate combinations [1–3]. Theoretical investigations of the SOS-effect [4, 5] have shown that its essential distinction from other switching mechanisms in semiconductor devices lies in the fact that the current breaking process develops not in the structure base but in its narrow heavily doped regions. During current breaking the base region is still filled with dense excess plasma, which results in a combination of high interrupted current density and nanosecond-range duration of current interruption.

Further experimental investigations of the SOS-effect have shown that the decisive factor in the current breaking process, other conditions being equal, is the initial doping profile of the semiconductor structure. It was found that an increase in the aluminum diffusion depth in the structure (larger p-region width and p–n junction depth x_p) leads to shorter current breaking times and lower power dissipation in the commutation process and, as well, in a higher overvoltage factor. This effect was used in a new semiconductor device, which became known as the SOS-diode, and represents a solid-state nanosecond-range opening switch of high-density currents intended not for rectification but for abrupt interruption of reverse currents in high-power pulse systems with inductive energy storage [6, 7]. The SOS-diodes are characterized by gigawatt levels of switched power at operating voltages in the range of hundreds of kilovolts and interrupted currents from a few to tens of kiloamperes. The current breaking time equals 5–10 ns and can be made shorter by increasing the rate of supply of reverse current to the opening switch. At reverse pumping times of 10 to 15 ns, the current breaking time is found in the subnanosecond range [8].

FORMULATION OF THE PROBLEM

Figure 1 shows a typical experimental plot of the overvoltage factor K_{ov} as a function of the p–n junction depth x_p measured in tests of various diodes under conditions of the SOS-effect. The quantity K_{ov} represents the ratio of the maximum voltage at the opening switch to the initial voltage at the pumping capacitor. Current region *1* corresponds to diodes in which the reverse voltage recovery regime is mild. Structures in the mild-recovery diodes have low values of x_p and more abrupt p–n junctions, which is achieved by introducing an epitaxial p^+-region with high acceptor concentration near the p–n junction [9]. Region *2* corresponds to conventional power rectifier diodes with stringent current breaking characteristics, in which the aluminum diffusion depth can be as great as 100–120 μm and the overvoltage in the recovery stage can rise threefold. The structures of SOS-diodes correspond to region *3* (x_p = 160–200 μm). At x_p in excess of 160 μm, the overvoltage is sixfold higher. According to the existing classification, SOS-diodes have a “superstringent” current interruption characteristic.

The purpose of this work is to explain the dependence in Fig. 1. To this end, processes involved in the dynamics of the electron–hole plasma in semiconductor structures operated in the SOS-effect mode with *p*–*n* junctions located at depths of 100–200 μm were investigated with the use of a physicomathematical model.

CALCULATION RESULTS

Dynamics of the Electron–Hole Plasma in the SOS-Diode

In calculations of the processes involved in the dynamics of the electron–hole plasma, a physicomathematical model was used which has been described in detail in [5]. The model takes account of the actual doping profile in a p^+–p–n–n^+ structure and of the following basic processes occurring in an electron–hole plasma: diffusion and drift of current carriers in strong electric fields, recombination via deep impurities and Auger recombination, and impact ionization in a dense plasma. The calculation of the electrical pumping circuit of the SOS-diode is based on the solution of Kirchhoff’s equations. Figure 2a is a schematic diagram of two-circuit pumping of the SOS-diode containing two capacitors of equal capacitance C_1 and C_2. Capacitor C_1, initially charged to voltage U_0 after the closing of key S^+, discharges onto capacitor C_2 via inductance L^+ and opening of switch SOS (key S^- being open). The equivalent capacitance value of the circuit for the direct pumping regime is $C^+ = C_1/2$. The direct pumping process in the opening switch is ended when the forward current has crossed the zero value. At this moment key S^+ opens, key S^- closes, and the reverse current starts to flow through the switch from capacitor $C^- = C_2$ via inductance L^-.

Figure 2b shows a typical SOS-diode structure with the *p*–*n* junction at a depth of $x_p = 180$ μm. In this structure, the p^+-region was formed by boron diffusion, the *p*-region by aluminum diffusion, and the *n*- and n^+-regions by phosphorus diffusion. As the *n*-base, *n*-type silicon of resistivity 50 Ω cm was used. The calculations were carried out for an SOS-diode consisting of 160 such structures in series connection of area 0.24 cm^2. With the opening switch short-circuited, the pumping circuit parameters were as follows: initial voltage at capacitor C_1, $U_0 = 32$ kV, the direct pumping time $t^+ = \pi(L^+C_1/2)^{1/2} = 400$ ns, and the reverse pumping time (equal to the discharge time of capacitor C_2) $t^- = 0.5\pi(L^-C_2)^{1/2} = 100$ ns. The calculated maximum densities of the forward and reverse currents through the opening switch were 2 and 8 kA/cm^2, respectively. The load resistance P_l was 200 Ω.

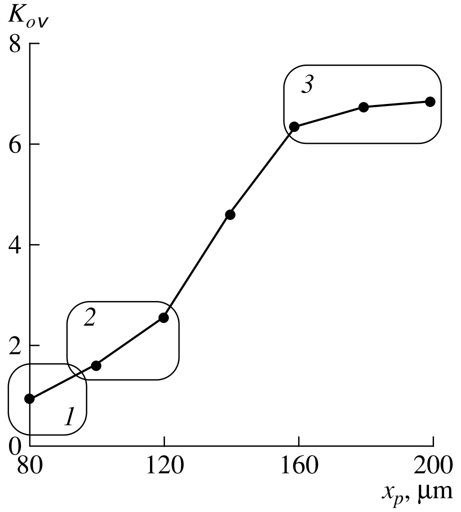


Fig. 1. The overvoltage factor as a function of the aluminum diffusion depth.

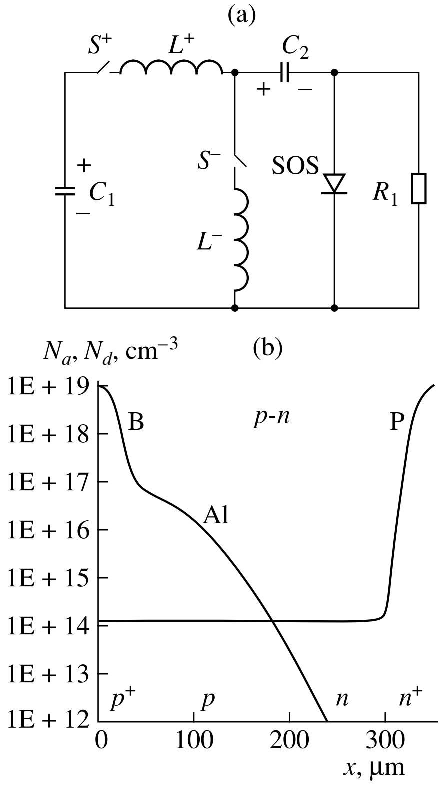


Fig. 2. (a) SOS-diode pumping circuit and (b) the doping profile of its structure.

The calculated curves for the current density $j(t)$ through the opening switch and voltage $U(t)$ are presented in Figs. 3a and 3b. Consider the dynamics of the electron–hole plasma as it progresses in the structure through the stages of direct pumping (time interval *1*–*3* in Fig. 3), reverse pumping (*3*–*4*), and the break of current (*4*–*6*). In analyzing the results of numerical calcu-

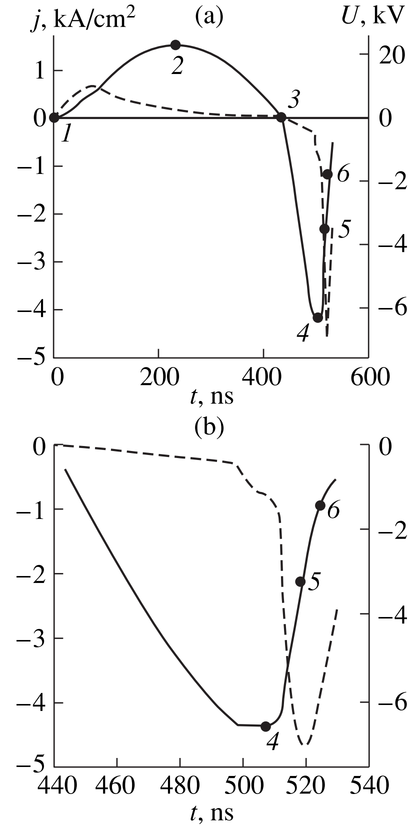


Fig. 3. Variation with time of the current density (solid curve) and voltage (dashed curve) in an SOS-diode.

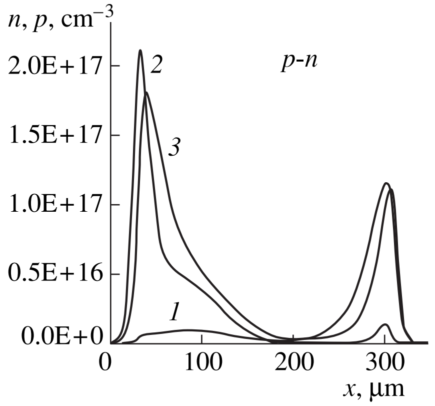


Fig. 4. Density distribution of the electron–hole plasma across the structure in the direct pumping stage: (*1*–*3*) time instances as in Fig. 3.

lations, approximate analytical expressions will be used as well.

Direct pumping. During direct pumping, double carrier injection occurs with holes penetrating into the n-region and the electrons into the p-region, where they drift under the action of an electric field. The injected carriers are compensated by the majority carriers of opposite sign, maintaining electrical quasi-neutrality, and plasma waves are generated propagating away from the p–n junction on both sides of it. Heavily doped regions drastically decelerate the plasma waves, which is revealed in characteristic density surges at the structure interfaces (Fig. 4). At forward current, the dominant contribution to the voltage drop comes from the lightly doped n-base. The rise of current proceeds at a higher rate than the filling of the base with plasma, which causes the emergence in the base of a region of strong electric field seen in the oscillograms as a burst of positive voltage (Fig. 3, time instant *1*). After the structure resistance has dropped to below the circuit characteristic wave impedance or load resistance, and until the current interruption, the opening switch operates as a generator, its current being determined by the external circuit instead of the parameters of the semiconductor structure.

Disregarding diffusion and recombination processes and also assuming that the carrier mobility is field-independent, the transport equations for the holes in the n-region and electrons in p-region can be written in the form

$$\begin{aligned}\dot{p} &= -\frac{j(t)b}{e((b+1)p+bf)^2}(fp' - pf'),\\ \dot{n} &= -\frac{j(t)b}{e((b+1)n-f)^2}(fn' - nf'),\end{aligned} \qquad (1)$$

where n, p are electron and hole concentrations, respectively; e is the electron charge; $b = \mu_n/\mu_p$ is the ratio of electron and hole mobilities; $f(x) = N_d(x) - N_a(x)$ is the difference in donor N_d and acceptor N_a concentrations at point x; $j(t)$ is the current density at a time moment t; dots and primes above a function denote its derivatives with respect to time and coordinate, respectively.

In distinction from [10], these transport equations were solved for an arbitrary distribution of donors and acceptors. Partial solutions of equation (1) have the form

$$\begin{aligned}p(x) &= \frac{b}{b+1}f\left(\left(\frac{Q(t)/b}{S_n(x)}\right)^{1/2} - 1\right),\\ n(x) &= \frac{b}{b+1}f\left(\left(\frac{Q(t)/b}{S_p(x)}\right)^{1/2} - 1\right),\end{aligned} \qquad (2)$$

where $Q(t) = \int_0^t j(t)dt$ is the charge that has passed the structure by the time moment t and

$$S_n(x) = e\int_{x_{p-n}}^{x} f(x)dx \quad \text{and} \quad S_p(x) = e\int_{x_{p-n}}^{x} f(x)dx$$

are integrals of $f(x)$ over the n- and p-regions taken from point x_{p-n} to point x of the p–n junction.

Analysis of equation (2) shows that the depth of plasma penetration into the n-region is limited by coordinate x_{nfr}

$$S_n = e\int_{x_{p-n}}^{x_{nfr}} f(x)dx = Q(t)/b$$

and into the p-region by coordinate x_{pfr}

$$S_p = e\int_{x_{p-n}}^{x_{pfr}} f(x)dx = Q(t)b. \tag{3}$$

Estimates based on relationships (3) show that the holes do not have enough time to traverse the n^+-region, whereas the electrons reach the boundary of the p^+-region and during the last 100 ns of the direct pumping stage escape via the contact. From equations (2) and (3), the charge accumulated in the n- and p-regions can be calculated:

$$\begin{aligned} Q_n(t) &= e\int_{x_{p-n}}^{x_{nfr}} p(x)dx = Q(t)/(b+1), \\ Q_p(t) &= e\int_{x_{p-n}}^{x_{pfr}} n(x)dx = Q(t)/(b+1). \end{aligned} \tag{4}$$

It follows from equations (4) that most of the charge accumulation occurs in the structure p-region due to the higher mobility of electrons. Lower plasma densities in heavily doped regions are obtained if recombination is taken into account in numerical calculations for the model. Figure 4 illustrates the lowering of the maximum density of the plasma accumulated in the p-region by the end of direct pumping (time instant *3*). Total charge loss due to recombination during direct pumping amounts to 30%.

Reverse pumping. When the current polarity is reversed, the direction of propagation of the plasma fronts changes to the opposite. Due to the fact that the ambipolar plasma velocity is inversely proportional to the density, the plasma front profiles sharpen. The minimum width of the fronts can be estimated from the condition of equality of the diffusion and drift currents at the plasma front boundary $\delta_n \approx D_n/V_{ns} = 0.04$ μm, $\delta_p \approx D_p/V_{ps} = 0.01$ μm, where D_n, D_p, V_{ns}, V_{ps} are the electron and hole diffusion coefficients and saturation velocities, respectively. Regions beyond the front boundaries remain free of injected charges, and there the current is maintained only due to majority carriers. Neglecting the diffusion terms, the condition of current flow through these regions can be written as follows:

$$\begin{aligned} j(t) &= eN_a(x)V_p(E(x)) \quad \text{for} \quad x < x_{pfr}, \\ j(t) &= eN_d(x)V_n(E(x)) \quad \text{for} \quad x > x_{nfr}, \end{aligned} \tag{5}$$

where E is the electric field and V_n and V_p are the electron and hole velocities, respectively. Because the field amplitude is a maximum where the carrier concentration is a minimum, negative field minima will appear at points near plasma fronts in the p- and n-regions, as well as in the base. With a further increase in reverse current, the plasma boundaries continue their movement along the profile towards the base until the velocities of majority carriers level off:

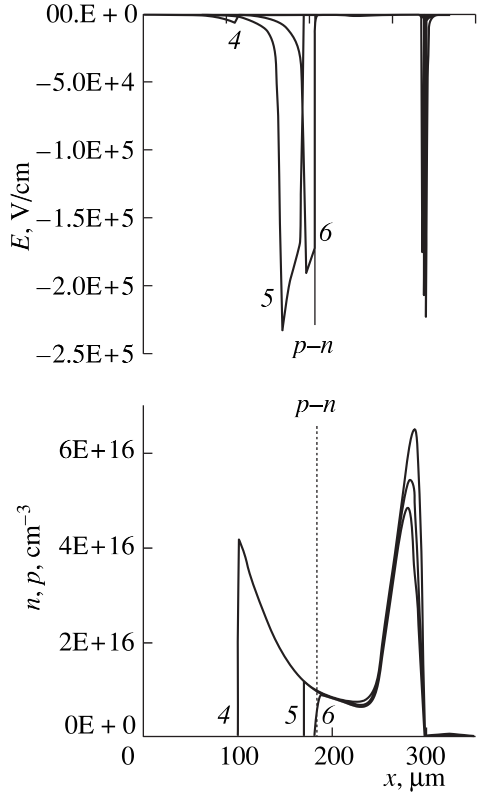


Fig. 5. Electron–hole plasma density and electric field distributions in the structure in the current breaking stage: (*4*–*6*) time moments as in Fig. 3.

$$j(t) = eN_a(x)V_{ps}, \quad j(t) = eN_d(x)V_{ns}. \tag{6}$$

The flow of the reverse current cannot then continue without generation of additional charge carriers. After conditions (6) have been fulfilled, a drastic increase in the field strength occurs beyond the plasma front leading to generation of electron–hole pairs as a result of impact ionization. Additional carriers provide the passage of current through the regions of low plasma density. Numerical calculations show that the field rise and the carrier generation at the end of the reverse pumping process begin first in the n-region and 15–20 ns later in the p-region. Emergence of the strong field in the n-region slows down the buildup of the reverse current, it becomes nearly time-independent, and a negative voltage forepulse forms at the opening switch and the load (time interval around point *4* in Fig. 3). Then, the strong field area spreads to the p-region and the current breaking process starts.

Current breaking stage. In the current breaking, stage, a characteristically strong field area with distinct boundaries (Fig. 5) emerges in the semiconductor structure *p*-region (region of avalanche breakdown). Right-hand boundary of this region coincides with the position of the plasma front, and the left-hand boundary corresponds to the point where condition (6) for current saturation is fulfilled. The breakdown field in this region is about 200 kV/cm and varies weakly at the current breaking stage. The overvoltage at the opening switch is largely determined by the breakdown region width. In the beginning of the breaking process, due to greater propagation velocity of the left-hand field boundary, the region is expanding and the voltage pulse forms at the opening switch (Fig. 3, time moments *4*, *5*). Then, over the time interval *5*–*6*, the velocities of the boundaries decrease; the velocity of the left-hand boundary becomes higher than that of the right-hand boundary, which results in a narrowing of the breakdown region and a decay of the voltage at the opening switch. Figure 6 (curves *a*, *b*) illustrates the variation of the velocity of the field boundary movement in the stages of reverse pumping and current breaking (numerical calculation).

Note that a characteristic breakdown region in the current breaking stage also exists in the right-hand part of the structure near the n–n^+-junction (Fig. 5). But it exerts practically no influence on the current breaking process, because the width of the strong field area here does not exceed a few μm.

Approximate analytical expressions for the field boundary velocities will be derived below for analysis of the current breaking process. In [11], propagation of the plasma front in a uniformly doped base was considered for fields below the breakdown threshold value. This approach is extended here to include the case of a plasma front propagating across a nonuniformly doped region of the structure and will take into account the avalanche multiplication of carriers beyond the excess plasma front. Applying the charge conservation law to the density differential, which is what the plasma front actually is, its propagation velocity can be estimated. For example, for a shift dx_{pfr} of the front in time dt in the *p*-region one can write

$$\Delta j_p dt = e dx_{pfr} \Delta(x_{pfr}), \tag{7}$$

where $\Delta j_p = j_p|_{x_{pfr}+\delta_p} - j_p|_{x_{pfr}-\delta_p}$ is the amount of change in the hole current at point x_{pfr}; $j_p|_{x_{pfr}+\delta_p} = j/(b+1)$ is the hole current in the plasma at high injection level; $j_p|_{x_{pfr}-\delta_p} = j - j_p$, the hole current in the region free of plasma; j_{pg}, the current due to holes generated under avalanche multiplication; j, the total current; and $\Delta(x_{pfr}) = p(x_{pfr}) - N_a(x_{pfr})$.

Hence, the velocity of the plasma boundary in the *p*-region is

$$V_{pfr} = \frac{dx_{pfr}}{dt} = \frac{1}{e\Delta(x_{pfr})}\left(\frac{jb}{b+1} - j_{pg}\right). \tag{8}$$

A similar formula can be derived for the velocity of plasma boundary in the *n*-region:

$$V_{nfr} = \frac{dx_{nfr}}{dt} = \frac{1}{e\Delta(x_{nfr})}\left(\frac{j}{b+1} - j_{ng}\right), \tag{9}$$

where j_{ng} is the electron current due to carriers generated by avalanche multiplication.

Analysis of the current breaking process based on the above analytical expressions revealed the following. Under well-developed avalanche multiplication, nearly the entire current in the breakdown regions is due to carrier generation. Therefore, for approximately equal saturation velocities of the electrons and holes, it can be assumed that the maximum current due to carriers generated near the plasma boundaries amounts to half the total current. Slowdown of the plasma front in the *p*-region because of the current of generated carriers causes a reduction in its velocity by a factor of 2 to 3 for $j_{pg} = j/2$, as follows from equation (8). This is illustrated in Fig. 6, which shows plots of the plasma front velocities obtained by numerical calculation (curve *a*) and calculated by equation (8) neglecting j_{pg} (curve *c*).

The plasma boundary in the *n*-region stops when $j_{ng} = j/b + 1$ (equation (9)). For $j_{ng} = j/b + 1$ and an exponential distribution of the donors, the maximum width of the breakdown region around point x_{nfr} is estimated at about 3 μm, in agreement with the numerical calculation results.

Thus, the current breaking process and the associated voltage rise at the structure are determined by the breakdown region dynamics near point x_{pfr} in the *p*-region. The width of this region is estimated as $W_p = x_{pfr} = x_{ppr}$, where x_{pr} is the point with coordinates corresponding to the left-hand boundary of the breakdown region, and defined by the velocity saturation condition (6): $j(t) = eV_{ps}N_a(x_{ppr})$. Therefore, the velocity of point x_{ppr} in the *p*-region is determined by the rate of change of the reverse current and the initial doping profile $N_a(x)$ of the structure. The velocity of point x_{pfr} belonging to the right-hand boundary of the strong field region is determined by the current magnitude and the concentration of injected carriers at the plasma front (equation (8)). At the beginning of current breaking, $x_{pfr} = x_{ppr}$; then, as long as the velocity of the right-hand boundary x_{pfr} exceeds that of the left-hand boundary x_{ppr}, a widening of the strong field region and voltage buildup take place. The maximum width of the region is achieved at the moment when the velocities of the left-hand and right-hand boundaries become equal (time instance *5* in Figs. 3, 5, 6), which is also the moment when the reverse voltage at the structure is a maximum. The subsequent drop in voltage at the structure (time instance *6*

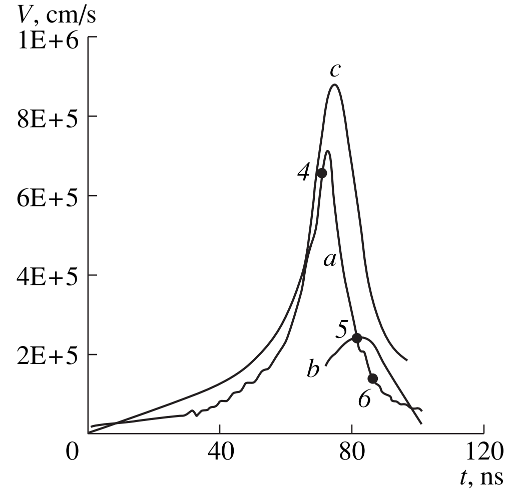


Fig. 6. Variation with time of the propagation velocity of breakdown region boundaries: (*a*) right-hand boundary with coordinate x_{pfr} (numerical calculation); (*b*) left-hand boundary with coordinate x_{ppr} (numerical calculation); (*c*) right-hand boundary, x_{pfr} by formula (8); (*4*–*6*) time instances as in Figs. 3 and 5.

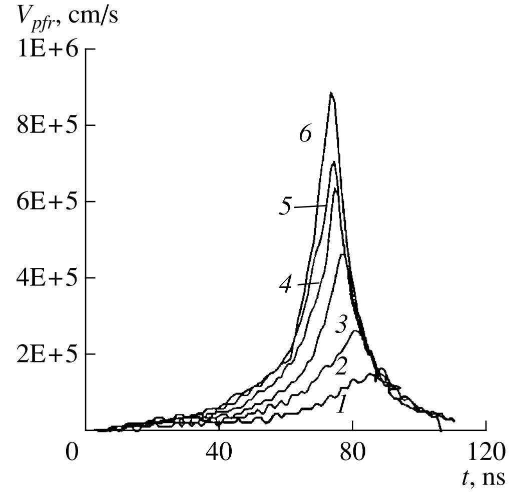


Fig. 7. Variation with time of the plasma front velocity in the *p*-region at the stages of reverse pumping and current breaking. Aluminum diffusion depth x_p, µm: (*1*) 100, (*2*) 120, (*3*) 140, (*4*) 160, (*5*) 180, (*6*) 200.

in Figs. 3, 5, 6) is caused by the narrowing of the breakdown region due to the higher velocity of point x_{ppr}.

Influence of the p-Region Width on the Current Breaking Process

To study the influence of the *p*–*n* junction width on the current breaking process by means of numerical calculations for this model, the semiconductor structures shown in Fig. 2b were employed. These structures differed in x_p values, which varied in the range from 100 to 200 µm in increments of 20 µm. Other parameters of the structures and the pumping circuit were identical. Calculation results are shown in Figs. 7 and 8.

Analysis of the processes of electron–hole plasma dynamics in the structures with different x_p values showed that with increasing aluminum diffusion depth, other conditions being equal, in the reverse pumping stage the velocity V_{pfr} of the excess plasma density front in the *p*-region rises (Fig. 7). This fact is explained by relationship (8), according to which the quantity V_{pfr} varies in proportion to current density *j* and in reverse proportion to the excess plasma density at the front $\Delta(x_{pfr})$. Increasing the diffusion depth leads to the following. First, energy dissipation in the structure in the direct pumping stage drops because of the shorter length of low-doped base. This is manifested in the smaller amplitude of the positive forepulse (Fig. 3, moment of time *1*) and larger direct current determining the charging voltage of the reverse pumping capacitor C_2. The large voltage across the capacitor C_2 causes the reverse current density to increase. Second, in the direct pumping stage, when charge accumulation takes place in the structure *p*-region, the average excess

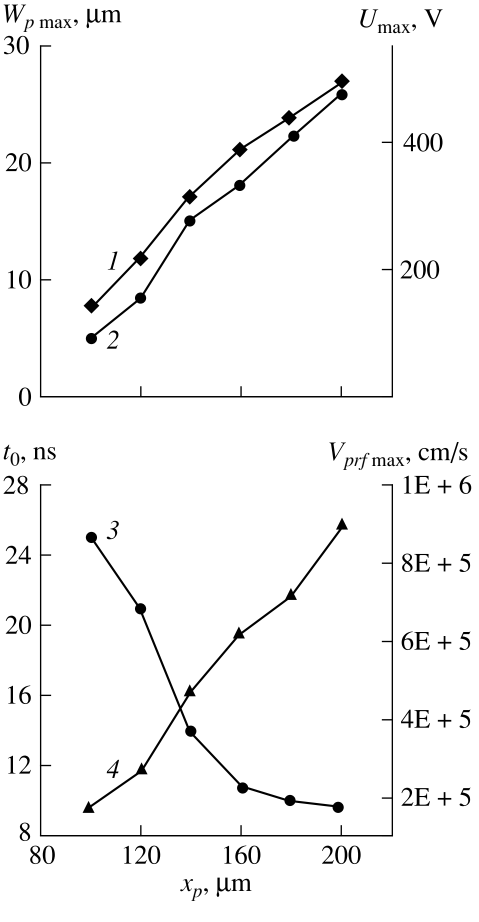


Fig. 8. Maximum voltage $U_{\max}$ across the structure (*1*), maximum width $V_{p\max}$ of the strong field region (*2*), current breaking time t_0 (*3*), and the maximum plasma front velocity $V_{pfr\max}$ (*4*) as functions of the aluminum diffusion depth x_p.

plasma density drops, because, in this case, the same accumulated charge is distributed over a thicker *p*-layer. This fact follows, in particular, from equation (2), according to which an increase in x_p corresponds to an increase in the integral $S_p(x)$ and the lowering of the electron concentration $n(x)$ in the structure *p*-region. Due to these factors, the maximum plasma front velocity $V_{pfr\max}$ reached prior to the initiation of current breaking process rises five- to sevenfold as the diffusion depth is increased from 100 to 200 μm (Fig. 8, curve *4*).

The average electric field strength in the breakdown region is practically independent of x_p and equals about 200 kV/cm. Therefore, the commutation characteristics of the structure as a current breaker are determined by the width of the strong field area and the rate of its expansion in the current breaking stage. Higher plasma front propagation velocity reached by the moment of current breaking in structures with larger x_p causes an increase in both the strong field area width and the voltage across the structure (Fig. 8, curves *1*, *2*). If x_p is increased from 100 to 200 μm, the maximum width of the strong field area increases from 5 to 26 μm and the voltage rises from 130 to 500 V per structure. With rising x_p, the current breaking time measured between 0.1–0.9 levels of the reverse current amplitude drops from 25 to 9 ns (Fig. 8, curve *3*).

CONCLUSIONS

Analysis of the dynamics of electron–hole plasma in semiconductor structures with wide *p*-regions in the ranges of the current density and the pulsed pumping duration characteristic of the SOS-effect have shown the following:

(1) Current breaking is determined by the processes taking place in a narrow, heavily doped *p*-region of the structure containing no excess plasma; its character is dynamic and not related to the extended space charge forming as a result of intensive avalanche carrier multiplication processes in a strong magnetic field. The current breaking process terminates before the excess plasma boundary could have reached the *p*-region plane. At the end of the current breaking process, the central part of the base still contains an electron–hole plasma of density two orders of magnitude higher than the base doping level. For this reason, the presence of the *p*–*n* junction has no effect on the current breaking process.

(2) The superstringent recovery regime (short current breaking time and large overvoltage) of the SOS-diode, which differs from conventional rectifier diodes in the greater extent of the *p*-region, is associated with the velocity buildup dynamics of the excess plasma density front propagating along the *p*-region in the reverse pumping stage. An increase in the aluminum diffusion depth x_p (width of *p*-region) from 100 to 200 μm causes a five- to sevenfold acceleration of the plasma front. Owing to the higher plasma front velocity, it takes less time for the strong field area to form and its width is larger, resulting in shorter current breaking times and higher overvoltages across the diode.

REFERENCES

1. Yu. A. Kotov, G. A. Mesyats, S. N. Rukin, *et al.*, in *Proceedings of IX IEEE Intern. Pulsed Power Conf., Albuquerque, 1993*, Vol. 1, p. 134.
2. S. A. Darznek, Yu. A. Kotov, G. A. Mesyats, *et al*., Dokl. Akad. Nauk **334**, 304 (1994).
3. G. A. Mesyats, S. N. Rukin, S. K. Lyubutin, *et al.*, in *Proceedings of X IEEE Intern. Pulsed Power Conf., Albuquerque, 1995*, Vol. 1, p.298.
4. S. A. Darznek, G. A. Mesyats, S. N. Rukin, *et al.*, in *Proceedings of XI Intern. Conf. on High-Power Particle Beams "BEAMS'96", Prague, Czech Republic, 1996*, Vol. 2, p. 1241.
5. S. A. Darznek, G. A. Mesyats, and S. N. Rukin, Zh. Tekh. Fiz. **67**, 64 (1997) [Tech. Phys. **42**, 1170 (1997)].
6. S. K. Lyubutin, G. A. Mesyats, S. N. Rukin, *et al.*, in *Proceedings of XI Intern. Conf. on High-Power Particle Beams "BEAMS'96", Prague, Czech Republic, 1996*, Vol. 1, p. 135.
7. S. K. Lyubutin, G. A. Mesyats, S. N. Rukin, and B. G. Slovikovskiĭ, in *Proceedings of XI IEEE Intern. Pulsed Power Conf., Baltimore, USA, 1997*, Vol. 2, p. 992.
8. S. K. Lyubutin, G. A. Mesyats, S. N. Rukin, *et al*., Dokl. Akad. Nauk **360**, 477 (1998).
9. V. A. Potapchuk and O. M. Meshkov, Élektrotekhnika **12**, 12 (1996).
10. V. M. Tuchkevich and I. V. Grekhov, *New Principles of High-Power Switching with Semiconductor Devices* (Nauka, Leningrad, 1988).
11. H. Benda and E. Spenke, Proc. IEEE **55**, 1331 (1967).

Translated by B. Kalinin

Technical Physics, Vol. 45, No. 4, 2000, pp. 443–452. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 63–72.
Original Russian Text Copyright © 2000 by Karmanenko, Semenov, Leonov, Bobyl', Dedoborets, Lunev, Nashchekin, Suris.

SOLID-STATE ELECTRONICS

Sources of Flicker Noise and the Technology of Superconducting Microstripes Based on Yttrium Barium Cuprate Films

S. F. Karmanenko,* A. A. Semenov,* V. N. Leonov, A. V. Bobyl',*** A. I. Dedoborets,*****
A. V. Lunev,* A. V. Nashchekin,*** and R. A. Suris*****

** St. Petersburg State Electrical Engineering University, St. Petersburg, 197376 Russia*

*** Vavilov State Optical Institute All-Russia Research Center, ul. Babushkina 36/1, St. Petersburg, 199034 Russia*

**** Ioffe Physicotechnical Institute, Russian Academy of Sciences, Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia*

Received April 1, 1999

Abstract—It has been shown by annealing simulations of $YBa_2Cu_3O_7$ epitaxial films that the predominant sources of flicker noise in superconducting microstripes in operating ranges of frequency and temperature are related to oxygen transitions in the vicinity of boundaries of the small-angle blocks. To make superconducting microstripes with a dimension of ~1 μm the techniques of magnetron sputtering, as well as chemical and ion beam etching, were applied. *© 2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Comparative analysis of the flicker noise level in film microstripes is usually carried out by means of the Hooge parameter (α) [1]. Epitaxial films of $YBa_2Cu_3O_7$ high-temperature superconductors have high values of the noise parameter $\alpha > 10^{-3}$ at 93 K which exceed α values for ordinary metals by 2–3 orders of magnitude [1–3]. The high values of α show that oxide superconductors have low competitiveness for cryoelectronic bolometer applications. Acceptable characteristics of the bolometric devices based on $YBa_2Cu_3O_7$ films are usually achieved only by taking advantage of the steep slope of the resistance versus temperature curve in the region of phase transition. Besides, the large values of the noise parameter are indicative of rapid degradation of the bolometer characteristics, since the high level of α indicates a considerable density of metastable defects and their intensive diffusion at low temperatures [4]. Improvement of structural perfection and reduction of defect density in the films is an important problem of superconductor bolometric device technology.

For a number of promising devices of modern cryoelectronics, it is necessary to achieve a submicron definition in film topologies [5–8]. Standard lithography processes developed for semiconductor microelectronics and based on chemical etching cannot ensure the necessary definition [6]. Straightforward adaptation of ion and laser etching techniques [9, 10] also does not produce satisfactory results because of intensive degradation of the oxygen sublattice and deterioration of the superconductive and noise parameters of the films, such as critical temperature (T_c), the width of the resistive transition (ΔT_c), the critical current density (j_c), α, and others [7, 8].

This study was aimed at improvement of the superconductor microstripe technology on the basis of model calculations and experimental investigations. At the initial stage, the nature of the flicker noise sources prevailing at the operating frequency f and temperature T was revealed. For this purpose, Monte Carlo simulations of the oxygen distribution in the basal CuO plane of $YBa_2Cu_3O_7$ were carried out. A complex approach to the problem allowed us to obtain the following results: Microstripes of size 1–3 × 1–5 μm were fabricated with submicron width definition at the structure edge of about 0.2 μm. The density of flicker noise sources or fluctuator defects in the energy range 0.2–0.6 eV was reduced approximately by a factor of 3 compared with the previous best results [6, 11, 12]. Record values of the Hooge parameter of approximately 10^{-4} at 93 K were achieved for $YBa_2Cu_3O_7$ microstripes.

MONTE CARLO SIMULATION AND NOISE SOURCES ON *f*–*T* DIAGRAM

The distribution of oxygen atoms over the two sublattices on the CuO plane was calculated by the standard Metropolis scheme of the Monte Carlo method in the framework of the so-called ANNN model [13–16], which takes into account repulsive interactions of nearest neighbor oxygen atoms and the attractive interaction of the oxygen chain atoms via orbitals of Cu atoms located between them. The annealing simulations started at a high temperature with a chaotic distribution of oxygen atoms among sublattices and ended at a low

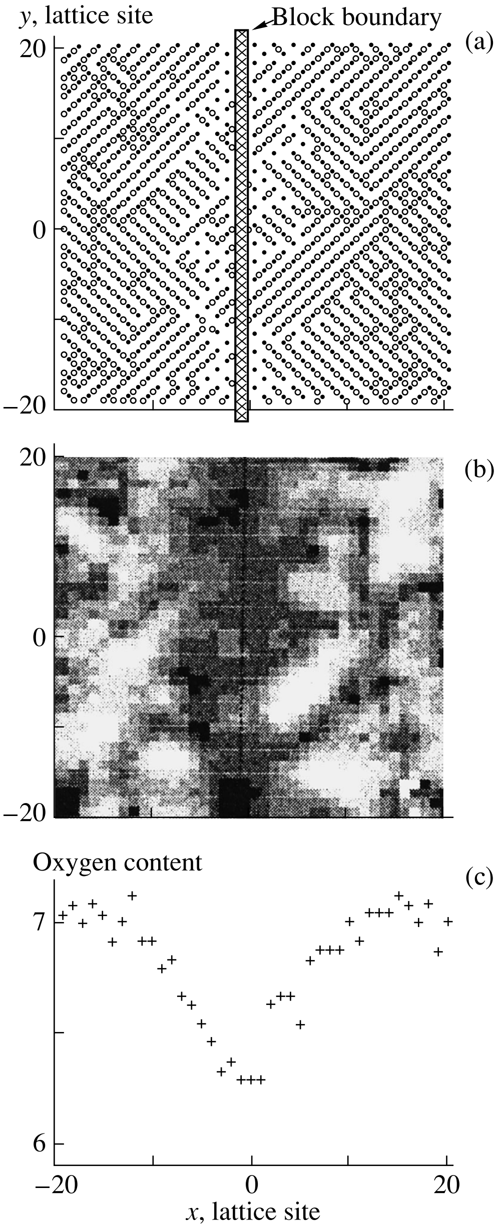


Fig. 1. (a) Monte Carlo simulation of the spatial distribution of oxygen atoms, (b) energy barriers at the CuO plane near the boundaries of the small-angle blocks in the $YBa_2Cu_3O_7$ films, and (c) the dependence of the oxygen content on the distance along the normal to a block boundary.

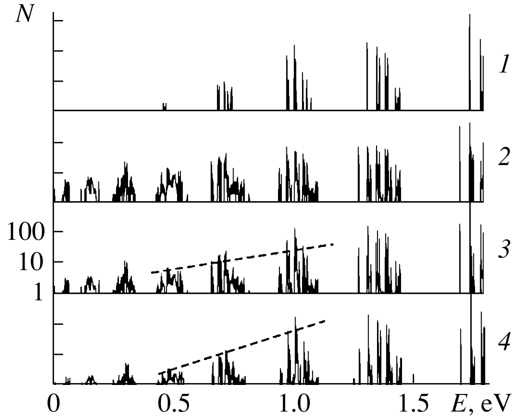


Fig. 2. Spectra of the fluctuator defects inside a block (*1*) and near the blocks (*2*), and simulation of real films with blocks of sizes 40 (*3*) and 200 nm (*4*).

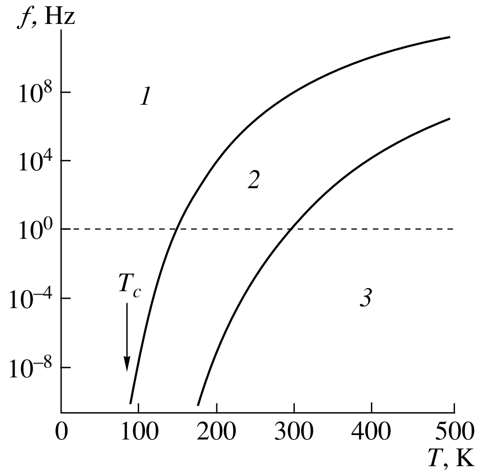


Fig. 3. Predominant sources of flicker noise on the *f*–*T* diagram. In regions *1*–*3* the fluctuator defects are oxygen atoms located near block boundaries, near defect chains, and in the ordered phase, respectively.

temperature when an ordered phase was formed. As shown in [3], epitaxial $YBa_2Cu_3O_7$ films consist of small-angle blocks and the block boundaries are the main channels for the flow of oxygen depending on post-growth regimes and storage atmosphere [4, 6, 17]. The proposed model takes into account the oxygen depletion at the block boundaries in the following way. Within a square of 44 × 44 sites, 860 atoms were distributed, which corresponded to oxygen nonstoichiometry of $\delta = 0$. As the temperature was reduced, after every four Monte Carlo steps per atom, the atoms that were found at the sites with coordinates $x = 0$ and $x = -1$, $y = -22 \div 22$, marked in Fig. 1a by the shaded band, were excluded from the calculation. The final distribution of atoms obtained after completion of the anneal, following 10^3 calculation steps and removal of a total of ~100 atoms, is shown in the same figure. The barrier height E for an atom transiting to the neighboring empty site was calculated using a harmonic potential approximation for the site well [14]. The spatial distribution of the barriers averaged over eight neighboring atoms shown in Fig. 1 corresponds to the distribution of oxygen atoms (the light and the dark circles denote oxygen and copper atoms, respectively) and the potential energy barriers at the CuO plane near the boundaries of small-angle blocks. In Fig. 1b, seven shades of grey are used to denote increments of the energy barriers ($\Delta E_b = 0.15$ eV per level), starting from the bottom (dark) level at $E_b = 0.5$ eV. A distribution of this kind within the block has a uniform or random character. The energy spectrum of the fluctuator defects shown in Fig. 2 was determined as a count of the number of barriers involved in the transition of an oxygen atom with energy E to a neighboring equilibrium site. Spectra *3* and *4* presented in Fig. 2 are obtained by appropriately combining spectra *1* and *2* taking into account the dependence of the fractions of fluctuator defects at the boundary and inside the block on the block size. The dashed lines in spectra *3* and *4* approximately correspond to the results of noise measurements presented in [3] for $YBa_2Cu_3O_{7-\delta}$ films on Si and MgO substrates which have similar block sizes. From Figs. 1 and 2, it can be seen that the oxygen atoms positioned near a block boundary, a defect chain, and those found in the ordered phase see the following energy barriers: less than 0.5; 0.5–1.4; and more than 1.4 eV, respectively. Using these values, as well as the known equation for the energy which an atom can gain by thermal activation over a time period $1/f$ at a temperature T [1],

$$E = kT\ln(2\pi f\tau_\partial), \qquad (1)$$

where τ_∂^{-1} is the Debye frequency ($\approx 10^{-13}$ s^{-1}), the regions on the f–T diagram of Fig. 3 can be designated, where the corresponding sources of flicker noise prevail. The dotted line separates the range of actual frequencies of signal registration (>1 Hz) from the frequency range of the degradation processes. In the devices operating near T_c at $f > 1$ Hz, the main sources of noise are block boundaries or other planar and topologically similar microdefects which cross the CuO plane. Among such planar defects are single dislocations and their conglomerations, as well as small-angle and twin boundaries, near which the activation energy for oxygen diffusion does not exceed 0.5 eV.

Thus, these simulations have shown that it is expedient to conduct special control of two-dimensional (planar) microdefects in the films, use technological steps which transform planar defects to surface three-dimensional or cationic defects, and maintain excess oxygen pressure during the heat treatment of $YBa_2Cu_3O_7$ films in order to suppress the flow of oxygen leaving the film near the planar defects.

GROWTH AND CHARACTERIZATION OF $YBa_2Cu_3O_7$ FILMS

The $YBa_2Cu_3O_7$ films were grown by *in situ* magnetron sputtering of a ceramic target ~90 mm in diameter [18–20]. The total pressure of the Ar + O_2 gas mixture was varied in a range p = 40–80 Pa with the ratio of partial pressures ~1 : 1. The area where the deposited film was uniform under these conditions was ~50 mm in diameter. Films produced by two standard technological processes were studied. In the first process [18], the film was grown at a rate V_g = 1.0–1.5 nm/min and the regime was not changed during the deposition.

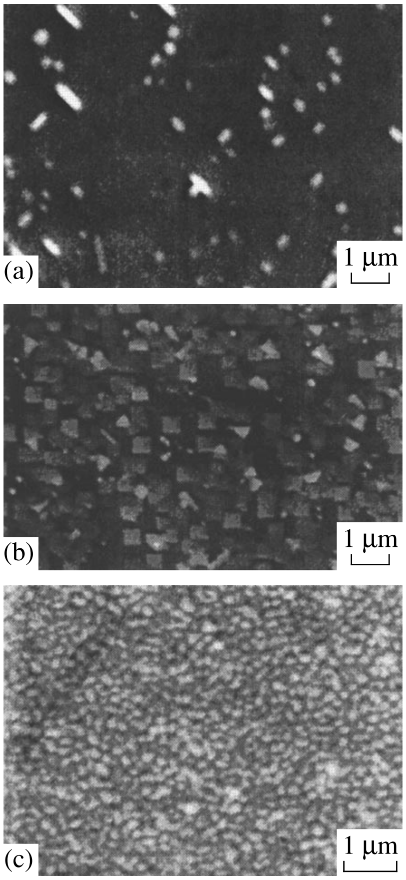


Fig. 4. The surfaces of $YBa_2Cu_3O_7$ films of different thickness grown on $LaAlO_3$ substrates.

Previous experimental and theoretical studies [3, 18–21] have demonstrated the need to reduce the growth rate, as well as to change the regimes of the initial stages of growth and heat treatment of the films. In the technological process, which is an improvement over that used in [3, 18, 19], during the initial period of deposition, the power of the magnetron discharge w was reduced to ~30% of the nominal value and then increased gradually to 4–5 W/cm^2 in 20–25 min. Over this period, the substrate temperature was lowered by 20–30 K. The substrate temperature in the established growth process was ~1000 K. The value of V_g was varied in a range 0.7–1.0 nm/min by controlling p and w. Thickness of the films was 200–1000 nm. The deposition was followed by heat treatment in an oxygen atmosphere. The temperature was reduced smoothly to ~750 K; and after a holding time of at least 20 min, the films were cooled at a constant rate for ~1 h. The heat treatment reduced the density of the planar small-angle boundaries formed in a tetra-*ortho*-structural transition [22].

As substrates in the growth of $YBa_2Cu_3O_7$, epitaxial film wafers of two types were used, which differed in

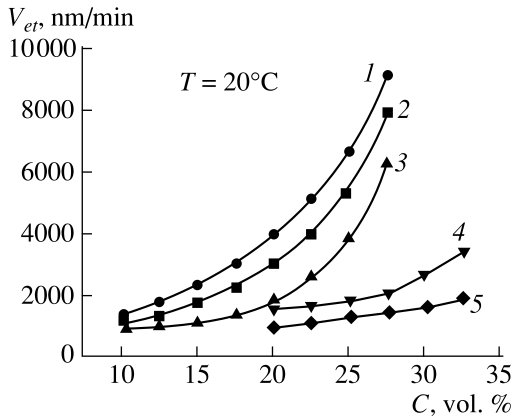


Fig. 5. Dependence of the etching rate of $YBa_2Cu_3O_7$ films on concentration of different acids and the Trilon-B compound: (*1*) HNO_3, (*2*) H_2SO_4, (*3*) H_3PO_4, (*4*) HCl, (*5*) Trilon-B. (The scale for V_{et} is greater by a factor of 30.)

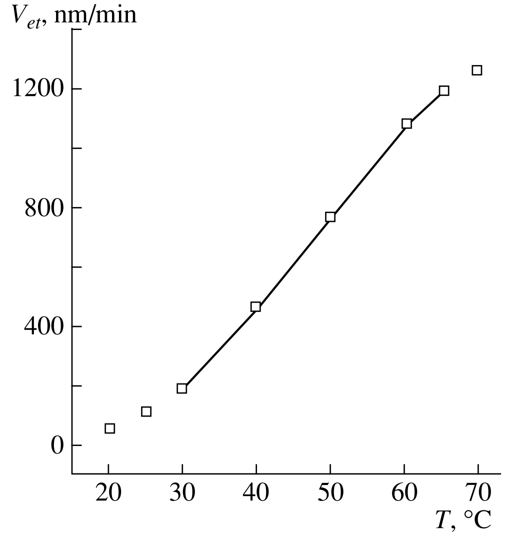


Fig. 6. The dependence of $YBa_2Cu_3O_7$ film etching rate in a 30% solution of Trilon-B on temperature.

the value of lattice mismatch parameter $\xi = 2(a_1 - a_2)/(a_1 + a_2)$, where a_1, a_2 are the lattice constants of the substrate and the film, respectively. The perovskite-type substrates of lanthanum aluminate $LaAlO_3$ and neodymium gallate $NdGaO_3$ have ξ = 1.0–2.0%; another substrate type used, namely, magnesium oxide MgO, had a much higher lattice mismatch ($\xi \sim 9\%$).

Raman spectroscopy was used for the structural analysis [18, 19]. The spectra were excited by an Ar^+ ion laser and measured using a spectrometer based on a double monochromator. The chemical composition was determined with a Camebax X-ray spectral microanalyzer [21] by measuring the intensities of the YL_a, BaL_a, CuK_a, CuL_a, and OK_a lines. The morphology of the film surfaces was monitored using a CamScan scanning electron microscope.

STRUCTURE AND MORPHOLOGY OF $YBa_2Cu_3O_7$ FILMS

It is known that the $c_\perp$-oriented structure is a metastable phase of $YBa_2Cu_3O_7$ [20, 22]. The formation of blocks with alternative $c_\perp$-orientation results in the formation of block boundaries, interference with oriented film growth, and deterioration of film characteristics. Low density of the flows of component atoms and increased temperature enhance surface diffusion and prevent the formation of nuclei of a competitive phase.

On the lattice-matched perovskite substrates, layer-by-layer growth (Frank–van der Merwe mechanism) is observed after the film becomes three $c_\perp$-monolayers thick [23]. On the MgO substrates, three-dimensional nuclei, spiral growth of islands, and their subsequent coalescence take place, with full surface coverage occurring at a thickness exceeding 10 $c_\perp$-monolayers [24]. Cationic and planar defects can stabilize the superconducting orthophase on substrates with low ξ. Thermodynamic calculations have shown [26] that the utilized deposition regimes do not include conditions for the equilibrium phase formation (123), but stabilization mechanisms of the perovskite compound are possible. An analysis of equilibrium in the $YO_{1.5}$–CuO–BaO system with the use of geometrical thermodynamics methods [25] gave an estimate of the Gibbs potential ΔG of the phase (123) that should be reduced with non-equilibrium defects by at least 50 kJ/g/atom, i.e., by 20–25%. One of the possible mechanisms of cationic defect formation is emergence of the defect-rich orthophase, supposedly of $\langle 235 \rangle$ composition [26]. This supposition is based on studies of the cationic composition of $YBa_2Cu_3O_7/LaAlO_3$ structures, which demonstrated that films with the highest parameters had insignificant deficits of barium and copper. This conclusion is in agreement with studies of $YBa_2Cu_3O_7/LaAlO_3$ structures in [27]. The deficit of barium and copper in $YBa_2Cu_3O_7$ films can be due to an on-axis deposition regime at a working gas pressure of 30–40 Pa. Substrates located opposite the target erosion zone are subjected to the glow-discharge plasma, which produces a slight deficit of the component most prone to reevaporation, i.e., barium [28]. Excess yttrium atoms can occupy barium sites and form substitutional defects of Y_{Ba} type, which contribute to stabilization of the $c_\perp$-oriented phase.

In Fig. 4, surfaces of the $YBa_2Cu_3O_7$ films of different thickness grown on $LaAlO_3$ substrates are shown. Figure 4a shows the fairly smooth morphology displayed by a film of ~600 nm thickness. Single light-colored crystallites (supposedly of $BaCuO_2$) appear on the surface of a $YBa_2Cu_3O_7$ film having a thickness of ~500 nm. It is known that the lanthanum aluminate substrates have high densities of twin and small-angle boundaries, which can move into the film during epitaxial growth [29]. The formation of small-angle boundaries results in a reduction of ΔG in the film and stabilization of the structurally perfect $YBa_2Cu_3O_7$ phase. It is seen from Fig. 4a that crystallites of the competitive phase are arranged on the film surface in rows along some lines, which are probably small-angle boundaries. Fixing of the competitive phase nuclei is an energetically favorable process, since the free surface energy along the line of imperfections is lower than elsewhere on the surface. The surface of a film of ~1200 nm thickness is shown in Fig. 4b. Some crystallites of a regular shape that were grown on a fairly smooth surface transform into pyramids composed of blocks during subsequent growth. The surface of a $YBa_2Cu_3O_7$ film of large thickness (about ~1.6 μm), representing an intergrowth of pyramidal microblocks, is shown in Fig. 4c.

Probably, the layer-by-layer growth mechanism of the $YBa_2Cu_3O_7$ films on $LaAlO_3$ substrates in the regime considered previously persists up to a certain thickness. As the film thickness reaches ~500 nm, crystallites of different composition can come to the surface and then, as the growth of the film proceeds, they are displaced to the surface and grow in size. The crystallites of the competitive phase represent a drain for excess cations. At a thickness of about 1 μm, screw dislocations come to the surface, which is revealed in the formation of isolated crystallites of regular shape and identical orientation. In this stage, layer-by-layer growth is superseded by growth of isolated crystallites having the shape of columns or pyramids around the dislocation axis. From Fig. 4c, the size of a single microblock can be estimated as 250–300 nm; i.e., the density of microblocks is $1–2 \times 10^9$ cm^{-2}, which is in agreement with the nucleation density of the $c_\perp$-oriented phase determined in [30].

PREPARATION OF MICROSTRIPES: PHOTOLITHOGRAPHY AND CHEMICAL ETCHING OF $YBa_2Cu_3O_7$ FILMS

Deposition of contact pads onto $YBa_2Cu_3O_7$ films was carried out through special masks by thermal evaporation in vacuum at pressure ~10^{-5} torr and sample temperature ~450 K. The contact layers were of silver coated with a thin layer of gold. The total thickness of the contact layer was ~500 nm. The films with contacts were subjected to a sizing treatment. The most critical stage of the process is sample etching through a photoresist mask.

Chemical etching is determined by the choice of etching solution, which should ensure precise reproduction of the topological pattern and conservation of the film parameters. As a chemical etch for $YBa_2Cu_3O_7$ films, solutions of acids such as sulfuric, hydrochloric, nitric, phosphoric, and acetic are usually used, which at low concentrations (less than 5 vol. %) show high etching rates (V_{et}) [9, 31, 32]. Linear dependence of V_{et} on solution concentration is characteristic of acid solutions with pH < 7. Etching in acid solutions has poor reproducibility; therefore, the possibility of etching with Trilon-B compound, which is a solution of disodium salt of ethylenediamine tetra-acetic acid (EDTA) was explored in this study. Trilon has an almost linear dependence of V_{et} on concentration, characteristic of base solutions with pH > 10 [9, 32].

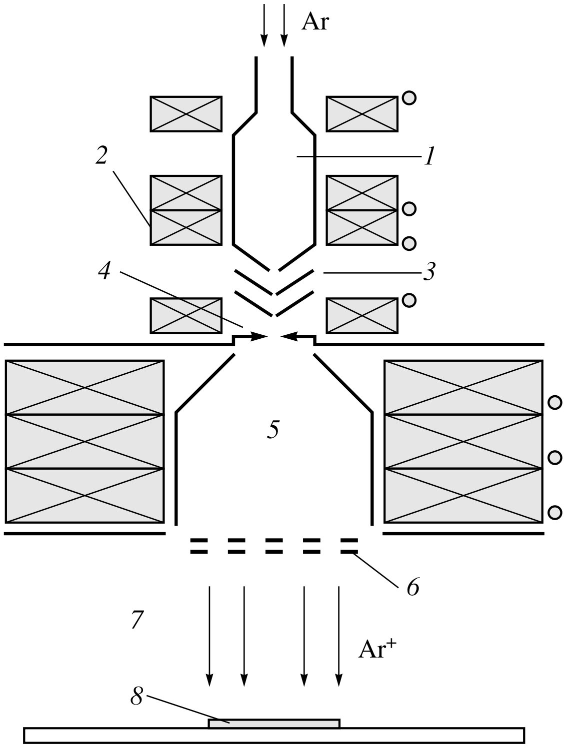


Fig. 7. The duo PIGatron source of ionized Ar atoms used for etching $YBa_2Cu_3O_7$ films: (*1*) ionization chamber; (*2*) magnets; (*3*) focal electrodes; (*4*) carbon diaphragm; (*5*) emission chamber; (*6*) emission electrodes; (*7*) vacuum chamber; (*8*) the sample.

The structures with ongrown films were coated with a layer of FP-383 or AZ photoresist. After drying in a thermostat at ~360 K, there followed exposure and development of the photoresist in KOH solution with

Table 1. The etching regimes of the $YBa_2Cu_3O_7$ films in a 30% solution of Trilon-B

Regime			Etching uncertainty, %
no.	*T*, °C	*t*, s	
1	20	150	21
2	45	17	–10
3	50	15	0
4	50	25	5
5	55	15	12
6	60	20	15
7	70	10	22

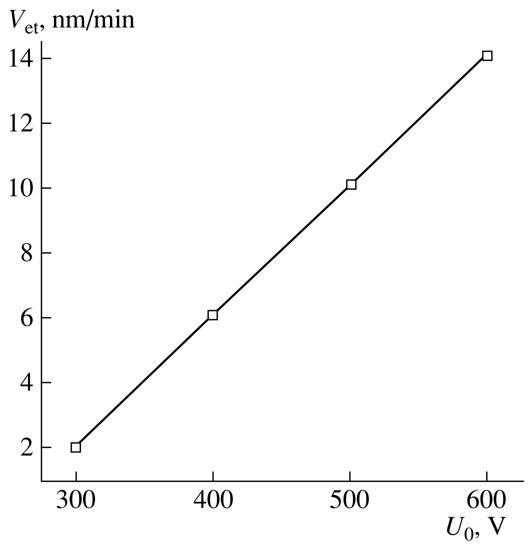


Fig. 8. The dependence of ion etching rate of $YBa_2Cu_3O_7$ films on accelerating voltage U_0.

subsequent rinsing in deionized water. During the succeeding photoresist hardening the temperature did not exceed 410 K. After preparation and heating of the etching solution, the film structure was dipped into the solution for a certain time period and rinsed in flowing water. The dependence of the etching rate of $YBa_2Cu_3O_7$ films on solution concentration is presented in Fig. 5. The V_{et} values for the Trilon-B solution are considerably lower than for acid solutions. The dependence of V_{et} for an aqueous Trilon-B solution of 30% concentration on temperature is shown in Fig. 6. The extent of film etching at the photoresist edge (etching uncertainty) was determined using a ratio $z = (w_0 - w_1)w_0$, where w_0 and w_1 are the stripe widths on the stencil and on the etched film, respectively. Experimental results are presented in Table 1, from which it can be concluded that the best results can be obtained with regime 3.

For technological experiments, samples with different T_c and structural perfection were selected. Control of the parameters of high quality films (i.e., with T_c close to 90 K and a smooth surface) during the steps of photoresist hardening, development, and the opening of the topological pattern did not show any effect on T_c. The films with a grained structure had T_c = 80–85 K and were sensitive to sizing treatment operations. Results of the measurements of parameters of the $YBa_2Cu_3O_7$ films for different etching regimes are presented in Table 2. Compared with the etching in acid solutions, the treatment in Trilon-B affects the film parameters less. The etching uncertainty after treatment in the Trilon-B solution is 0–6%, whereas for the H_3PO_4 solution it is 12–27%. It has thus been found that for $YBa_2Cu_3O_7$ films, chemical etching in the Trilon-B solution under regime 3 described in Table 2 is preferable.

After completion of the steps of micropatterning of film topology and removal of the remaining photoresist mask from the surface, the microstripes were subjected to a heat treatment in an oxygen atmosphere similar to that carried out after film deposition. The heat treatment eliminated defects induced in the film during the sizing treatment and reduced the ohmic resistance of the contact pads. Acceptable definition and insignificant impairment during the sizing treatment process indicate that chemical etching is a promising technique for fabrication of superconductor integrated devices.

ION ETCHING OF $YBa_2Cu_3O_7$ FILMS

The process of ion etching was carried out using a VUP-5 setup where the specially made duoPIGatron ion source [5, 33] was installed. The vacuum chamber was pumped out to a pressure 10^{-6} torr, and during the etching the pressure did not rise higher than 10^{-3} torr.

Table 2. Electrical parameters of the $YBa_2Cu_3O_7$ films after chemical etching

Sample number	Etching regime	T_{c0}, K		ΔT_c, K		$r = R_{300}/R_0$	
		before PL	after PL	before PL	after PL	before PL	after PL
1.1	HCl (20%), 20°C	88.1	<77	6.2	–	1.6	1.3
1.2		88.7	88.6	2.1	2.7	2.7	2.5
1.3		90.7	91.2	1.0	0.7	2.9	2.9
2.1	H_3PO_4 (10%), 20°C	87.1	85.7	2.2	3.1	2.1	1.9
2.2		89.2	89.0	1.5	1.6	2.5	2.5
3.1	Trilon-B, 50°C	89.1	89.0	2.3	2.5	2.7	2.5
4.1	Trilon-B, 60°C	86.7	86.4	4.0	5.1	2.3	2.0
		90.2	90.5	0.7	0.6	2.7	2.8

Note: T_{c0} values correspond to zero sample resistance; ΔT_c is the width of the resistive transition; $r = R_{300}/R_0$ is the ratio of the resistance of the sample at 300 K to that near the beginning of the transition.

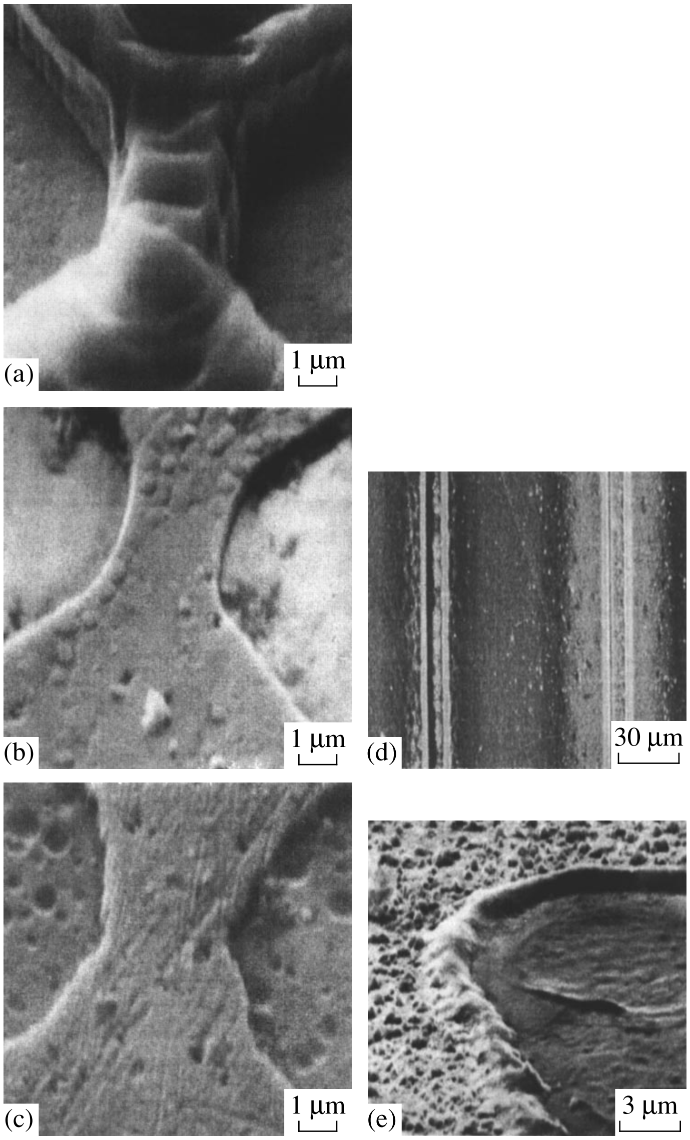


Fig. 9. Scanning electron microscope images of the structures: (a) ion etching before photoresist removal, (b) after photoresist removal, (c) chemical etching in the Trilon-B solution (regime 3 in Table 1), and (d), (e) structures after laser etching under different magnifications.

The ion source formed a neutralized beam of argon ions of diameter 35-mm and homogeneity ~95%. The ion energy was varied in a range of 50–1500 eV at current densities below 1 mA/cm^2.

The duoPIGatron sputtering device shown schematically in Fig. 7 is a modification of a duoPIGatron ion source in which dense low-temperature plasma is produced by gas discharge plasma emitters which have large surfaces of ion emission. Design of the device allows us to obtain ion emission currents of tens of milliamperes per 1 cm^2 and to form a point source with low flow rate of the cool gas from the discharge chamber to the accelerating gap. To create a beam of accelerated particles with wide aperture and high density, it is necessary to use a plasma extender held under the potential of the anode and anticathode and maintaining the regime of discharge with oscillating electrons. The gas discharge plasma of the duoPIGatron consists of two regions: cathode plasma and anode plasma separated by a double layer. Application of a longitudinal magnetic field results in nonhomogeneity of the radial distribution of $j+$ and the appearance of a radial electric field in the plasma. Introduction, alongside the longitudinal field, of a transverse multipolar magnetic field at the periphery of the anode region reduces losses of the plasma ions and the electrons and considerably improves homogeneity of the $j+$ distribution over the emitting surface. The transverse magnetic field can be created by a permanent magnets of alternating polarity arranged along the element of the cylindrical anode.

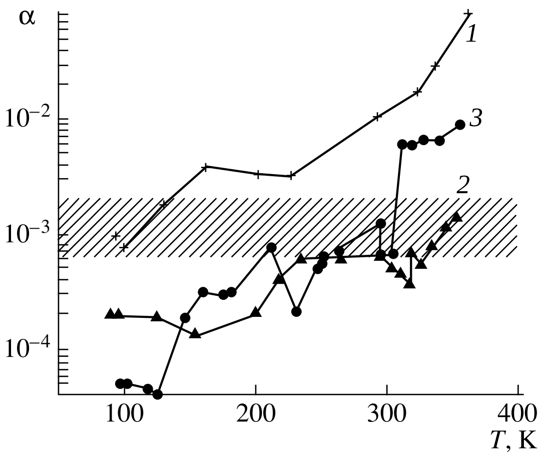


Fig. 10. The dependence of the Hooge parameter for $YBa_2Cu_3O_7$ microstripes of 2.3–3.0 μm width on temperature: (*1*) and (*2*) samples obtained by chemical etching in the Trilon-B solution before and after improvement of microstripe preparation technology, respectively; (*3*) samples obtained by ion etching. The shaded area shows the best Hooge parameter values achieved previously.

The field at the anode surface is about 0.1 T and rapidly decays toward the axis of the discharge chamber.

In order to control the etching process duration, so-called witness samples of $YBa_2Cu_3O_7$, films with deposited contact pads screened from the ion beam were used. The film resistance was measured during etching. The process was terminated when the resistance increased sharply as a result of complete removal of the film. The dependence of the ion etching rate of $YBa_2Cu_3O_7$ films on the accelerating voltage U_0 in the range of 300–600 eV (at the ion current density $j \approx$ 0.3 mA/cm^2) is shown in Fig. 8. The temperature of the film surface during the etch did not exceed 350 K.

The sample dimensions and state of the surface were determined using a CamScan scanning electron microscope having a resolution better than 0.1 μm. Figure 9 shows different microstripes produced by different etching techniques. It can be seen from the images that chemical and ion etching ensure a high definition of 0.2 μm or better as determined by the width of the topology edge. Microstripes produced by ion etching have smoother edge and better definition, but it should be taken into account that chemical etching does not involve irradiation producing additional defects.

LASER ETCHING OF $YBa_2Cu_3O_7$ FILMS

In laser etching an ME-551B, stencil retouching unit and an LGI-505 nitrogen laser were used. The laser parameters were wavelength 0.37 μm, pulse duration 7 ns, power density on the sample up to 4 J/cm^2, beam diameter on the sample 5 μm, repetition rate up to 1 kHz, and the beam displacement between pulses 1 μm.

Figure 9e shows a film edge after the laser etching process. An insulator ridge about 3 μm wide is seen, which is formed by the melt ejected over the laser spot edge. According to X-ray spectral microanalysis data, the ridge contains excess Y and Ba atoms (less than 10–15%). Probably the size of this region is determined by the intense heating depth of the film edge (about 3 μm), which is necessary for adhesion of the ejected melt [7, 34]. The drops at the surface indicate that the melt is displaced to the film edge by the laser plasma jet. Since, at the specified laser wavelength, the absorption depth is ~0.1 μm [7] and the film thickness is 200–500 nm, all the energy is absorbed in the $YBa_2Cu_3O_7$ film, and after the first several pulses the film is completely removed. However, the observed image nonhomogeneity over the treated surface shows the need to increase the number of laser pulses, which will result in additional thermal load and probably excess flicker noise.

In Fig. 9d, light-colored (compared with the starting image) stripes about 30 μm wide are seen, which are close to the film edges and result from surface damage. Two lines in this image correspond to different laser power levels *U*: 165 W (left line) and 220 W (right line). Improvements in the laser treatment process relating to modernization of optical equipment and ensuring flexible adjustment of laser beam power, allowed a reduction in the damaged area.

Thus, the laser etching technique is inferior to the chemical or ion etching. However, this technique is very convenient for forming contact pads and correcting the topology of superconductor structures more than 100 μm in size.

NOISE CHARACTERIZATION OF $YBa_2Cu_3O_7$ MICROSTRIPES

Comparative analysis of noise characteristics of the samples obtained by chemical and ion etching, as well as elucidation of the nature of flicker noise sources, was carried out in the study.

The microstripes were mounted in a special holder inside the vacuum section of a priming metal cryostat. The size of the microstripes allowed measurements of the spectral density of the noise voltage S_V in a wide range $S_V = 10^{-19}$–10^{-8} V^2 Hz^{-1} at frequencies $f = 1$–10^5 Hz [11]. The Hooge parameter was calculated using the equation

$$\alpha(f, T) = N_e f S_V(f, t) V^{-2}, \qquad (2)$$

where N_e is a number of charge carriers in a sample and V is a sample volume.

The results of measurements and corresponding calculations using equation (2) are presented in Fig. 10. The noise level in the samples obtained by the improved technique dropped considerably, as can be concluded from the comparison of curves *1* and *2*. The comparison of curves *2* and *3* shows an increased noise level at a high temperature (more than 300 K) in samples treated by ion etching compared with chemical etching. At the same time, the normal phase noise at temperatures near 93 K in strips produced by ion etching drops to very low values ($\alpha < 10^{-4}$), which is evidence of suppression of the noise sources created in the film during ion treatment. Probably, the greatest damage is produced in oxygen chains of the CuO plane crossing the planar defects. The relationship between the noise levels observed at different temperatures shows that at low temperatures, the oxygen defects in the CuO plane located near boundaries of the small-angle blocks are stabilized to a degree.

CONCLUSIONS

The improvement in the growth technique and sizing treatment of the $YBa_2Cu_3O_7$ films has allowed a considerable reduction in the flicker noise level in the films. The simulation of annealing of the films has shown that the prevailing sources of the flicker noise in epitaxial structures are related to the oxygen transitions near small-angle and twin boundaries, which are the predominant planar defects in the films. Probably, the planar defects are transformed into three-dimensional defects at the film surface or into cationic defects. Substitutional defects of the $YBa_2Cu_3O_7$ type reduce the thermodynamic potential of the superconductive phase, bring about structural stabilization, and reduce the noise parameter of the films. The main factors in the improvement of the growth technique of the $YBa_2Cu_3O_7$ films by magnetron sputtering are the reduction of the growth rate (to below ~0.8 nm/min) and the application of the "smooth" approach to the assigned technological regime.

Defects are induced in the film topologies during sizing treatment as well. To produce superconducting microstripes of about 1 μm in size, chemical and ion etching of $YBa_2Cu_3O_7$ films has been applied. To minimize damage to the films and the size uncertainty, it is preferable to use chemical etching in solutions of Trilon-B, which is the salt of EDTA organic acid. At the same time, only etching with a low-energy argon beam from a duoPIGatron source has ensured submicron definition (~0.2 μm) of the topology edge in $YBa_2Cu_3O_7$ films without degrading the high parameters of the superconducting state of the films. The developed technique of micropatterning has allowed considerable reduction in the noise level in superconducting microstripes and to obtain a record value, 10^{-4}, of the Hooge parameter at a temperature of 93 K and frequency of 10 Hz.

ACKNOWLEDGMENTS

Authors acknowledge assistance in the work and helpful discussions with I.A. Khrebtov, the help with chemical etching experiments of T.I. Anokhina, and morphology studies of the films made by M.E. Gaevskiĭ.

This work was supported by the Russian State Program "Superconductivity" (Projects nos. 98031 and 98055).

REFERENCES

1. Sh. Kogan, *Electronic Noise and Fluctuations in Solids* (Cambridge University Press, Cambridge, 1996).
2. J. Pelz and J. Clarke, Phys. Rev. B **36**, 4479 (1987).
3. A. V. Bobyl, M. E. Gaevskiĭ, S. F. Karmanenko, *et al.*, J. Appl. Phys. **82**, 1274 (1997).
4. S. J. Rothman, J. L. Routbort, U. Welp, *et al*., Phys. Rev. B **44**, 2326 (1991).
5. S. F. Karmanenko, V. T. Barchenko, A. I. Dedyk, *et al*., Supercond. Sci. Technol. **11**, 284 (1998).
6. A. V. Bobyl, M. E. Gaevskiĭ, S. G. Konnikov, *et al*., Scanning Microsc. **10**, 679 (1996).
7. W. Eidelloth, W. J. Gallagher, and R. P. Robertazzi, Appl. Phys. Lett. **59**, 1257 (1991).
8. R. Barth, B. Spangenberg, and W. Langheinrich, J. Vac. Sci. Technol. A **10**, 3411 (1992).
9. W. Moro, *Microlithography*, Part 2 [in Russian] (Mir, Moscow, 1990).
10. *VLSI Electronics Microstructure Science. Plasma Processing for VLSI*, Ed. by N. G. Einspruch and D. M. Brown (Academic, Orlando, 1984), Vol. 8.
11. I. A. Khrebtov, V. N. Leonov, S. F. Karmanenko, *et al.*, Sverkhprovodimost: Fiz., Khim., Tekh. **6**, 786 (1993).
12. M. Fardmanesh, A. Rothwarf, and K. J. Scoles, J. Appl. Phys. **79**, 2006 (1996).
13. D. Fontaine, L. T. Wille, and S. C. Moss, Phys. Rev. B **36**, 5709 (1987).
14. A. G. Khachaturyan and J. W. Morris, Phys. Rev. Lett. **61**, 215 (1988).
15. M. Goldman, C. P. Burmester, L. T. Wille, and R. Gronsky, Phys. Rev. B **50**, 1337 (1994).
16. A. A. Berezkin, A. V. Bobyl', A. I. Dedoborets, *et al*., Fiz. Tverd. Tela (S.-Peterburg) **41**, 957 (1999) [Phys. Solid State **41**, 870 (1999)].
17. D. V. Kulikov, R. A. Suris, and Yu. V. Trushin, Fiz. Tverd. Tela **36**, 2975 (1994) [Phys. Solid State **36**, 1583 (1994)].
18. S. F. Karmanenko, M. V. Belousov, V. Yu. Davydov, *et al*., Supercond. Sci. Technol. **6**, 23 (1993).
19. S. F. Karmanenko, V. Yu. Davydov, A. P. Mitrofanov, *et al*., Pis'ma Zh. Tekh. Fiz. **22** (23), 69 (1996) [Tech. Phys. Lett. **22**, 982 (1996)].
20. S. F. Karmanenko, Supercond. Sci. Technol. **12**, 36 (1999).
21. N. A. Bert., A. V. Bobyl', *et al*., Physica C **280**, 121 (1997).
22. A. V. Vishnyakov, *High-Temperature Superconductivity. Basic and Applied Reserch* [in Russian] (Nauka, Leningrad, 1990), pp. 377–404.
23. N. Savvides and A. Katsaros, Physica C **226**, 23 (1994).

24. V. V. Afrosimov, R. N. Il’in, S. F. Karmanenko, *et al.*, Fiz. Tverd. Tela (S.-Peterburg) **41**, 588 (1999) [Phys. Solid State **41**, 527 (1999)].
25. E. M. Sokolovskaya and L. S. Guzeĭ, *Metallochemistry* [in Russian] (Mosk. Gos. Univ., Moscow, 1986).
26. S. F. Karmanenko, A. A. Svishchev, A. A. Semenov, *et al.*, Pis’ma Zh. Tekh. Fiz. **25** (15), 79 (1999) [Tech. Phys. Lett. **25**, 626 (1999)].
27. P. Schneider, G. Linker, R. Schneider, *et al.*, Physica C **266**, 271(1996).
28. S. F. Karmanenko, A. P. Mitrofanov, K. F. Nyakshev, *et al.*, Zh. Tekh. Fiz. **65**, 64 (1995) [Tech. Phys. **40**, 34 (1995)].
29. A. N. Morozov, O. Yu. Morozova, N. M. Ponomarev, Sverkhprovodimost: Fiz., Khim., Tekh. **5**, 388 (1992).
30. C. W. Nieh, L. Anthony, J. Y. Josefowicz, and F. G. Krajenbrink, Appl. Phys. Lett. **56**, 2138 (1990).
31. L. M. Shirokova and N. A. Petrenko, Izvestiya LETI **437**, 46 (1991).
32. S. Ginley, C. I. H. Ashby, T. A. Plut, *et al.*, Appl. Phys. Lett. **63**, 2429 (1993).
33. W. L. Stirling, C. C. Tsai, and P. M. Ryan, Rev. Sci. Instrum. **48**, 533 (1977).
34. P. Vase, S. Yueqiang, and T. Freltoft, Appl. Surf. Sci. **46**, 61 (1990).

Translated by M.V. Lebedev

Technical Physics, Vol. 45, No. 4, 2000, pp. 453–457. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 73–77.
Original Russian Text Copyright © 2000 by Andrianov, Balakin, Daishev, Kisun'ko, Murzakhanov, Skochilov.

OPTICS, QUANTUM ELECTRONICS

Stabilization of the Different Frequency of Reference and Signal Resonators in the Dulkyn Gravitational-Wave Detector

S. N. Andrianov, A. B. Balakin, R. A. Daishev, G. V. Kisun'ko,
Z. G. Murzakhanov, and A. F. Skochilov

Dulkyn Scientific Center of Gravitational-Wave Research,
Tatarstan Academy of Sciences, Kazan, 420503 Russia

Received May 15, 1998

Abstract—It is shown that stabilization of the different frequency of optical radiations generated in the reference and signal resonators of the Dulkyn laser-interferometric gravitational-wave detector results in an efficient reevaluation of basic parameters describing the line widths of the laser radiation and the passive resonator, as well as the minimal laser power required to separate the valid signal from noises caused by natural fluctuations. *© 2000 MAIK “Nauka/Interperiodica”.*

INTRODUCTION

The complex of problems coupled with the elimination of external interferences, minimization of internal fluctuations, and separation of a weak signal from noise is a key obstacle for detecting gravitational waves (GW). For more than one decade, world gravitational-wave community has spent a huge intellectual effort and material resources on the investigation of these problems. For each presently known type of GW detectors, interferences and noises are cataloged and their main classes are detailed theoretically, simulated on computer systems, and tested experimentally. However, the hierarchy of interferences and noises is highly individual and reflects design features of the GW detector; therefore, any new technical proposals require revision and reassessment of the role of one or other noises. It is such revision that is required when we consider the noise immunity of the Dulkyn compact two-resonator pentagonal laser-interferometric GW detector developed at the Kazan Scientific Center of Gravitational-Wave Research [1, 2]. There are several reasons for such a revision.

First, since the active medium is situated inside the interferometer, the Dulkyn GW detector is compact and is mounted on a single base, which differs it substantially from long-base passive laser interferometers. Second, the basic concept of the Dulkyn project is coupled with detecting periodic low-frequency GW signals and relies on the long-term coherent acquisition of these signals. Third, in order to increase the signal-to-noise ratio, the Dulkyn GW detector uses a self-regulation system to control the different frequency of optical radiations generated in the reference and signal resonator [2, 3]. The third circumstance is the main subject of discussion in this paper.

There are opposite views on the possibility of utilization of the frequency and phase control systems in GW detectors. One of extreme opinions is that the self-regulation system eliminates the detected GW signal in the observed interferometric picture. From our viewpoint, this judgment was formed in the process of development of long-base laser-interferometric GW detectors. These detectors are aimed at the reception of individual pulsed GW signals (supernova outbursts, etc.); therefore, their detection requires that the signal-to-noise ratio must exceed unity. This fact excludes the possibility of utilization of the active regulating system to stabilize the interferometer end mirrors at the frequency of the expected GW signal.

In the Dulkyn GW detector considered in this paper, another situation takes place: the instantaneous signal-to-noise ratio can be below unity, because the detected GW signal is continuous and periodic and the final signal-to-noise ratio required for detection is achieved after a long-term coherent acquisition. Our experimental investigations have shown that, for the Dulkyn two-channel laser-interferometric GW detector, a regulating system can be created such that does not eliminate the useful periodic signal under the condition that its amplitude is at least two orders of magnitude lower than the so-called “regulation threshold” (a report about these results was submitted to the *Pis'ma Zh. Tekh. Fiz.* journal).

However, we see the problem of not only utilization of the self-regulation system in the GW detector, but also the effect of the aforementioned system on the estimation of the signal-to-noise ratio and on the minimum laser power required for isolation of the valid signal from noises caused by natural frequency and phase fluctuations of optical radiation.

In calculations, we use the following considerations. Equations describing generation of optical radiation in a compact pentagonal two-resonator laser interferometer involve the evolution equation for the density matrix, Maxwell equations, and conditions fixing the boundary operating mode [2, 4].

The main channel of action of the GW field is that the GW efficiently changes the refractive index along the optical path [2], so that GW parameters are involved in the Maxwell equations in the form of slowly varying coefficients of the electric field vector and its derivatives.

The frequency and phase control system responds to any displacements of the interferometer mirrors and thus, introduces its own corrections to the boundary operating mode, changing eigenfrequencies $\Omega_{1,2}(t)$ of the signal and reference resonators.

The following time scale exists for processes running inside the interferometer: $\tau_a = 10^{-8}$ s is the relaxation time for elements of the density matrix (it determines the natural width of the line radiated by atoms of the active medium); $\tau_r = 10^{-8}$ s is the photon travel time along the interferometer perimeter; $\tau_g = 10^{-6}$ s is the transient period of the stationary generation in the ring laser; $\tau_0 = 10^{-4}$ s is the time constant in the regulating system for most aggressive interferences; and $\tau_{GW} = T_g = 10^3$ s is the period of a gravitational wave.

The existence of the aforementioned scale allows us to uncouple the basic equation system: we obtain the density matrix and then the microscopic polarization of atoms under the assumption that the electromagnetic field is given; we solve the Maxwell equations with account for the obtained macroscopic polarization for a fixed boundary operating mode and specified instantaneous values of $\Omega_1(t)$ and $\Omega_2(t)$; and the boundary operating mode is stabilized for a fixed value of the slowly varying gravitational field.

The regulating system partially compensates for fluctuating displacements of mirrors and efficiently lowers rms fluctuations, which is equivalent to an increased Q-factor of the optical circuit.

A natural question arises, which quantities should be used as the linewidth of the laser system radiation and the Q-factor of the optical circuit during the operation of the regulating system stabilizing the boundary mode.

STABILIZATION BEYOND THE SYNCHRONIZATION RANGE

If we proceed from the model [2] describing the two-wave generation mode, then, for operation near the synchronization range, electromagnetic equations for phase difference Φ of the optical radiation in the signal and reference resonators can be represented as

$$\dot{\Phi} = [\Delta\Omega(t) + \Delta\nu(t) + \Delta\omega_{GW}] - M\sin\Phi, \qquad (1)$$

where M is the coefficient describing the linear coupling between orthogonal optical modes and $\Delta\omega_{GW} = h(t)\omega_0\zeta$ is the gravitationally induced shift of the generated frequency; $h(t)$ is the dimensionless amplitude of GW; ω_0 is the central frequency of the gain loop of the atoms of active medium; ζ is the factor depending on the resonator shape ($\zeta = \sqrt{5} - 2$ for a pentagon [1, 2]); $\Delta\nu(t)$ is the total noise source equal to the instantaneous value of fluctuations of the difference frequency caused by all types of interferences and noises; and $\Delta\Omega(t) = \Omega_1(t) - \Omega_2(t)$ is the instantaneous difference between eigenfrequencies of the signal and reference resonator, which is determined by instantaneous values of resonator perimeters and involves changes introduced by the regulating system.

Note that the expression in square brackets is the source of the phase deviation; this source contains a deterministic periodic part (third term), stochastic part (second term), and controlled part (first term). Of course, such a representation is purely conventional, since, only theoretically, we can explicitly separate consequences of the fluctuating displacement of mirrors from compensating displacements caused by the regulating system.

Evidently, the sum in the square brackets cannot be zeroed even in the most perfect regulating system; it can only be minimized to a certain threshold Π (regulation threshold). Consider the following chain of inequalities:

$$\begin{gathered}|\Delta\Omega(t) + \Delta\nu(t)| \le \Pi \ll \max|\nu(t)|, \\ |\Delta\omega_{GW}| \ll \Pi.\end{gathered} \qquad (2)$$

Here, $\max|\Delta\nu(t)|$ is the upper threshold of nonreduced noise (for operation without a regulating system). The regulating system allows us to fix sum $\varepsilon(t) \equiv \Delta\Omega(t) + \Delta\nu(t)$ at a level of $|\varepsilon(t)| \le \Pi$, with the upper threshold for reduced noise ε being several orders of magnitude lower than $\max|\Delta\nu(t)|$. Finally, $|\Delta\omega_{GW}| \ll \Pi$; i.e., the rms value of the reduced noise is much higher than the amplitude of GW signal.

Consider the case when $M \ll \Pi$, i.e., the laser system operates beyond the synchronization range. Spectral density $S_\varepsilon(\omega)$ of reduced noise $\varepsilon(t)$ is determined by spectral density $S_\nu^{(0)}(\omega)$ of initial noise $\Delta\nu(t)$ and by the effective spectral density of internal noises of regulating system $S_\nu^{(n)}(\omega)$ [5, 6]

$$S_\varepsilon(\omega) = S_\nu^{(0)}(\omega)\frac{1}{|1 + G(\omega)|^2} + S_\nu^{(n)}(\omega)\left|\frac{G(\omega)}{1 + G(\omega)}\right|^2, \qquad (3)$$

where $G(\omega)$ is the gain factor of the frequency autotuning (FAT) system.

The gain factor can be written as [6]

$$G(\omega) = \frac{K}{1 + i\omega\tau_0}, \qquad (4)$$

where K is the transfer factor at low frequencies and τ_0 is the loop transfer constant.

Substituting expressions (4) into formula (3), we obtain

$$S_\varepsilon(\omega) = S_\nu^{(0)}(\omega)\frac{1+(\omega\tau_0)^2}{(1+K)^2+(\omega\tau_0)^2} + S_\nu^{(n)}(\omega)\frac{K^2}{(1+K)^2+(\omega\tau_0)^2}. \tag{5}$$

The fluctuation-induced phase increment arising in time τ can be obtained from equation (1) at $M = 0$,

$$\delta\phi = \int_0^\tau \Delta\nu(\xi)d\xi. \tag{6}$$

Then, for the rms phase increment,

$$\chi(\tau) = \frac{1}{2}\langle|\delta\phi|^2\rangle$$

we obtain

$$\chi(\tau) = \int_0^\tau (\tau-\xi)R_\nu^{(2)}(\xi)d\xi, \tag{7}$$

where

$$R_\nu^{(2)}(\tau) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} R_\nu^{(1)}[t, t+\tau]dt \tag{8}$$

is the correlation function of the second kind, which, for stationary processes, coincides with the correlation function of the first kind

$$R_x^{(1)}[t_1, t_2] = \langle x(t_1)x(t_2)\rangle. \tag{9}$$

Taking into account that the correlation function of the second kind can be expressed in terms of noise spectral density $S_\varepsilon(\omega)$ by the formula

$$R_\nu^{(2)}(\tau) = \int_{-\infty}^{\infty} S_\varepsilon(\omega)\cos(\omega\tau)d\omega, \tag{10}$$

expression (7) can be rewritten after partial integration as

$$\chi(\tau) = \int_{-\infty}^{\infty}\frac{1-\cos(\omega\tau)}{\omega^2}S_\varepsilon(\omega)d\omega = 2\int_{-\infty}^{\infty}\frac{S_\varepsilon(\omega)}{\omega^2}\sin^2\left(\frac{\omega\tau}{2}\right)d\omega. \tag{11}$$

In the case of white noise, we obtain from formula (11) with account for expression (5) that

$$\chi(\tau) = \frac{D_0/2}{(1+K)^2}\left\{|\tau| + \frac{K(K+2)}{K+1}\left[1-e^{-\frac{\tau}{\tau_0}(1+K)}\right]\right\} + \frac{D_nK^2/2}{(1+K)^2}\left\{|\tau| - \frac{\tau_0}{1+K}\left[1-e^{-\frac{\tau}{\tau_0}(1+K)}\right]\right\}, \tag{12}$$

where D_0 and D_n are the constants of the laser white noise and the FAT circuit, respectively.

Hence, for large K, fluctuations of the radiation phase difference in the loop of the laser interferometer with FAT are limited by the intrinsic noise of the FAT circuit, which, in its turn, may be low, if one uses a quartz-crystal oscillator as a reference.

Taking into account that $D_0 = \Delta\omega$, where $\Delta\omega$ is the spectral width of laser radiation, neglecting the intrinsic noise of the FAT system, and assuming that $K \gg 1$ and $|\tau| = T \gg K\tau_0$, we obtain from expression (12) that

$$\langle|\delta\phi|^2\rangle = \Delta\omega^* T, \tag{13}$$

where $\Delta\omega^* = \Delta\omega/(1+K)^2$.

Expression (13) represents a well-known law for the phase diffusion caused by the finite width of the line of laser radiation. Quantity $\Delta\omega^*$ in (13) plays a role of the linewidth; it can be called an effective linewidth of laser radiation during the operation of the regulating system. Thus, in order to lower phase fluctuations, it is necessary to use either a laser system with a very narrow oscillating line or a FAT system. In this case, if one measures in the experiment only phases and there is not any other information, it is impossible in principle to find out which of the two aforementioned reasons caused the observed reduction of the phase noise. According to the well-known formula [7],

$$\Delta\omega = \frac{4\hbar\omega_0}{P}\gamma^2, \tag{14}$$

where P is the power of laser radiation, $\hbar$ is the Planck constant, $\gamma = cp/L$ is the intrinsic linewidth of a passive resonator with perimeter L, and loss per pass p (c is the speed of light).

Therefore, formula (13) can be written as

$$\langle|\delta\phi|^2\rangle = \frac{4\hbar\omega_0}{P}(\gamma^*)^2 T. \tag{15}$$

Here we introduced quantity $\gamma^* = \gamma/(1+K)$, which can be considered, similarly to $\Delta\omega^*$, as an effective linewidth of a passive resonator during the operation of the regulating system. Since, for larger K, quantity $\gamma^* \ll \gamma$, it follows that utilization of the FAT system is equivalent to the narrowing of the line of the passive resonator, i.e., to the rise of its Q-factor. Thus, in order to take into account operation of the regulating system, for example, in calculation of the signal-to-noise ratio or minimal power of laser radiation, one can use well-known formulas with $\Delta\omega$ and γ replaced with $\Delta\omega^*$ and γ^*.

Estimate the power of laser radiation required in the Dulkyn GW detector for the reliable separation of the GW signal from noises caused by spontaneous radiation of atoms of the active medium. Works [8, 9] present formulas that can be used to estimate the minimal power of laser radiation $P_{\min}$

$$P_{\min} \simeq \frac{\hbar\omega_0}{(\Delta\omega_{GW})^2 T}\gamma^2. \qquad (16)$$

In accordance with the above considerations, for the case of the operating regulating system, one should replace γ with γ^*. For $\omega_0 = 3 \times 10^{15}$ rad/s, $L = 3.5$ m, $p = 10^{-2}$, $h = 10^{-22}$, $T = 10^7$ s, and $K = 10^3$, we obtain $P_{\min} \simeq 10^{-4}$ W. If stabilization is absent ($K = 0$), then $P_{\min} \simeq 10^2$ W. Thus, the Dulkyn GW detector with a regulating system stabilizing the difference frequency requires the laser radiation power of the order of several milliwatts. This value is typical for helium–neon lasers, which are assumed to be used.

STABILIZATION WITHIN THE SYNCHRONIZATION RANGE

If condition $M \gg \Pi$ is satisfied, the regulating system brings the laser system into the synchronization range [2]; in this case,

$$\Phi = \pi n + \delta\phi(t) \qquad (17)$$

(where n is an integer) and equation for $\delta\phi(t)$ takes the form

$$\frac{d\delta\phi}{dt} = h(t)\omega_0 - M\delta\phi + \varepsilon(t). \qquad (18)$$

Assume that

$$h(t) = h\cos(\Omega_g t), \qquad (19)$$

where Ω_g is the frequency of gravitational radiation.

Solving this equation, we obtain

$$\Delta\phi = \phi_s + \delta\phi, \qquad (20)$$

$$\phi_s = \frac{h\omega_0 M}{M^2 + \Omega_g^2}[\cos(\Omega_g\tau) - e^{-M\tau}] + \frac{h\omega_0\Omega_g}{M^2 + \Omega_g^2}\sin(\Omega_g\tau),$$
$$\delta\phi = \int_0^\tau dt e^{-Mt}\varepsilon(\tau - t), \qquad (21)$$

where ϕ_s is the signal and $\delta\phi$ is the noise of the recording process.

Hence, for the rms fluctuation of the phase difference, we have

$$\langle|\delta\phi|^2\rangle = \int_0^\tau\int_0^\tau e^{-M(\xi+\eta)}R_\nu(\xi - \eta)d\xi d\eta, \qquad (23)$$

where $R_\nu(\tau) = \langle\varepsilon(t + \tau)\varepsilon(t)\rangle$ is the correlation function of frequency fluctuations.

Using substitutions $\eta' = \eta$ and $\xi' = \xi - \eta$, we obtain

$$\langle|\delta\phi|^2\rangle = \frac{1}{2M}\int_{-\tau}^{\tau}\left[e^{-|\xi|M} - e^{-(2t-|\xi|)M}\right]R_\nu(\xi)d\xi. \qquad (24)$$

Passing to the spectral density of frequency fluctuations $S_\varepsilon(\omega)$, we obtain

$$\langle|\delta\phi|^2\rangle = \int_{-\infty}^{\infty}\frac{S_\varepsilon(\omega)}{M^2 + \omega^2}[1 + e^{-2M\tau} - 2\cos(\omega\tau)e^{-M\tau}]d\omega. \qquad (25)$$

In accordance with formula (5), the spectral density of the free-of-noise FAT system takes the form

$$S_\varepsilon(\omega) = S_\nu^{(0)}(\omega)\frac{1 + (\omega\tau_0)^2}{(1 + K)^2 + (\omega\tau_0)^2}. \qquad (26)$$

For the case of white noise $S_\nu^{(0)} = D/(2\pi)$, substituting expression (26) into formula (25), we obtain the following formula

$$\langle|\delta\phi|^2\rangle = \frac{D}{2M}(1 - e^{-2M\tau})\frac{1 - M^2\tau_0^2}{(1 + K)^2 - M^2\tau_0^2} + \frac{D\tau_0}{2(1 + k)}$$
$$\times\left(1 + e^{-2M\tau} - 2e^{-M\tau}e^{-(1+K)\frac{\tau}{\tau_0}}\right)\frac{K(K + 2)}{(1 + K)^2 - M^2\tau_0^2}. \qquad (27)$$

With no FAT system ($K = 0$ and $\tau_0 = 0$), we obtain a law typical for the synchronization range

$$\langle|\delta\phi|^2\rangle = \frac{D}{2M}(1 - e^{-2M\tau}). \qquad (28)$$

For $K \gg 1$, $M\tau_0 \ll 1$, and $M\tau \gg 1$, formula (27) is substantially simplified

$$\langle|\delta\phi|^2\rangle = \frac{D}{2MK^2}(1 + MK\tau_0). \qquad (29)$$

If, in this case, $\Omega_g t \ll 1$, $M \gg \Omega_g$, and $MK\tau_0 \gg 1$, then the instantaneous signal-to-noise ratio is

$$q_{\text{inst}} = \frac{h\omega_0}{M}\sqrt{\frac{2K}{D\tau_0}}. \qquad (30)$$

From this formula, for $h\omega_0 = 10^{-7}$, $M = 10$ Hz, $K = 10^4$, $\tau_0 = 10^{-4}$ s, and $D = 10^{-2}$ Hz, we obtain that $q_{\text{inst}} = 10^{-3}$ (remind that, for this value, the useful signal is not eliminated by the regulating system). In order to obtain a finite value of the signal-to-noise ratio q_{fin}, one should multiply q_{inst} by $\sqrt{N}$: $q_{\text{fin}} = q_{\text{inst}}\sqrt{N} \simeq 3$, where $N = 10^7$ samples with durations of one second each. Therefore, a reliable recording of the gravitational wave requires an acquisition time of the order of 10^7 s.

CONCLUSION

Analysis of the discussed problem allows us to answer a particular question posed at the end of the introduction as well as to formulate a principle for evaluation of basic parameters of the optical subsystem of the Dulkyn GW detector using a regulating system to stabilize the difference frequency. The initial point for evaluation of the parameters is well-known formula (3), which shows that the spectral density of frequency fluctuations is lowered during the operation of the regulating system. Calculations based on this formula have shown that the linewidth of the laser radiation caused by spontaneous fluctuations and the linewidth of a passive resonator can be expressed by the following working formula:

$$\Gamma^* = \frac{\Gamma}{(1+K)^{\Theta}}, \tag{31}$$

where $\Theta = 2$ and 1 in the first and second case, respectively.

Extending this idea, we assume that, for the case of stabilization of the difference frequency, any considered parameter Γ^* can be expressed in terms of Γ with the help of formula (31) with parameter Θ depending on the type of the investigated parameter (if we take into account nonlinear effects, parameter Θ can be fractional and even irrational).

Since the catalogue of the Dulkyn GW detector noises is not exhausted by natural quantum noises, we have to perform a systematic analysis of the role of other interferences. We are sure that the substantiated simple and evident method for the revaluation of parameters under the conditions of stabilization of the difference frequency will be good phenomenological support in this work. In the cases when parameter Θ cannot be calculated theoretically, it can be estimated with the help of experimental data obtained on the Dulkyn pentagonal laser interferometer.

ACKNOWLEDGMENTS

The authors are grateful to Ph. Tourrenc for communication and fruitful discussions that stimulated this work.

REFERENCES

1. A. B. Balakin, G. V. Kisun’ko, Z. G. Murzakhanov, *et al*., Dokl. Ross. Akad. Nauk **341**, 37 (1996) [Phys. Dokl. **41**, 19 (1996)].
2. A. B. Balakin, Z. G. Murzakhanov, and A. F. Skochilov, Gravitation and Cosmology **3** (1(9)), 71 (1997).
3. V. B. Konstantinov, Z. G. Murzakhanov, and A. F. Skochilov, Izv. Vyssh. Uchebn. Zaved., Fiz., No. 2, 22 (1998).
4. A. B. Balakin, Z. G. Murzakhanov, and A. F. Skochilov, Opt. Spektrosk. **76**, 671 (1994) [Opt. Spectrosc. **76**, 602 (1994)].
5. L. S. Cutler and C. L. Searle, Proc. IEEE **54**, 136 (1966).
6. M. R. Kaplanov and V. A. Levin, *Frequency Autotuning* (Gosénergoizdat, Moscow, 1956).
7. L. Allen and D. G. C. Jones, *Principles of Gas Lasers* (Plenum, New York, 1967; Nauka, Moscow, 1979).
8. M. O. Scully and J. Gea-Banacloche, Phys. Rev. A **34**, 4043 (1986).
9. A. Brillet and Ph. Tourrenc, in *NATO ASI Series Gravitational Radiation*, Ed. by N. Deruelle and T. Piran (North Holland, Amsterdam, 1983), pp. 475–484.

Translated by A. Kondrat’ev

Technical Physics, Vol. 45, No. 4, 2000, pp. 458–461. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 78–81.
Original Russian Text Copyright © 2000 by Sorokin, Chaĭkovsky.

OPTICS, QUANTUM ELECTRONICS

Measurements of the Power and Energy Spectrum of Radiation of the Plasma Liners

S. A. Sorokin and S. A. Chaĭkovsky

Institute of High Current Electronics, Siberian Division, Russian Academy of Sciences, Tomsk, 634055 Russia

Received November 6, 1998

Abstract—A diagnostic complex based on bolometers and X-ray diodes is developed and tested. The complex is designed for measurements of the power and the energy spectrum of the pulses of soft X-rays in a quanta energy range of 70–1500 eV. Both thin films and filters formed by pulsed gas-puff in the tube of the diagnostic channel were used as X-ray filters. Experiments were carried out on a high current generator under a load current up to 2 MA. Soft X-ray pulses with a power of about 1 TW and a duration of about 40 ns were formed by the implosion of krypton liners. © *2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Magnetic implosion of cylindrical plasma liners is used widely as a source of high-power soft X-ray (SXR) pulses [1, 2]. The efficiency of conversion of the generator electric energy, fed to an X-ray diode, into the energy of radiation can be as high as 50%. The main part of the radiation energy is concentrated, as a rule, in a range of the quanta energy of 100–700 eV. X-ray diodes (XRD) [3] and foil bolometers [4] are used most widely for measurements of power and the total yield (per pulse) of radiation. Spectral response of the detection channel is determined by the X-ray transmission curve of the filter and by the quantum efficiency of the photocathode (for XRD). A bolometer without filter exhibits a spectrally constant response function and can be used for measuring the total (over the spectrum) radiation power. In general, the signal of one detection channel provides information neither on the total power, nor on the energy spectrum of the radiation. The application of an array of detectors with different, partially overlapping spectral response functions makes it possible to unfold the shape of the emission spectrum [5–7]. The problem of the deconvolution of a spectrum on the basis of the signals from an array of detectors is a mathematically ill-posed problem. The deconvolution accuracy depends drastically on the selection of the combination of the materials of the photocathodes (XRD) and filters and on the accuracy of determination of the mass thickness of the filters. The films with a thickness less then 1 μm and a high transmission in the range below 100 eV are not commercially available. The production of such films meets the problems of the control of their thickness and uniformity and their safety in course of assembling and maintenance. The requirements of the simplicity and low cost of the production of thin films impose limitations on the properties of the materials. As a rule, these are carbon-containing films with low transmission in a range of 290–400 eV (above K-edge of carbon), which makes it difficult to unfold the corresponding part of the spectrum. The application of gas filters allows one to find the way out. Gas filters make it possible to substantially increase the number of combinations of the materials of the cathode and the filter. Being once constructed and adjusted, the system of the formation of the gas filter allows one to get rid of the procedures of manufacturing, thickness measurement, and maintenance of thin (0.1–1 μm) films. Mass thickness of the gas filter can be varied easily and can be virtually infinitely small; i.e., the sensitivity of XRD can be easily extended to lower energies.

In this work, we used a bolometer made from copper foil and an array of five XRDs for measurements of the conversion efficiency, power, and energy spectrum of SXR in a range of 70–1500 eV. We used both thin film and gas filters. The latter were formed by pulsed gas-puff in the tube of the diagnostic channel with the use of a fast electrodynamic valve. The experiments were performed on a MIG high-current generator [8] at a load current of 2 MA. The leading edge of the current was about 80 ns. SXR pulses with a duration of 30–40 ns and a power of ~1 TW were formed by implosion of krypton liners. The initial liner with an initial diameter of 20 mm and a length of 35 mm was formed with the use of a fast opening gas puff valve and Laval nozzle.

DIAGNOSTIC EQUIPMENT

Copper strips with a length of 14 mm, width of 1 mm, and thickness of 2 and 15 μm were used as resistive elements. The bolometer is assembled on the basis of a standard SR-75-167 connector. A reference current with an amplitude of 100–200 A and duration of ~10 μs is fed to the resistive element. Bolometer with the foil of 15-μm thickness makes it possible to measure only the total (per pulse) yield of radiation because the characteristic time of the temperature equalizing over the foil thickness due to heat conduction is longer than the pulse duration. The temperature equalizing time for a 2-μm foil is about 10 ns, which allows one to trace the radiation absorption dynamics of the foil at the radiation pulse duration longer than 10 ns.

XRD is made on the basis of a SR-75-167 connector as well as the bolometer. The detector design is simple, cheap, and user friendly. In addition to the elements of the connector, one needs only a disk of the photocathode with a diameter of 14 mm, a Teflon insulator that supports the photocathode, and the anode metal grid. A negative voltage of 1 kV is applied to the photocathode via a signal cable inside a shielded enclosure for recording. The signal from the cable is fed to an oscilloscope via a high-frequency coupling capacitor. The time of a double signal passage along the cable (~300 ns) is substantially longer than the duration of the radiation pulse, which allows one to avoid signal distortions caused by reflection from the coupling capacitor. Such an XRD has a subnanosecond time resolution. For an actual diode design, the density of the saturation current (saturation by the volume charge) is ~100 A/cm^2. Quantum efficiency of the photocathodes is such that the requirement of the linear dependence of the XRD current on the radiation intensity imposes limitation on the radiation intensity on the surface of the photocathode: $W < 10^5$ W/cm^2. If one works with an SXR source with the power $P > 10^{12}$ W, an XRD must be removed from the source by $r > (P/4\pi W)^{0.5} \approx 10^3$ cm. Long distance between the XRD and the source necessitates mounting of long vacuum diagnostic channels, which is not always possible because of finite sizes of the laboratory building. The application of intensity attenuators makes it possible to diminish the length of the measuring lines. In this work, the radiation flux was attenuated by means of a slit mounted inside the tube of the measuring channel orthogonally to the pinch axis. The radiation transmitted through the slit illuminates a rectangular area on the XRD photocathode. The sizes of the rectangle are $l_1 = Lb/a$ and $l_2 = h(a + b)/a$, where L is the pinch length; a and b are the distances from the slit to the pinch and photocathode, respectively; and h is the height of the slit. Attenuation of the intensity at the surface of the photocathode is

$$\frac{W_2}{W_1} = \frac{a+b}{b}\frac{\Delta}{L},$$

where W_1 and W_2 are the intensities with and without the slit and Δ is the thickness of the slit.

This formula was derived under the assumption that the emissivity of the pinch does not depend on the z-coordinate. The finite diameter of the pinch was not taken into account when the size l_2 was determined.

Figure 1 shows a schematic of the diagnostic channel with the gas filter. The pulse valve is at the center of the tube. The gas-puffing time must be shorter than the time of gas expansion inside the tube to the SXR detector and the edge of the tube. In this case, the mass thickness of the filter can be determined easily from the known gas mass inside the tube and its cross section. As the diameter of the tube is much smaller than its length, the gas distribution over the cross section of the tube is virtually uniform. The mass of the gas is determined and controlled by the forevolume of the gas valve V and the gas pressure in it p, because the valve empties completely in each operation cycle. Mass thickness of the gas filter is

$$D = m/S = m_a nV/S.$$

Here m is the gas mass in the forevolume of the valve, S is the cross section of the tube, m_a is the mass of the

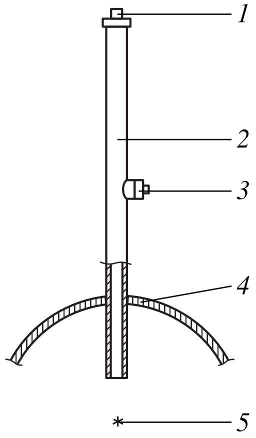


Fig. 1. Schematic diagram of the diagnostic channel with a gas filter: (*1*) detector; (*2*) tube of the detection channel; (*3*) pulse valve; (*4*) wall of the vacuum chamber; and (*5*) radiation source.

Table 1

Channel	Filter	Thickness	Photocathode
1	Nitrogen	10^{-3} kg/m^2	Aluminum
2	Polystyrene	0.4 μm	Graphite
3	Mylar	2.5 μm	Aluminum
4	″	2.5 μm	″
	Aluminum	0.4 μm	″
5	″	0.8 μm	″

Table 2

Channel	Filter	Thickness	Photocathode
1	Acetylene	1.2×10^{-3}	Aluminum
2	Nitrogen	6×10^{-4}	″
3	Oxygen	3×10^{-3}	Graphite
4	Neon	6×10^{-3}	Aluminum

gas molecule, and n is the gas density. For nitrogen at $V = 1.1$ cm^3 and $S = 18$ cm^2, we have

$$D\ [\mu\text{g/cm}^2] = 76.7p\ [\text{atm}].$$

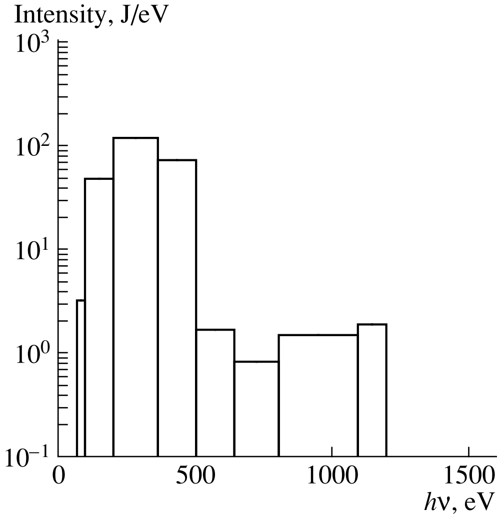


Fig. 2. Histogram of the emission spectrum of the krypton liner.

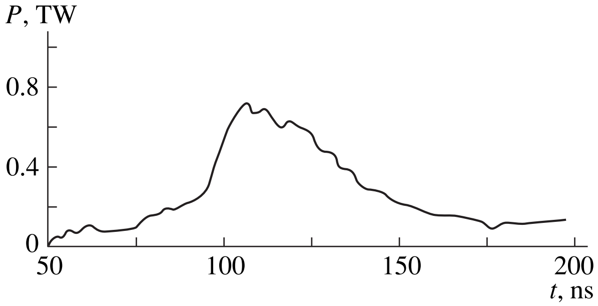


Fig. 3. Signal of XRD with quasi-constant response function in a range of 70–900 eV.

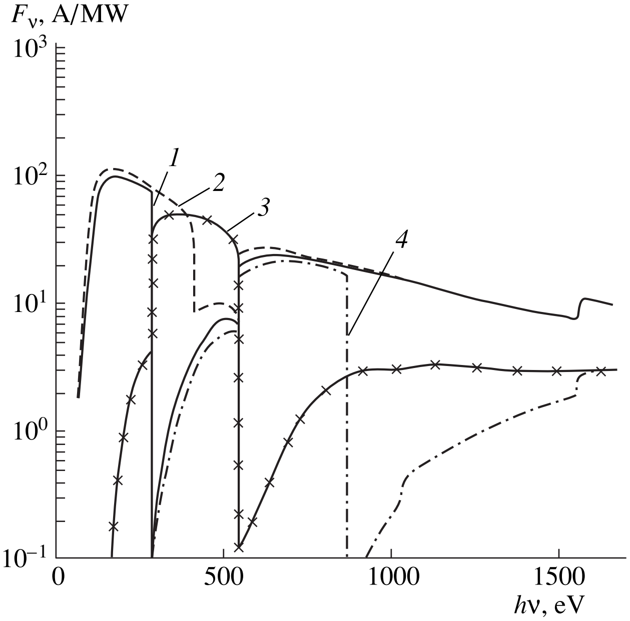


Fig. 4. Response function F_ν of four XRDs with gas filters for spectral measurements in a range of 100–870 eV.

RESULTS AND DISCUSSION

Table 1 presents an array of five XRDs used in this work. Response function of XRD was calculated based on the data on the quantum efficiency of the photocathodes [9] and the transmission of the filters [10, 11]. XRD signals were recorded with a Tetronix 640A digital oscilloscope. Digital signals substantially simplify the procedure of unfolding of the spectrum and the radiation power. Iterative algorithm [7] was used for unfolding the spectral data. Figure 2 shows a histogram of the time-averaged spectrum for one shot of the krypton liner. The total yield of radiation over the spectrum is 34 kJ. Constant spectral density of the radiation was chosen as the initial approximation in the unfolding procedure. The shape of the spectrum undergoes only minor changes if the Plank spectrum with a temperature of 100 eV is used as the initial approximation. The changes are within 1% in a spectral interval of 200–350 eV, in which half of the radiation energy is concentrated. The spectral function changes by no more than 6% in the adjacent spectral intervals (100–200 and 350–500 eV). The changes of the radiation yield are also less than 1%. In this shot, the reading of the copper-foil bolometer with 2-μm thickness and screened by a nitrogen filter with a mass thickness of 3.8×10^{-4} kg/m^2 corresponded to the radiation energy of 20 kJ. Such a level of radiation behind the filter for the spectrum shown in Fig. 2 corresponds to a total energy of 27 kJ. The lower level of radiation yield obtained, with the use of the bolometer, is possibly related to partial cooling of the foil due to heating of the dielectric substrate that supports the foil.

XRD with a graphite cathode and filter made from polystyrene with a width of 0.4 μm exhibits a slightly varying response function in a range of 70–900 eV and allows direct measurements of the radiation power with an accuracy of ~30%. The radiation yield in the considered shot, measured by using of the averaged response function of this detector in a range of 70–600 eV, was 32 kJ. Figure 3 shows the radiation power. The accuracy of measurements with this detector can be improved substantially by its calibration in one of the shots with the use of power measurements by an alternative method. It must be also assumed that the radiation spectrum changes negligibly from one shot to another.

Note that the use of gas filters allows one to unfold the spectrum based on the signals of XRDs with quasi-constant response functions in certain spectral intervals. Figure 4 shows response functions F_ν of four XRDs for spectral measurements in a range of 100–870 eV. Mass thicknesses of the filters were slightly unbalanced in order to avoid overlapping of the response functions and to make the graphical presentation more clear. The parameters of the filters and the photocathodes of these XRDs are presented in Table 2. The power of radiation in a certain spectral interval can be determined either directly from the detector signal or by subtraction of the signals of two detectors. For example, the subtraction of the signal of the fourth channel from the signal of the first channel yields the power of radiation in an interval of 100–290 eV. The radiation powers measured in spectral intervals can be used as initial approximations for the iterative unfolding algorithm. We did not use this method of spectral measurements, because in our experiments we could install only one measuring channel with a gas filter.

CONCLUSION

We measured the power, the total yield of radiation per pulse, and the energy spectrum of SXR in experiments on the implosion of plasma liners. The experiments were carried out in a MIG-high current generator at a load current of up to 2MA. The radiation power and spectrum in a range of 70–1500 eV were unfolded with the use of the Tarasko iterative algorithm based on the signals of five XRDs with different, partially overlapping spectral response functions. A construction of the diagnostic channel with a gas filter was developed and tested. The application of gas filters allows one to increase substantially the number of possible combinations of photocathodes and filters and to extend the spectral response function of the XRD to lower quanta energies. Slit attenuator made it possible to establish such a level of intensity at the XRD photocathode that provides a linear dependence of the XRD current on the intensity. The power and the total yield of radiation per pulse were measured also by a foil bolometer and XRD with a quasi-constant spectral response function. We considered an array of XRDs with gas filters that exhibit quasi-constant response functions in spectral intervals that span a range of 100–870 eV. Such an array of the detectors ensures direct measurements of the power of radiation in certain spectral intervals without using the iterative unfolding procedure.

REFERENCES

1. P. J. Turchi and W. J. Baker, J. Appl. Phys. **44**, 4936 (1973).
2. N. R. Pereira and J. Davis, J. Appl. Phys. **64** (3), R1 (1988).
3. H. N. Kornblum and U. W. Slivinsky, Rev. Sci. Instrum. **49**, 1204 (1978).
4. J. H. Degnan, Rev. Sci. Instrum. **50**, 1223 (1979).
5. E. J. T. Burns, Adv. X-Ray Anal. **18**, 117 (1974).
6. J. Bailey, A. Fisher, and N. Rostoker, J. Appl. Phys. **60**, 1939 (1986).
7. N. G. Basov, Yu. A. Zakharenkov, A. A. Rupasov, *et al.*, *Diagnostics of Dense Plasma* (Nauka, Moscow, 1989).
8. A. V. Luchinsky, N. A. Ratakhin, V. F. Fedushchak, *et al.*, Izv. Vyssh. Uchebn. Zaved., Fiz. **40** (12), 67 (1997).
9. R. H. Day, P. Lee, E. B. Saloman, *et al.*, J. Appl. Phys. **52** (11), 6965 (1981).
10. W. J. Veigele, At. Data **5**, 52 (1973).
11. B. L. Henke and E. S. Ebisu, Adv. X-Ray Anal. **17**, 150 (1974).

Translated by A. Chikishev

Technical Physics, Vol. 45, No. 4, 2000, pp. 462–466. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 82–86.
Original Russian Text Copyright © 2000 by Libenson, Sarnakov, Chuĭko, Shandybina.

OPTICS, QUANTUM ELECTRONICS

Laser-Induced Instabilities of a Continuous Thermal Trace on a Surface

M. N. Libenson, S. M. Sarnakov, V. A. Chuĭko, and G. D. Shandybina

Institute of Fine Mechanics and Optics (State Technical University),
St. Petersburg, 197101 Russia

Received November 27, 1998; in final form, October 6, 1999

Abstract—Results are presented that corroborate the important role played by laser-induced vapor-phase oxidation in the onset of an instability of the irradiation trace and formation of self-organized structures. *© 2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Irradiation of surfaces by intense, sharply focused, periodically pulsed scanning laser beams may be accompanied by quasi-periodic changes in the resulting traces [1]. Generally, they are observed when the optical properties of the laser-irradiated medium change at locations approached by the scanning light spot. As a typical example, consider a situation when irradiation gives rise to resonance electromagnetic disturbances (surface electromagnetic waves or waveguide modes) propagating in the scanning direction [2, 3]. Absorption of these waves results in a stronger heating of the surface ahead of the spot, and the effectiveness of irradiation increases as the spot moves along (both width and depth of the trace increase). The process is disrupted if the degree of surface overheating is so high that the dynamic relief against which the surface modes are generated is destroyed.

A major role in the onset of instabilities along the laser-beam scanning path may also be played by the reduction–oxidation reactions involved in the process. It is well known [4] that the solid-phase oxidation mechanism characterized by oxygen adsorption on the surface and subsequent diffusion of ions through the oxide layer changes to the vapor-phase (ablation-controlled) mechanism when the pulse width is reduced. In the latter mechanism, the vaporized particles react in air and the reaction products partially settle back onto the surface.

In this paper, we present some results of an experimental study and a descriptive model of the ablation processes associated with vapor-phase oxidation and condensation of the reaction products on the irradiated surface that lead to instabilities of a rectilinear laser-induced trace.

EXPERIMENTAL TECHNIQUE AND MEASUREMENT RESULTS

A YAG : Nd laser beam with a pulse width of 250 ns was focused onto a spot 50 μm in diameter on a target surface and used to uniformly scan a 5 × 5 mm area. A mode was singled out for irradiation by passing a laser beam through an aperture. The signal was received by a photodiode in the rear cavity mirror and was used to tune the mode to the generation regime characterized by the highest stability. The radiative power output was attenuated by turning the Fresnel mirrors.

Aluminum and titanium films deposited on silicon and glass substrates, as well as bulk aluminum and titanium slabs, were used as specimens.

The irradiated areas, their boundaries, and the adjoining portions of the films were examined by means of optical microscopy, scanning electron microscopy, and Auger microscopy.

Figure 1 shows the electron spectra of irradiated titanium and aluminum films on silicon substrates. Here, the radiant flux of the laser beam corresponds to the regime of intense vaporization of a metal film. The relative change in the elemental composition of the films in the areas directly exposed to laser light and in their vicinity is illustrated by Figs 1a–1d. It is clear that both titanium and aluminum are almost completely removed from the substrate in the irradiated areas (see Figs. 1a and 1c). However, the metals are characterized by different behavior in the vicinity of the irradiated areas: a substantial increase in the oxygen content is observed for titanium (Fig. 1b) while a minor change (in terms of a statistical average) takes place in the case of aluminum (Fig. 1d). The excess oxygen in titanium

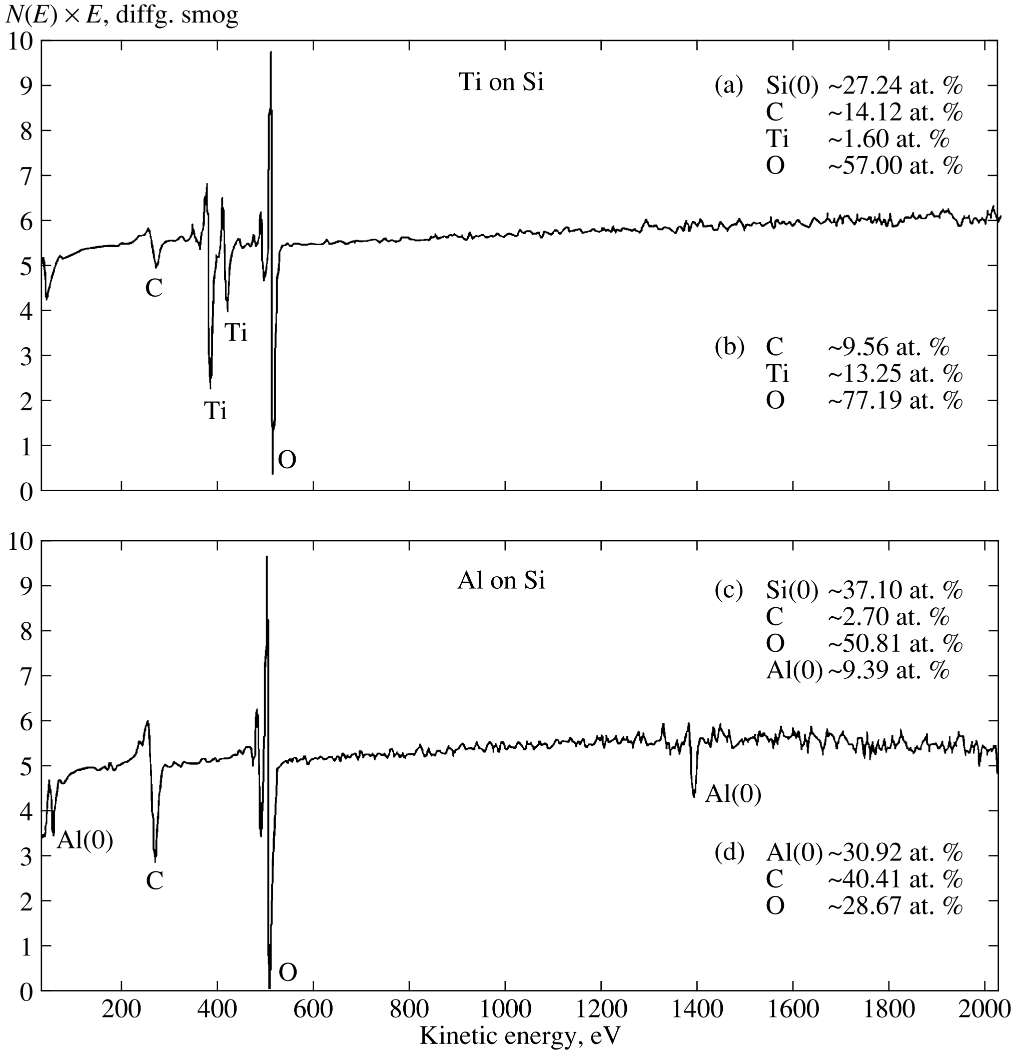


Fig. 1. Electron spectra of irradiated titanium (a, b) and aluminum (c, d) films on silicon substrates and their relative elemental compositions in directly irradiated zones (a, c) and in their vicinity (b, d). Spectra for cases (a) and (c) are not shown.

may be attributed to the formation of its oxides. Note that the area of excess oxygen content is larger than the irradiated area.

Figure 2 shows distributions of elements along an electronic scanning line that crosses the boundary between the laser-irradiated and nonirradiated areas for titanium (Fig. 2a) and aluminum (Fig. 2b) on silicon substrates. The distributions were derived from secondary electron emission data. They demonstrate that the boundary is 2–2.5 times wider for titanium film as compared to aluminum film and some silicon manifests itself in the former case. This suggests that the impact of a laser beam was sufficiently effective to vaporize not only the metal film, but also a part of the silicon substrate. The resulting backward flow of condensing particles is much more intense for titanium film as compared to aluminum film.

To evaluate the amount of substance that remains on the surface after it has been scanned by a laser beam, experiments were conducted with titanium films deposited on glass substrates. The specimens were irradiated by periodically pulsed beams with constant radiant flux at various pulse recurrence frequencies. After the irradiation, optical microscopy was used to evaluate the area occupied by the microscopic particles that had remained or condensed on the substrate (in terms of percentage of the irradiated area) as a function of microparticle size. Figure 3 shows histograms obtained for titanium films irradiated at various pulse recurrence frequencies. It demonstrates that the concentration of relatively large microscopic particles in the irradiated area increases with the pulse recurrence frequency. Moreover, the amount of metal that has remained on the substrate increases as well. This experimental result does not fit into the framework of the destruction mechanism controlled by evaporation only, suggesting that the particles tend to evaporate, react in air, and then condense back on the surface.

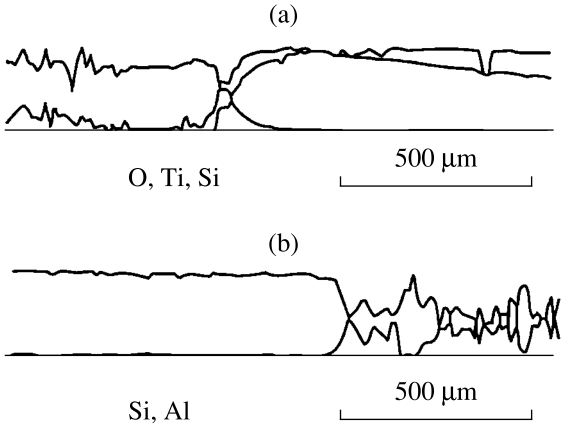


Fig. 2. Electron scanning of film surfaces at the boundary of directly irradiated zones; (a) Ti–Si; (b) Al–Si.

Indeed, the formation of chemical compounds in a gas (e.g., in air) reduces the volume occupied by their molecules as compared to the case of separate molecules, resulting in a pressure drop above the target and thus stimulating the backward flow of particles [4].

These studies provide additional evidence of the high efficiency of the vapor-phase oxidation induced by laser-beam irradiation of titanium and the much lower efficiency of a similar mechanism of aluminum oxidation.

As noted above, the thermal trace of a laser beam observed in certain irradiation and scanning regimes has a quasi-periodic structure typical for self-organization phenomena (see Fig. 4). The spatial period of the structure lies within 250–300 μm for specimens of various types. The quasi-periodic structure formation manifests itself by oscillations of trace widths on thin films (Fig. 4a) and by the varying destruction depth in bulk materials. As a result, periodically spaced craters appear and the vaporization trace is rimmed by an oxide layer on the titanium surface (Fig. 4b), whereas the trace is periodically disrupted and no oxide fringe is observed on aluminum targets (Fig. 4c).

DISCUSSION

A self-consistent explanation of the results presented above can be suggested within the framework of the qualitative model of laser-induced trace formation proposed in [1]. In what follows, we give a more detailed description of the model for thin films and bulk materials. The model relies on the fact that laser-induced ablation triggers various processes with multiple-loop feedbacks on the surface and in the near-surface plasma. In their dynamics, a major role is played by vapor-phase oxidation (in a more general case, by gas-phase chemical reaction), as well as by the nonuniform (say, Gaussian) distribution of radiant flux in the focal spot.

Indeed, when the radiant flux is slightly higher than the evaporation threshold of a material, intense evaporation takes place only around the spot center. The vaporized substance chemically reacts in air and partially condenses on the heated surface around the spot center. The condensed substance is characterized by increased adhesion, and its nonmetal-type conductivity enhances the absorptivity of the surface exposed to a subsequent pulse. Since the next scanning pulse acts not only on the original surface, but also on a part of the oxidized surface characterized by a higher absorptivity, the destruction efficiency increases, and the trace widens as the spot moves on. Thus, an instability of the trace geometry develops through a spatially distributed positive-feedback loop involving the absorptivity and the condensation-area width. Obviously, this may occur only within a certain range of scanning speeds and pulse recurrence frequencies.

However, it should be noted that, as the vaporized area grows, every new spot contains not only “new” portions characterized by higher absorptivity, but also an progressively increasing fraction of portions that have already been affected by scanning. In the case of a thin-film specimen, the film has already evaporated

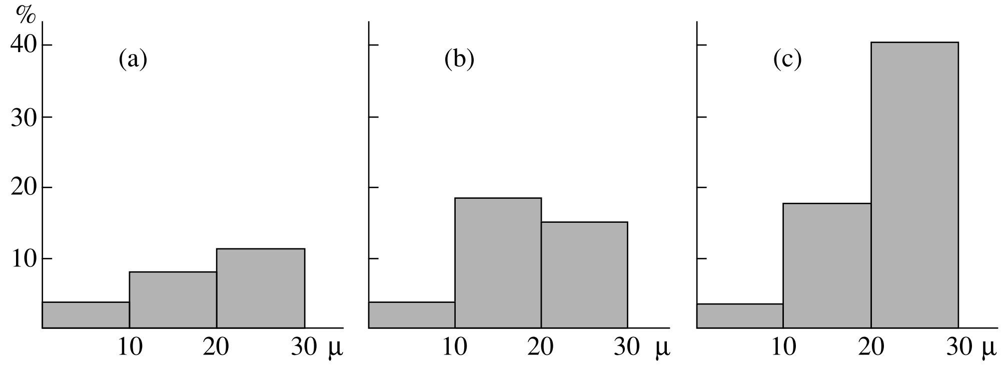


Fig. 3. Histograms of numerical density of particles as a function of particle size for titanium films on glass substrates irradiated by laser pulses at a frequency of (a) 250, (b) 500, and (c) 1 kHz.

from these portions and they are characterized by the low absorptivity of the transparent substrate. In effect, this leads to the gradual development of a negative-feedback loop involving the aforementioned parameters.

The resultant contributions and signs of feedback loops can be rigorously shown to depend, in particular, on the radiant-flux distribution across the spot and on the irradiation timing (these effects will be analyzed in a separate paper). In principle, when the scanning laser beam is continuous, both effects due to spatially distributed positive and negative feedback loops and the trace width may somehow reach a quasi-steady state. An essentially different pattern develops when a material is exposed to a periodically pulsed beam. Then, the "balanced" state is skipped because of the irradiation intermittency, and the resulting dynamics is always controlled by a feedback loop of a definite sign. It is obvious that negative feedback must play a dominant role when the vaporized-area width equals the focal-spot diameter. Accordingly, as the spot "jumps" further along the scanning path, the vaporized-trace width will tend to decrease until the spot center leaves the previously vaporized area and local conditions become similar to the initial conditions. The cycle is then repeated. Thus, a specific instability may develop along the scanning path for periodic pulsed radiation, manifesting itself by the formation of self-organized quasi-periodic structures on metal (in a more general case, absorptive) films.

Note that an increase in absorptivity may in certain cases occur as a result of a residual heating caused by a pulsed laser beam. However, simple theoretical estimates show that the residual heating is negligible in regimes characterized by formation of quasi-periodic structures (when the pulse period-to-width ratio is greater than 1000).

A totally different pattern of laser-induced trace structure formation is characteristic of bulk materials (see Figs. 4b, 4c). Nonetheless, the experimental results presented above show that laser-induced chemical processes play a major role in this case as well. Below, we describe a model of laser-induced trace formation on metal surfaces that is similar to that outlined above, but is adapted to bulk specimens.

For metals that are easily oxidized in condensed and gaseous states (such as titanium), the processes induced by short-pulse laser beams involve vapor-phase oxidation and partial oxide condensation on the surface [4]. These processes increase absorptivity and thereby enhance the thermal effect, and an instability of the laser-induced trace eventually develops through the mechanism described above. However, this instability is manifested in a different manner for a bulk target: a drastic increase trace depth, rather than a change in the trace width, is observed starting from a certain pulse. This manifestation of the instability is quite consistent with its typical physics. It is much more difficult to find an unambiguous interpretation of the subsequent trace disruption prior to the impact by the next pulse, which results in a jump of the laser-induced spot (see Fig. 4b). The most plausible explanation appears to lie in the fact that condensation of the oxidation products (which is required to maintain the positive feedback loop involving scanning efficiency and increase in absorptivity) is affected by processes that take place in the laser-induced flare. Observations have shown that the formation of a relatively deep crater on titanium surface involves a bright flash. This suggests that a "destructive" pulse induces an absorption burst inside the flare, resulting in a wider scatter of the destruction products. In the context of the present study, this means that the density of the condensed products drops around the area affected by the "destructive" pulse and the absorptivity returns to its initial value before the next pulse hits the surface. Then, the instability develops once again, and the resulting trace will consist of deep craters periodically forming along a fairly uniform background thermal trace.

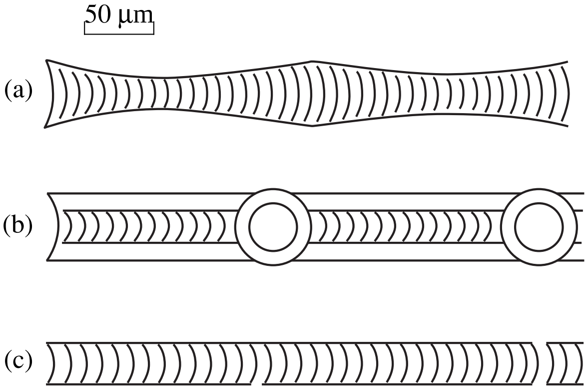


Fig. 4. Typical quasi-periodic variations of structure: (a) thin metal films, (b) titanium slabs, and (c) aluminum slabs.

Certain materials (e.g., aluminum alloys) are less easily affected by vapor-phase oxidation processes. The corresponding laser-induced trace is a narrow strip whose width is comparable to the "active" spot diameter (see Fig. 4c). Under these conditions, a gradual buildup of induced absorptivity may result in the destruction of a surface layer by a single "purging" pulse, which reduces the surface absorptivity so that no visible damage will be incurred when the next pulse hits the surface. The process repeats after the spot has jumped to a new location. As a result, the thermally induced trace will have the periodically discontinued form illustrated by Fig. 4c.

CONCLUSION

The descriptive model presented here is more detailed as compared to that set out in [1] and provides a unified qualitative explanation of structure formation in the thermal trace induced by the impact of a scanning periodically pulsed laser beam on bulk materials and thin films.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 97-02-18289. We thank V.I. Beklemyshev and S.V. Murashov for measurements performed by means of electron microscopy.

REFERENCES

1. V. Chujko, M. Libenson, A. Lysenko, and G. Shandybina, Proc. SPIE **3093**, 324 (1997).
2. V. V. Bazhenov, A. M. Bonch-Bruevich, M. N. Libenson, and V. S. Makin, Pis'ma Zh. Tekh. Fiz. **12**, 1104 (1986).
3. V. V. Bazhenov and V. S. Makin, Pis'ma Zh. Tekh. Fiz. **16**, 20 (1990).
4. A. G. Akimov, A. M. Bonch-Bruevich, A. P. Gagarin, *et al.*, Pis'ma Zh. Tekh. Fiz. **13**, 1093 (1987).

Translated by A. S. Betev

Technical Physics, Vol. 45, No. 4, 2000, pp. 467–469. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 87–89.
Original Russian Text Copyright © 2000 by Karpukhin, Malikov.

OPTICS, QUANTUM ELECTRONICS

Nonlinear Frequency Conversion of Copper Vapor Laser Radiation Using Convergent and Parallel Laser Beams

V. T. Karpukhin and M. M. Malikov

Associated Institute for High Temperatures, Russian Academy of Sciences, Moscow, 127412 Russia

Received December 23, 1998

Abstract—Results of experiments are presented on the comparison of the efficiency of radiation generation at the sum frequency ($\lambda = 0.271$ μm) of copper vapor laser (CVL) in a nonlinear DKDP crystal using either convergent or parallel laser beam. An intensity of UV radiation of 0.34 W and a conversion efficiency of 6% obtained under conditions of strong focusing occur to be substantially lower as compared to those obtained using a parallel laser beam of the same power (0.73 W and 12%, respectively). Experimental data on the structure of a CVL beam generated with an unstable resonator are presented, and physical reasons for the limitation of the efficiency of nonlinear frequency conversion are analyzed. *© 2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Medium-power (5–20 W) pulsed copper vapor lasers (CVLs) operating in the visible range ($\lambda_1 = 0.51$ μm and $\lambda_2 = 0.578$ μm) demonstrate quite high efficiency (of 1–3%) and high repetition rate (of 10–20 kHz) and are rather simple in design. This makes CVLs attractive for the generation of radiation at the second harmonic and the sum frequency (SF) in the UV range [1]. A UV generator based on a CVL may be of significant interest for a variety of practical applications [2].

However, there are certain difficulties in achieving a high efficiency of frequency conversion of CVL radiation in nonlinear crystals, because the conversion efficiency depends strongly on both the laser intensity and the divergence of the laser beam. Unfortunately, the peak power of CVLs ($W \sim 10^4$–10^5 W) is fairly low as compared, e.g., to solid lasers.

Possibly, this is the reason why, in most of the experiments (see, e.g., [1, 3]), the CVL radiation was focused into nonlinear crystals (BBO, KDP, DKDP, and others) with a lens or more complicated optical system in order to increase as much as possible the intensity of laser radiation and thus to achieve a high conversion efficiency. However, a convergent laser beam is known to display some disadvantages in comparison with a parallel one [4], because, in the former case, the conversion efficiency is strongly affected by the increased divergence of the beam in the waist, diaphragm and diffraction losses, and nonuniform heating of a crystal.

In our previous papers [5–8], we described the experiments on the SF ($\lambda_3 = 0.271$ μm) generation with a parallel laser beam of 1–2 mm in diameter that was formed with a telescopic collimator and directed into a crystal. Here, we present the results of experiments carried out with strong focusing of a laser beam for the same CVL and crystal parameters.

EXPERIMENTAL RESULTS

The nonlinear DKDP crystal was 4 cm in length and 1 cm in diameter. The temperature of the crystal was stabilized at a level of 333 K with the help of a thermostat. The optical layout of the experimental setup is shown in Fig. 1. The CVL was a commercial CL-201 laser tube placed in a telescopic unstable resonator with magnification of $M = 200$ or $M = 5$ (mirrors *2*, *3*). An intracavity Glan prism served as a polarizer (in the case of $M = 5$, the resonator was not supplied with a prism and laser radiation was not polarized). The diameter D of the output laser beam was 20 mm. The average output CVL power was 13 or 20 W for unstable resonators with $M = 200$ or $M = 5$, respectively. With the use of

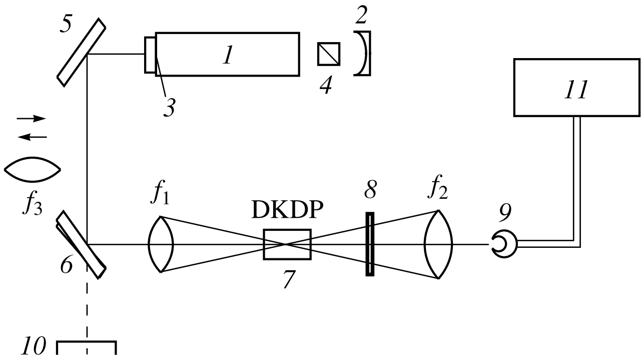


Fig. 1. Optical layout of the experiment: (*1*) CVL, (*2*, *3*) resonator mirrors, (*4*) Glan prism, (*5*, *6*) deflecting mirrors, (*7*) nonlinear crystal, (*8*) UFS5 optical filter, (*9*) sensing element of the power meter, (*10*) screen, (*11*) IMO-ChS power meter, (f_1) focusing lens, (f_2) CaF_2 lens, and (f_3) long-focus lens.

1063-7842/00/4504-0467$20.00 © 2000 MAIK "Nauka/Interperiodica"

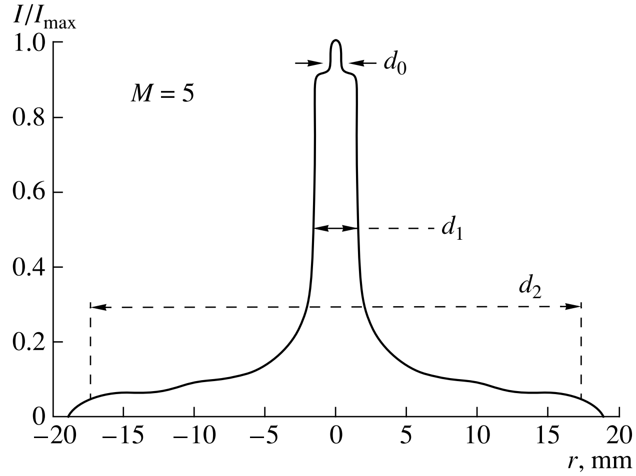


Fig. 2. Radial distribution of the intensity of CVL radiation in the focal spot of the lens for $f_3 = 10$ m and $M = 5$.

plane mirrors and lens f_1, the CVL radiation was directed and focused into the crystal. The center of the beam waist was located approximately in the middle of the crystal on its optical axis. Laser radiation with wavelengths λ_1 and λ_2 was absorbed by optical filter *8*. UV radiation with a wavelength $\lambda_3 = 0.271$ μm passed through the filter and was focused with lens f_2 made of CaF_2 onto a sensing element of a power meter. The focal length of lens f_1 was varied.

The divergence and structure of the CVL beam in the far-field zone was measured using lens f_3 with a focal length of 10 m. The radial distribution of the laser intensity in the focal plane of the lens (screen *10*) was obtained by scanning the sensing element of the power meter along the radius of the focal spot. The input aperture of the sensing element was limited with a 0.2-mm-diameter diaphragm. The calibrated-diaphragm method was applied to evaluate the diameters of the focal spots when lenses with smaller focal lengths were employed.

Table 1

$M = 200$		$M = 20$	
φ, mrad	Δ*E*/*E*, %	φ, mrad	Δ*E*/*E*, %
7.7	100	7.5	100
2.0	50	1.5	70
0.25	25	0.3	70
0.07	10	–	–

Table 2

	M	*P*, W	W_1, kW	W_2, kW	P_3, W	η, %
Parallel beam	200	6.2	20	15	0.75	12
	5	12.0	40	28	0.2	3.3
Strong focusing	200	6.1	22	16	0.34	5.9
	5	16.3	54	38	0.16	2.0

Four concentric spots were resolved by the gradation of the laser intensity *I* in the far-field zone, on screen *10* in the case of the $M = 200$ unstable resonator. Only three spots could be resolved when the $M = 5$ unstable resonator was used. Figure 2 shows the radial distribution of the intensity *I* of laser radiation. The values of the angular divergence of the laser beams corresponding to these spots and the fractions $\Delta E/E$ of the laser pulse energy (summed over λ_1 and λ_2) falling on these laser beams are presented in Table 1. Such a distribution is typical of lasers with unstable resonators using an active medium with a short (of ~30–40 ns) lifetime of inverted population (in this case, the laser beams are formed during the first three-four passages through the resonator and are delayed by the one-passage time with respect to each other).

Figure 3 shows the spot diameters d_0, d_1, and d_2 measured in the focal planes of the lenses with f_1 varying from 3 to 160 cm. These focal spots correspond to laser beams in the far-field zone and contain 10% ($\varphi = 0.07$ mrad), 25% ($\varphi = 0.25$ mrad), and 50% ($\varphi = 2.0$ mrad), respectively, of the total energy of a laser pulse produced in the resonator with $M = 200$. Similarly, d_1 and d_2 represent the laser beams containing 30 and 70%, respectively, of the laser pulse energy in the case of $M = 5$. It is these laser beams that took part in the SF generation, because a substantial fraction of the beam with a higher divergence of ~7–8 mrad is cut off by the aperture of the focusing lenses and the crystal. Note, that the minimum value of $\varphi = 0.07$ mrad obtained using the $M = 200$ resonator is twice as high as the diffraction divergence of the laser beam. For $M = 5$, the diffraction core of the laser beam was less pronounced (Fig. 2) and the fraction of the laser pulse energy in it was low (1–2%).

Figure 4 shows the SF generation efficiency η, which was defined as the ratio of the average power of UV radiation P_3 to the average power of the focused two-frequency CVL radiation *P* at the input aperture of the crystal (or to the half of the CVL power in the case of $M = 5$, because, in this case, laser radiation was not polarized).

For comparison, Table 2 lists the maximum values of η and P_3 achieved in our experiments with strong focusing of CVL radiation and those obtained using a parallel laser beam [5–8] for almost the same values of the average (*P*) and peak (W_1 and W_2) CVL powers at the input aperture of the crystal (for λ_1 and λ_2, respectively).

DISCUSSION AND CONCLUSIONS

With strong focusing of the CVL radiation into a crystal, laser beams with different divergences have focal spots with substantially different diameters

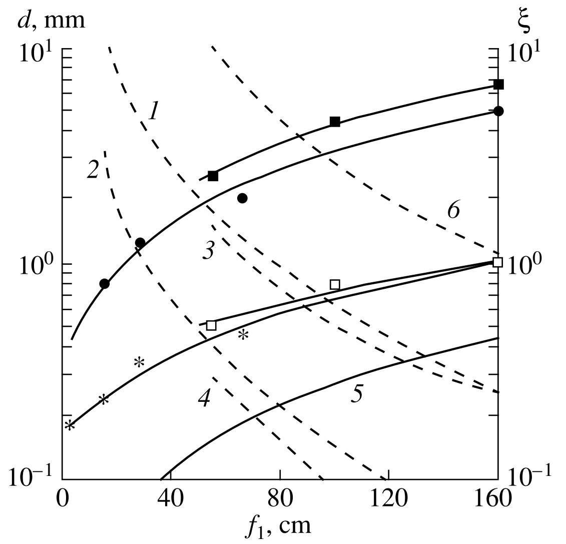


Fig. 3. The diameters of focal spots, d_0, d_1, and d_2, and the focusing parameter ξ as functions of the focal length of the lens for $M = 5$ [(∗) d_1, (●) d_2, (*1*) $\xi(d_1)$, and (*2*) $\xi(d_2)$] and $M = 200$ [(□) d_1, (■) d_2, (*3*) $\xi(d_1)$, (*4*) $\xi(d_2)$, (*5*) $2d_0$, and (*6*) $\xi(d_0)$].

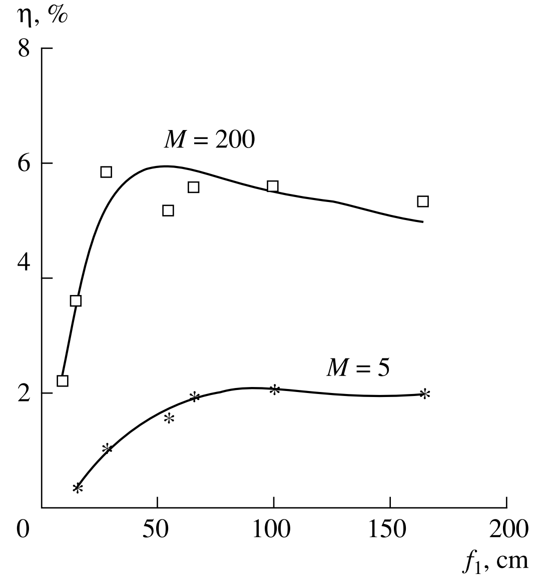


Fig. 4. Efficiency of the SF generation under conditions of strong focusing of CVL radiation into the crystal.

(d_0, d_1, and d_2; Fig. 3) and laser intensities and different locations of the beam waists on the optical axis of the crystal. It is known [4], that the conversion efficiency obtained using a focused beam depends on the laser intensity in the focal spot and on the focusing parameter ξ. The location of the focal spot in the middle of the crystal is optimum. Figure 3 shows the focusing parameter $\xi = LD/2f_1d$ as a function of the focal length f_1 for laser beams with different values of the diameter d (and, accordingly, different values of φ) corresponding to the experimental values of d_0, d_1, and d_2. It is seen that the values of ξ for laser beams with different divergences are different; therefore, they can not take an optimum value simultaneously. Thus, for the conversion efficiency $\eta = 6\%$, $M = 200$, and $f_1 = 550$ mm (see Fig. 2), the value of ξ changes from 0.3 for d_2 to 10 for d_0.

In our experiments, the maximum values of the UV radiation power ($P_3 = 0.34$ W) and the conversion efficiency ($\eta = 6\%$) under conditions of strong focusing were achieved for $M = 200$ (see Table 2). Presumably, this is due to the fact that, for high values of M, a beam with divergence close to the diffraction divergence appears, the fraction of the laser beam energy contained in low-divergence beams increases and, therefore, the laser intensity rises in the center of the focal spot.

In the experiments with parallel beams [5–8], up to 1.5–2 times higher values of η and P_3 were achieved (see Table 2) using almost the same energetic parameters of the CVL beam at the input of the crystal. Note, that the diameter of the parallel beam (1–2 mm) is comparable with the diameter of the beam waist ~d_2 under conditions of strong focusing, whereas the radial and longitudinal distributions of the laser intensity differ substantially for parallel and convergent beams.

A comparison of the results obtained allows us to conclude that, at least under conditions of our experiments, the efficiency of nonlinear frequency conversion of the CVL radiation is higher for a parallel laser beam than in the case of strong focusing into a crystal. Apparently, in both cases, an increase in the conversion efficiency can be achieved with a CVL operating in a one-beam mode, when most of the laser pulse energy is concentrated in the diffraction cone. It is also necessary to increase the peak CVL power keeping in the average power at the same level.

REFERENCES

1. J. A. Piper, *Pulsed Metal Vapor Lasers* (NATO ASI Series), Ed. by C. E. Little and N. V. Sabotinov (Kluwer Academic, London, 1996), p. 277.
2. E. K. Illy and J. A. Piper, Photonics Spectra, No. 3, 106 (1998).
3. D. W. Coutts and D. J. W. Brown, IEEE J. Sel. Top. Quantum Electron. **1**, 768 (1995).
4. V. G. Dmitriev and L. V. Tarasov, *Applied Nonlinear Optics* (Radio i Svyaz’, Moscow, 1982).
5. V. T. Karpukhin, Yu. B. Konev, and M. M. Malikov, Proc. SPIE **2502**, 172 (1995).
6. V. T. Karpukhin, Yu. B. Konev, and M. M. Malikov, Opt. Atmos. Okeana **8**, 1652 (1995).
7. V. T. Karpukhin and M. M. Malikov, Opt. Atmos. Okeana **11**, 181 (1998).
8. V. T. Karpukhin, Yu. B. Konev, and M. M. Malikov, Kvantovaya Électron. (Moscow) **25**, 809 (1998).

Translated by A. P. Lytkin

Technical Physics, Vol. 45, No. 4, 2000, pp. 470–475. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 90–94.
Original Russian Text Copyright © 2000 by Bratman, Glyavin, Goldenberg, Savilov.

RADIOPHYSICS

The Spread of the Initial Energy of Electrons in a Gyrotron Due to the Negative-mass Instability Developing in a Magnetron–Injector Gun

V. L. Bratman, M. Yu. Glyavin, A. L. Goldenberg, and A. V. Savilov

Institute of Applied Physics, Russian Academy of Sciences,
Nizhniĭ Novgorod, 603600 Russia

Received June 3, 1997; in final form, February 22, 1999

Abstract—The formation of an electron beam in a magnetron-injector gun of a gyrotron is investigated in the case when it is affected by the negative-mass instability due to the Coulomb repulsion and nonisochronous cyclotron rotation of particles. A technique is proposed for calculating the spread of the initial energy of electrons caused by the instability, which develops as the electron beam moves in the presence of a nonuniform magnetostatic field of the magnetron-injector gun. It is demonstrated that this instability can be one of the main factors providing the energy spread in electron guns of gyrotrons. *© 2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

In gyrotrons [1–3], a spread of the particle energy is observed in the region where the electron beam forms and drifts to the operating resonator. One of the possible factors which cause this spread is the negative-mass instability (NMI) [1, 4]. The NMI, which was first found in cyclic charged-particle accelerators [5, 6] and has long been used for explaining the effects of the RF space charge in the operating region of high-current gyrotrons (see, e.g., [7–12]), in principle, can convectively amplify perturbations and provide an energy spread in the electron beam during its formation and motion from the cathode to the operating resonator. The NMI seems to have a more universal nature because it develops regardless of the form of electron-optical system and the presence of scattered electrons in the formation region. The NMI is due, on the one hand, to the non-isochronous cyclotron rotation of electrons in the presence of a magnetostatic field and, on the other hand, to their Coulomb repulsion [5, 6]. In fact, a fluctuation of the electron density repulses neighboring particles, increasing or decreasing their oscillation velocity, depending on their position with respect to the fluctuation (Fig. 1). Thus, the linear velocity of the particles following the fluctuation decreases. Simultaneously, their angular velocity increases, because the cyclotron rotation frequency

$$\omega_B = \frac{eBc}{E} \qquad (1)$$

is in the inverse proportion to relativistic electron energy E. At the same time, the angular velocity of the particles moving in front of the fluctuation decreases. Thus, the Coulomb repulsion of particles oscillating in the magnetostatic field provides their efficient rotation-phase bunching and, hence, increases the fluctuations of the electron density. The spread in the particle energy increases simultaneously with the fluctuation.

In this paper, we investigate the NMI developing in the formation region of the electron beam in a gyrotron and give a theoretical interpretation for the measurements of the spread of the initial electron energy [13]. A technique is proposed for calculating the spread of the initial electron energy due to the NMI developing when the electron beam moves in a nonuniform magnetostatic field of the gyrotron magnetron-injector gun (MIG).

THE NMI IN AN ELECTRON BEAM MOVING IN A UNIFORM MAGNETIC FIELD

Here and below, the state of an electron beam, in which particles are uniformly distributed in the initial rotation phases φ on each Larmor orbit, is referred to as an unperturbed state. Due to periodic behavior with

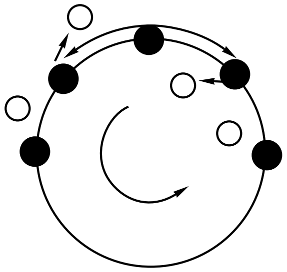


Fig. 1. The NMI mechanism [5, 6].

respect to φ, an arbitrary initial perturbation of rotation phases $\vartheta(\varphi)$ (a perturbation of the electron density in the proximity of the cathode) can be represented in the form of the Fourier series

$$\vartheta(z=0,\varphi) = \sum_{n=1}^{+\infty} \vartheta_n \cos(\phi_n + n\varphi). \tag{2}$$

In this series, the nth term corresponds to the nth perturbation cyclotron harmonic with the initial perturbation amplitude ϑ_n and phase ϕ_n. At the linear stage of the instability, when particle phases slightly differ from the unperturbed values $\vartheta(z) \ll \pi$, perturbation harmonics grow independently of each other. When the guiding magnetostatic field is uniform, the perturbation cyclotron harmonics increase with the longitudinal coordinate as [4]

$$\vartheta_n(z) = \frac{1}{2}\vartheta_n(0)[\exp(\Gamma_n z/r_c) + \exp(-\Gamma_n z/r_c)]$$

with the constant increment

$$\Gamma_n = g_c\sqrt{sL_n}. \tag{3}$$

Here,

$$s = \frac{eI}{mc^3}\frac{r_c}{R_c}\frac{2}{\pi\beta_{z_c}}$$

is the parameter of the current;

$$L_n = \frac{4n^2}{\pi(4n^2-1)};$$

r_c, R_c, and $g_c = v_{\perp c}/v_{z_c}$ are the electron Larmor radius, the beam radius, and the electron pitch factor in the proximity of the cathode; and $\beta_{z_c} = v_{z_c}/c$. The variance of the electron energy

$$D = \sqrt{\langle(1 - v^2/v_0^2)^2\rangle_\varphi}$$

(which is coupled with the measured spread $\delta\varepsilon$ [13] by the relationship $\delta\varepsilon = 2.56D$ in the case of the Gaussian particle energy distribution) is determined by the sum

$$D_{\text{lin}}(z) = \frac{\sqrt{2}}{g_c\beta_0^2}\sum_{n=1}^{+\infty}\vartheta_n(0)\Gamma_n\sinh(\Gamma_n z/r_c), \tag{4}$$

where $\beta_0^2 = (v_0/c)^2 = 2eU_0/mc^2$ is the normalized unperturbed energy of the particle.

For an ideal beam, increments Γ_n slightly depend on the harmonic number. However, taking into account the initial energy and pitch-factor spreads, as well as the position spread of particle guiding centers, yields a rapid decrease in the increment with the increasing harmonic number [14, 15]. Thus, perturbations at the fundamental harmonic of the cyclotron frequency are the governing factor for an actual beam.

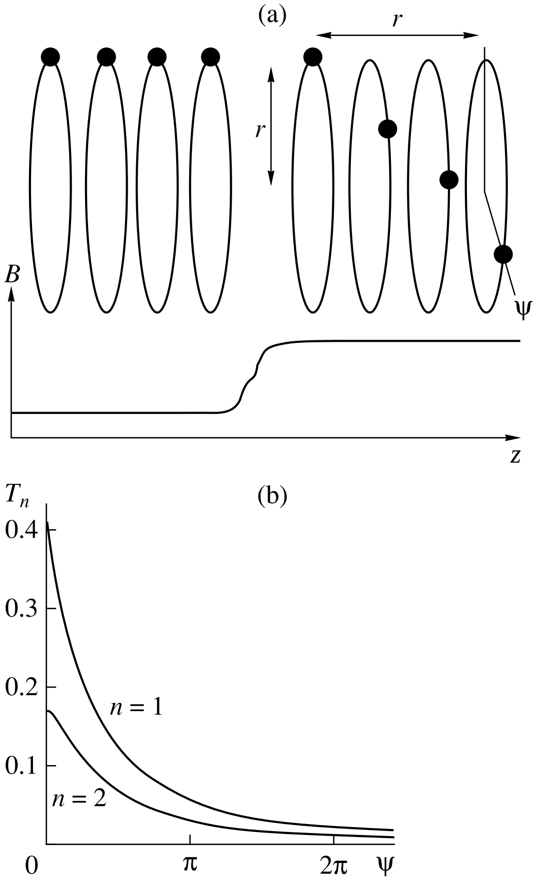


Fig. 2. (a) The spatial distribution of rotation phases of electron bunches formed due to the developed NMI and the transformation of a uniform electron density perturbation into a nonuniform one caused by a nonuniform magnetic field; (b) cyclotron harmonics T_n of the effective Coulomb field vs. nonuniformity parameter ψ of an electron density perturbation.

Consider the dependence of the spread due to the NMI at the linear stage of its development on the electron current and guiding magnetic field in the case of a fixed cathode voltage. When the pitch factor is small in the proximity of the cathode ($g_c \ll 1$), (4) implies

$$D_{\text{lin}} = A_1\frac{\sqrt{I}}{B}\sinh(A_2\sqrt{I}z) \approx \frac{A_1}{B}\left(A_2 Iz + \frac{A_2^3}{6}I^2z^3\right), \tag{5}$$

where $A_{1,2}$ are dimensional coefficients.

Thus, the energy spread is in inverse proportion to the magnetic induction and increases approximately in direct proportion to the current, which agrees with the experimental results [13] (see Fig. 3 in this paper and Figs. 2 and 4 in [13]).

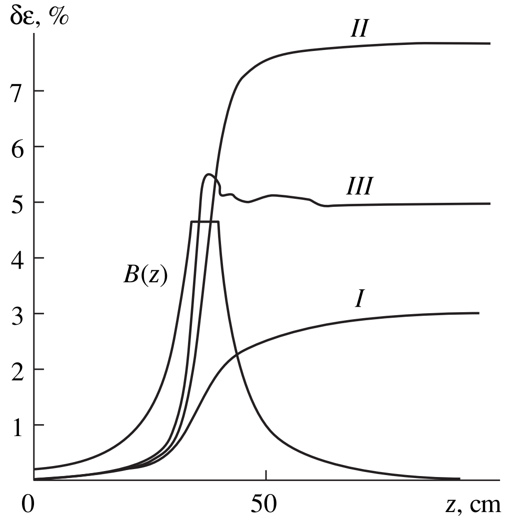


Fig. 3. Energy spread calculated for the gyrotron investigated experimentally in [13]. Magnetic field $B(z)$ and current energy spread vs. coordinate; (*I*) the modeling mode, (*II*) the operating mode, and (*III*) the operating mode of the gyrotron with the operating frequency twice as large as the frequency of the experimental gyrotron.

Next, we investigate the energy spread (due to the NMI) affected by the varying magnetic field, voltage, and current when the modeling mode is replaced by the operating one, and the corresponding values are coupled by the similitude relationships [13]

$$B_{\text{mod}} = B_{\text{op}}/k^{1/2}, \quad U_{\text{mod}} = U_{\text{op}}/k, \quad I_{\text{mod}} = I_{\text{op}}/k^{3/2}.$$

Formula (4) implies the following relationship for spread in the operating mode

$$D_{\text{lin}}^{(\text{op})}(z) = \frac{C_1}{\sqrt{k}}\sinh(C_2\sqrt{k}z) \approx D_{\text{lin}}^{(\text{mod})}(z) + (k-1)\frac{C_1 C_2^3}{6}z^3, \tag{6}$$

where $C_{1,2}$ are dimensional coefficients.

Thus, at the linear stage of the NMI development, spread $D_{\text{lin}}^{(\text{op})}$ in the operating mode is greater than spread $D_{\text{lin}}^{(\text{mod})}$ in the modeling mode. However, the spreads in both modes are close to each other as long as the length of the drift region is small enough.

THE NMI IN AN ELECTRON BEAM MOVING IN A NON-UNIFORM MAGNETIC FIELD

The results obtained above also hold when the variation of the magnetic field with the longitudinal coordinate is taken into account, i.e., when $B = B(z)$. However, in this case, the NMI increments become functions of the coordinate which is due to two effects. First, the unperturbed characteristics of the electron beam (the space charge density, the cyclotron frequency of the particle rotation, and the pitch factor) vary with length. Second, the space structure of the electron density perturbations evolves (i.e., the degree of nonuniformity along the longitudinal coordinate increases) as the beam moves through the drift region [15, 16]. This effect is attributed to the magnetic field nonuniformity. In fact, in the presence of a uniform magnetic field, uniform perturbations (in which the electron bunches due to the NMI have equal rotation phases of their centers in each cross section of the electron beam) are characterized by the maximum increments [17]. However, when the beam moves along a nonuniform magnetic field, at each moment, the frequencies of the cyclotron rotation prove to be different for electrons located in different cross sections of the beam. As a result, the uniform perturbations observed near the cathode become nonuniform (Fig. 2a).

When the electron beam moves along the magnetic field varying with the coordinate, formula (4) can be modified to the form [16]

$$D_{\text{lin}}(z) = \frac{\sqrt{2u(z)}}{\beta_0^2 b(z)}\sum_{n=1}^{+\infty}\vartheta_n\Gamma_n(z)\sinh\left(\int_0^z\Gamma_n(z)dz/r_c\right). \tag{7}$$

Here, $b(z) = B(z)/B_c$ is the longitudinal component of the magnetic field normalized to the initial (observed in the proximity of the cathode) value and $u(z) = v_z/v_{\perp_c} = \sqrt{g_e^{-2} + 1 - b(z)}$ is the normalized longitudinal velocity of particles.

An instantaneous increment value (observed in a given cross section of the electron beam) is specified by the expression

$$\Gamma_n(z) = \sqrt{sb^2(z)\frac{u_c}{u^3(z)}}\sqrt{nT_n(\psi)}. \tag{8}$$

The first factor depending on the coordinate describes the variation of the unperturbed beam characteristics caused by the nonuniformity of the magnetic field. This factor reaches its maximum when the magnetic field has the maximum value. The evolution of the space structure of electron density perturbations (which occurs while the beam moves through the system) is described by the nth cyclotron harmonic of the effective Coulomb field varying with the coordinate (Fig. 2b) [16, 17]

$$T_n(\psi) = \frac{1}{2\pi}\int_0^{+\infty}d\zeta\int_0^{\pi}d\theta\cot\left(\frac{\theta}{2}\right)\left[1 - \frac{\zeta}{\sqrt{\zeta^2 + 4\sin^2(\theta/2)}}\right] \times \sin(n\theta)\cos(n\psi\zeta).$$

Here, the parameter $\psi = \pi l_z/r$ is the ratio of the characteristic longitudinal size of nonuniform electron density perturbations to the current Larmor electron radius

[15–17]. For stationary (time-constant) initial perturbations of the electron density, the nonuniformity factor coincides with the current electron pitch factor [16], i.e., $\psi(z) = g(z)$.

In the proximity of the cathode, where the pitch factor is small, the perturbations of the electron density are nearly uniform and $\psi \ll \pi$. Therefore, in this region, the NMI increments specified by (8) are identical to the increments observed in the uniform magnetic field

$$nT_n(\psi = 0) = L_n.$$

The magnetic field growing with the coordinate increases the nonuniformity of electron density perturbations, and the cyclotron harmonics of the Coulomb field decrease (Fig. 2b). This decrease becomes very substantial when the non-uniformity factor reaches the value $\psi = \pi$, which corresponds to the nonuniformity size equal to the Larmor electron radius.

Thus, at the linear stage of the instability development, increment (8) increases with magnetic field $B(z)$, which is caused by the increase in the oscillation component of the velocity, the increase in the space charge density due to the decrease in the cross-section area of the beam (the decrease in the Larmor and beam radii), the decrease in the longitudinal velocity of particles, and, correspondingly, a longer time interval during, which the Coulomb field of the beam affects the particle. This growth of the increment is compensated in part, by decreasing cyclotron harmonics T_n of the effective Coulomb field. At the same time, the maximum variance of the electron energy observed at the stage of nonlinear instability saturation slightly depends on the magnetic field in the saturation region and the profile of the magnetic field. The maximum variance is determined by [16]

$$D_{\text{sat}} \approx \frac{\sqrt{s}}{\beta_0^2}. \qquad (9)$$

Taking into account that perturbations increase mainly in the region where the magnetic field has its maximum (Fig. 3), and comparing (7) and (9), we can estimate the characteristic length of this region which provides the stage of the NMI nonlinear saturation,

$$\Delta z \approx \frac{r_c}{\Gamma_1} \ln\left[\frac{1}{\vartheta_1}\sqrt{\frac{u}{2u_c T_1}}\right], \qquad (10)$$

where values Γ_1, T_1, and u correspond to the maximum magnetic field.

Naturally, this length depends on initial perturbation level ϑ_1. Note that, when passing from the modeling to operating mode, the spread observed at the saturation stage decreases while the rate at which the instability approaches the nonlinear mode increases,

$$D_{\text{sat}}^{(\text{op})} = D_{\text{sat}}^{(\text{mod})}/\sqrt{k}, \quad \Delta z^{(\text{op})} = \Delta z^{(\text{mod})}/\sqrt{k}. \qquad (11)$$

CALCULATION OF THE ELECTRON ENERGY SPECTRUM

Using the equations describing the NMI development in an electron beam, which moves along a nonuniform magnetic field [16], we calculate the electron energy spread due to the NMI. This energy spread increases as the electron beam moves in the gyrotron [13]. In calculations, we take into account the effect of the guiding nonuniform magnetic field on the unperturbed characteristics of the beam and spatial structure of electron density perturbations. The electron density perturbation at the first harmonic is specified in the form

$$\theta_0 = \varphi + \vartheta_1 \sin\varphi, \quad \varphi \in [0.2\pi),$$

and, then, the electron energy spread observed at the structure output is calculated as a function of the electron current for different magnetic fields.

It follows from estimates (9) and (10) that, in the modeling mode, the characteristic spread corresponding to the nonlinear saturation stage is $\delta\varepsilon = 23\%$ for the electron current $I = 30$ A, which substantially exceeds the experimental values of the output energy spread [13]. This means that the instability develops linearly along the entire length of the system. In this case, the spread depends on the initial (observed near the cathode) level ϑ_1 of the electron density perturbations, which is, generally speaking, a priori unknown. Thus, the characteristic initial noise should be determined by providing an agreement of calculations with experimental results. Computations performed for different guiding magnetic fields (Fig. 3) show that the experimental results fit well to the dependences calculated for the level of initial perturbations $\vartheta_1 = 0.1$. The constancy of this level for the varying magnetic field is consistent with the assumption that the energy spread is caused by the NMI.

Figure 3 shows the current energy spread versus the coordinate along which the electron beam moves in the structure. The energy spread is calculated for the modeling and operating modes when the initial perturbation level and current are, respectively, $\vartheta_1 = 0.1$ and $I = 30$ A. According to expression (8) for the current increment, the maximum increase in the spread is observed in the resonator where the magnetic field reaches its maximum. It follows from estimate (10) that the nonlinear stage of the instability is provided in the modeling mode when the characteristic size of the region where the magnetic field reaches its maximum is $\Delta z^{(\text{mod})} \approx 40$ cm, which substantially exceeds the corresponding measured value. However, in the operating mode, this length is much smaller than $\Delta z^{(\text{op})} \approx 10$ cm for the same values of the electron current and initial perturbation level. This means that, in the operating mode, this length of the structure is sufficient for the instability to reach the saturation stage. According to estimate (9) and calculations (Fig. 4), the expected

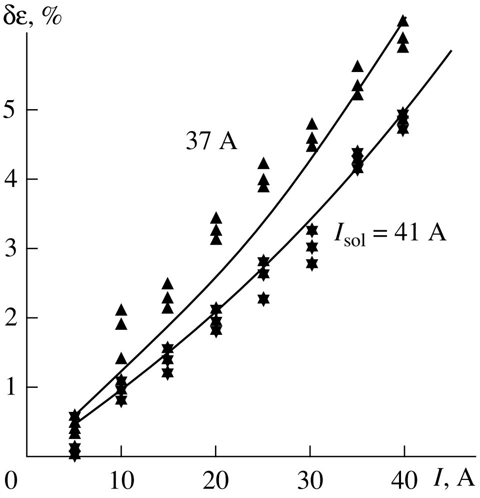


Fig. 4. Electron energy spread vs. current in the gyrotron operating in the modeling mode obtained for different solenoid currents. Comparison of experimental results [13] with calculations.

spread observed at the saturation stage is $\delta\varepsilon$ = 7–8% for the currents I = 30–40 A.

The spread governs the efficiency of the interaction of electrons with the RF field in the gyrotron resonator. As far as this influence is concerned, the spread observed at the resonator input (in the proximity of the maximum magnetic field) is a more important characteristic than the spread near the collector. In the experimental gyrotron, the spreads observed at the resonator input in the operating and modeling modes are almost identical (Fig. 3) according to formula (7) and their values are less than two percent. This spread, virtually, does not deteriorate the gyrotron performance. However, in the case of a smooth profile of the magnetic field when the region in which the electron beam drifts from the cathode to the resonator is rather long, the spread observed at the resonator input may increase up to the value $\delta\varepsilon$ = 7–8% corresponding to the NMI saturation stage.

Note that, in shorter-wavelength gyrotrons, the electron-beam formation region is usually longer. In this case, obviously, the relative length measured in inverted increments r_c/Γ_1 also increases. Correspondingly, the electron energy spread observed at the resonator input may noticeably increase. Thus, Fig. 3 displays (curve *III*) the spread versus the coordinate for the operating mode of the gyrotron with the magnetic field twice as large as the magnetic field in the device treated in this work (the rest parameters of these gyrotrons are identical), which models a change of the operating frequency to a value of about 170 GHz. Formula (9) yields the spread observed in this device at the instability saturation stage which is less by a factor of $\sqrt{2}$ than the spread in the experimental gyrotron. However, formulas (3) and (4) imply that the NMI increment Γ_1/r_c at the linear stage is greater by a factor of $\sqrt{2}$ in the gyrotron operating at the higher frequency. This effect increases the spread at the resonator input almost by a factor of two (as compared with to experimental gyrotron).

CONCLUSION

The NMI theory enables one to estimate the energy spread of particles in the region where the gyrotron electron beam is formed. The electron energy spreads are measured [13] and calculated using the NMI theory as functions of the current and guiding magnetic field. The experimental data are in a qualitative agreement with the results obtained by calculations, which allows us to suppose that the electron beam instability is unlikely to be attributed to other possible factors (at least in the modeling mode). The NMI theory demonstrates that the electron energy spectrum continuously extends as the beam moves towards the collector even in the absence of RF oscillations. A comparison of the theory and experiment yields a characteristic initial fluctuation of the electron density which has been unknown so far. This makes it possible to estimate the NMI effect in other gyrotrons in different sections of the electron beam.

For the gyrotron investigated in [13], this estimate shows that the NMI extends the electron energy spectrum observed at the resonator input by no more than 2%. This extension slightly affects the gyrotron efficiency. However, in principle, the extension of the electrons, drift length from the cathode to the resonator may increase the spread at the resonator input up to 7–8%, corresponding to the NMI saturation. This relative (with respect to the characteristic electron orbit) increase in the drift region is observed in gyrotrons designed for higher frequencies and larger radius of the main solenoid where the NMI can substantially impair the gyrotron performance.

REFERENCES

1. A. V. Gaponov, M. I. Petelin, and V. K. Yulpatov, Izv. Vyssh. Uchebn. Zaved., Radiof. **10**, 1414 (1967).
2. V. A. Flyagin and G. S. Nusinovich, Proc. IEEE **76**, 644 (1988).
3. *Gyrotrons*, Ed. by V. A. Flyagin (IPF AN SSSR, Gorki, 1989), p. 216.
4. V. L. Bratman, Zh. Tekh. Fiz. **46**, 2030 (1976) [Sov. Phys. Tech. Phys. **21**, 1188 (1976)].
5. C. Nielsen and A. M. Sessler, Rev. Sci. Instr. **30**, 80 (1959).
6. A. A. Kolomenskiĭ and A. N. Lebedev, At. Énerg. **7**, 549 (1959).
7. I. S. Kovalev, A. A. Kuraev, S. V. Kolosov, *et al.*, Dokl. Akad. Nauk BSSR **17**, 416 (1973).
8. V. L. Bratman and M. I. Petelin, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **18**, 1538 (1975).

9. V. I. Kanavets and V. A. Cherepenin, Radiotekh. Élektron. (Moscow) **20**, 2539 (1975).
10. J. L. Hirshfield, Int. J. Infrared Millimeter Waves **2**, 695 (1981).
11. R. G. Kleva, T. M. Antonsen, Jr., and B. Levush, Phys. Fluids **31**, 375 (1988).
12. V. L. Bratman and A. V. Pylin, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **32**, 1408 (1989).
13. N. P. Venediktov, M. Yu. Glyavin, A. L. Goldenberg, *et al.*, Zh. Tekh. Fiz. **70**, 95 (2000).
14. V. L. Bratman and A. V. Savilov, Zh. Tekh. Fiz. **64**, 154 (1994) [Tech. Phys. **39**, 602 (1994)].
15. V. L. Bratman and A. V. Savilov, Phys. Plasmas **2**, 557 (1995).
16. A. V. Savilov, Phys. Plasmas **4** (6), 2276 (1997).
17. A. V. Savilov, Int. J. Infrared Millimeter Waves **15**, 1819 (1994).

Translated by I. G. Efimova

Technical Physics, Vol. 45, No. 4, 2000, pp. 476–479. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 95–98.
Original Russian Text Copyright © 2000 by Venediktov, Glyavin, Goldenberg, Zapevalov, Kuftin, Postnikova.

Measurements of the Spread in the Initial Electron Energy in a Gyrotron

N. P. Venediktov, M. Yu. Glyavin, A. L. Goldenberg, V. E. Zapevalov,
A. N. Kuftin, and A. S. Postnikova

Institute of Applied Physics, Russian Academy of Sciences, ul. Ul'yanova 46, Nizhniĭ Novgorod, 603600 Russia

Received June 3, 1997; in final form, February 22, 1999

Abstract—An analyzer intended for measuring electron energy in gyrotrons is described. Electron energy spectra are measured in various operating modes of an experimental gyrotron. It is shown that, even when no microwave field is generated, the spread in electron energy due to the space-charge effect can be as high as several percent. *© 2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

It is well known [1–3] that gyrotrons exhibit a low sensitivity to the spread in the translational (along the guiding magnetic field) electron velocity. This property is attributed to quasi-transverse propagation of the operating wave, which ensures that the Doppler spread of cyclotron resonance is small. At the same time, it is clear that, since the dependence of the electron gyrofrequency on energy E,

$$\omega_B = \frac{eBc}{E} \tag{1}$$

is relativistic, the weakly relativistic gyrotron, in whose operating space electrons usually perform a great number ($N \gg 1$) of gyrations, must be very sensitive to the spread in the initial electron energy. A simple estimate $\Delta E/E \ll 1/N$, which immediately follows from the requirement that the cyclotron resonance be maintained for all of the electrons is confirmed, e.g., by calculations [4], which show that a several-percent spread in the initial energy significantly degrades gyrotron efficiency. There are a number of effects that can cause a noticeable spread in the electron energy before they enter the gyrotron operating space, among them, the excitation of spurious oscillations and the onset of instabilities in a dense beam in the beam shaping space. Probably, it is these effects that lead to the reduction of the efficiency of high-power gyrotrons as compared to the theoretically predicted value.

This paper is concerned with the measurements of the energy spread of the electron beam in a gyrotron.

MEASUREMENT TECHNIQUE AND EXPERIMENTAL SETUP

The electron energy spectrum can be determined experimentally by the decelerating-field method, allowing one to find the translational electron velocity distribution from the beam current that hits the target as a function of the decelerating voltage. Decelerating-field analyzers mounted in the cavity [4] allow one to estimate the spread in the velocity components under the assumption that the beam electrons have the same energy. Usually, such measurements can be performed only in a gyrotron operating in a special (modeling) mode, i.e., at a decreased magnetic field B, beam voltage U, and beam current I. These parameters are chosen based on the similitude relationships,

$$B_{\text{mod}} = B_{\text{op}}/k^{1/2},\ U_{\text{mod}} = U_{\text{op}}/k,\ I_{\text{mod}} = I_{\text{op}}/k^{3/2}, \tag{2}$$

which provide the invariance of the unperturbed electron trajectories [i.e., trajectories calculated with no regard of the negative-mass instability (NMI)]. Here, k is the modeling factor, the "mod" subscript stands for the modeling mode, and the "op" subscript stands for the normal operating mode, in which a gyrotron with the given electron gun has the design output power and efficiency. Distribution of the electron velocity components in a helical gyrotron beam is characterized by the relative spread in the rotational velocity,

$$\delta v_\perp = \Delta v_\perp/\bar{v}_\perp,$$

where $\bar{v}_\perp$ is the average rotational velocity and $\Delta v_\perp$ is the velocity range that does not include the regions corresponding to the lowest and highest electron velocities (these regions carry 10% of the total beam current each).

In high-power gyrotrons, the spread $\delta v_\perp$ is about 20% or a little higher, which degrades the efficiency from 50–60% (which can be achieved when there is no velocity spread) to 30–40%.

When the measurements are carried out in the cavity, the gyrotron should operate in the modeling mode, because, for the beam parameters corresponding to the normal operating mode (even in the pulsed regime), it is impossible to prevent the target from breakdown or destruction due to small dimensions of the analyzer components and gaps between them. In contrast, an analyzer proposed in [5] and employed in this study allows measurements not only in the modeling mode but also at the operating beam voltages (on the order of several tens of kilovolts) and is also capable of measuring the spread in the full electron velocity.

The analyzer (Fig. 1) is placed in a tube branched off from the collector of a gyrotron described in [6, 7] at an angle equal to the inclination angle of the magnetic field lines. The electrons enter this tube through a narrow longitudinal slot in the collector and move toward the target placed perpendicular to the magnetic field lines (in this region, the magnetic field is much lower than in the collector and the more so than in the cavity). As the electrons travel, they are decelerated by the electric field; the dependence of the target current on the decelerating voltage is used to calculate the electron velocities.

An analysis of trajectories shows that electron motion in the beam is sufficiently accurately described in terms of the adiabatic approximation until the electrons meet the analyzer target. In the collector slot and near the target, respectively, the magnetic field is by a factor of 25 and 400 weaker than in the cavity. Using the adiabatic invariant $v_\perp^2/B$ = const, we obtain that, in the decelerating field of the analyzer, electron rotational velocities are about 20 times lower than the translational velocities. The decelerating-field method is thus capable of determining the full electron velocity (energy) at which the electrons enter the collector. The electron energy distribution function is calculated by differentiating the collector current with respect to the voltage of the decelerating grid. We define the energy spread as

$$\delta\varepsilon = \frac{U_{0.1} - U_{0.9}}{U_0},$$

where $U_{0.9}$ and $U_{0.1}$ are the grid voltages at which the target current is, respectively, 0.9 and 0.1 times its maximum value and U_0 is the total beam voltage, which determines the maximum electron energy.

The analyzer target was fabricated in two versions: either of metal or of glass covered with a luminophor layer. The collector slot was 3-mm wide and transmitted about 1% of the total current into the analyzer. The slot width was determined by the requirement that the signal from the target substantially exceed the noise in the automated data processing system.

Gyrotron's operating magnetic field was produced by a superconducting solenoid placed in a cryostat. In order to direct the electrons approximately to the target center, we used auxiliary coils in the collector and on the way of the electron beam, as well as small permanent magnets that corrected the magnetic field near the analyzer target. Nevertheless, since the magnetic field was low and was therefore liable to disturbing external fields, we could not prevent the beam from touching the analyzer tube, which, as measurements showed, caused a systematic error in the obtained energy spread. We used a glass target for the initial adjustment. A metal target covered with a zirconium layer in order to reduce the secondary electron emission due to bombardment by the electron beam was employed in the electron energy measurements.

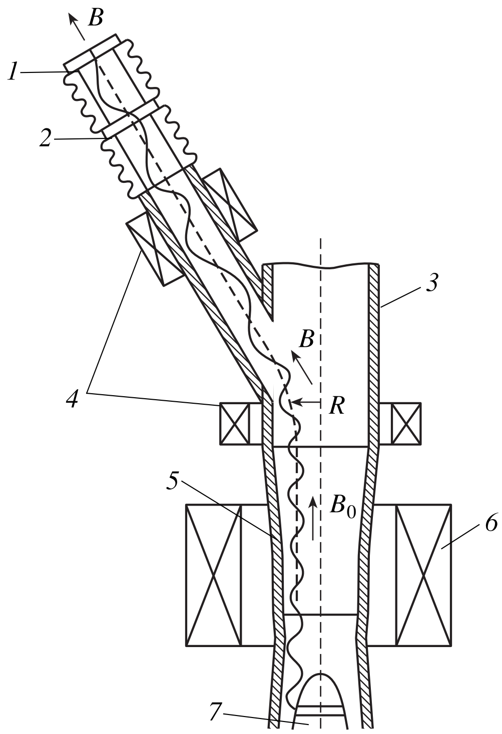


Fig. 1. Schematic of the experimental setup: (*1*) analyzer target, (*2*) grid, (*3*) collector, (*4*) auxiliary solenoid, (*5*) gyrotron cavity, (*6*) primary solenoid, and (*7*) electron gun.

EXPERIMENTAL RESULTS

Although the analyzer was designed to operate at the decelerating voltage of up to 70 kV and higher and was tested at this voltage, most of the measurements of the electron energy spectrum were carried out in the modeling mode, i.e., at a reduced beam voltage and current. This allowed us to avoid distortions in the electron energy spectrum that occur when high-power microwave oscillations are excited in the operating space of the gyrotron. The performance parameters of the studied short-pulse gyrotron are the following: the oscillation mode is $TE_{11.3}$, the operating frequency is 83 GHz, the output power is 1 MW, the pulse duration is 100 μs, the accelerating voltage is 70 kV, the maximum beam current is 40 A, the pitch factor is 1.3, and the magnetic field is 30 kOe.

In the modeling mode, the modeling factor k was 10 and the basic parameters were U_{mod} = 7 kV, $I_{\text{mod}} \approx$ 1 A,

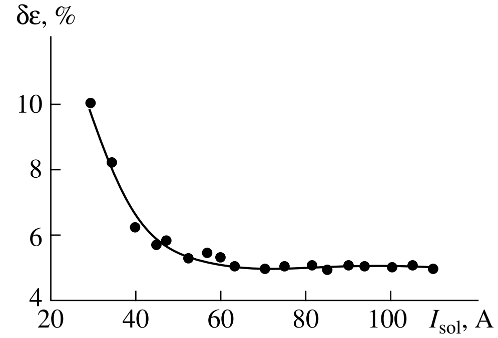


Fig. 2. Electron energy spread vs. solenoid current (magnetic field) in the modeling mode at a low electron current.

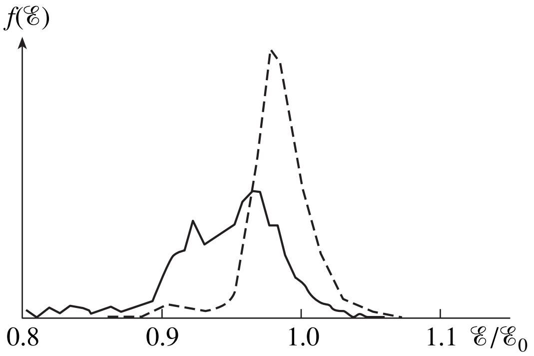


Fig. 3. Electron energy distribution at a low (solid line) and operating (dashed line) magnetic fields.

and $B_{mod} \approx 10$ kOe. In the experiment, we measured the magnetic field and the beam current.

Velocities of the beam electrons that entered the analyzer were measured as follows. A rectangular pulse was applied to the gyrotron cathode, and a triangular pulse of the same polarity with the amplitude 20% greater than the amplitude of the cathode voltage pulse was applied to the analyzer grid. Signals proportional to the target current and decelerating voltage were applied to the automated data acquisition and processing system. In our experiments, a system based on a CAMAC module and HiCom(Dec) and IBM PC computers was used. Automation of the data acquisition and processing provided both a great number of points on the cutoff curve and a high measurement speed. This allowed us to decrease errors associated with the instability of the power supply and obtain a large body of data, which provided more reliable measurements.

Target currents and grid voltages were written to respective storage buffers at a given rate. The data were written to a hard disk for the secondary processing and displayed on a monitor screen. In order to improve the measurement accuracy and avoid noise interference, the ADC range was varied to conform to the signal level. The accuracy of the measured beam parameters significantly depended on the quality of the beam positioning in the drift tube and on the calibration accuracy, which determined the signal scaling factors. The ADC was calibrated by applying a calibration signal to the computer and oscilloscope and choosing an appropriate scaling factor. The accuracy was determined by the sensitivity of the oscilloscope, which was about 2%. The required relative measurement accuracy was 1–2%. The accuracy was validated at low beam currents at which the energy spread was expected to be negligible.

One of the main sources of errors could be strays induced in the instrumentation when a high-voltage accelerating pulse was shaped and applied to the gyrotron. In order to suppress interference, the signal generated in the absence of the emission current was subtracted from the processed signal. As a result, the measurement error associated with electric strays was about 0.5%, which was tested against calibration signals.

As was noted above, one more source of errors was associated with touching the wall of the analyzer tube by electrons, which apparently occurred due to magnetic field distortions in this region. We failed to correct the electron trajectories by applying weak local fields using permanent magnets. The above effect resulted in an about 5% systematic error in the measured energy spread even when the space charge was sufficiently small (see Fig. 2). Note that the rise in the measured spread at week magnetic fields was caused by the instability associated with trapping the electrons in a magnetic trap as they drifted from the cathode toward the gyrotron cavity. When processing the experimental results, this systematic error was subtracted from the calculated energy spread. Since the systematic error associated with touching the analyzer tube wall by the beam depends on the initial electron energy, all measurements were carried out at a fixed beam voltage.

Dependence of the current on the decelerating voltage at the analyzer grid has the form of a step whose slope is determined by the energy spread. Normalized derivatives of this function with respect to the decelerating voltage, which characterize the electron energy distribution, are plotted in Fig. 3 for two modes. The narrow distribution function (dashed line) is obtained at a low beam current (20% of the current corresponding to the operating current in the modeling mode) and high magnetic field, which ensured the number of reflected electrons to be small, space-charge density to be low, and, consequently, distortion of the beam due to its self-field to be low. The mentioned systematic measurement error caused about a 5% broadening of the distribution and a corresponding shift of the maximum of the distribution function toward lower energies by 2–2.5%. Except for this shift, the spectrum broadening at different currents and magnetic fields $B \geq B_0$ (where B_0 is the modeling-mode field corresponding to the optimum generation regime) occurred symmetrically toward higher and lower energies, evidence that energy exchange existed between the beam electrons. A wide

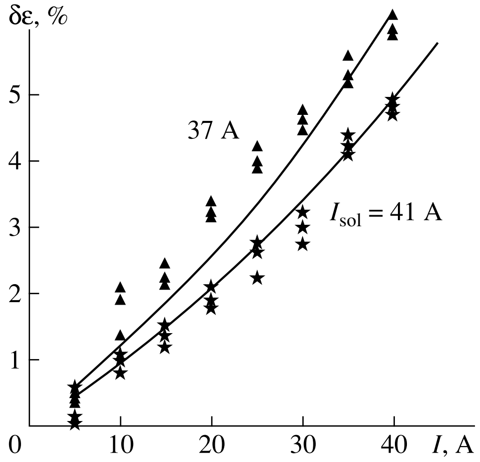


Fig. 4. Electron energy spread vs. current for two values of the solenoid current in the modeling mode of the gyrotron.

energy distribution function was measured in the modeling mode at a high current corresponding to the maximum gyrotron current and at a magnetic field close to the magnetic mirror mode ($B < B_0$), when a significant fraction of electrons was reflected from the magnetic mirror and spurious oscillations occurred, which was evidenced by an additional spectrum broadening and a decrease in the average electron energy due to beam radiative loss.

Figure 4 plots the energy spread versus beam current for two values of the magnetic field, B_0 and $0.9B_0$ (the gyrotron parameters—the magnetic field, beam voltage, and beam current—are referred to the normal operating mode). For the rated magnetic field B_0, the energy spread increases as the current changes from 5 to 40 A (the latter value corresponds to about a 1-MW power and efficiency higher than 40% [6, 7]). For the field $0.9B_0$, which differs significantly from the operating field, the spread increases insignificantly as compared to that observed at the same currents and the rated magnetic field B_0.

CONCLUSION

The above technique for measuring the energy spread of the gyrotron electron beam allowed us to experimentally determine the electron energy spectra in various modes of gyrotron operation. The results obtained show that, in the modeling mode, the energy spread of the electron beam produced with a standard electron gun in the absence of interaction between particles and microwave field increases with increasing the space charge (beam current) and, for the operating current, amounts to several percent at the output of the system. The most probable cause of this effect is the onset of the NMI typical of gyrotrons [8, 9]. Basically, this instability can result in a convective amplification of perturbations and appearance of the energy spread as the electron beam is formed and travels from the cathode toward the gyrotron's operating cavity. A detailed analysis of the mechanism for the onset of the NMI and interpretation of experimental results are given in [10]. As the magnetic field decreases, the spectrum becomes wider. When the decrease in the magnetic field is relatively small, this effect can be attributed to the perturbation amplification caused by the NMI. As the magnetic field decreases further and the fraction of electrons reflected from the magnetic mirror increases, the spectrum keeps broadening. This broadening is associated with the growth of the space-charge field due to a greater number of reflected electrons and the onset of the instability in the drift space between the cathode and mirror.

ACKNOWLEDGMENTS

This work was supported by the AOZT NPP GIKOM.

REFERENCES

1. A. V. Gaponov, M. I. Petelin, and V. K. Yulpatov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **10**, 1414 (1967).
2. V. A. Flyagin and G. S. Nusinovich, Proc. IEEE **76**, 644 (1988).
3. *Gyrotrons*, Ed. by V. A. Flyagin (Inst. Applied Physics, USSR Acad. Sci., Gorki, 1989).
4. E. G. Avdoshin and A. L. Goldenberg, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **16**, 1605 (1973).
5. A. L. Goldenberg, in *Digest of 19th International Conference on Infrared and Millimeter Waves, Sendai, 1994*, p. 349.
6. N. P. Venediktov, V. E. Zapevalov, and A. N. Kuftin, in *Gyrotrons*, Ed. by V. A. Flyagin (Inst. Applied Physics, USSR Acad. Sci., Gorki, 1989), p. 12.
7. V. A. Flyagin, A. N. Kuftin, A. G. Luchinin, *et al.*, *in Proceedings of Joint IAEA Technical Committee Meeting on ECE and ECRH (EC-7 Joint Workshop), Hefei, China*, p. 355.
8. S. Y. Cai, T. M. Antonsen, Jr., G. Saraph, *et al.*, Int. J. Electron. **72**, 759 (1992).
9. V. L. Bratman, Zh. Tekh. Fiz. **46**, 2030 (1976) [Sov. Phys. Tech. Phys. **21**, 1188 (1976)].
10. V. L. Bratman, Zh. Tekh. Fiz. **70** (4), 90 (2000) [Tech. Phys. **45** (4) (2000)].

Translated by A. D. Khzmalyan

Technical Physics, Vol. 45, No. 4, 2000, pp. 480–485. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 99–105.

RADIOPHYSICS

Nonstationary Processes in a Diffraction-Output Orotron

N. S. Ginzburg, N. A. Zavol'skiĭ, V. E. Zapevalov, M. A. Moiseev, and Yu. V. Novozhilova

Institute of Applied Physics, Russian Academy of Sciences,
ul. Ul'yanova 46, Nizhniĭ Novgorod, 603600 Russia

Received January 11, 1999

Abstract—The nonlinear dynamics of an orotron with diffraction output is studied. The evolution of the longitudinal field distribution is described by means of a parabolic equation subject to the condition of zero reflection at the collector end of the interaction space. For stationary regimes, the regions of high efficiency are delineated on the parameter plane, with the parameters being the departure from the critical frequency and the reduced length of the interaction space. From numerical results, the parameter plane is divided into the regions of stationary operation, periodic self-modulation, and stochastic self-modulation. It is demonstrated that self-modulation oscillating modes can be obtained most easily if the effective bounce frequency slightly exceeds the cutoff frequency, with the former being the blinking frequency of the dipole comprising an electron and its reflection in the regularly corrugated slow-wave guide. *© 2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Oscillators with diffraction output constitute a widely employed class of high-power microwave devices. Their merit is good selectivity, which in turn allows them to use advanced configurations of the cavity. Among them are gyrotrons, orotrons, and ubitrons, to name but a few. Orotrons and ubitrons are remarkable for their operation near the critical frequency. The dynamics of such oscillators can be studied by means of a parabolic equation for the evolution of the longitudinal field distribution, as suggested in [1]. Nonstationary processes in gyrotrons were extensively examined in the context of the parabolic-equation approach [2–4]. It has been shown that stationary operation gives way to periodic and then stochastic self-modulation if the parameters surpass their starting values significantly.

This paper deals with nonstationary processes in the orotron; i.e., a Cherenkov oscillator operating near the critical frequency. The problem is interesting in several respects. First of all, it should be pointed out that the nonlinear theory of relativistic orotrons with diffraction output is still in its infancy. In particular, stationary operation was addressed only under the assumption of a fixed field distribution in the longitudinal direction [5]. However, the longitudinal distribution may seriously be affected by the electron stream because of an extremely small Q factor. Nonstationary equations seem to be an efficient technique to attack a stationary boundary-value problem. Furthermore, the nonstationary approach enables one to delineate the stability boundaries of stationary oscillating modes and to examine self-modulation, which arises when the generation threshold is surpassed to a sufficient extent.

The recent surge of interest in nonstationary operation is also inspired by the idea to apply stochastic self-modulation to producing noise-like radiation. In this connection, it is worth mentioning the experiment reported in [6], during which periodic and stochastic self-modulation were observed in a carcinotron providing some 100 kW. Since the output frequency was fairly close to the critical frequency, the nonstationary operation can be simulated in terms of the orotron model that is studied below.

OROTRON MODEL AND ITS BASIC EQUATIONS

Consider an orotron with an annular electron beam (Fig. 1). The cavity is designed as a section of an axially symmetric waveguide with a faintly corrugated wall. The cavity tapers to a below-cutoff waveguide at the input end and is smoothly matched to the output waveguide at the collector end.

Nonstationary processes in an orotron can be described with a self-consistent system comprising parabolic excitation equations and equations of electron motion in a synchronous wave. Assume that the electrons inside the regularly corrugated slow-wave guide are in synchronism with the first spatial harmonic of an eigenmode whose frequency is close to the critical frequency (Fig. 2)

$$\omega_c \approx \bar{h} v_0, \quad (1)$$

where $\bar{h} = 2\pi/d$ with d being the pitch of the corruga-

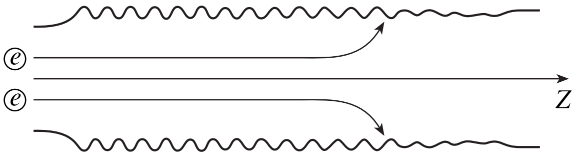


Fig. 1. Schematic drawing of the orotron cavity with an electron beam.

tions, v_0 is the initial longitudinal velocity of the electrons, and ω_c is the critical frequency.

Time-dependent excitation equations will be derived from the wave equation

$$\Delta\mathbf{E} - \frac{1}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} = -\frac{4\pi}{c^2}\frac{\partial\mathbf{j}}{\partial t}. \qquad (2)$$

Let us seek for a solution of (2) in the form

$$\mathbf{E} = \mathrm{Re}(f(z,t)\mathbf{E}_s(\mathbf{r}_\perp, z)\exp(i\omega_c t)), \qquad (3)$$

where $f(z, t)$ is a slowly varying amplitude; $\mathbf{E}_s(\mathbf{r}_\perp, z)$ is an eigenfunction of a cold cavity; and the carrier frequency is chosen to be equal to the cutoff frequency.

Since the corrugations are regular, $\mathbf{E}_s(\mathbf{r}_\perp, z)$ can be expanded in terms of spatial harmonics

$$\mathbf{E}_s(\mathbf{r}_\perp, z) = \sum_{l=-\infty}^{+\infty}\mathbf{E}_{sl}(\mathbf{r}_\perp)\exp(-il\bar{h}z), \qquad (4)$$

where $\mathbf{E}_{sl}$ meets the equation

$$\Delta_\perp\mathbf{E}_{sl} - g_{sl}^2\mathbf{E}_{sl} = 0 \qquad (5)$$

with $g_{sl}^2 = (l\bar{h})^2 - \omega_c^2/c^2$ being the transverse wave number.

Substituting (3) and (4) into (2) and averaging over fast oscillations, in view of (5), yield

$$\sum_l \mathbf{E}_{sl}\exp(-il\bar{h}z)\left(\frac{\partial^2 f}{\partial z^2} - 2il\bar{h}\frac{\partial f}{\partial z}l\bar{h} - 2i\frac{\omega_c}{c}\frac{\partial f}{\partial t}\right) = \frac{i\omega_c 4\pi}{c^2}\mathbf{j}_{\omega_c}. \qquad (6)$$

Here,

$$\mathbf{j}_{\omega_c} = \frac{1}{\pi}\int_0^{2\pi}\mathbf{j}\exp(-i\omega_c t)d(\omega_c t)$$

is the current-density harmonic at the carrier frequency. Multiply (6) by $\mathbf{E}_s^*(\mathbf{r}_\perp, z)$ and integrate it over the cross section. Since field energy is concentrated mainly in the harmonic with $l = 0$, the result is

$$\frac{\partial^2 f}{\partial z^2} - \frac{2i\omega_c}{c^2}\frac{\partial f}{\partial t} = \frac{i\omega_c}{2W_s c^2}\int_{S_\perp}\mathbf{j}_{\omega_c}\mathbf{E}_s^*\,dS_\perp, \qquad (7)$$

where

$$W_s = \frac{1}{8\pi}\int_{S_\perp}|\mathbf{E}_{s0}|^2 dS_\perp$$

is field energy per unit waveguide length.

Now, assume that the guiding magnetic field is so strong that the current density has the only a nonzero component in the longitudinal direction. Furthermore, take into account only the slowly varying sum term when dealing with the integral in (7) so that only the term corresponding to the first synchronous harmonic will be left on the right-hand side of (4). Bearing this in mind, multiply the sides of (7) by $(E_{s1}^*)_z$ to obtain

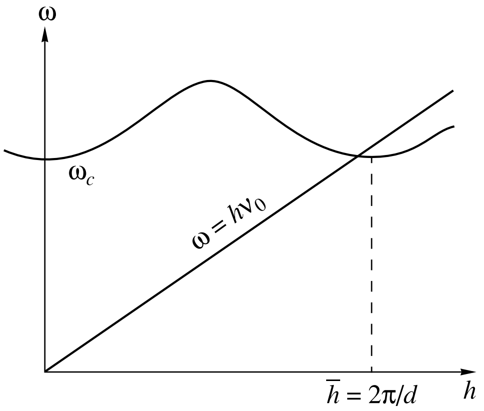


Fig. 2. Dispersion curve of orotron interaction.

$$\frac{\partial^2 A}{\partial z^2} - \frac{2i\omega_c}{c^2}\frac{\partial A}{\partial t} = J\frac{ik_c^3\varepsilon I_0}{c\beta_0^2}. \qquad (8)$$

Here, $A(z, t) = f(z, t)(E_{s1})_z$ is the amplitude of the spatial harmonic that is synchronous with the beam; $J = (1/\pi)\int_0^{2\pi}\exp(-i\vartheta)\,d\vartheta_0$ is the Rf-current density; $\vartheta = \omega_c t - hz$ is the instantaneous phase of the synchronous harmonic; ϑ_0 is the initial phase at $z = 0$; I_0 is the beam current; $k_c = \omega_c/c$; and $\beta_0 = v_0/c$. Equation (8) is derived using the charge-conservation law $j_z dt = j_0 dt_0$ (the subscript 0 indicates the instant at which the electron enters the interaction space). The coupling parameter can be defined as

$$\bar{\varepsilon} = |E_{s1}|^2 v_0^2/(2W_s\omega_c^2). \qquad (9)$$

Let us find concrete expressions for $\bar{\varepsilon}$ with respect to E- or H-waves in the cavity of interest. In [7], the following formula for the coupling impedance of a faintly corrugated circular waveguide was derived:

$$Z = |E_z|^2/2h^2P_s. \qquad (10)$$

Here, P_s is the energy flux of the mode. As is known, $P_s = W_s v_{gr}$, where W_s is the stored energy per unit waveguide length and $v_{gr} = h_0c^2/\omega_c$ is the group velocity of the wave, with h_0 being the longitudinal wave number of the fundamental harmonic. According to [7], the coupling impedance for E-waves is

$$Z = l_0^2\frac{(g_0^2 - h_0\bar{h})^2}{h^2\omega_c h_0 r_0^2}\frac{J_m^2(g_1 r_b)}{J_m^2(g_1 r_0)} \qquad (11a)$$

and that for *H*-waves is

$$Z = l_0^2 \frac{\omega_c \bar{h}^2 m^2}{h^2 c^2 h_0 r_0^2 (g_0^2 r_0^2 - m^2)} \frac{J_m^2(g_1 r_b)}{J_m^2(g_1 r_0)}. \tag{11b}$$

Here, r_0 is the waveguide radius, r_b is the beam radius, l_0 is the corrugation depth, I_m is the modified Bessel function of index *m*, and $h = h_0 + \bar{h}$ is the wave number of the slow harmonic. Formulas (11) imply that *Z* rises to infinity near the cutoff frequency, since h_0 is small. The indeterminacy is eliminated by using the coupling parameter as in (9). Thus, due to (1), (10), and (11), the coupling parameter for *E*-waves is

$$\bar{\varepsilon} = \frac{l_0^2 k_c^2 I_m^2(g_1 r_b)}{r_0^2 \bar{h} I_m^2(g_1 r_0)} \tag{12a}$$

and that for *H*-waves is

$$\bar{\varepsilon} = \frac{l_0^2 m^2}{r_0^2 (k_c^2 r_0^2 - m^2)} \frac{I_m^2(g_1 r_b)}{I_m^2(g_1 r_0)}. \tag{12b}$$

Excitation equation (8) is a parabolic equation, since electrons synchronously interact with both the copropagating and the counterpropagating wave. The RF-current amplitude appearing on the right-hand side of (8) comes from electron-motion equations. Consider the modes where inertial bunching dominates (relative particle-energy variations being small) so that the motion equations can be written in a universal form

$$\frac{d^2\vartheta}{dz^2} = \frac{ek_c}{mc^2(\gamma_0^2 - 1)^{3/2}} \operatorname{Re}(A \exp(i\vartheta)) \tag{13}$$

subject to the boundary conditions

$$\left.\frac{d\vartheta}{dz}\right|_{z=0} = \frac{\omega_c}{v_0} - \bar{h}, \quad \vartheta|_{z=0} = \vartheta_0 \in (0.2\pi).$$

The boundary conditions are dictated by the configuration of the cavity. The input waveguide being in a below-cutoff condition, the field is zero at the input cross section

$$A|_{z=0} = 0. \tag{14}$$

At the output cross section, reflection is zero and electron–field interaction ceases, since corrugations die away there, so that the condition

$$A(z_{\text{out}}, t) + \frac{c}{\sqrt{\pi i}} \int_0^t \frac{1}{\sqrt{\omega_c(t - t')}} \left.\frac{\partial A(z, t')}{\partial z}\right|_{z = z_{\text{out}}} dt' = 0 \tag{15}$$

(z_{out} being the length of the interaction space) must be used as a boundary condition for radiation [1–3].

In order to simplify the equations, assume that the time of electron transit through the interaction space is much less than the time of diffraction field decay

$$z_{\text{out}}/v_0 \ll T_{\text{rad}} = Q_{\text{dif}}^{\min}/\omega_c, \tag{16}$$

where $Q_{\text{dif}}^{\min} = (z_{\text{out}}\omega_c/c)^2$ is the minimum diffraction *Q* factor.

As long as condition (14) is met, the field amplitude remains unchanged during the electron transit time so that the time derivative in (13) can be neglected. Let us switch to the normalized variables

$$\zeta = k_c B z, \quad \tau = \omega_c t B^2/2, \quad \alpha = \frac{ieA}{mc^2 k_c (\gamma_0^2 - 1)^{3/2} B^2},$$

$$\Delta = \frac{\bar{h} v_0/\omega_c - 1}{\beta_0 B}, \quad B = \left(\frac{eI_0 \bar{\varepsilon} \gamma_0^2}{mc^3 (\gamma_0^2 - 1)^{5/2}}\right)^{1/4}. \tag{17}$$

Then (8) and (13) are brought to a form with as few independent parameters as possible:

$$\frac{\partial^2\alpha}{\partial\zeta^2} - i\frac{\partial\alpha}{\partial\tau} = J, \quad \frac{\partial^2\vartheta}{\partial\zeta^2} = \operatorname{Im}(\alpha \exp(i\vartheta)) \tag{18}$$

the boundary conditions being

$$\left.\frac{\partial\vartheta}{\partial\zeta}\right|_{\zeta=0} = -\Delta, \quad \vartheta|_{\zeta=0} = \vartheta_0 \in (0.2\pi),$$

$$\alpha(\zeta_{\text{out}}, \tau) + \frac{1}{\sqrt{i\pi}} \int_0^\tau \frac{1}{\sqrt{\tau - \tau'}} \left.\frac{\partial\alpha}{\partial\zeta}\right|_{\zeta = \zeta_{\text{out}}} d\tau' = 0, \tag{19}$$

where $\zeta_{\text{out}} = k_c B z_{\text{out}}$ is the reduced length of the interaction space.

Notice that Δ can be interpreted as the normalized initial departure of the effective bounce frequency ($\omega_b = \bar{h} v_0$) from the cutoff frequency, with the former being the blinking frequency of the dipole comprising an electron and its reflection in the regularly corrugated slow-wave guide.

Oscillator efficiency is

$$\eta = \frac{B\gamma_0^2\beta_0^3}{1 - \gamma_0^{-1}} \bar{\eta}, \quad \bar{\eta} = \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{\partial\vartheta}{\partial\zeta} + \Delta\right) d\vartheta_0. \tag{20}$$

The energy-conservation law for (18) is

$$\frac{d\bar{W}}{d\tau} + \bar{P} = \bar{\eta}, \tag{21}$$

where

$$\bar{P} = \frac{1}{2} \operatorname{Im}(\alpha \partial\alpha^*/\partial\zeta)|_{\zeta = \zeta_{\text{out}}}, \quad \bar{W} = \frac{1}{4} \int_0^{\zeta_k} |\alpha|^2 d\zeta.$$

In the original notation, (21) becomes

$$\frac{I_0}{e}mc^2\langle(\gamma_0-\gamma)\rangle_{\vartheta_0} = P_0\big|_{z=z_{\text{out}}} + \frac{dW_0}{dt}, \qquad (22)$$

where $P_0 = W_s v_{gr}\big|_{z=z_{\text{out}}}$ is the radiation power at the output cross section and

$$W_0 = \int_0^{z_k} W_s dz$$

is the field energy stored in the interaction space [see (7)].

NUMERICAL RESULTS

The system of equations (18) has two independent parameters, namely, the reduced interaction-space length ζ_{out} and the frequency mismatch Δ. A simulation of (18) enabled us to delineate regions in the Δ–ζ_{out} plane in terms of operation modes, as shown in Fig. 3. The figure also includes lines of equal efficiency for stationary operation. As ζ_{out} increases (due to an increased length of the interaction space or a stronger beam current), stationary single-mode operation gives way to periodic and then stochastic self-modulation. The threshold of self-modulation depends on Δ.

First, consider the case of a large positive Δ, with which ω_b exceeds the cutoff frequency. Then the electrons are synchronous with the counterpropagating wave (the interaction with the copropagating wave) being negligible. Consequently, the solution of (18) is

$$\alpha = (2\Delta)^{-2/3} a \exp(i\Delta\zeta + i\Delta^2\tau),$$
$$\vartheta = -\Delta\zeta - \Delta^2\tau + \psi(\zeta, \tau).$$

Furthermore, the condition $\Delta \gg 1$ allows one to transform (18) into well-known time-dependent equations for the carcinotron [8, 9]:

$$\frac{\partial a}{\partial \xi} - \frac{\partial a}{\partial \tau} = -iJ, \quad \frac{\partial^2 \psi}{\partial \xi^2} = \operatorname{Im}(a\exp(i\psi)), \qquad (23)$$

where $\xi = \zeta(2\Delta)^{-1/3}$, and $\hat{\tau} = \tau(2\Delta)^{2/3}$.

Under this approximation, the oscillating mode is controlled by a single parameter, namely, the reduced oscillator length $\xi_{\text{out}} = \zeta_{\text{out}}(2\Delta)^{-1/3}$. Self-excitation occurs at $\xi_{\text{out}} > 1.98$ and self-modulation arises at $\xi_{\text{out}} > 2.9$. Returning to (18), we obtain the following formulas for bifurcation curves with $\Delta \gg 1$:

$$\zeta_{st} = 1.98(2\Delta)^{1/3}, \quad \zeta_{sm} = 2.9(2\Delta)^{1/3}. \qquad (24)$$

Relations (24) are depicted by the dashed curves in Fig. 3. They were corroborated by the simulation of the transitions between the oscillating modes in terms of (18). Figures 4–6 show the amplitude variations and the

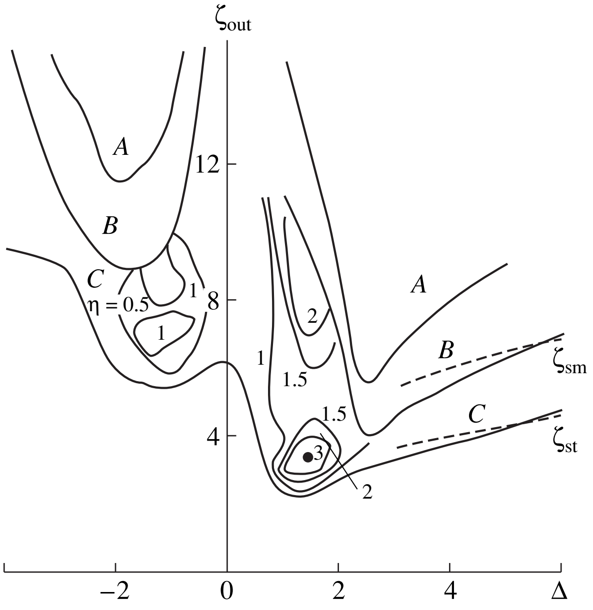


Fig. 3. Regions of operation: (*A*) stochastic self-modulation, (*B*) periodic self-modulation, and (*C*) stationary operation. The regions of stationary operation include lines of equal efficiency (the value is indicated). The dashed curves represent asymptotic relations (24).

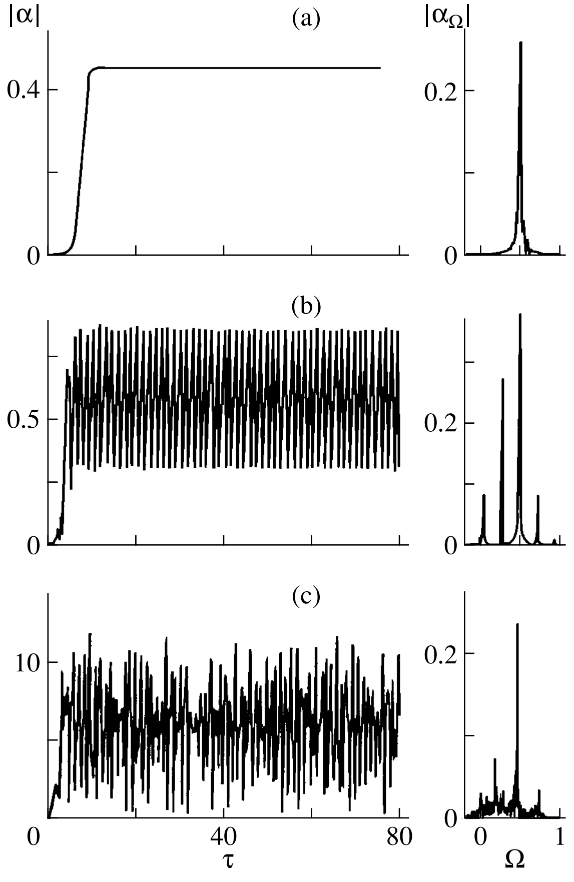


Fig. 4. Amplitude variation and spectrum of the output radiation at $\Delta = 4$ and ζ_{out} = (a) 4.2, (b) 5.8, and (c) 9.

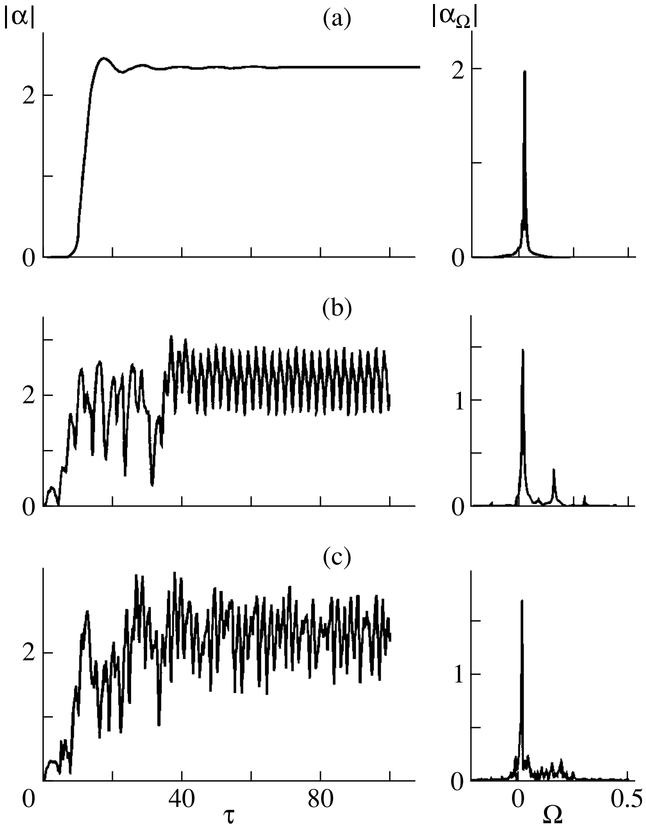


Fig. 5. Amplitude variation and spectrum of the output radiation at $\Delta = 2$ and ζ_{out} = (a) 3, (b) 8, and (c) 11.

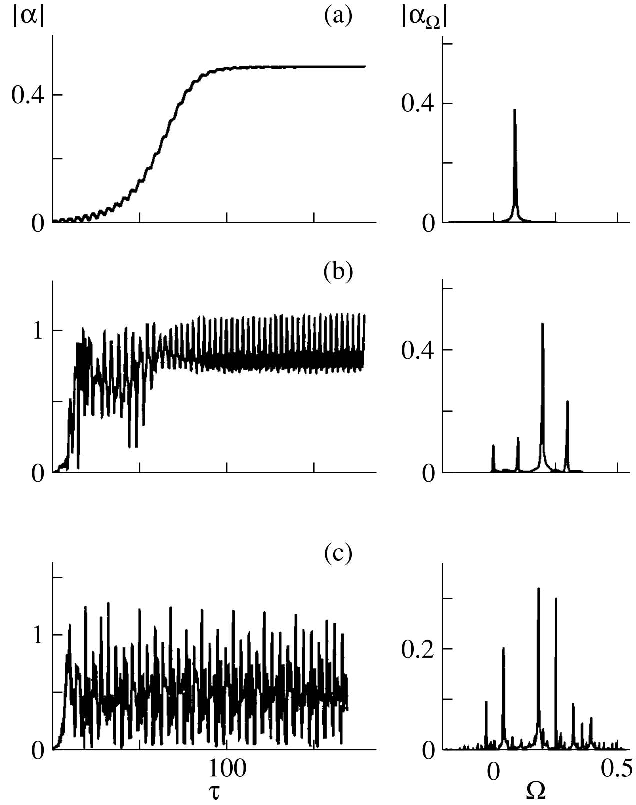


Fig. 6. Amplitude variation and spectrum of the output radiation at $\Delta = -2$ and ζ_{out} = (a) 6.5, (b) 9.5, and (c) 12.

spectra at the output for the stationary mode and for periodic and stochastic self-modulation, with

$$\Omega = \frac{2(\omega - \omega_c)}{\omega_c B^2}$$

being the fractional departure of the operating frequency from the cutoff frequency. Figure 4 depicts the operation under the synchronism with the counterpropagating wave at $\Delta = 4$.

In accord with (24), Fig. 3 indicates that the self-excitation and self-modulation thresholds lower with a decrease in the absolute value of Δ (i.e., in the distance to the cutoff frequency). This stems from a decreased group velocity in the wave and from enhanced electron–wave coupling. The minimum starting length, $\zeta_{out} = 2.2$, is attained at $\Delta = 1.2$. Self-modulation arises most readily at $\Delta = 2.5$, when the ζ_{out} value for the bifurcation to the stochastic regime, $\zeta_{out} = 5.6$, exceeds the starting value by a factor smaller than 2 (so that the beam current must exceed its starting value by a factor close to 16). Figure 5 depicts the operation at $\Delta = 2$. Contrasting Figs. 4 and 5 with each other suggests that a decrease in Δ is accompanied by (1) an increase in the self-modulation period, due to a lower group velocity of the wave; (2) an increase in the duration of the transient to self-oscillation; and (3) a shift in the operating frequency toward the cutoff frequency.

In the stationary regime, the reduced efficiency attains its maximum value, $\bar{\eta} \approx 3$, at $\Delta = 1.75$ and $\zeta_{out} = 3.3$. The corresponding longitudinal distributions of the field amplitude and phase are shown in Fig. 7. Approximately, the length of the interaction space accommodates one variation of the RF-field. If the efficiency is maximum, the electrons synchronously interact with both the counterpropagating and the copropagating wave (the corresponding electron transit angles are about 2π). A further decrease in $|\Delta|$ is accompanied by a rise in the starting and the bifurcation currents.

In the region $\Delta < 0$, the electrons are predominantly in synchronism with the copropagating wave. Self-excitation (feedback) in this parameter domain results from the partial reflection of the wave from the step in permittivity at $z = z_{out}$, where electron-field interaction ceases, as in gyrotrons [10]. By and large, the bifurcations with $\Delta < 0$ are similar to those with $\Delta > 0$. The operation at $\Delta = -2$ is depicted by Fig. 6. Note that the duration of the transient to self-oscillation, the period of self-modulation, and the departure of the operating frequency from the critical one are comparable to those obtained for $\Delta = 2$. As Δ decreases further, the reflection from the permittivity step must be more and more

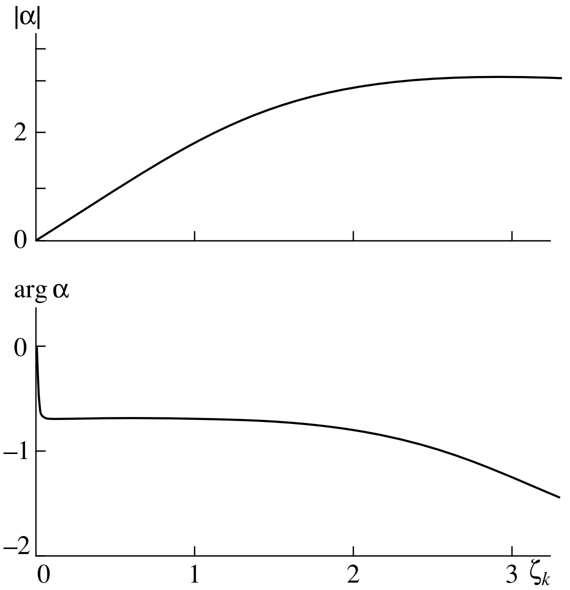


Fig. 7. Longitudinal distribution of the field amplitude and phase at maximum efficiency (ζ_{out} = 3.3 and Δ = 1.75).

attenuated. That is why the starting current was observed to rise more rapidly than at positive Δ.

In conclusion, let us apply the simulation results to the experiment reported in [6], where self-modulation was observed in a Cherenkov oscillator producing some 100 kW in the centimeter-wave band. The parameters of the corrugated cavity were d = 1.7 cm, l_0 = 0.2 cm, and ζ_{out} = 45 cm. The annular-beam radius was 0.7 cm, the beam current varied from 3 to 35 A, and the accelerating voltage was 70 kV. The operating mode was chosen to be TE_{11}. For those conditions, our computation demonstrated that $4.6 < \zeta_{out} < 5.2$ and $2.15 < \Delta < 2.5$, varying within the limits. It follows from Fig. 3 that such values of ζ_{out} and Δ are close to the optimal self-modulation region. Obviously, as mentioned in the above, the nearer the operating frequency to the cutoff frequency, the larger the coupling parameters, which in turn favors the transition to self-modulation at a given beam current. On the other hand, bifurcation currents grow steeply in the vicinity of the cutoff frequency, seemingly because of strong dispersion. Consequently, the field amplitude becomes much more uniform along the interaction space. This equalizes the delay effect, which contributes to self-modulation in carcinotrons (and so does the nonlinearity, namely electron rebunching [8]).

ACKNOWLEDGMENTS

We are grateful to G. S. Nusinovich for helpful discussions and comments.

This study was supported by the Russian Foundation for Basic Research, project no. 97-02-1761.

REFERENCES

1. N. S. Ginzburg, N. A. Zavol’skiĭ, G. S. Nusinovich, *et al.*, Radiotekh. Élektron. **29**, 106 (1984).
2. N. S. Ginzburg, G. S. Nusinovich, and N. A. Zavolsky, Int. J. Electron. **61**, 881 (1986).
3. N. S. Ginzburg, N. A. Zavol’skiĭ, G. S. Nusinovich, *et al.*, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **29**, 106 (1986).
4. N. A. Zavol’skiĭ and G. S. Nusinovich, Radiotekh. Élektron. **36**, 135 (1991).
5. N. I. Zaĭtsev and N. F. Kovalev, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **29**, 229 (1986).
6. N. S. Ginzburg, N. I. Zaĭtsev, E. V. Ilyakov, *et al.*, Pis’ma Zh. Tekh. Fiz. **24** (10), 816 (1998) [Tech. Phys. Lett. **24**, 816 (1998)].
7. N. F. Kovalev, Élektron. Tekh. Ser. 1: Élektronika SVCh, No. 3, 102 (1978).
8. N. S. Ginzburg, S. P. Kuznetsov, and T. N. Fedoseeva, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **21**, 1037 (1978).
9. N. S. Ginzburg and S. P. Kuznetsov, in *Relativistic RF Electronics: Raising Output Power and Frequency* (IPF AN SSSR, Gorki, 1981), p. 101.
10. V. L. Bratman, M. A. Moiseev, and M. I. Petelin, in *Gyrotron* (IPF AN SSSR, Gorki, 1981), pp. 122–145.

Translated by A. A. Sharshakov

Technical Physics, Vol. 45, No. 4, 2000, pp. 486–489. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 106–110.
Original Russian Text Copyright © 2000 by Lazarev, Babulevich, Varfolomeev.

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Analysis and Monitoring of the Stressed State in Solids by Gas Sensors

S. D. Lazarev, N. E. Babulevich, and A. E. Varfolomeev
Kurchatov Institute Russian Research Center, pl. Kurchatova 1, Moscow, 123182 Russia

Received January 11, 1999

Abstract—Gas release from solids under mechanical stressing of specimens up to failure was studied. The amount of hydrogen being released vs. strain developed in a specimen under step-by-step loading was measured. The hydrogen evolved in strained metals (iron, nickel, copper, aluminum, etc.) and some epoxy-resin-based composites was detected by chemical sensors in a gas atmosphere at normal atmospheric pressure. The amount of hydrogen was shown to grow with the amount of plastic strain and peak upon failure. A method for strain analysis in solids was worked out. It was applied to analyze the stressed state of the rotor material of a centrifuge during its operation. © *2000 MAIK “Nauka/Interperiodica”*.

INTRODUCTION

Gaseous impurities penetrate into solids either from the environment or from the process atmosphere where materials are synthesized. In metal crystals, foreign gas atoms are, as a rule, segregated at lattice defects: vacancies, interstitials, or dislocations. When combined, they may produce bubbles. In ceramics and glasses, gaseous impurities may fill pores. Hydrogen dominates over other gases in solids. The small size and the high mobility of hydrogen atoms allow them to readily penetrate into the crystal lattice and migrate there. When a solid material is mechanically stressed or fails, the impurities may evolve. Measurements of evolved H_2 were reported [1–3]; however, methods for detecting small hydrogen quantities were very sophisticated, and their practical use in studying the stressed state of solids seems to be problematic. With the advent of a new class of sensors, chemical sensors, the situation has radically changed. These sensors offer high sensitivity to a particular gas and can operate both in a vacuum and in different gas environments under normal atmospheric pressure [4, 5]. The sensitivity of semiconductor SnO_2 oxide sensors and Pd–SiO_2–Si sensors to the H_2 content in a gaseous atmosphere is as high as 10^{-5} vol. %. In the oxide sensors, H_2 present in the atmosphere changes the conductivity of a sensing element, and in the metal–insulator–semiconductor (MIS) sensors, H_2 changes the capacitance of a Pd–SiO_2–Si capacitor or, when the capacitance is fixed, the bias voltage of the *C*–*V* curve.

In this work, we studied gas release from solids within a wide range of applied mechanical loads. The main goal was to find a correlation between the stressed state of a specimen material and the amount of the gas evolved on loading.

EXPERIMENTAL RESULTS

Specimens were prepared from metal and polymers. In most experiments, the amount of hydrogen evolved from specimens under deformation was measured by a Pd–SiO_2–Si sensor, and sometimes a SnO_2 sensor was used.

An experimental setup is shown in Fig. 1. A specimen is placed into a sealed chamber and subjected to uniaxial stretching or torsion. Hydrogen being released from the specimen under deformation is entrained by a circulating carrier gas (helium) to a measuring chamber equipped with a sensor. A sensor signal is recorded with an *x*–*y* recorder or an oscilloscope. To calibrate the

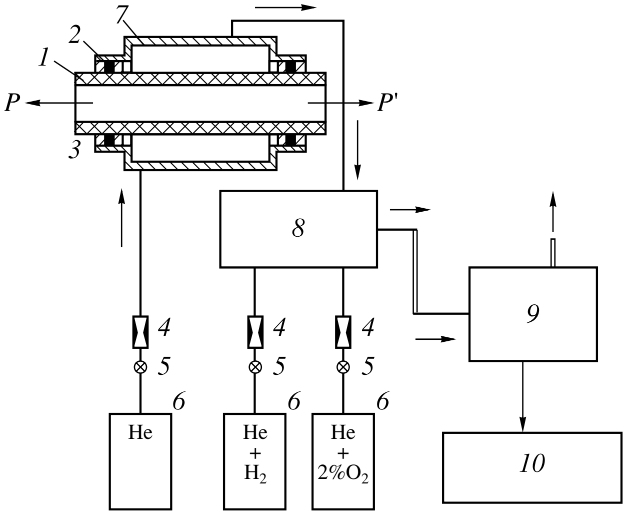


Fig. 1. Schematic representation of the setup for diagnostics of the stressed state in solids: (*1*) specimen; (*2*) seal; (*3*) deforming device; (*4*) flow-rate controller; (*5*) pressure regulator; (*6*) gas container; (*7*) sealed chamber; (*8*) gas-flow switch; (*9*) chamber with sensor; and (*10*) recorder.

setup and enhance the rate of H_2 desorption from the sensing element, we used He + H_2 mixtures of known concentrations of the components and a helium–air (10%) mixture. The gas mixtures were supplied to the chamber from gas containers through flow-rate controllers.

The leading edge of a sensor signal was taken as an indicator of the content of H_2 delivered to the chamber. The calibrating curve reflected the dependence of the time derivative of the sensor signal at the instant of signal formation (at $t \longrightarrow 0$) on the hydrogen amount in the chamber. With this (dynamic) approach, the measuring time is significantly cut in comparison with signal amplitude measurements. Also, a dynamic range of the measured H_2 amount can substantially be extended [5]. Most specimens were rectangular plates measuring $18 \times 8 \times 2$ mm. Some experiments were performed with 150-mm-long tubes with an outer and an inner diameter of 7 and 6 mm, respectively. Uniaxial stretching experiments were carried out with a conventional breaking machine. Torsional deformation was produced on a special installation. The amount of H_2 being released was measured at fast step-by-step loading.

Experiments with the metals (Fe, Ni, Cu, Al, etc.) and alloys subjected to step-by-step loading have shown that H_2 starts to evolve under plastic deformation. Its amount grows with increasing the applied load and peaks when a specimen fails. At small elastic strains, the detected hydrogen is close to zero (its amount is comparable to the sensitivity of the setup). Typical experimental results are shown in Fig. 2.

Two groups (*A* and *B*) of identical stainless steel specimens with different strength properties were studied. The surface of group-*A* specimens was intentionally scratched. Scratches played the role of stress concentrators. In this way, the mechanical strength of group-*A* specimens was reduced, and they failed at a lower (than *B*-group specimens) tensile force. A time dependence of the stress developed in the specimen is shown in Fig. 2a, and Fig. 2b depicts a time dependence of the actual sensor signal when the strain is changed (dashed curves stand for H_2 desorption from the sensitive element of the sensor when the He + 10% air mixture is applied to the measuring chamber). Figure 2c shows a sensor signal vs. time curve for group-*B* specimens. The curves in Fig. 2 imply that group-*A* specimens evolve H_2 prior to those of group-*B*. The less the strain at which a specimen fails, the less the strain at which H_2 emission starts. Data for the metals and the alloys indicate that gas release coincides with the development of plastic strains in a specimen and can be thought of as an indicator of forthcoming failure.

The effect of gas release under polymer failure is well known [6]. It has been observed in a high vacuum by means of a mass spectrometer. Evolved gases contain various volatile products due to the breakdown of chemical bonds in a strained specimen. Reactive free radicals, which initiate secondary chemical reactions, play an essential part in this process. Note that the radicals are hard to recombine in a high vacuum. Therefore, gaseous products detected under normal atmospheric pressure by means of chemical sensors may differ from those determined with a mass spectrometer. Experiments with some polymers and epoxy-resin-based composites have shown that hydrogen is among gases being released under specimen deformation and failure. We did not use chemical sensors to identify other gases.

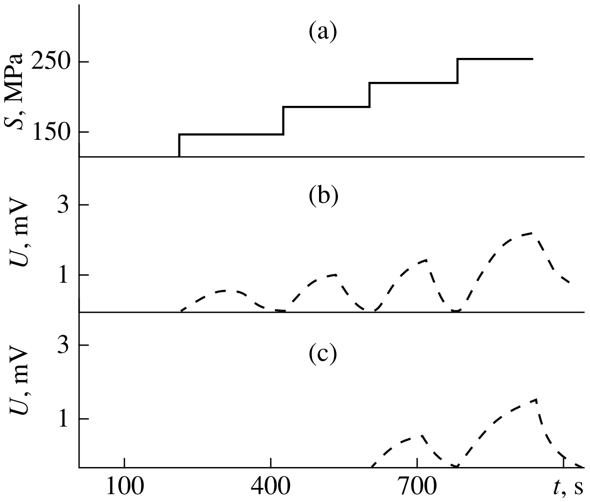


Fig. 2. H_2 evolution from solid materials at step-by-step loading of the specimen. Dashed line, sensor signal upon applying an He +2% O_2 mixture to the chamber (“desorption curve”).

A subject of investigation was also graphite filaments held together by epoxy resin. Such specimens show strength anisotropy, since the filaments are aligned with the specimen geometric axis. The tensile strengths along the filaments and in the transverse direction usually differ by several orders of magnitude. The specimens were subjected to torsional deformation, the axis of torsion being coincident with or normal to the filament orientation. The hydrogen evolved under deformation and failure of the specimens was detected with a sensor. Figure 3 demonstrates the amount of the evolved hydrogen vs. the torque for specimens whose orientation coincides with (specimen *1*) and is normal to (specimen *2*) the torsion axis.

From Fig. 3, it is seen that H_2 in specimen *1* evolves at a higher moment and its amount is more than 10 times greater than for specimen *2*. As for the metals, data for polymer specimens suggest that gas release under deformation means the onset of failure. The smaller the strain at which the gas is released, the smaller the breaking stress.

It seems inviting to apply the described method for diagnostics of the prefailure state of parts and structures immediately during their operation. One possible appli-

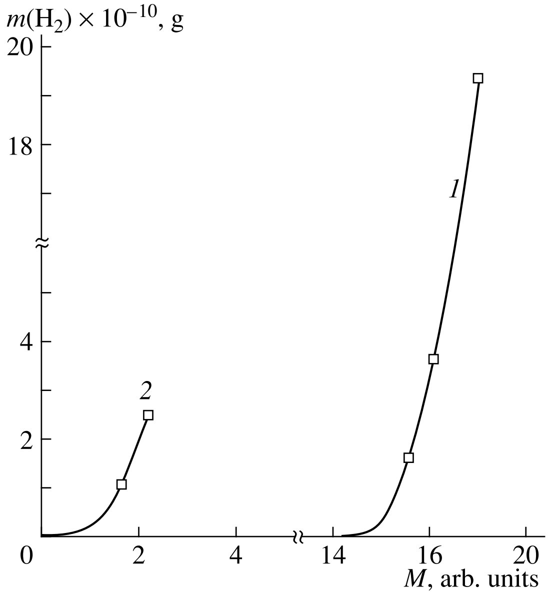


Fig. 3. Amount of H_2 being released from the epoxy-resin-based plastic vs. torque applied to the specimen.

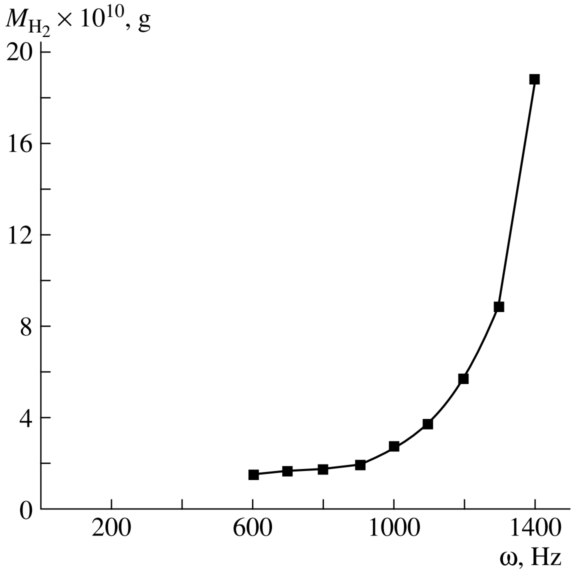


Fig. 4. Amount of H_2 being released from the centrifuge material vs. frequency of rotor rotation.

cation is strain detection in rails during train movement. Another project involves the detection of the most strained areas of flying vehicles in flight. The realization of the projects, however, requires high financial costs. In this work, we have attempted to apply our method for detecting the prefailure state of parts of a centrifuge during its operation. Forces due to the rotation of the centrifuge rotor cause strains in the materials. At very high rates of rotation, the rotor may fail. We tried to predict this event, analyzing experimental data for the H_2 evolution as a function of the rate of rotor rotation. Small amounts of H_2 in the vacuum chamber with the centrifuge were measured with a Pd–SiO_2–Si sensor.

Experiments were carried out on a centrifuge test bed. Ring-shaped specimens, parts of the centrifuge body, were made of a special plastic. The ring sizes were 130 mm (outer diameter), 124 mm (inner diameter), and 18 mm (height). A specimen was place into a ring-shaped groove (mated with the specimen) on a cylindrical table. The table rested on a needle and was driven by an electrical generator. The whole assembly was mounted in a steel chamber evacuated to 10 torr with a vacuum pump. When the pump was switched off for 5–7 min, the pressure in the chamber remained practically unchanged. The frequency of specimen rotation, specified by the generator, was changed stepwise with a step of 100 Hz. The time taken to change the rotation frequency by 100 Hz was several seconds. The capacitive MIS (Pd–SiO_2–Si) sensor, used to measure the hydrogen amount, offered high sensitivity and selectivity to H_2 in an oxygen-free atmosphere. Its temperature was kept at 180 ± 0.1°C. Dynamic measurements were provided by placing the sensor into a separate measuring chamber, which communicated with the main chamber via a valve. The sizes of the valve and the communications were selected so that, when the valve opens, the pressures in both chambers equalize for a time small in comparison with the characteristic time of the sensor τ_s (at small partial pressures of H_2, τ_s is of the order of tens of seconds). When the valve opens, a concentration pulse with a rise time $\tau \ll \tau_s$ is applied to the sensitive element, so that measurements can be taken in the dynamic mode. The dependence of the time derivative of a sensor signal at the initial time instant on the H_2 pressure in the chamber is linear in the pressure range between 10^{-4} and 5×10^2 torr. The amount of H_2 being released upon stressing was then determined with the calibration curve. The measurements were taken for every 100-Hz step of the rotation frequency with the pump switched off. After the measurements, H_2 was pumped off from the chambers.

The frequency dependence of the evolved H_2 for one of the specimens is shown in Fig. 4. The evolution of H_2 begins at 500 Hz; the amount of the gas sharply grows with frequency, and the specimen fails at 1400 Hz.

Similar measurements were made with a lot of 20 identical specimens. The evolution of the gas at different frequencies of rotation was observed only for four specimens. All of them failed at a frequency of 1400 Hz. In the rest of the specimens, gas evolution was absent up to 1500 Hz, and they did not fail within this frequency range. Similar results were obtained for other lots. This allowed us to suggest that the evolution of H_2, the amount of which grows with the frequency of rotation, is an indicator of forthcoming failure. Some specimens are likely to have crystal defects, which cause mechanical stresses and reduce the material strength. These specimens fail at low frequencies of rotation. Gas evolution well before complete failure of a specimen may be associated with failures of local regions of the specimen. This proposal calls for further investigation.

The method for detecting the prefailure state of materials by hydrogen chemical sensors can be helpful in monitoring the strain state of centrifuge parts during fabrication and, after a slight modification, operation.

CONCLUSION

Gas release from solids under loading proceeds in a similar way in metals and polymers. Results obtained may be useful in creating a sensory system for monitoring the stressed state of solids. In principle, the use of chemical sensors makes it possible to monitor the stressed state of parts operated both in a vacuum and in various gaseous atmospheres under normal pressure. Such an approach is based on gas emission from materials under mechanical loading and failure and seems to be especially useful when data obtained by conventional methods are insufficiently reliable. Chemical sensors are cheap, have small sizes (~1 mm^3), and are easy in service. The suggested technique for measuring the amount of gas evolving from stressed materials allows for measurements under high acoustic and electromagnetic noise. Chemical sensors may operate without direct physical contact between a sensor and a specimen and are hence applicable to monitoring the stressed state of moving mechanical parts immediately during operation. Due to their small sizes, chemical sensors may, in principle, solve the problem of locating mechanical damage in materials.

REFERENCES

1. D. A. Donovan, Metall. Trans. A **7**, 1677 (1976).
2. C. A. Wert, *Hydrogen in Metals*, Ed. by G. Alefeld and J. Volkl (Springer, Berlin, 1978), p. 369.
3. V. S. Sevast’yanov, B. K. Zuev, N. E. Babulevich, *et al.*, Fiz.-Khim. Mekh. Mater., No. 4, 47 (1990).
4. I. Lundström, Sens. Actuators, No. 1, 403 (1981).
5. A. Babichev, N. Babulevich, S. Lazarev, *et al.*, Sens. Actuators, No. 13–14, 362 (1993).
6. V. R. Regel’, A. I. Slutsker, and É. E. Tomashevskiĭ, Usp. Fiz. Nauk **106**, 193 (1972).

Translated by V. A. Isaakyan

Technical Physics, Vol. 45, No. 4, 2000, pp. 490–496. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 111–117.
Original Russian Text Copyright © 2000 by L. Veresov, O. Veresov, Litvinov.

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Investigation of an Ion Source Intended for Beam Technologies

L. P. Veresov, O. L. Veresov, and P. A. Litvinov

Sukhumi Physicotechnical Institute, Academy of Science of Abkhaziya, Sukhumi, 384914 Abkhaziya

Received December 28, 1998

Abstract—An ion source with a cold cathode is described, which generates ion beams of various gases and solids of circular cross sections for technological purposes. © *2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

A growing interest in beam technologies stimulated us to design and investigate a source generating heavy ions. The source initially operated in a pulse mode (τ_{imp} = 0.1–1 ms, F = 1–50 Hz), and subsequently, it was changed to a steady-state mode. The source generates mixed beams, because they consist of ions of the working gas and the sputtered solid substance. Therefore, depending on the requirements imposed upon a technological process, both mixed and separated beams were used. The source design ensured a beam energy variation within a wide range of 0.2–100 keV. The beam current depended on the source operating conditions and the substances constituting the beam. In the steady-state source operating mode, the current of mixed beams was 0.1–10 mA. In the pulse operating mode, the source yielded beams with currents of tens of milliamperes.

ION SOURCE DESIGN

The source (its design is shown in Fig. 1) consists of two basic units: a plasma generator and an ion-optics system. The plasma generator is a cylindrical gas-discharge chamber located between a flat cathode unit and a cone-shaped anode with a central emission hole. This gas-discharge cell (GDC) is placed in a magnetic field with a fan-shaped cross section and a mirror ratio of 2000/120 Gs. Its lines of force diverge from the cathode to the anode. The cathode unit *2* includes a container–holder *3*, which is located at its center and advanced into the discharge chamber, a disk (washer) of the sputtered working substance *5* inserted into the container, and a system *4* for leaking the working gas in and measuring the vacuum in the source chamber. The insulator *1* that insulates the cathode from the anode is protected against films of the sputtered substances by a shield *6*.

The adjusted ion-optics system for acceleration and deceleration (IOSAD) is joined to the anode (emitting electrode) *7* of the GDC. The anode and the first electrode *9* of the IOSAD form the accelerating high-voltage gap. The three-electrode IOSAD used consists of four elements in the form of metallic disks assembled on dielectric rods *8*. The first and second disks are the accelerating and decelerating electrodes *9* and *10*, respectively. The latter electrode also serves as the first edge electrode of a single electrostatic lens. The third and fourth disks are, respectively, the central focusing electrode *11* and the other edge electrode *12* of this lens.

Such an IOSAD allows one to vary the beam particle energy in a wide range of 10^3–10^6 eV without appreciable changes in the beam current, focus the ion beam, and separate the gas-discharge plasma from the beam plasma.

A shield *13*, located between the last electrode *12* of the lens and the target, divides the vacuum chamber into two regions in order to reduce the effect of reverse gas and sputtered-particles flows. The use of the shield significantly reduced the intensity of breakdowns in the ion-optics channel. Figure 2 shows a source circuit diagram and voltage charts.

DIAGNOSTICS

The following measuring and diagnostic devices were used for measuring the parameters of the beam extracted from the ion source. The beam current was

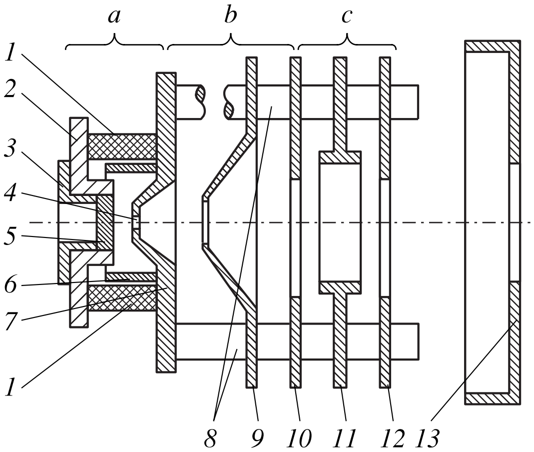


Fig. 1. Ion source design: (*a*) gas-discharge cell; (*b*) ion-optics system for acceleration and deceleration; and (*c*) lens.

measured with a Faraday cup (FC) with a 60-mm diameter of the entrance aperture. The radial distribution of the beam current density was measured by moving a small Faraday cup in the transverse direction. Depending on the beam diameter, the small Faraday cup with a variable entrance aperture diameter of 11, 5, and 1.1 mm was used.

In order to determine the ratio of the beam components represented by ions of gases and solids, we employed an indirect measurement technique by using an FC with a 3 × 2-mm rectangular entrance slit and a mass indicator, which selected a jet of the same cross section in the beam. The mass indicator, with a transverse magnetic field $B = 0.49$ T and a transit length of 80 mm, separated the components of the mixed beam; thus making it possible to visualize the traces of the separated components on a doped quartz glass [1] and then to measure their currents. Measurements of the currents of the beam jets in two Faraday cups allowed us to determine the ratio of components in a mixed beam.

Inert gases used in the source as working gases had impurities, which could be taken for ions of solids. This made it doubtful that indirect measurements were correct. Therefore, we designed and applied a compact multichannel mass analyzer that made it possible to determine the mass composition of the beam. A feature of the device is that it has a constant magnetic field, and ten magnetic tracks with given radii and individual receiving collectors are selected in this field. Therefore, only two quantities, the particle mass and energy, are variable. By changing the particle energy, we can receive a beam jet and measure its current in different collectors; thus, determining the masses of the beam particles or seeking for its components with different masses in different collectors at a fixed beam energy. This allowed us to analyze the beam composition without changing the operating conditions of the ion source under study. This device ensured precise measurements of the components of gas and solid ions.

EXPERIMENTAL RESULTS

When working with a plasma source intended for production of negative ions, intense sputtering of a local region of one of electrodes was revealed. In this case, the entire process proceeded at a comparatively low discharge current (a few or several tens of milliamperes). This was a starting point for beginning purposeful work on a new ion source for technological uses, which has a simple design, but the physical processes proceeding in it occurred to be complex. It was revealed that, in the beam extracted from the source, there are ions of the sputtered material, which coated a beam-bombarded glass target with a film.

The following working gases were used in experiments: H_2; N_2; O_2; the atmospheric air; all inert gases, except for Rn; and B, C (graphite), Al, Ti, Fe, Ni, Cu, Y, Zr, Nb, Mo, Ta, W, and Pb served as sputtered substances.

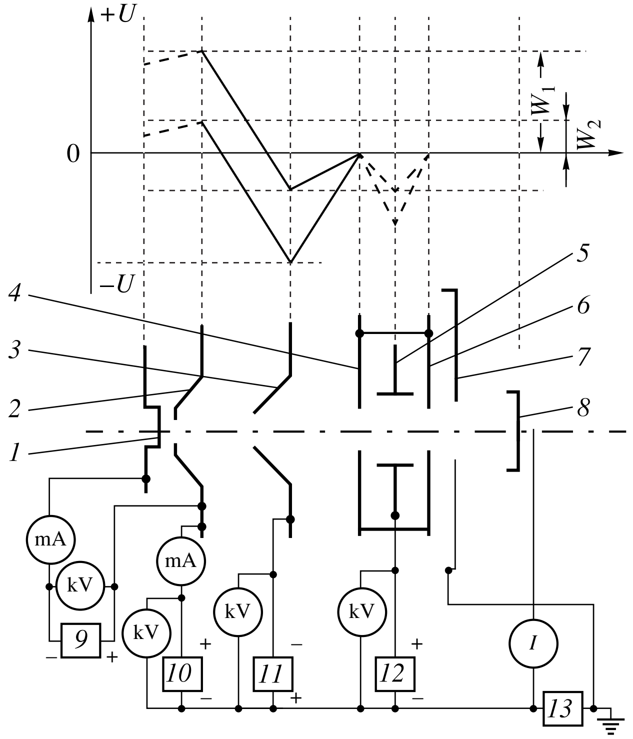


Fig. 2. Electric circuit diagram for the source connections and voltage charts: (*1*) cathode unit; (*2*) anode (emitting electrode); (*3*) accelerating electrode; (*4*) decelerating electrode (the first electrode of the electrostatic lens); (*5*) central (focusing) electrode of the electrostatic lens; (*6*) the last electrode of the electrostatic lens; (*7*) shield; (*8*) collector; (*9*) discharge power supply unit (BP-100); (*10*) high-voltage power supply unit, $U = 110$ kV; (*11*) high-voltage power supply unit, $U = 50$ kV; (*12*) high-voltage power supply unit, $U = 50$ kV; and (*13*) current-relay coil. At equal extraction (accelerating) voltages, beams of different energies ($W_1 > W_2$) are obtained.

1. Source operation in the pulse mode ($\tau_{imp} = 0.1$–1 ms and $F = 1$–50 Hz). The pulse operation mode of the source was achieved through pulse gas admission and pulse power supply of the discharge. The accelerating, decelerating, and focusing voltages were constant.

Gas ion beams with currents of hydrogen, helium, argon, xenon, nitrogen, and oxygen—up to 60, 40, 30, 10, 20, and 20 mA, respectively—were extracted from the source with a 3-mm-diameter emission hole at an accelerating voltage close to 40 kV. Depending on the type of the working gas, sputtered substance, and source operating mode, the ratio of the flow of the sputtered solid-substance ions to the total beam current was 1–30%.

For example, the results obtained were as follows. In source *1*, tantalum was a sputtered material, and helium was a working gas. The currents of helium, tantalum, and heavy-particle ions measured in the beam were 16, 1.5, and 2 mA, respectively. We believe that

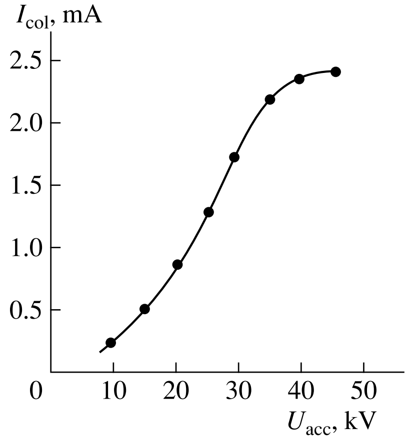


Fig. 3. Collected ion current as a function of the accelerated voltage in Ar at constant U_{ch} = 3 kV and $p = 3 \times 10^{-5}$ torr.

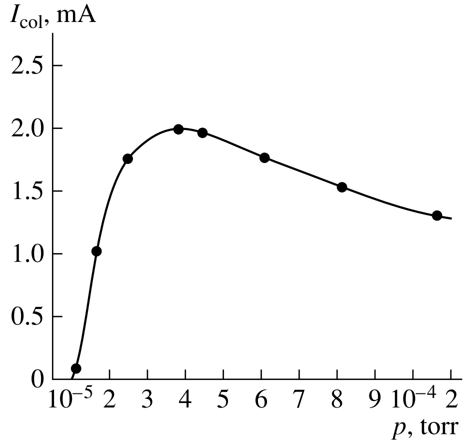


Fig. 4. Collected ion current as a function of pressure at constant U_{ch} = 3.4 kV and U_{acc} = 40 kV in Ar.

heavy particles are charged groups (clusters) of Ta atoms. It can be concluded that the content of Ta atoms in the beam was 7%. If its fraction is added to that of heavy particles, then the ratio of the sputtered particles in the beam to the gas ones is 17%. In the source *2*, molybdenum is a sputtered material, and hydrogen and helium are working gases. In the case of hydrogen, the total beam current was 20 mA, and the currents of Mo and heavy-particle ions were 4 and 2 mA, respectively. Thus, the ratios of the Mo ion current and the current of the sputtered particles to the total beam current were 20 and 30%, respectively. When helium was used as a working gas, the total beam current was 30 mA, and the currents of Mo ions and heavy particles were 8–10 and 2 mA, respectively. Hence, the ratios of the Mo ion and sputtered-particles currents to the total beam current were 27–33 and 33–40%, respectively.

In source *3*, copper was a sputtered material, and helium was a working gas. The total beam current was 28 mA, the copper ion and heavy-particles currents being 2 and 3.5 mA, respectively. Hence, the ratios of the Cu ion and sputtered-particles currents to the total beam current were 7 and 19%, respectively.

In the *4*, titanium was a sputtered material, and hydrogen and argon were working gases. The total beam current obtained for hydrogen was 10 mA, the titanium ion current was 1.2 mA, and the heavy-particles current was not observed. Hence, the ratio between the Ti ion and hydrogen currents was 12%. When argon was a working gas, the heavy-particles current was not observed, and the total beam and Ti ion currents were 2–3 and 0.4–0.6 mA, respectively, yielding a ratio between the Ti and Ar currents of 20%.

The aforementioned total beam currents extracted from the source differ significantly from each other not only because of different working gases and sputtered materials used. The difference between the currents is mainly due to the fact that the source operated under different conditions determined by the gas pressure, discharge parameters, etc. Note that the pulse operating mode of the source is important, because the treated sample is heated to a much lower temperature than in the case of the continuous operating mode.

2. Source operation in a steady-state (continuous) mode. In technological experiments, a continuous source operation mode was preferred over a pulse mode, because the dose of the implanted ions is accumulated more rapidly in the former case. Therefore, the source was switched to a continuous mode, in which its full-scale studies were carried out. Since the source was not cooled, the extracted beam current was restricted to 3 mA, but it could be raised up to 10 mA for short time intervals.

We do not present such an experimental parameter as the discharge current in the source, because the discharge in it and the beam formation and extraction in the accelerating gap represent a unified process, in which a complex current of positive and negative particles flows in three circuits: cathode–anode, anode–accelerating electrode, and cathode–accelerating electrode. Therefore, the total beam current was taken for the main parameter.

Figure 3 shows a typical dependence of the total collected ion current on the accelerating voltage, which does not virtually differ from the corresponding characteristics of plasma ion sources (e.g., a duoplasmatron). It has a long rising section (close to a linear one) smoothly changing to a saturated value.

Figure 4 shows the total collected ion current as a function of the pressure in the chamber measured near the accelerating gap. The initial section of the characteristic with a large slope demonstrates the plasma density increase owing to the build-up of ionization processes in the plasma. The characteristic then transforms into a fairly broad optimum region with a further gradual decrease, which is determined by a reduction of ionization processes and intensified near-wall particle recombination.

Figure 5 shows the total collected ion current as a function of the discharge voltage in the GDC at various pressures and accelerating voltages. The superposition of discharge currents in the GDC and the accelerating gap leads to a complex plasma generation process, in which one current may prevail over another. As a result, it becomes difficult to interpret the observed processes. Curves *1* and *2* (for differing accelerating voltages (21 and 25 kV) and equal pressures of 5×10^{-5} torr; which correspond to the extraction of low-current beams) rise smoothly and have slight curvatures. Curves *3* and *4*, which were measured at equal accelerating voltages of 30 kV and differing pressures (4×10^{-5} and 7×10^{-5} torr), have profiles indicating a discharge-intensity build-up and an increase in the collected current. Curve *5*, measured at a 45-kV accelerating voltage and an optimal pressure of 2.2×10^{-5} Torr, indicates more intense beam current collection. Curves *3*–*5* have typical humps characterizing the regions of intense source operation. Operating modes with two and three humps were observed. The dips between them in characteristics were no deeper than a half of the beam current at the corresponding maximum. This phenomenon can be explained by the variable intensity of the high-voltage discharge in the accelerating gap. The high and low discharge intensities are characterized by humps and dips, respectively.

The radial current-density distributions in a focused (curve *1*) and defocused (curve *2*) beam (see Fig. 6) were measured by using the FC with a slit of 2 × 3 mm. The FC moved in the horizontal direction transverse to the beam, being at a distance of 15 cm from the last ion-optics electrode. The total beam current measured by the large FC was 0.48 mA.

The most accurate measurements of the ratios between the currents due to ions of gases and solids in the beam were performed with a mass analyzer and, depending on the type of the working gas, sputtered material, and source operation conditions, yielded 1–10%.

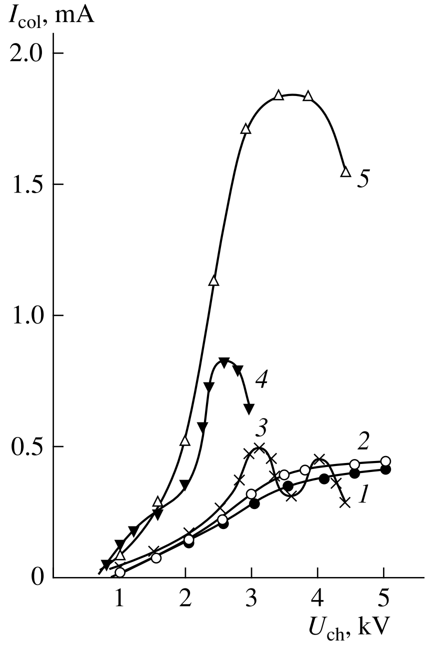


Fig. 5. Collected ion current as a function of the discharge voltage in the gas-discharge cell.

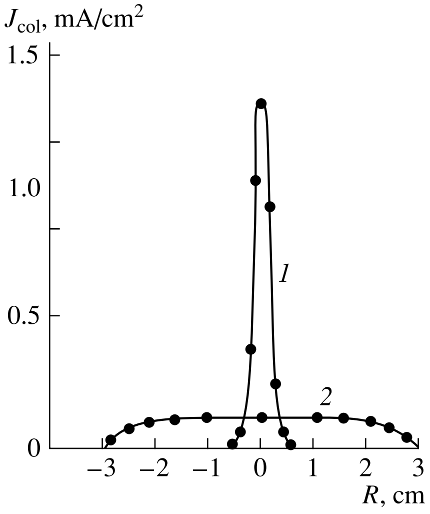


Fig. 6. Radial distribution of the current density in a (*1*) focused and (*2*) defocused beam.

Interesting results were obtained with oxygen as a working gas and copper as a sputtered material for variable discharge voltages, with other source parameters being unaltered. At a 1-kV discharge voltage, the copper ion current obtained for a 1-mA total beam current was 66 μA; thus, yielding a gas-to-solid ion current ratio of 6.6%. At a 2-kV discharge voltage, the copper ion current decreased to 30 μA, the total current increased up to 1.5 mA, and the corresponding ratio was 2%. This effect is caused by different degrees of ionization of ions of gases and solids. It was noticed that, at lower discharge voltages, metal ions are ionized more intensely, and, as the discharge voltage increases, the degree of ionization of ions of solids falls and that of gas ions rises abruptly. When the discharge voltage was within 1 kV, the gas inflow into the source was at a minimum, and ion beams with low currents no higher than 250 μA were extracted; the fraction of ions of solids in these beams reached 43%. Note that difficultly sputtered metals were ionized more easily than easily sputtered metals. Tantalum, niobium, and molybdenum were compared to copper. For example, Ta ion current reached 110 μA. Nonmetals, such as boron and carbon, are difficultly ionized, and the ratio of the number of these ions to the number of gas ions in the beam was ~1%.

An important fact is that, due to low discharge currents, molecular ions constituted 70% of the gas component of the beams extracted from this source.

In the pulse source operation mode, beam currents with ions of solids were much higher than in the continuous mode. This is determined by a higher intensity in the pulse operating mode and better conditions for ionization of sputtered particles.

In order to increase the yield of the ion current extracted from the source, we tested multiaperture emitting electrodes (anodes) with three and five identical holes with 3-mm diameters. This source version exhibited stable operation but consumed a much larger amount of gas. Note that the beam jets ejected from each emission hole, which were clearly observed visually, converged on the axis, and, at a distance of 1–2 cm from the emitting electrode, the jets merged into a common beam with a crossover section approximately equal to the diameter of a single emission hole. This source version was not studied.

A slit-geometry source was also tested. The slit dimensions were 10 × 2 mm. A film was deposited on a glass target in the pulse source operation mode with hydrogen as a working gas and aluminum as a sputtered material. The selection of the cathode–anode and anode–accelerating electrode distances in this source determines the configuration of the electric and magnetic fields, being the principal factor in the optimal source operation. The values of these distances were determined experimentally.

3. Operation of an ion source with a single power supply unit. Since small discharge currents comparable to the beam current were recorded in the source, it was interesting to switch on the source without a discharge-supplying rectifier (BP-100), by using a single high-voltage power supply unit that supplies the energy for discharge initiation and maintenance. This was accomplished by connecting the cathode to the ground through a ballast resistor. In this case, the entire process of the ion source operation did not qualitatively differ from the operating mode with two power supply sources connected to the source. In this source operation mode, which was much more stable, a sort of discharge self-regulation occurred. The discharge initiation and maintenance were shifted to lower voltages. The discharge voltages were controlled in a wide range of 500 V to 3 kV by replacing the ballast resistors. Note that at low discharge voltages, the yield of ions of solids increases in the beam, and at high voltages, it is the fraction of gas ions that increases. In the operating conditions with a single high-voltage power supply unit, higher beam currents were extracted from the ion source than in the operation with two supply units. This is determined by better matching of the discharge currents in the GDC and the accelerating gap during beam formation and extraction in the case when a single high-voltage power supply unit is connected to the ion source.

PHYSICAL PROCESSES OCCURRING IN THE SOURCE

When an accelerating voltage is initially applied between the source's anode and the accelerating electrode of the IOSAD at a pressure of 10^{-4}–10^{-6} torr, a low-current high-voltage discharge is initiated. The electron flow originated from this discharge is guided by a narrowing magnetic-field "funnel" and streams to the central part of the anode. A significant fraction of the electron stream flies through the emission hole in the anode and moves in the direction opposite to that of the electric field applied between the cathode and anode in the GDC. Despite its deceleration, electrons reach the central region of the cathode. The decelerated electrons simultaneously interact with the fan-shaped magnetic field of the source, acquiring a complex rotational motion; thus, elongating their path. A part of electrons from the high-voltage discharge does not reach the cathode, because they have a refection point in the magnetic trap. The contribution of these electrons to ionization is especially great. The secondary electron emission from the cathode and primary electrons from the discharge in the accelerating gap produce efficient ionization in the GDC, initiating a high-voltage magnetron-type glow discharge. This discharge, in the form of a plasma bunch shaped as a truncated cone, turns with its smaller base near the sputtered cathode–disk. The ions from this discharge bombard the cathode and lead to efficient sputtering of the disk material in the form of ions, atoms, and clusters. As a glow discharge is initiated, the electrons being in the GDC of the source begin to ionize the sputtered particles of the substance and increase their ion concentration and the gas-kinetic pressure of the discharge plasma consisting of gas and solid fractions.

Local heating of the central disk region, which is produced by the primary electrons from the accelerating gap, to a temperature close to the melting point for the substance (of which the disk is manufactured) should be considered as a process promoting the particle sputtering from the disk cathode. Local evaporation from a small central region (a crater) of the cathode is also possible. This can be inferred from the appearance of the sputtered region on the cathode. This region takes the form of a circle with a crater at the center. The cross section of the sputtered region usually has a shape of a third-order curve called a witch of Agnesi, and as the intensity of sputtering increases, the curve's profile becomes more pronounced. The crater diameter is 1.2–2 times larger than the diameter of the emission hole, and its depth is 0.3–0.7 of the crater diameter. The diameter of the sputtered region changed from 1 to 12 mm, depending on the specified conditions and the discharge mode. The specified conditions are as follows: the configuration of the magnetic field in the gas-discharge cell; the configuration of the electric field in the cell; the diameter of the emission hole; the diameter of the accelerating electrode aperture; and various sputtering coefficients of the materials of which disk cathodes were manufactured. The dimensions of the disk sputtering region mentioned above were reached after 5–6-h operation of the source with copper used as a working substance.

A truncated cone-shaped plasma bunch of a magnetron-type high-voltage glow discharge, which burns in the GDC, embraces the anode with its large base. Positively charged particles are collected from the plasma region near the emission hole and are initially shaped in the form of a charged plasma stream directed to the accelerating electrode. At low-intensity discharge modes, the plasma-stream boundary is near the emission hole, but as the discharge intensity and gas-kinetic pressure of the plasma increase, the plasma stream protrudes out of the emission hole as a well observed "plasma tongue". At a further increase in the gas-kinetic pressure of the plasma in the GDC, the plasma tongue grows in length and its diameter expands hardly. The gas-kinetic plasma pressure can be controlled by the working-gas inflow and the discharge-voltage variation. Note that, when the discharge voltage changes in a wide range from 0.6 to 5 kV, the discharge current changes insignificantly.

The elongation of the plasma tongue is also strongly affected by the build-up of the voltage in the accelerating gap, which shifts the tongue to the accelerating electrode. When they come in contact, the high-voltage gap is broken down. The elongation of the plasma tongue in this source with increasing accelerating voltage differs from the corresponding processes in other types of sources, in which the plasma is repelled by the electric field to the emitting electrode [2]. In our opinion, this effect is explained by the fact that the accelerated electrons of the beam plasma enter the accelerating gap from the ion-optics channel and intensify the high-voltage discharge in it. The magnetic-field funnel in the accelerating gap does not allow this discharge to expand, but allows it to advance in the longitudinal direction.

The plasma-tongue elongation leads to an increase in the beam current, which can be explained by the following factors: an increase in the ion concentration in both the GDC and the discharge of the accelerating gap (plasma tongue); an increase in the emitting area of the plasma surface; and narrowing of the accelerating gap d (plasma tongue-accelerating electrode), where d is determined by the Child's equation $J = \chi U^{3/2}/d^2$.

All these factors ensure the high emissivity of the system despite an unfavorable shape (convex and long) of the emitting surface for beam shaping. However, the experiment has shown good beam formation and focusing, which can apparently be explained by a favorable magnetic-field distribution and value in the accelerating gap.

When a two-electrode IOS (emitting electrode–accelerating electrode) was used in intense operating modes, the source exhibited unstable operation, because the control over the plasma tongue was poor, and its contact with the accelerating electrode resulted in a breakdown of the high-voltage gap. At a beam current of even 1–2 mA, the process was already uncontrollable. In a two-electrode IOS, the drifting electrons of the beam plasma are accelerated in the high-voltage gap and, moving to the source, substantially impair its performance characteristics. The change to a three-electrode IOS (emitting electrode-accelerating and electrode-decelerating electrode) ensured an improved control over the plasma tongue; thus, stabilizing the source operation and allowing us to extract higher beam currents from it. In such an IOS, the source plasma is separated from the beam plasma due to the formation of a potential barrier in the ion-optics channel for beam-plasma electrons, which drift to the source. The three-electrode IOS forms a virtual-cathode region from the electrons that compensate for a positive space charge of the ion beam.

In order to focus the ion beam, electrostatic focusing was used, which allowed the injection of a parallel beam into the mass analyzer. Focusing was performed by a single electrostatic lens described above in the first section.

CONCLUSION

The following features of the ion source should be highlighted.

(1) The source operates in the pulse and continuous modes. In the latter case, the source requires no cooling at a beam current below 3 mA due to small discharge currents. As the beam current rises up to 3–5 mA, the emitting electrode (the source's anode) requires cooling, since it bears the chief thermal load being exposed to electron bombardment from the source discharge and to the back accelerated electrons from the high-voltage gap.

(2) The source has high emissivity.

(3) The source has a magnetron-type cold cathode that ensures the operation with arbitrary gases.

(4) Ions of solids (metals and nonmetals) were obtained from this source.

(5) Ions of solids can be obtained without their preliminary evaporation.

(6) The time of uninterrupted source operation is determined by the time of sputtering of approximately 70–80% of the cathode substance and amounts to hundreds of hours.

(7) The ion-optics system used in this study allowed us to control the beam energy from 1 to 100 keV without an appreciable beam current fluctuation and to focus the beam.

(8) A discharge can be initiated and maintained in the source by using only one high-voltage power supply unit. Its operation becomes more stable, beams with

higher currents can be extracted from the source in this case, and its power supply system is simplified.

(9) Ions of solids cannot be obtained from the source without a working gas. In this respect, its behavior is similar to that of a classical sputtering source.

This source was used in technological experiments on implantation of ions of gases and solids, surface cleaning, and deposition of thin films; for example, for doping the ends of cylinders of diesel engines with a mixed beam of zirconium and oxygen ions, and for doping the interior surfaces of diesel nozzles with a mixed beam of carbon and nitrogen ions. Experiments on modification of the surfaces of aircraft-engine blades were also carried out.

ACKNOWLEDGMENTS

We are grateful to S. V. Grigorenko, V. G. Kuznetsov, and V. P. Sidorov for attention and valuable comments.

REFERENCES

1. E. A. Meleta, L. P. Veresov, V. N. Kizhin, *et al.*, Fiz. Khim. Stekla **17** (3), 77 (1991).
2. O. F. Poroshin and Zh. Zh. Kutan, *Formation and Study of Intense Beams of Hydrogen Ions* (Izd. FTI GK IAE SSSR, Moscow, 1963).

Translated by A. Seferov

Technical Physics, Vol. 45, No. 4, 2000, pp. 497–500. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 118–122.
Original Russian Text Copyright © 2000 by Veselov, Dzhumaliev.

BRIEF COMMUNICATIONS

Synthesis of Ordinary- and Inclined-Texture Zinc Oxide Films in an Inhomogeneous Gas-Discharge Plasma

A. G. Veselov and A. S. Dzhumaliev
Institute of Radio Engineering and Electronics (Saratov Branch), Russian Academy of Sciences,
Saratov, 410019 Russia

Received March 15, 1999

Abstract—Zinc oxide films having ordinary and inclined texture were grown on extended amorphous substrates. It was shown that a desired structure can be provided by controlling charged particle fluxes with the substrate remaining parallel to the target plane. Experiments were performed in a dc magnetron sputterer in a mixture of argon and oxygen. The textured films were studied by X-ray diffraction and by exciting longitudinal and shear bulk acoustic waves. © *2000 MAIK “Nauka/Interperiodica”.*

INTRODUCTION

Thin films of zinc oxide are widely used for exciting bulk and surface acoustic waves (SAWs). These films offer high electromechanical coupling and are compatible with any planar process. Moreover, the fabrication technology is energy- and time-saving; hence, ZnO-based devices are inexpensive.

The orientation of texture is a crucial issue in obtaining piezoactive layers. In most works, crystallographic properties of the films are related to growth conditions: film deposition rate, substrate temperature, electrode geometry, argon-to-oxygen ratio in the mixture, direction and energy of sputtered particles, and magnetic field strength and configuration [1–6]. In [5, 6], composite structures consisting of variously textured zinc oxide films are viewed as a single possibility to advance acoustics to a mm-wave range.

Sputtered inclined-texture zinc oxide films where shear waves are excited are of particular interest. Such films are obtained by varying the substrate position with respect to the plane parallel to the target plane. Films where the texture axis makes an angle of 40° with the silicon substrate plane were reported [7]. Here, the substrate was positioned at an angle of 15° to the magnetron plane. Similar results were obtained in [8], where the effect of various sputtering parameters on the texture orientation in the films was examined. Inclined-texture ZnO films produced in a planar magnetron where a silicon substrate and the target plane made an angle of 40° were reported [6]. When the substrate is inclined, the inclination of texture depends on the film position over the erosion zone on a target.

Inclined texture is always formed at tangential delivery of a sputtered matter to the substrate [6–8]. Under these conditions, grown films are nonuniform in thickness and feature a large misorientation of the texture axis.

Complex processes taking place in a gas-discharge plasma and a great variety of process parameters influencing the quality of the films make it difficult to separate reasons and mechanisms for the formation of texture in the films. It is, therefore, necessary to find factors that specify the texture orientation in ZnO films. Based on the results of the works cited above, we studied textures formed with fluxes incident normally and at an angle to the substrate. The fluxes were formed by varying the configuration of the negative glow region (NGR) (or plasma region).

In this work, we studied the texture orientations in ZnO films deposited on extended amorphous substrates in a glow-discharge plasma generated in a dc planar magnetron sputterer. It is shown that the formation of an inclined texture region and the extension of an ordinary texture region depend on the NGR configuration. X-ray data for the textures are compared with the excitation of shear and longitudinal bulk acoustic waves (BAWs) by the textures.

EXPERIMENTAL

Zinc oxide films were prepared in a dc planar magnetron sputterer where a zinc target was subjected to ion-plasma sputtering in an argon + oxygen mixture (Fig. 1). Such equipment is common in the industry [9]. In our system, the NGR is easy to control by varying the magnetic field inhomogeneity.

A disk made of reagent-grade zinc (100 mm in diameter) was used as a target. To improve the thermal contact, the disk was soldered to a water-cooled copper cylinder with annular magnets inside. The thickness of the target in the sputtering site was 6 mm. The ring-shaped aluminum anode 100 mm in diameter was placed 15 mm away from the target surface. The anode voltage was +300 V (the magnetron assembly is grounded) at a current of 200 mA. The pressure of the

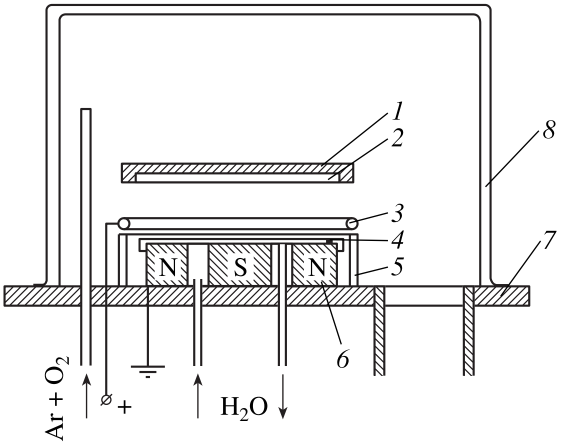


Fig. 1. Planar magnetron sputterer: (*1*) substrate holder; (*2*) substrate; (*3*) anode; (*4*) target; (*5*) quartz cup; (*6*) magnets; (*7*) vacuum system base; and (*8*) cap.

gas mixture (75% Ar + 25% O_2) in the sputtering chamber was maintained between 0.4 and 1.33 Pa by passing the gases through a precision leak.

A 11-mm thick fused-quartz substrate 60 mm in diameter was used for BAW initiation to study piezoelectric properties of the films (along with the X-ray texture examination). The substrate temperature during sputtering, 620 K, was kept with a VRT-3 high-precision temperature controller. The target–substrate distance was 35 mm.

The above process parameters were found to be optimal in terms of the film thickness reproducibility and piezoactivity. The obtained films were optically transparent, and their growth rate was 0.7–0.8 nm/s. In the experiments, the only variable parameter was the NGR shape in the chamber. Figure 2 shows the most typical NGR shapes. In a planar magnetron sputterer, the NGR shape is changed by varying the magnetic field strength and configuration. It should be noted that characteristic erosion zones differ for the different NGR shapes. In Fig. 2a, the NGR (type I) has the form of a flame of diameter ~40 mm, which narrows away from the target. The erosion zone is a circle whose diameter is also ~40 mm. In the second case (Fig. 2b), the glow region is a torus pressed against the target edge (NGR of type II). The erosion zone, in this case, is a ring with outer and inner diameters of ~80 and ~40 mm, respectively.

The texture orientation was tentatively (qualitatively) estimated by exciting BAWs [10]. With the substrate thickness mentioned above, the surface of the piezoactive film generates BAWs, and echo-signals from longitudinal (inclined texture) and shear (ordinary texture) BAWs can be detected separately. Time delays for the echo-signals are ~3 and ~7 μs, respectively. To generate and receive BAWs in the reflection mode, we used an interdigital transducer made of a metal foil on a separate insulating substrate. The pitch of the structure was 150 μs; the finger length, 2 mm; and the number of fingers, 10. When exciting SAWs, this transducer operates at a frequency of ~10 MHz. In our case, the operating frequency was 400 MHz. Under such conditions, several planar-front BAWs are generated, and echo-signals can be detected. The small grid sizes and mobility of the transducer, as well as the possibility of exciting waves of different types due to the presence of nonuniform electric fields, allow it to be thought of as a probe suitable for rapidly and effectively estimating the piezoactivity of the film and the texture orientation.

The relative intensity of the first echo-signal can then be taken as a measure of the piezoactivity. In Figs. 3 and 4, this intensity is plotted on the vertical axis.

Within our approach, the texture orientation can be most simply estimated (qualitatively) if the film is grown on a metal sublayer. In this case, BAWs are initiated only by the normal component of the electric field. All the ZnO films studied were deposited onto an aluminum sublayer preevaporated onto the quarts substrate.

Figure 3 demonstrates the BAW excitation efficiencies for samples obtained in the gas-discharge plasma with the NGR shapes of type I (Fig. 3a) and type II (Fig. 3b). Longitudinal BAWs are excited within a circular zone at the center of the substrate (Fig. 3a). Its diameter increases in going from the NGR of type I to that of type II. Next to this zone is a zone of shear BAWs, which is most pronounced for the type-I NGR.

The X-ray study of the film structure performed with a DRON-2 diffractometer confirms the above qualitative results. For the type-I NGR, the zone of (002) reflections has a diameter of 10–15 mm and coincides with the zone of longitudinal BAWs. This zone increases to ~35 mm in diameter for the type-II NGR. In the zone of shear BAWs, (100) reflections are observed. In the transition region, (002), (100), and (101) reflections are seen, which is an indication of the mixed texture. The (002) and (100) reflections correspond to ordinary and inclined textures, respectively.

In a planar magnetron sputterer, the magnetic field strength and configuration have a decisive effect on NGR formation. Methods for creating nonuniform magnetic fields are fairly simple and well known [8]. Therefore, we give only the shape of the flame, influencing the film texture. Note that the shape of the flame depends not only on the magnetic field but also on the pressure in the chamber. For better visualization of the processes occurring in the NGR, the potential distribution over the target during sputtering was measured with a probe. The results shown in Fig. 4 point to the existence of charged particle fluxes over the substrate. This presumably results in the growth of inclined-texture films.

For the type-I NGR, a considerable cathode drop of the potential in the central zone of the target (the circular erosion zone) leads to the formation of a charged

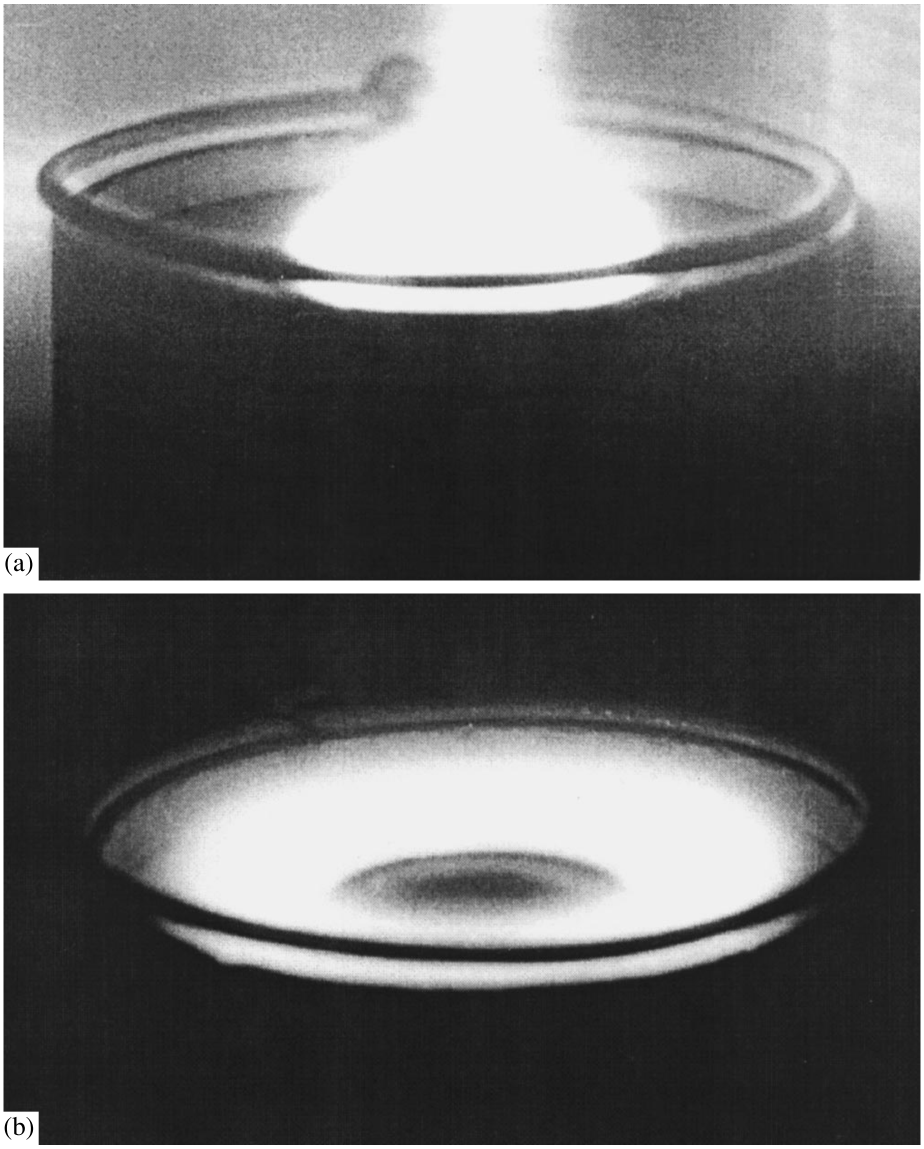


Fig. 2. Negative glow regions for different magnetic systems: (a) type I and (b) type II.

particle flux in this zone. These particles recombine on the way to the substrate. The density of the flux falls toward the target edge in a bell-like manner [11]. This irregular-in-density flux of particles bombarding the substrate produces a temperature gradient along its radius and also causes adatoms to migrate from an area with an elevated density of the incident particles to that where the density is reduced. This is likely to explain the formation of inclined-texture regions. For the NGR of type II, the target is sputtered from the circular erosion zone. The distribution of sputtered particles is also nonuniform in this case. It has a bell-like shape with a maximum nearly at the center of the erosion zone. Thus, there exist two zones of bell-like sputtering on the target diameter. The combined effect of two diametrically opposite sputtering zones might result in the situation that particle fluxes, when added, form a uniform central zone where the particles are incident to the substrate at right angles. Under such conditions, the particles do not produce the temperature gradient in the central zone, and the driving force for adatom migration is absent. As a result, ordinary-texture films grow. A further increase in the diameter of the erosion zone brings us to a conventional sputtering system, where the substrate is practically removed from the recombination zone. It is well known that only ordinary texture forms in this case.

Note that the simplicity of creating a large (10 × 30 mm) region of shear texture suggests that the films can be applied to generating SAWs. In this case, there is no need for the sublayer upon using the interdigital transducer, and the design is simplified. This has made

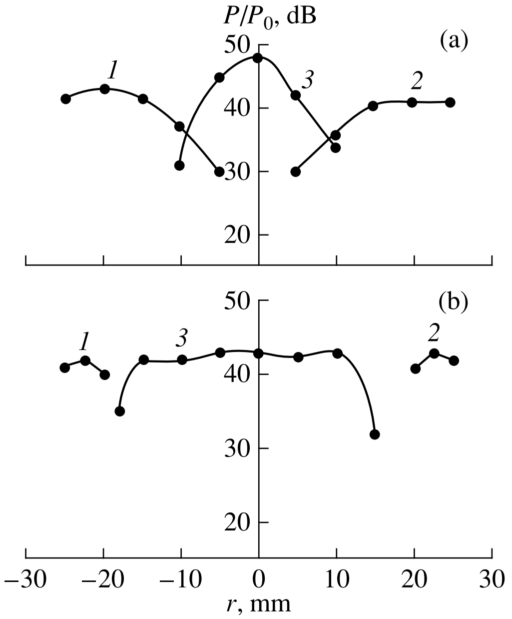


Fig. 3. Efficiency of generating (*1*, *2*) shear and (*3*) longitudinal BAWs vs. distance from the substrate center: (a) type I and (b) type II.

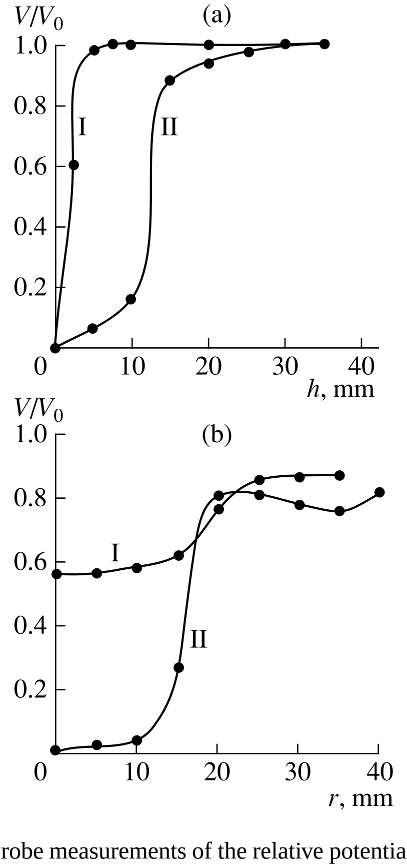


Fig. 4. Probe measurements of the relative potential against (a) height and (b) radius for the type-I and type-II NGRs.

it possible to produce prototypes of sublayer-free ZnO–Si structures with a high SAW excitation efficiency.

CONCLUSION

The feasibility of controlling the texture orientation in ZnO films obtained in a planar magnetron sputterer was shown. The orientation depends on whether the incidence of particles is normal or tangential and is also related to the potential gradient over the substrate. Our findings may be helpful in fabricating films with a desired texture on large substrates. ZnO composite structures with alternating ordinary- and inclined-texture layers can be produced by simply displacing the substrate off the chamber axis without changing the inclination of the substrate with respect to the target. Of great practical value is also the possibility of effectively generating SAWs in inclined-texture ZnO films.

ACKNOWLEDGMENTS

This work was supported by the Federal Program “Integration” (grant no. 696.3).

REFERENCES

1. G. R. Aita, Ultrason. Symp. Proc. (1980), p. 795.
2. K. Tominaga, S. Iwamura, I. Fujita, *et al.*, Jpn. J. Appl. Phys. **21**, 999 (1982).
3. A. C. Anderson and D. E. Oates, Ultrason. Symp. Proc. 329 (1982).
4. M. Miura, Jpn. J. Appl. Phys. **21**, 264 (1982).
5. B. Hadimioglu, L. J. La Comb, D. R. Wright, *et al.*, Appl. Phys. Lett. **50**, 1642 (1987).
6. D. Howell, L. Goddard, and B. T. Khuri-Yakub, Ultrason. Symp. Proc. (1987).
7. J. S. Wang and K. M. Lakin, Ultrason. Symp. Proc. (1982), p. 480.
8. S. V. Krishnaswamy, B. R. McAvoy, and W. J. Takei, Ultrason. Symp. Proc. (1982), p. 476.
9. B. S. Danilin and V. K. Syrchin, *Magnetron Sputtering Systems* (Radio i Svyaz’, Moscow, 1982).
10. A. G. Veselov and A. S. Dzhumaliev, Élektron. Tekh., Ser. 7: Tekhnol. Organiz. Proizv. Oborudov., 71 (1988).
11. K. Tominaga, N. Ueshiba, Y. Shintani, *et al.*, Jpn. J. Appl. Phys. **20**, 519 (1981).

Translated by V. Isaakyan

Technical Physics, Vol. 45, No. 4, 2000, pp. 501–504. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 123–127.
Original Russian Text Copyright © 2000 by Lobastov, Shudegov, Chudinov.

BRIEF COMMUNICATIONS

Computer Experiments on Aluminum Single Crystal Deformation

A. I. Lobastov, V. E. Shudegov, and V. G. Chudinov

Udmurt State University, Krasnoarmeĭskaya ul. 71, Izhevsk, 426034 Russia

Received July 15, 1998

Abstract—Two series of computer experiments on the evolution of the atomic structure of aluminum single crystals were performed. Specimens were subjected to pulse loading and high plastic strains. Atomic interaction was described by the pseudopotential (in the former case) and the empirical potential (in the latter). The system was strained step by step and, after each step (2%), was relaxed to a new state at 300 K with a molecular dynamic method. The final strain was 70%. The plastic behavior of aluminum differs when described by the different potentials. In both cases, however, deformation proceeds largely by the twin mechanism. A changeover to another plastic deformation mechanism or scheme depends on interatomic forces, specimen size, and density of defects in a specimen. © *2000 MAIK “Nauka/Interperiodica”*.

INTRODUCTION

Plastic deformation of solids is known to be associated with various kinds of defects. In spite of the fact that dislocations and other elementary “carriers” of deformation can be observed by many direct methods, plastic deformation mechanisms, particularly those of impact-induced large plastic deformation, have not yet completely understood. A loss of information can be compensated for by computer-aided mathematical simulation of related processes. Over the last years, computer simulation has become a powerful method for studying plastic deformation mechanisms. For example, a mechanism of transition from the defect-free state of a uniaxially deformed argon crystal to the defect one was studied in computer experiments in terms of a quasi-three-dimensional quasi-static model [1]. In [2, 3], the behavior of α- and γ-Fe single crystals under instantaneous plastic deformation was examined with a molecular dynamics method. This allowed researchers to directly observe the origination and evolution of defects. In this work, which is an extension of [2], we consider the structure of face-centered crystals under plastic deformation at different interatomic potentials.

COMPUTING TECHNIQUE

Computer experiments were performed on face-centered cubic aluminum single crystals consisting of 2048 atoms. The strain ε was built up step by step with repeated boundary conditions at $T = 300$ K. At each step, a single crystal was compressed by 2% along the 0Z-axis (negative ε_{zz} strain) and stretched along the 0X- and 0Y-axes so that its volume remained unchanged (the strains ε_{xx} and ε_{yy} are positive). Two series of the computer experiments were carried out. In the former case, atomic interaction was described by the Heine–Abarenkov–Animalu pseudopotential [4], while in the latter, by the Baskes empirical potential [5]. After each deformation step, the system was brought up to a new equilibrium state with the molecular dynamics method [6]. During the experiments, the atomic radial distribution function (ARDF), temperature, system energy, and other parameters were monitored. The arrangement of atoms in different planes was analyzed.

RESULTS AND DISCUSSION

Every experiment was conventionally subdivided into several stages. The beginning and the end of a stage were determined from the ARDF. At these instants, it comprised peaks inherent in a face-centered lattice. During each stage, the ARDFs either heavily diffused or had extra peaks. In the case of the pseudopotential, for example, the deformation process begins with polymorphic transformation [7]. In our experiments, a regular body-centered cubic (bcc) cell was observed at $\varepsilon \approx 21\%$, when a unit face-centered cubic (fcc) cell is compressed along the 0Z-axis by 0.356 to 0.286 nm and is stretched along the 0X- and 0Y-axes by 0.356 to 0.399 nm. The associated ARDFs are represented in Fig. 1. During this process, the potential energy of the system linearly rose with strain up to 14% and then slowly decreased till the end of the phase transformation. Subsequently, it decreased only because of the formation of crystal defects. At $\varepsilon \approx 24\%$, a displacement of the {100}' planes was observed.[1] At this stage, these planes were the close-packed {111}

---

[1] Hereafter, the prime means that the Miller indexes are determined with respect to the initial coordinate system.

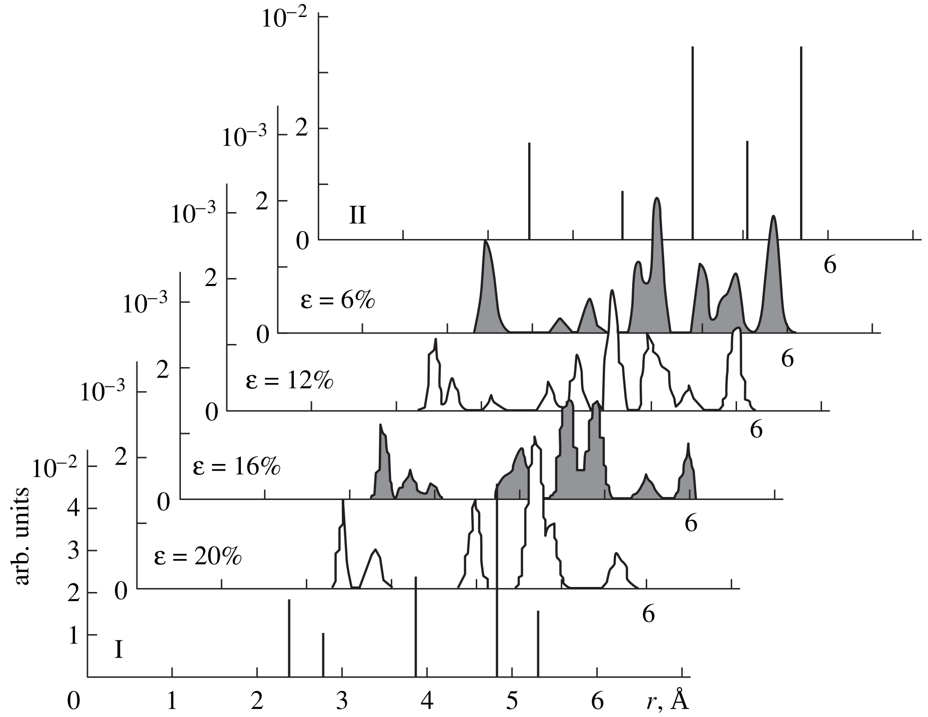


Fig. 1. ARDF modifications: I, bcc lattice; II, fcc lattice.

planes in the [010]' and $[0\bar{1}0]'$ directions. Under these conditions, a defect structure having both stacking faults and twins forms. Figure 2 shows the (110){(001)'} plane of the crystal strained 30%. It is seen that interatomic and interplanar distances completely coincide with the same distances in the starting fcc single crystal.

The next stage of plastic flow is governed by a dislocation-free mechanism. The successive atomic arrangements depending on the strain are given in Fig. 3. Due to collective movement of atoms at ε = 40%, the close-packed {(120)'} planes make an angle α = 63.435° with the related planes of the crystal strained 30%. Unlike the previous state, the resulted defect state has more extended regular face-centered regions. From this time on, the crystal deforms due to partial dislocations moving over different slip planes. At this stage of deformation, Lomer–Cottrell barriers arise, which totally block slip planes, reducing crystal ductility [8]. Figure 4 demonstrates the atomic arrangement in the (110) plane at ε = 60%. At ε = 70%, the barriers disappear, the crystal structure consists of twins only, and the ARDF again becomes typical of the fcc lattice. The close-packed planes, in this case, have indexes (3,5,10)'. At this value of strain, the crystal size along the 0*Z*-axis is so small that plastic deformation mechanisms become impossible to identify.

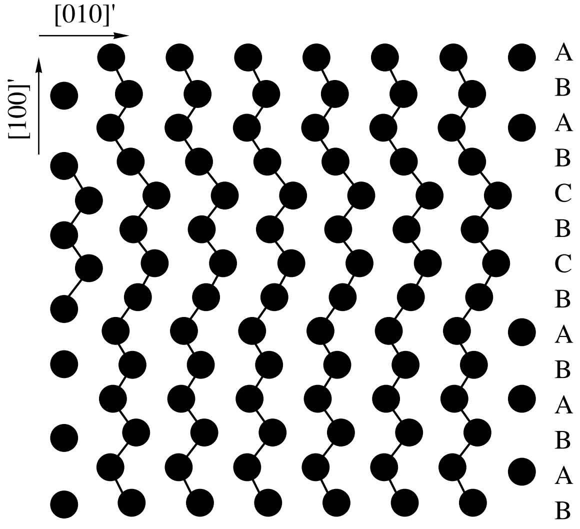


Fig. 2. Atomic arrangement in the (110) plane. ε = 30%.

When atomic interactions are described by the Baskes potential, deformation initially proceeds following the Saks scheme [9]. Starting from ε = 10%, successive displacements of the (111) planes cause twin origination and growth due to partial dislocations. At ε = 24%, the slip direction and in the slip planes start to change, and at ε = 32%, the (010)' and (001)' planes of the initially perfect lattice transform into the (111) and (110) planes, respectively. Alternating close-packed planes are shown in Fig. 5. At the next stage, ε = 38%, due to multiple slipping along different planes, the middle part of the crystal (between the dashed lines) has two (111) planes more. Subsequently, however, the

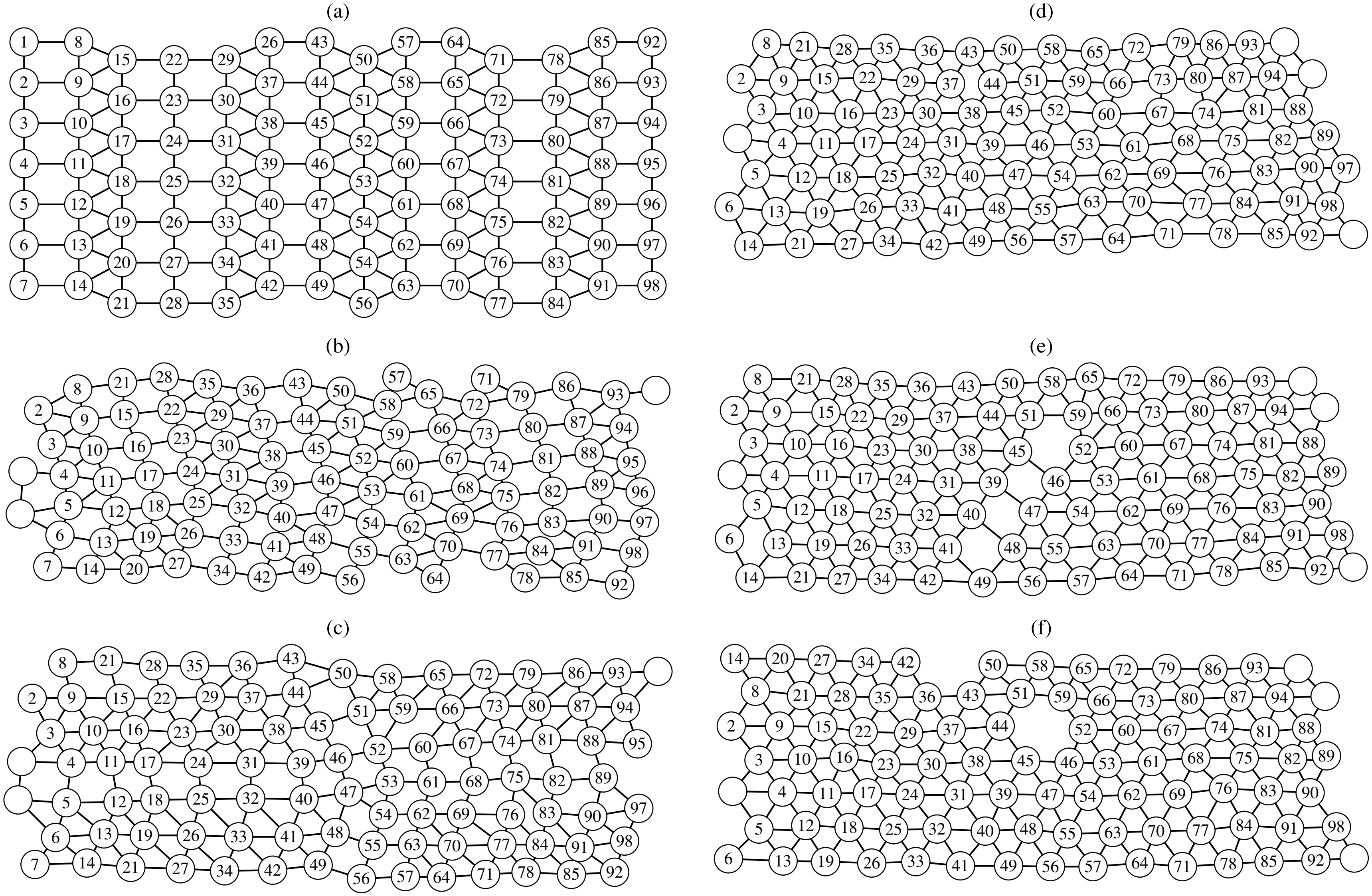


Fig. 3. Atomic arrangement in the (120)' plane. ε = (a) 32, (b) 34, (c) 36, (d) 38, and (e, f) 40%.

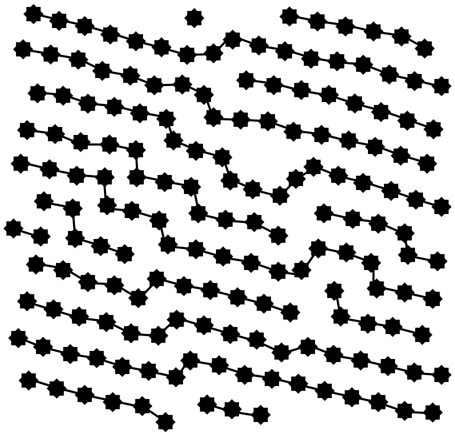


Fig. 4. Atomic arrangement in the (110) plane. ε = 60%.

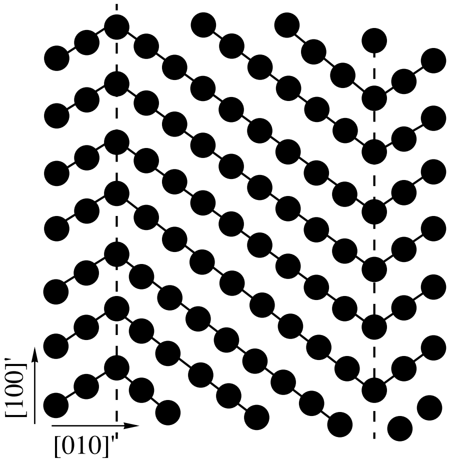


Fig. 5. Atomic arrangement in the (110) plane. ε = 32%.

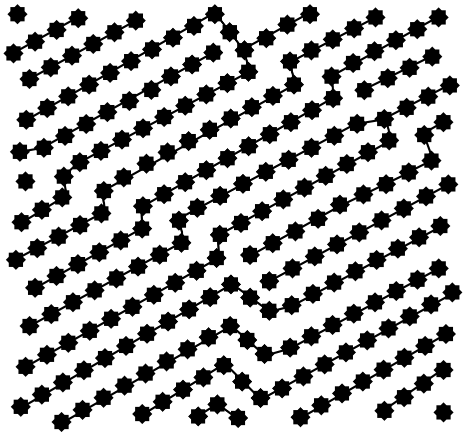


Fig. 6. Atomic arrangement in the (110) plane. ε = 68%.

single slip system is again observed, and at ε = 52%, the lattice becomes free of defects but rotated about the initial coordinate system.

At higher strains, the process becomes still more complicated because of a large number of arising defects that interact with each other. This is demonstrated with the (110) plane in Fig. 6. At ε = 70%, the crystal acquires the same structure as in Fig. 4. The difference is that the crystal strained 70% has only three close-packed planes between the dashed lines.

Thus, despite the different plastic behaviors of the crystals subjected to the different potentials, deformation in both cases proceeds largely by the twin mechanism. A change-over to another deformation mechanism or scheme depends on interatomic forces, crystal size, and density of defects in a crystal.

To conclude, computer-aided simulation proves to be an efficient tool for studying plastic deformation mechanisms.

REFERENCES

1. S. V. Dmitriev, A. A. Ovcharov, M. D. Starastenkov, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **38**, 1805 (1996) [Phys. Solid State **38**, 996 (1996)].
2. V. E. Shudegov, A. I. Lobastov, V. A. Likhachev, *et al.*, Zh. Tekh. Fiz. **65** (6), 94 (1995) [Tech. Phys. **40**, 571 (1995)].
3. A. I. Lobastov, V. E. Shudegov, and V. G. Chudinov, Zh. Tekh. Fiz. **67** (12), 100 (1997) [Tech. Phys. **42**, 1460 (1997)].
4. V. G. Chudinov, E. B. Dolgusheva, and A. A. Yur'ev, Sverkhprovodimost: Fiz., Khim., Tekh. **4**, 2086 (1991).
5. M. I. Baskes and C. P. Melius, Phys. Rev. B: Condens. Matter **20**, 3197 (1979).
6. V. M. Dyadin, V. G. Chudinov, L. I. Gondyreva, *et al.*, Available from VINITI (Moscow, 1991), No. 1537-B-91.
7. I. I. Novikov and K. M. Rozin, *Crystallography and Lattice Defects* (Metallurgiya, Moscow, 1990).
8. M. P. Shaskol'skaya, *Crystallography* (Vysshaya Shkola, Moscow, 1976).
9. V. E. Panin, Yu. V. Grinyaev, T. F. Elsukova, *et al.*, Izv. Vyssh. Uchebn. Zaved., Fiz. **25** (6), 5 (1982).

Translated by V. Isaakyan

Technical Physics, Vol. 45, No. 4, 2000, pp. 505–506. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 128–129.
Original Russian Text Copyright © 2000 by Saranin.

BRIEF COMMUNICATIONS

Bifurcation of the Equilibrium Droplet Geometry in Electric Field

V. A. Saranin
Korolenko State Pedagogical Institute, Glazov, Udmurtia, 427600 Russia

Received October 19, 1998

Abstract—It is shown that neutral and charged liquid droplets placed in an external electric field admit two branches of stable ellipsoidal shapes (corresponding to energy minima) with hysteretic hard transitions between them. © *2000 MAIK “Nauka/Interperiodica”.*

Studies of the behavior of charged and neutral liquid droplets in electric fields are important because such droplets play a key role in various technological and natural processes [1]. It is well known that the spherical geometry of a charged conductive droplet becomes unstable when its charge reaches a critical value [2]. Moreover, the equilibrium geometry of a conductive droplet having the shape of a prolate ellipsoid of revolution becomes unstable when the field strength reaches a critical value [3]. A similar analysis of the behavior of a magnetic liquid droplet placed in an external magnetic field has shown that, apart from the instability of certain ellipsoidal geometries, there exists a family of stable geometries. In other words, the equilibrium droplet shape bifurcates and exhibits a hysteretic behavior [4]. In what follows, it is shown that both neutral and charged particles placed in an electric field can also have equilibrium ellipsoidal geometries that bifurcate and exhibit a hysteretic behavior.

First, consider the case of a dielectric droplet placed in a uniform electric field. Suppose that the droplet is a prolate ellipsoid of revolution and its semimajor axis is parallel to the field lines. Gravity effects are assumed to be negligible.

To determine admissible equilibrium ellipsoidal droplet geometries, let us calculate the potential of an ellipsoid placed in an external electric field. If the field is uniform, then the electrical component of energy is (see [5])

$$W_e = -\frac{1}{2}E_0 P_x = -\frac{2\pi\varepsilon_0 ab^2 E_0(\varepsilon-1)}{3(1+(\varepsilon-1)n_x)}, \qquad (1)$$

where P_x is the projection of the ellipsoid's dipole moment onto the direction of the electric field; ε is the droplet-to-ambient permittivity ratio; a and b are the semimajor and semiminor axes of the ellipsoid, respectively; and n_x is the depolarization coefficient, which can be expressed in terms of eccentricity e as follows:

$$n_x = \frac{1-e^2}{2e^3}\left(\ln\left(\frac{1+e}{1-e}\right)-2e\right).$$

Using the expression for the surface area of an ellipsoid of revolution, one can write the potential energy of the droplet as

$$W_\alpha = 2\pi\alpha R_0^2\left(\beta^{-2/3}+\frac{\beta^{4/3}\arctan\sqrt{\beta^2-1}}{\sqrt{\beta^2-1}}\right) \equiv 2\pi\alpha R_0^2 S_x. \qquad (2)$$

Here, α is the surface tension and R_0 is the radius of the sphere of volume equal to that of the ellipsoid ($R_0^3 = ab^2$). The ratio of semiaxes, $\beta = a/b$, is related to the eccentricity by the formula $e = \sqrt{\beta^2-1}/\beta$. Using $2\pi\alpha R_0^2$ as an energy unit, one can represent the dimensionless total energy of the ellipsoid as

$$W = S_x - \frac{4(\varepsilon-1)N^2}{3(1+(\varepsilon-1)n_x)}, \quad N^2 \equiv \frac{\varepsilon_0 E_0^2 R_0}{4\alpha}. \qquad (3)$$

Admissible equilibrium ellipsoidal geometries must satisfy the conditions $\partial W/\partial\beta = 0$ and $\partial^2 W/\partial\beta^2 > 0$. A numerical minimization of the energy gives the logarithm of the ratio of the ellipsoid's semiaxes at points of W minimum as a function of the dimensionless field strength squared (more precisely, the product $E_0^2 R_0$) as graphically shown in Fig. 1. The stable ellipsoidal geometries corresponding to min$W$ are represented by branches *1* and *2* for $\varepsilon = 81$ and branches *3* and *4* for $\varepsilon = 40$. It is clear that there exist families of unstable equilibrium ellipsoids, for example, in the interval $1.83 < \beta < 12.8$ for $\varepsilon = 81$. Hard transitions (indicated by arrows) may occur between the two stable branches of equilibrium ellipsoidal geometries. They are characterized by hysteresis: transition to more prolate shapes

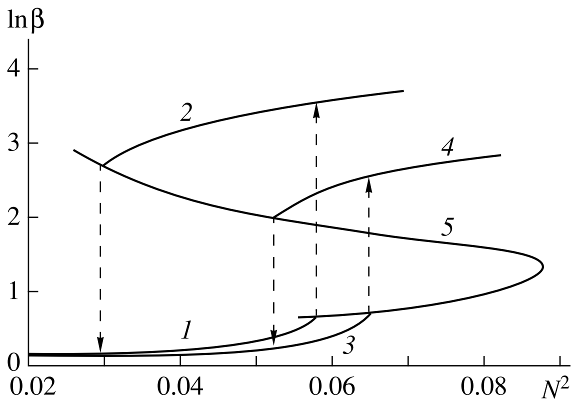


Fig. 1. Logarithm of the ratio of stable-ellipsoid semiaxes versus the dimensionless external-field strength squared for $\varepsilon = 81$ (*1*, *2*) and $\varepsilon = 40$ (*3*, *4*); *5* – envelope of the domain of unstable geometries.

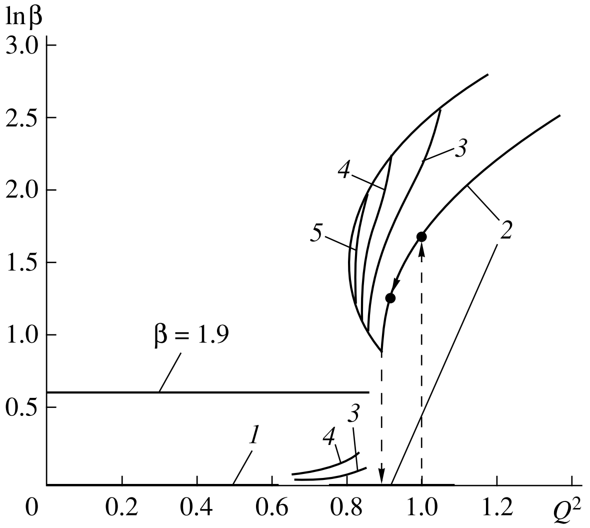


Fig. 2. Logarithm of the ratio of stable-ellipsoid semiaxes of conductive droplets versus the dimensionless droplet charge squared for $N^2 = 0$ (*1*, *2*), 0.002 (*3*), 0.003 (*4*), and 0.0036 (*5*).

takes place at greater values of N^2 than does the reverse transition. The domain of parameters where the ellipsoids are in unstable equilibrium depends on the droplet-to-ambient permittivity ratio as shown by curve *5* in Fig. 1. The extreme right point of this branch corresponds to $\varepsilon \simeq 20$. When $\varepsilon < 20$, any prolate liquid ellipsoid of revolution placed in an external electric field can reach a corresponding minimum of potential energy and is stable in this sense [6].

Consider now the case of a charged conductive droplet in an external electric field. Again, the droplet is assumed to be a prolate ellipsoid of revolution. In this case, the electrical component of the droplet's potential energy is the sum of its intrinsic electric energy and the energy of a polarized body placed in an external field (the energy of interaction between the droplet and the external field increases the droplet's kinetic energy):

$$W_e = \frac{q^2}{2C} - \frac{E_0 P_x}{2}. \tag{4}$$

Here, the second term is given by (1) with $\varepsilon \longrightarrow \infty$, which corresponds to a conductive droplet. Using the expression for the capacity of a prolate ellipsoid of revolution [5], one obtains the dimensionless total potential energy of a droplet:

$$W = S_x + \frac{4Q^2}{C_x} - \frac{4N^2}{3n_x}, \quad Q^2 \equiv \frac{kq^2}{16\pi\alpha R_0^3},$$
$$\frac{1}{C_x} = \frac{\beta^{1/3}}{\sqrt{\beta^2 - 1}} \ln\left(\frac{\beta + \sqrt{\beta^2 - 1}}{\beta - \sqrt{\beta^2 - 1}}\right), \quad k \equiv 1/4\pi\varepsilon_0. \tag{5}$$

Figure 2 shows the logarithm of the ratio of semiaxes of a stable equilibrium ellipsoid as a function of the dimensionless droplet charge Q^2 squared for various values of the dimensionless field parameter N, obtained by minimizing W. The case of $N = 0$ corresponds to a charged droplet in zero field. In this case, one of the equilibria is a sphere ($\beta = 1$), that becomes unstable at $Q^2 = 1$ [2]. The reverse transition from a prolate ellipsoid to a sphere takes place at $Q^2 \approx 0.88$. It is clear that the upper branch of equilibrium ellipsoids drastically reduces with increasing field strength. For example, when $N^2 = 0.003$, this branch corresponds to the interval $3 \lesssim \beta \lesssim 10$. At $N_*^2 \approx 0.0045$, it vanishes completely. The corresponding limit values are $Q_*^2 \approx 0.80$ and $\beta_* \approx 4.4$. The line $\beta = 1.9$ corresponds to limit equilibrium geometries of neutral conductive droplets in external electric fields [3].

Finally, note that higher order unstable modes can develop against a background ellipsoidal mode in real experimental conditions [1], but their instability can be damped for high-viscosity liquids or small droplets. In particular, experiments on hard transitions and hysteresis in magnetic fields were conducted with very small droplets of magnetic liquids ($R_0 = 15$ μm) [4].

REFERENCES

1. A. I. Grigor'ev and S. O. Shiryaeva, Izv. Akad. Nauk, Mekh. Zhidk. Gaza, No. 3, 3 (1994).
2. Lord Rayleigh, Philos. Mag. **14**, 184 (1882).
3. G. J. Taylor, Proc. R. Soc. London, Ser. A **280**, 383 (1964).
4. É. Ya. Blum, M. M. Maĭorov, and A. O. Tsebers, *Magnetic Fluids* (Zinatne, Riga, 1989).
5. L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Nauka, Moscow, 1982).
6. K. J. Cheng and J. B. Chaddock, Phys. Lett. A **106**, 51 (1984).

Translated by A. S. Betev

Technical Physics, Vol. 45, No. 4, 2000, pp. 507–508. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 130–131.

BRIEF COMMUNICATIONS

On the Distribution of a Current in Hall Medium with Metallical Inclusions

V. E. Arkhincheev

Burjat Research Center, Siberian Division, Russian Academy of Sciences, Ulan-Ude, 670047 Russia

Received August 18, 1998; in final form, October 12, 1999

Abstract—Hall ($\sigma_{xx} = 0$, σ_{xy} = const, where σ_{ik} is the conductivity tensor) medium with metallic inclusions is considered. The current (fields) distribution in such medium is established. It is shown that the electric field vanishes in the inclusion of a metallic phase and consequently the electric current streamlines miss these inclusions. The effective characteristics of the medium are calculated. *© 2000 MAIK "Nauka/Interperiodica".*

In the present paper, the limiting case of a two-dimensional conducting medium, with inclusions of another conductive phase inserted in a perpendicular magnetic field, is considered. This medium with the external magnetic field is described by means of a conductivity tensor with only one nonzero Hall component ($\sigma_{xx} = 0$, σ_{xy} = const, where σ_{ik} is the conductivity tensor), and an inclusion is approximated by spherical metallic drops of radius R. The current and field distributions in such a medium are determined. It is established that the electric field vanishes in metallic inclusions. This means that the current does not penetrate into metallic inclusions and flows along the phase boundary. Therefore, the presence of impregnation of a metallic phase in a Hall host medium does not affect the current; thus the effective characteristics of a nonuniform medium with metallic inclusions appear to be precisely equal to corresponding values of the Hall phase:

$$\sigma_{xx}^{e} = 0, \quad \sigma_{xy}^{e} = \sigma_{xy}^{(1)}. \tag{1}$$

For the solution of a problem, we take advantage of its two dimensional character and proceed to the plane of a complex variable $z = x + iy$. Let us introduce the generalized current and electric field as complex analytical functions:

$$j(z) = j_x(z) - i j_y(z), \quad E(z) = E_x(z) - iE_y(z), \tag{2}$$

which are interrelated by Ohm's law as [1]

$$j(z) = \frac{\sigma}{1 - i\beta}E(z). \tag{3}$$

Here, $\beta = \tau eH/mc$ is the Hall factor. For a purely Hall phase, one obtains:

$$j(z) = i\frac{\sigma}{\beta}E(z).$$

Then it is possible to ascribe an extended sense to standard boundary conditions:

$$j_{1n} = j_{2n}, \quad E_{1t} = E_{2t}. \tag{4}$$

In our case one gets:

$$\begin{gathered} \mathrm{Re}[tj(t)]_1 = \mathrm{Re}[tj(t)]_2, \\ \mathrm{Im}\left[ti\frac{\sigma}{\beta}j(t)\right]_1 = \mathrm{Im}\left[t\frac{1 + i\beta}{\sigma}j(t)\right]_2. \end{gathered} \tag{5}$$

Here, $t = \cos\theta + i\sin\theta$ is a variable coordinate of the boundary contour, index 1 corresponds to values in the Hall medium, and index 2 denotes the quantities in metallic inclusions. It follows from the boundary conditions (3) at the interface between the Hall phase and metal that the electric field inside inclusions should vanish,

$$E_2(z) = 0, \quad |z| < R. \tag{6}$$

This result can qualitatively be explained as follows. The flowing of a current through any medium should be accompanied by minimum dissipation of Joule's heat. The Hall phase is a nondissipative one (the current is perpendicular to the applied field); therefore, the flowing only through the Hall phase is more preferential to a current. However, it is impossible to suspect any other reason for which the current would not penetrate into inclusions of a metallic phase arranged at random, except for the vanishing of a field in these inclusions.

The same result follows from the full solution of a problem concerning the current distribution in a conducting medium which includes sites of another conductivity in a perpendicular magnetic field. Thus, the pure Hall phase is reached through a passage to the limit

$$\sigma_1 \longrightarrow \infty, \quad \beta_1 \longrightarrow \infty, \quad \frac{\sigma_1}{\beta_1} = \text{const}. \tag{7}$$

It follows from analytic properties of functions $j_1(z)$ and $j_2(z)$, that in the domains of their definition, they can be expanded into convergent power series as follows:

$$j_1(z) = A_0 + \frac{A_1}{z} + \frac{A_2}{z^2} + \ldots, \quad |z| > R,$$
$$j_2(z) = B_0 + B_1 z + B_2 z^2 + \ldots, \quad |z| < R. \tag{8}$$

Further, it is essential to take into account that at infinity, the field and current induced by it are uniform ones; i.e., $j_1(\infty) = \langle J \rangle$, and the current inside an inclusion should not have poles and also boundary conditions (5), as it is considered in more detailed form for example, as in [2]. A particularly interesting case for us is (6) which, after employing indispensable calculations, we obtain the following results:

$$j_1(z) = \langle J \rangle - \langle \bar{J} \rangle \frac{R^2}{z^2}, \quad j_2(z) = 0. \tag{9}$$

Here, $\langle J \rangle$ is the current in the medium and $\langle \bar{J} \rangle$ is a complex conjugate current. Let us discuss the obtained formulas (9). The problem concerning current or field distributions in a conducting medium with inclusions of another conductive phase is a classical one [3]. Nevertheless, in the limit considered by us, a slightly paradoxical result is obtained: the current cannot percolate through metallic inclusions because of the vanishing of an electric field in them. Repulsion of an electric field from a metallic phase and bending of a Hall current around metallic inclusions along the interface result in a negative electric dipole moment of a metallic circle. An analogy between an inclusion in electric field and superconducting circle in an external magnetic field is present here. The magnetic field is popped out from a superconducting phase, and the electric dipole moment of a circle in a magnetic field also becomes negative.

The obtained results are in accord with those of paper [4], where, by using another way within the framework of Dykhne's approach for two-phase randomly nonuniform media, the effective characteristics were determined and macroscopic formulas similar to (1) were obtained.

ACKNOWLEDGMENTS

The author acknowledges Yu. P. Emets for the possibility to be introduced to the book [2].

REFERENCES

1. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **81**, 665 (1981).
2. Yu. P. Emets, *Electric Characteristics of Composite Materials With Regular Structure* (Naukova Dumka, Kiev, 1986); Zh. Tekh. Fiz. **44**, 916 (1974).
3. L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Nauka, Moscow, 1987; Pergamon Press, Oxford, 1984).
4. V. E. Arkhincheev and É. G. Batyev, Solid State Commun. **12**, 1059 (1989).

Translated by V. A. Voĭtenko

Technical Physics, Vol. 45, No. 4, 2000, pp. 509–511. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 132–134.
Original Russian Text Copyright © 2000 by Avetyan, Arakelyan, Ancharakyan, Patvakanyan.

BRIEF COMMUNICATIONS

X-Ray-to-Optic Image Converter for Visualization of Crystal Structure Defects

K. T. Avetyan, M. M. Arakelyan, S. A. Ancharakyan, and A. G. Patvakanyan
Erevan State University, Erevan, 375049 Armenia

Received August 19, 1998

Abstract—The designed luminescent screen for an X-ray-to-optical image converter showed a high spatial resolution and a high efficiency in X-ray absorption and conversion. © *2000 MAIK “Nauka/Interperiodica”.*

INTRODUCTION

In the direct X-ray diffraction topographic techniques, an image of crystal structure imperfections is converted to a video signal of a camera tube (a vidicon) with an X-ray sensitive photoconductive layer. However, for various reasons, these vidicons show a low sensitivity and can do nothing but at very high X-ray intensities [1, 2]. The vidicons for optical spectrum display a high sensitivity. To apply them to the topographic investigations, an intermediate element, namely, a screen converting an X-ray image of defects to a visible one, is needed. This screen should of high efficiency in X-ray absorption and conversion and be of a high spatial resolution for image details, but these requirements are competitive factors. Calculations reveal that, with due regard to Rayleigh region scattering (in the case of an uniform phosphor layer transparent to the visible light), the closest separation of two luminous points, whose images can be resolved, yields $\Lambda \approx 2z_0$, where z_0 is a phosphor layer thickness. When $\Lambda \approx 10$ μm is required, the layer thickness $z_0 \approx 5$ μm provides only limited absorption of X-rays. An actual layer is still thinner. For example, in accordance with the experimental data from [3], $\Lambda \approx (5–6)z_0$.

FORMULATION OF THE PROBLEM

We devised a converting screen providing the required spatial resolution at a much thicker layer. In this case, the high efficiency in the conversion and, as a result, the high luminosity are ensured. Both these competitive factors could be raised together through combining a compact uniform phosphor layer with the material of a fiber-optics disk (FOD). At this combination, a luminous point image at the exit of the FOD is formed by emission within a minor solid angle given by

$$\Omega = \frac{\pi(n_f^2 - n_e^2)}{n_{ph}^2},$$

where n_f and n_e are the refractive indexes of a light-guiding fiber and an envelop, respectively; and n_{ph} is that of a phosphor layer.

The radiation that occurred beyond this solid angle does not penetrate into the FOD because of the TIR condition breaking down at the light-guiding-fiber–envelope boundary. It is this condition that defines the critical angle of incidence (the Brewsterian angle) on the luminophor-FOD boundary as

$$\sin\gamma = \frac{\sqrt{n_f^2 - n_e^2}}{n_{ph}}$$

and the minimum separation resolved given by

$$\Lambda = \frac{z_0}{n_{ph}}\sqrt{n_f^2 - n_e^2}.$$

The light flux entering one of the light-guiding fibers experiences a multiple reflection and shows a uniform distribution within a fiber cross section as it leaves the fiber; therefore, the integrated flux at the exit will be of a discreet character. This is an important point, since the minimum separation resolved is of the order of the fiber diameter. There is one more special feature of the phosphor-layer-FOD combination. It is easily seen that, at any separation

$$z \le z_0 = \frac{\Lambda n_{ph}}{\sqrt{n_f^2 - n_e^2}}$$

between a luminous point and the FOD surface, the distribution of a light flux as it leaves a fiber is kept constant, because the number of fibers involved in the formation of an image and the flux inside of each fiber are the same. The FOD resolving power is determined by the fiber diameter d and equals one-half of the number of fibers per unit length [4]; thus, the minimum separation resolved is $\Lambda \approx 2d$. Thus, the phosphor-layer thickness providing the maximum resolution is

$$z_0 = \frac{2dn_{ph}}{\sqrt{n_f^2 - n_e^2}}.$$

The luminosity of a phosphor-layer-FOD combination is governed by the light flux penetrating into the FOD and depends on phosphor parameters, namely, the X-ray and optical absorption coefficients μ_X and μ_o, the conversion factor η, the layer thickness z, and on the FOD parameters n_f and n_e.

To select the parameters ensuring the peak luminosity at a given resolving power, the analytical expression $\Phi(\eta, \mu_X, \mu_o, z, n_f, n_e, n_{ph})$ for the light flux as it leaves an FOD is required.

Under the normal incidence of a narrow X-ray beam on the layer with a thickness z_0, the radiation absorbed within a thickness dz at a depth $z_0 - z$ (at the layer exit, $z = 0$) is represented as $dl(z) = I_0\mu_X\exp[-(z_0 - z)\mu_X]dz$, where I_0 is an incident X-ray intensity. A visible radiation flux, appearing within this layer per solid angle Ω, can be represented at the entry into the layer as

$$d\Phi = I_0\mu\eta\frac{\Omega}{4\pi}\exp[(-z_0 - z)\mu_X]\exp(-\mu_0 z)dz.$$

The total light flux penetrating into an FOD is

$$\Phi = I_0\eta\frac{n_f^2 - n_e^2}{4n_{ph}}\frac{\mu_X}{\mu_X - \mu_0}[\exp(-\mu_0 z) - \exp(-\mu_X z)]$$

(thereafter, z is the layer thickness).

The peak value of the function

$$F(\mu_X, \mu_0, z) = \frac{\mu_X}{\mu_X - \mu_0}[\exp(-\mu_0 z) - \exp(-\mu_X z)]$$

is a unity under the total absorption of X-ray radiation ($\mu_X z \gg 1$) and zero optical radiation absorption ($\mu_0 = 0$).

As discussed below, there is no point in using transparent layers. Let us deal with layers of $\mu_0 \neq 0$. Denoting $\mu_X + \mu_0 = 2\mu$, $\mu_X - \mu_0 = 2\Delta\mu$, we represent the function $F(\mu_X, \mu_0, z)$ by

$$F(\mu, \Delta\mu, z) = \mu z\exp(-\mu z)\frac{\sinh\Delta\mu z}{\Delta\mu z}\left(1 + \frac{\Delta\mu}{\mu}\right).$$

The factor $\mu z\exp(-\mu z)$ peaks at $\mu z = 1$. At $\Delta\mu z \ll 1$; the factor

$$\frac{\sinh\Delta\mu z}{\Delta\mu z} = 1$$

and is somewhat over unity at $\Delta\mu z \approx 1$ (note that $\Delta\mu \leq \mu$). Therefore, at $\mu z \approx 1$, the value of $F(\mu, \Delta, \mu, z)$ is determined by the factor $(1 + \Delta\mu/\mu)$ ranging between 1 and 2.

EXPERIMENT AND RESULTS

To peak up the parameters η, n_{ph}, μ_X, μ_0, and z for a phosphor layer (and those of n_f, n_e, and d for a FOD), one starts from the requirement for the maximum luminosity at a given resolving power. We have already worked out that the resolving power is independent of the layer thickness given by the formula

$$z \leq \frac{2dn_f}{\sqrt{n_f^2 - n_e^2}}.$$

For an FOD aperture $A = \sqrt{n_f^2 - n_e^2} = 0.54$, $d = 8$ μm, $n_{ph} = 2.2$, and $z \approx 60$ μm. At such a thickness, for almost all compact transparent luminophors, $\mu_X z < 1$ is true; and for MoK_α, radiation with $\lambda = 0.71$ Å. Besides, it is significant that the conversion factor η of the thick luminophors is rather low (1–2%), but is considerably in excess of this value for activated powdery phosphors with 3–5-μm grains (20–22%) [5]. We have adopted a powdery terbium-activated Gd_2O_2S-based phosphor. All the conventional powdery techniques for layer deposition suffer from a substantial disadvantage. In the process of layer deposition, a great many pores are bound to occur, and their size is several times more than that of grains. These pores are as blind regions of a screen; its structure is superimposed on an image. At a low thickness, it is especially evident. Besides, at a friable layer, one is forced to increase its thickness to achieve the needed X-ray absorption.

We have devised a procedure for the deposition of reasonably uniform and compact phosphor layers [6]. The point of the procedure is as follows. The powdery phosphor and a liquid polymerizing binder are mixed together in a volume ratio of 1 : 10. The mixture is deposited on the carefully cleaned FOD surface and subjected to vibration at a frequency of 50–100 Hz. In so doing, it shows an essential rise in its density and the stiff paste spreads over the FOD surface. The excess liquid emerging from the paste onto its surface is a sign of the mixture thickening. Once the layer solidifies, it is mechanically polished to the thickness when $\mu_X z = 1$. The measurements reveal that the thickness z, of a layer produced from powdery Gd_2O_2S (for which $\mu_X z = 1$ is true), is about 50 μm (for MoK_α radiation). If the refractive index of a binder equals that of phosphor grains (for Gd_2O_2S, $n_{ph} = 2.2$), the layer would be of the peak transparency. We failed to pick up such a binder. OP-1-0.28 epoxy resin was considered to be a suitable one. Moreover, the assortment of FODs is poor. Despite this, when coupled with a LI-702 supervidicon, the transducer screen meets the requirements on visualization of X-ray diffraction topographic images of crystal structure imperfections.

The resolving power was measured with the use of a fan of tungsten filaments 10 μm in diameter. At layers 50–60 μm thick, the resolving power is 30–35 line pairs per 1 mm which agrees with the rated one. As for the layer density, $\rho_1 \approx (0.6–0.7)\rho_2$ (for Gd_2O_2S, $\rho_2 = 7.2$ g/cm^3).

In addition, the deposition of a phosphor on an FOD is well suited, as an optical image at emergence is obtainable almost without a loss in intensity and can be transferred in its true shape onto the supervidicon photocathode due to the immediate optical contact with the entrance window of the supervidicon (the window also involves optical fibers). At any other technique of the image transfer onto the photocathode, severe losses and distortions are bound to occur.

The experiments with screens of aperture $A = 0.54$, deposited on the FODs, have been performed. Since the luminosity is proportional to A^2 and the resolving power is proportional to $1/A$, an increase in the luminosity by $1/(0.54)^2$ times and a decrease in the resolving power by 0.54 times were expected after changing an FOD of $A = 0.54$ to that of $A = 1$. However, after this replacement, the layer thickness drops below that of the best resolving power, and the condition $\mu z = 1$ breaks down. Thus, the FOD of aperture $A = 0.54$ and the Gd_2O_2S-based phosphor layer $z \sim 50–60$ μm are best suited for the initial conditions of $\lambda = 0.71$ Å and resolving the power of 30 line pairs per 1 mm.

REFERENCES

1. J. Chikawa and I. Fujimoto, Appl. Phys. Lett. **13**, 387 (1968).
2. A. N. Chester and F. B. Koch, Adv. X Ray Anal. **12**, 165 (1969).
3. I. I. Lobanova, M. B. Provotorov, and S. S. Galaktionov, USSR Inventor’s Certificate No. 1222129 (1986).
4. V. B. Veĭnberg and D. K. Sattarov, *Optics of Light Guides* (Mashinostroenie, Leningrad, 1977).
5. A. M. Gurvich, *X-ray Phosphors and Luminescent Screens* (Atomizdat, Moscow, 1976).
6. K. T. Avetyan, T. K. Melkonyan, A. P. Dzhotyan, *et al.*, USSR Inventor’s Certificate No. 1697548 (1991).

Translated by B. A. Malyukov

Technical Physics, Vol. 45, No. 4, 2000, pp. 512–514. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 135–137.
Original Russian Text Copyright © 2000 by Chizhov.

BRIEF COMMUNICATIONS

Extracting Model Parameters from Experimental Data Using the Minimum Information Principle

A. V. Chizhov

Ioffe Physicotechnical Institute, Russian Academy of Sciences,
ul. Politekhnicheskaya 26, St. Petersburg, 194021 Russia

Received November 12, 1998

Abstract—A method is described for extracting model parameters of a physical process from statistical experimental data based on the minimum information principle. Potentialities of the method are studied by examples of regular and chaotic processes simulated by the transport and diffusion equation and by the logistic mapping. *© 2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

When fitting mathematical models to experimental data, it is sometimes necessary to solve the inverse problem of extracting unknown coefficients of the model equations from the measured data. The experimental data are statistical and one should construct the most likely model of a phenomenon. Problems of this kind include spectrometry and geological exploration problems, finding transport coefficients of hydrodynamics equations, turbulent viscosity retrieval, and fitting models of biological systems. Examples of such inverse problems and methods for solving them can be found in review [1]. This paper demonstrates the principle behind the reconstruction of unknown parameters of the partial differential equations that describe transport and diffusion processes and of the logistic mapping. The reconstruction method relies on the minimum information principle detailed by Haken [2], in particular, as applied to the problem of model finding. In the case of our concern, the desired model is an equation that describes the observed process.

EXPERIMENTAL DATA

Consider a nonstationary process u under the assumption that it is described by the equation

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} - \eta\frac{\partial^2 u}{\partial x^2} = s \tag{1}$$

with unknown coefficients a, η, and s.

Also assume, that one can experimentally measure u at three closely located points of the region Ω. Namely, each measurement marked by index n suggests that $u_0^n = u(t_n, x) + \xi_1^n$, $u_-^n = u(t_n, x - \Delta x) + \xi_2^n$, and $u_+^n = u(t_n, x + \Delta x) + \xi_3^n$ are determined at the instant t_n and, additionally, $\tilde{u}^n = u(t_n + \tau, x) + \xi_4^n$ at the central point at the instant $t_n + \tau$, where ξ_i^n $(i = 1, 4)$ are the random errors, $|\xi_i^n| < \varepsilon$, and Δx and τ are given. Points on the (x, t) plane, where the measurements are taken, comprise a four-point pattern as shown in Fig.1.

Consider auxiliary quantities

$$q_1^n = \tilde{u}^n - u_0^n, \quad q_2^n = u_-^n - u_0^n, \quad q_3^n = u_+^n - u_0^n. \tag{2}$$

When processing data of a series of N measurements, statistical moments are calculated which comprise the vector $\bar{\mathbf{q}}$ and matrix Q

$$\bar{\mathbf{q}} = \begin{pmatrix} \overline{q_1} \\ \overline{q_2} \\ \overline{q_3} \end{pmatrix}, \quad Q = \begin{pmatrix} \overline{q_1^2} & \overline{q_1q_2} & \overline{q_1q_3} \\ \overline{q_1q_2} & \overline{q_2^2} & \overline{q_2q_3} \\ \overline{q_1q_3} & \overline{q_2q_3} & \overline{q_3^2} \end{pmatrix}, \tag{3}$$

where the overbar means averaging over the series of experiments:

$$\overline{f_j} = \frac{1}{N}\sum_{n=1}^{N} f_j^n.$$

These are the source data for extracting coefficients a, η, and s.

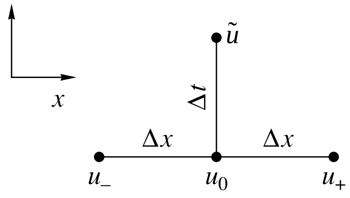


Fig. 1.

PROBABILITY DISTRIBUTION FUNCTION

The principal stage in solving the problem consists in reconstructing the simultaneous probability distribution function $P(q_1, q_2, q_3)$ such that

$$\iiint f_j(q_1, q_2, q_3)P(q_1, q_2, q_3)dq_1dq_2dq_3 = \overline{f_j}; \quad j = 1, 9, \qquad (4)$$

where f_j are the quantities whose average values $\overline{f_j}$ constitute vector $\bar{\mathbf{q}}$ and matrix Q.

A set of functions $P(\mathbf{q})$ satisfy condition (4). In order to introduce the minimum amount of information in addition to that present in relationships (4), of all these functions, we select the one that contains the minimum information on $\mathbf{q}$, i.e., minimizes the functional

$$\int P(\mathbf{q})\ln P(\mathbf{q})d\mathbf{q} = \min. \qquad (5)$$

It can be shown in [1] that this probability distribution function has the form

$$P(\mathbf{q}) = c\exp\left\{-\frac{1}{2}(\mathbf{q}-\bar{\mathbf{q}})^T G(\mathbf{q}-\bar{\mathbf{q}})\right\} = c\exp\{V(\mathbf{q})\}, \qquad (6)$$

where c is a constant factor and $G = (Q - \bar{\mathbf{q}}\bar{\mathbf{q}}^T)^{-1}$.

3. RECONSTRUCTION OF THE NUMERICAL SCHEME

The next stage consists in constructing the numerical scheme that describes process $u(t, x)$ with the given pattern (Fig. 1). In order to do this, we should calculate the most probably value of the quantity q_1, which is responsible for the change in field $u(t, x)$ as a result of going over to the new time section with q_2 and q_3 given on the old time section; i.e., we must require that $\partial P/\partial q_1 = 0$. Substituting expression (6) yields

$$\frac{\partial P}{\partial q_1} = (G_{11}(q_1-\overline{q_1}) + G_{12}(q_2-\overline{q_2}) + G_{13}(q_3-\overline{q_3}))c\exp\{V\} = 0,$$

i.e., $q_1 = a_1q_2 + a_2q_3 + a_3$, where $a_1 = -G_{12}/G_{11}$, $a_2 = -G_{13}/G_{11}$, and $a_3 = \overline{q_1} - a_1\overline{q_2} - a_2\overline{q_3}$, or

$$\tilde{u} - u_0 = a_1(u_- - u_0) + a_2(u_+ - u_0) + a_3. \qquad (7)$$

Using the Taylor series expansion for function u about the origin of the pattern, we obtain

$$\frac{\partial u}{\partial t} + (a_1 - a_2)\frac{\Delta x}{\tau}\frac{\partial u}{\partial x} = \frac{\Delta x^2}{2\tau}(a_1 + a_2)\frac{\partial^2 u}{\partial x^2} + \frac{a_3}{\tau} + O(\Delta x^3, \tau). \qquad (8)$$

Since equation (1) is the zero differential approximation for scheme (7), a comparison of (8) and (1) gives

$$a = \frac{\Delta x}{\tau}(a_1 - a_2), \quad \eta = \frac{\Delta x^2}{2\tau}(a_1 + a_2), \quad s = \frac{a_3}{\tau}.$$

Note that one can arrive at the same result following the procedure used in [2, Chap. 9], which determines the conditional probability distribution $P(q_1|q_2, q_3)$ under the assumption that the process is Markovian, and then calculates the drift coefficients for the Fokker–Plank equation that corresponds to this conditional probability. These coefficients define the Langevin equation and, consequently, desired parameters a, η, and s.

4. EXAMPLES

1. Parameters a, η, and s were extracted for the process described by equation (1) with exact values $a = 1.3$, $\eta = 0.4$, and $s = 2$; the zero initial condition in region Ω; and boundary conditions

$$u|_{x=0} = \sin 5t + t, \quad u|_{x=1} = 1.$$

The pattern with $\Delta x = 0.033$ and $\tau = 0.001$ was used to take $n = 400$ measurements contaminated by random Gaussian errors with an amplitude no higher than 0.001 (for reference, $\overline{q_1} = 0.0039$) and standard deviation $\sigma = \varepsilon/\sqrt{3}$. The extracted coefficients were found to be $a_r = 1.2866$, $\eta_r = 0.3996$, and $s_r = 2.031$.

Figure 2 plots the relative extraction error as a function of the noise amplitude $\varepsilon(a)$, measurement pattern scale $\Delta x(b)$ (the time parameter was controlled so that the pattern's Courant number $a\tau/\Delta x + \eta\tau/\Delta x^2 = 0.8$ was kept constant), and number of measurements. The remaining parameters are the same as in Section 1. The accuracy of extracting the coefficient a_r improves as the measurement error ε decreases and the total number of measurements N increases. The low accuracy at a small spatial pattern scale is attributed to the fact that, in the corresponding small time increment τ, the change in u becomes smaller than the noise amplitude.

2. Consider a steady-state process described by the essentially nonlinear logistic mapping

$$x_{i+1} = rx_i(1 - x_i), \quad 0 < r \le 4, \qquad (9)$$

whose bifurcation pattern is shown in Fig. 3a. The mapping law was assumed to be $x_{i+1} = a_1x_i + a_2x_i^2 + a_3$ with unknown coefficients a_1, a_2, and a_3. To extract these coefficients, information minimum principle (5) was employed as this was done in the case of equation (1) (here, we measured $q_1^n = x_{i+1} + \xi_1^n$, and $q_2^n = x_i + \xi_2^n$, $q_3^n = x_i^2 + \xi_3^n$ with random errors $\xi_j^n : |\xi_j^n| < \varepsilon$).

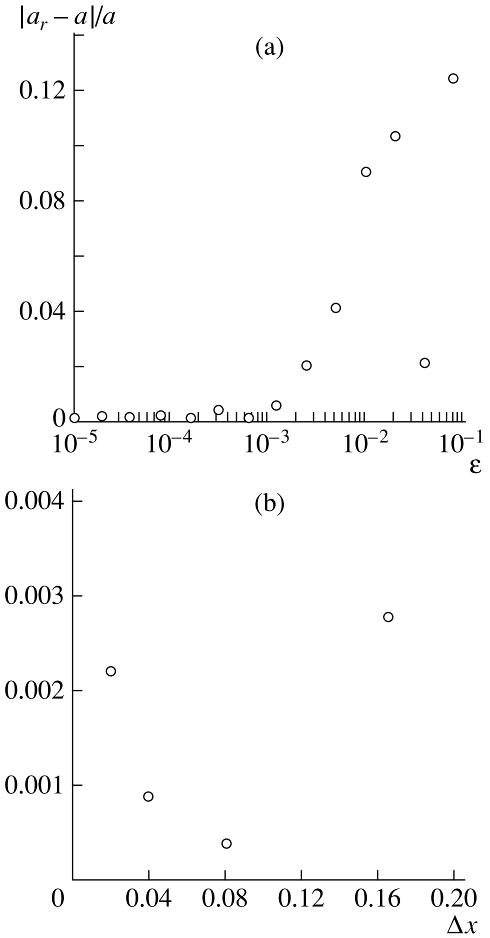


Fig. 2.

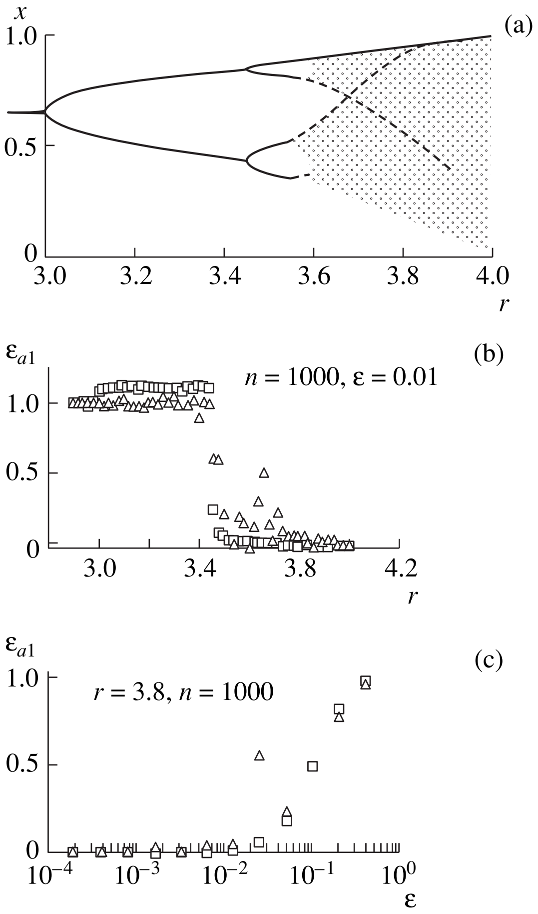


Fig. 3.

Squares in Fig. 3b plot the relative error of extracting the first coefficient $\varepsilon_r = |a_1 - r|/r$ versus parameter r of mapping (9) at the noise amplitude $\varepsilon = 0.01$. It can be seen that the coefficients are successfully extracted when the steady-state solution is sufficiently nontrivial; i.e., when $r \geq 3.5$, in particular, when the solution is described by a strange attractor. Figure 3c illustrates the effect of noise. For the sake of comparison, triangles plot results obtained by the least squares method proposed for solving problems of this kind in [3]. When the noise amplitude is sufficiently high, the minimum information principle reveals its advantages. In particular, it allows one to extract the coefficients when parameter r ranges approximately from 3.5 to 3.8.

CONCLUSION

The principal possibility of extracting unknown equation parameters from statistical experimental data using the minimum information principle is thus demonstrated. The extraction was found to be successful for a process without nonlinear effects (1) and for essentially nonlinear processes (9), including the deterministic chaos. Advantages of the proposed method over the least squares methods are revealed. It also has a better physical justification [2]. Therefore, this method can be used to fit general mathematical models to a studied phenomenon based on experimental data.

REFERENCES

1. V. F. Turchin, V. P. Kozlov, and M. S. Malkevich, Usp. Fiz. Nauk **102**, 345 (1970) [Sov. Phys. Usp. **13**, 681 (1970)].
2. H. Haken, *Information and Self-Organization* (Springer, Berlin, 1988; Mir, Moscow, 1991).
3. G. L. Baker, J. P. Gollub, and J. A. Blackburn, Chaos **6**, 528 (1996).

Translated by A. D. Khzmalyan

Technical Physics, Vol. 45, No. 4, 2000, pp. 515–518. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 138–140.
Original Russian Text Copyright © 2000 by Gavrilov.

BRIEF COMMUNICATIONS

Transition through the Critical Velocity for a Moving Load in an Elastic Waveguide

S. N. Gavrilov

Institute of Problems in Machine Science, Russian Academy of Sciences, St. Petersburg, 199178 Russia

Received December 7, 1998

Abstract—Phenomena that accompany the transonic transition experienced by a load moving along a string on a deformed base are studied. A solution in the form convenient for a qualitative analysis of the wave processes is proposed. The cases of the acceleration and deceleration of the load are considered. © *2000 MAIK “Nauka/Interperiodica”.*

INTRODUCTION

This paper describes the phenomena that accompany the transonic transition experienced by a load moving along a string on a viscoelastic base (Fig. 1). The moving load is modeled by the Dirac delta-function. This problem had already been studied earlier [1], and the asymptotics of the solution for the acceleration tending to zero had been derived. Later [2], it was found that, at the instant of the transition through the critical velocity, a clearly defined wavefront is formed, and this wavefront moves along the string with the sound velocity; for the solution in the vicinity of the wavefront, the asymptotic behavior at long time intervals had been determined. This paper presents a detailed analysis of the wave pattern formed in the string at the instant of the transition through the critical velocity. (Some of the results obtained below were mentioned in [1], yet without any appropriate theoretical justification.) A solution is obtained in the form that is convenient for analyzing the singularities, which occur in the expression for the slope of the string at the instant of the transonic transition. It is shown that, in the framework of the model under study with an accelerated concentrated load, the latter can exceed the sound velocity only under an infinite driving force, even in the presence of dissipation in the deformed base. The cases of accelerated and decelerated loads are studied, and a distributed load is considered.

FORMULATION OF THE PROBLEM

The equation of the string motion and the corresponding initial conditions have the form

$$u'' - \frac{1}{c^2}\ddot{u} - 2\gamma\dot{u} - ku = \chi(t)\delta(x - l(t)), \tag{1}$$

$$u(x, t)|_{t=0} = 0, \quad \dot{u}(x, t)|_{t=0} = 0. \tag{2}$$

Here, $u(x, t)$ is the displacement of the point x of the string at the instant t; c is the sound velocity; k and γ are the coefficients of elasticity and viscosity of the deformed base, respectively; $l(t)$ is the load coordinate; and $\chi(t)$ is the load intensity. We assume that $\chi(t)$ is a sufficiently smooth function. We consider the Cauchy problem: the displacement is $u \equiv 0$ if $|x|$ is sufficiently large for a given $t < \infty$.

GENERAL REPRESENTATION FOR THE SOLUTION

It is easy to derive the general formula for the solution to problem (1), (2). Apply the Fourier transform with respect to the x coordinate to equation (1); express the solution to the resulting differential equation through the Duhamel integral; and, then, apply the inverse Fourier transform. By changing the order of integration in the resulting expression and calculating the inner integral, we obtain

$$\begin{aligned} u = {} & -\frac{c}{2}\int_0^t \chi(\tau)e^{-c^2\gamma(t-\tau)}\theta\left(c - \frac{|x - l(\tau)|}{t-\tau}\right) \\ & \times J_0\left(\sqrt{k_*(c^2(t-\tau)^2 - (x - l(\tau))^2)}\right)d\tau, \end{aligned} \tag{3}$$

$$k_* = k - c^2\gamma^2, \tag{4}$$

where $\theta(t)$ is the Heaviside function.

We assume that the viscosity γ is sufficiently small, so that $k_* > 0$. Formula (3) is valid for arbitrary func-

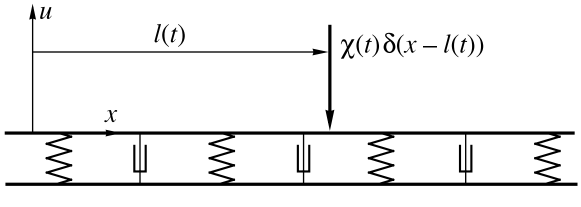


Fig. 1. String on a deformable base.

1063-7842/00/4504-0515$20.00 © 2000 MAIK “Nauka/Interperiodica”

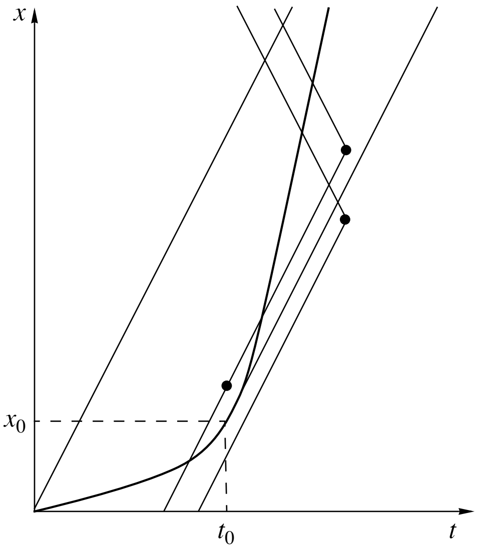


Fig. 2. Determination of the instant t'.

tions $l(t)$ and $\chi(t)$. (This formula was derived in paper [1] for the case of a motion with constant acceleration.) From formula (3), we can calculate the slope of the string u':

$$u' = \Psi_1 + \Psi_2. \tag{5}$$

For limited x and t, the function Ψ_2 is a continuous and bounded one; this function appears as a result of the differentiation of the smooth integrands in expression (3). For Ψ_1, the following representation is valid ($v = \dot{l}$):

$$\Psi_1 = \frac{1}{2}\sum_{i=1}^{N}\chi(t_i')e^{-c^2\gamma(t-t_i')} \times\left(\frac{\theta(x-l(t_i'))}{|1-v(t_i')/c|}-\frac{\theta(l(t_i')-x)}{|1+v(t_i')/c|}\right), \tag{6}$$

where the instants of time t_i' are the solutions to the equation

$$t' + \frac{|x-l(t')|}{c} = t. \tag{7}$$

The function Ψ_1 appears as a result of the differentiation of the Heaviside function in expression (3). In the absence of a deformed base, we have $u' = \Psi_1$. In the general case, representation (5)–(7) is convenient, because all singularities that occur in u' are determined by the function Ψ_1. The solution to the problem in the form (5)–(7) resembles the Lienard–Wiechert potentials in electrodynamics.

The determination of the instant t' is illustrated in Fig. 2. The thick line shows the motion of the load. To calculate the instants t' corresponding to a given point of the plane, it is necessary to draw two rays with the origin at this point and with the slopes c and $-c$. The instants t' will be the abscissas of the points of intersection of these rays with the curve representing the load motion. It is easily seen that, if the motion occurs with the velocity below c, only one point of intersection is possible. In the presence of a region of supersonic motion, the instant t' may be nonunique; in this case, either a finite number of instants t_i' will be observed, or they will form a discrete spectrum. The solutions to equation (7) can form a continuous spectrum only in the presence of a time interval within which the motion occurs with the velocity c; however, in this case, according to [3], the solution is discontinuous, and formula (5) is meaningless.

TRANSITION THROUGH THE CRITICAL VELOCITY

If a load moves along the string with the sound velocity within some interval of time, the solution to the corresponding problem will be discontinuous [3]. One can prove the converse: the existence of such an interval of time is a necessary condition for the discontinuity of the solution. This can easily be done by using representation (3) for the solution and applying the theorem on the continuity of the integral with respect to the parameter. Our main interest here is with the wave processes that occur in the string as a result of an instantaneous transition through the critical velocity. We study them numerically by using formula (3). First, we consider the case of an accelerated load. For the load motion, we use the only assumption that $\ddot{l}(t_0) = a \neq 0$ and $\dot{l} < c$ for $t < t_0$, where t_0 is the instant of the transition through the critical velocity. It turns out that, at the instant t_0, the "well" under the moving load begins to lag behind the load in the manner shown in Fig. 3 (ξ is the coordinate in the moving coordinate system fixed to the load and $t_1 - t_0$ is a short time interval). With the passage of time, this "well" is transformed into a clearly defined wavefront moving with the velocity c along the string. The asymptotics of the solution in the vicinity of this wavefront was obtained in [2].

Now, we analyze the transition through the critical velocity by using formulas (5)–(7). We consider the behavior of the string at $t = t_0$ and $x = x_0 + 0$, where $x_0 = l(t_0)$. Using Fig. 2, one can easily see that, in this case, the first term in formula (6) yields $u' \longrightarrow +\infty$. Then, we obtain $u' \longrightarrow +\infty$ at $x = x_0 + ct + 0$ for $t > t_0$, which corresponds to the value at the wavefront formed at the instant of the transonic transition. For $x = x_0 - 0$ at $t = t_0$, we obtain $u' = O(1)$. Thus, at the instant the load velocity becomes equal to the critical one, the slope of the string before the load becomes vertical, and this singularity moves along the string before the load with the velocity c; the slope of the string behind the load remains finite. The question now arises: What driving force should be applied to the load to exceed the critical velocity? Based on the model allowing for the inertia of the load [4], a formula was obtained for the longitudinal drag force $\mathcal{F}$ acting on a moving load (see also [5]):

$$\begin{aligned}\mathcal{F} &= -\frac{\rho}{2}(c^2 - v^2)(u'^2)\Big|_{\xi=-0}^{\xi=+0} \\ &= -\frac{T\chi(t)}{2}(u'|_{\xi=-0} + u'|_{\xi=+0}).\end{aligned} \tag{8}$$

Here, ρ is the line density of the string and T is the tension of the unperturbed string. From formula (8), we obtain $\mathcal{F} = -\infty$ for $t \longrightarrow t_0 - 0$. Thus, for a concentrated load to exceed the critical velocity, an infinitely large driving force is necessary. We note that this result was obtained in the presence of viscosity in the deformed base.

Now, we consider the transition through the critical velocity for a decelerated load (Fig. 4). In this case, we can prove in a similar way that $u' \longrightarrow \infty$ at $t = t_0 - 0$ and $x = x_0 - 0$, while at $x = x_0 + 0$ we have $u' = O(1)$. Correspondingly, we obtain a similar result: $\mathcal{F} \longrightarrow -\infty$, but in this case an infinite force acting from the side of the string promotes the transition through the critical velocity.

The results obtained above suggest that a consideration of a nonlinear formulation of the problem should be advantageous. The differences between the linear and nonlinear problems will be related to the presence of a coupling of the transverse and longitudinal vibrations. The phenomena caused by the transition through the critical velocity should lead to considerable changes in the string tension before the load, i.e., to a considerable deformation of the string in the longitudinal direction, which is not taken into account in the linear problem.

A DISTRIBUTED LOAD

Let us consider a distributed load. We assume that at $t = 0$ the load density is $\chi_0(x)$; for definiteness, we assume that $\chi_0(x) > 0$ for $x \in [-\alpha; 0]$ and $\chi_0(x) = 0$ for $x \notin [-\alpha; 0]$. The density of the moving load will be $\chi(x, t) = \chi_0(x - l(t))$. For u', the following representation is valid:

$$u' = \int_{-\alpha}^{0} \chi_0(\sigma)(\Psi_1(x-\sigma, t) + \Psi_2(x-\sigma, t))d\sigma. \tag{9}$$

First, we consider the transition through the critical velocity in the case of acceleration. We calculate the singular part of u' for $t = t_0$ and $x = x_0 + 0$; it is determined by the integral

$$\frac{1}{2}\int_{-\alpha}^{0} \frac{\chi_0(\sigma)d\sigma}{1 - v(t'(x_0 - \sigma, t_0))/c}. \tag{10}$$

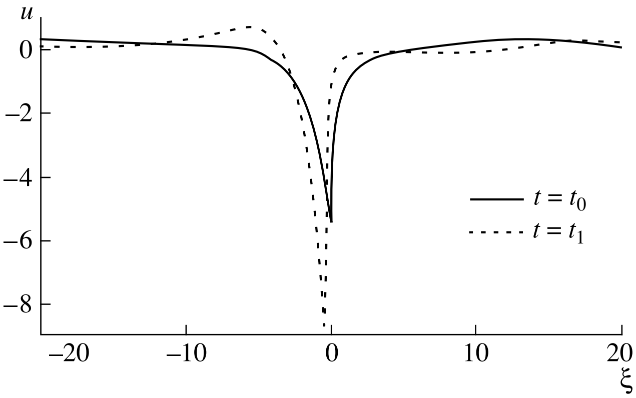


Fig. 3. "Well" under the load.

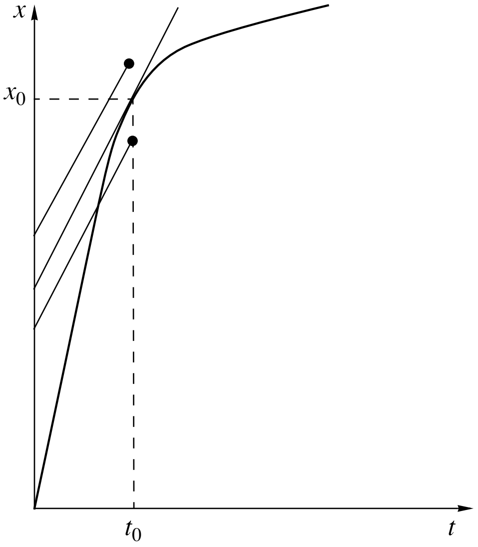


Fig. 4. Transition through the critical velocity in the case of the deceleration.

Expanding the denominator of the integrand (with the use of equation (7)) in powers of σ, one easily obtains that the character of the singularity is determined by the integral

$$\frac{c}{2\sqrt{2}}\int_{-\alpha}^{0} \frac{\chi_0(\sigma)d\sigma}{\sqrt{|\sigma|a}}; \tag{11}$$

i.e., the singularity is absent. This fact is illustrated in Fig. 2. The physical mechanism underlying the appearance of singularities at the transonic transition is the "accumulation" of disturbances in the vicinity of the load because of the equality of the load velocity and the velocity of the wave propagation. To determine the contribution of the load portion, which at $t = t_0$ has the coordinate $x_0 - \sigma$, to integral (9), it is necessary to determine the point of intersection of the ray characterized by the slope c and originating from the point $(t_0; x_0 + \sigma)$ with the curve $l(t)$. However, we have $v(t_0) = c$; therefore, the distance to the point of intersection is a quantity of the order of $O(\sqrt{\sigma})$, and, hence, we obtain $v(t') - c = O(\sqrt{\sigma})$. In terms of physics, this result means that, at $t = t_0$, the waves from the farther edge of the load arrive with a considerable delay (the delay is the quantity $O(\sqrt{\sigma})$), and, therefore, the disturbance carried by them is insufficient for the singularity formation.

ACKNOWLEDGMENTS

I am grateful to P.A. Zhilin and D.A. Indeĭtsev for the attention given to this work and for valuable advice.

The work was supported in part by the Russian Foundation for Basic Research (grant no. 99-01-00693) and by the Government of St. Petersburg.

REFERENCES

1. Yu. D. Kaplunov and G. B. Muravskiĭ, Mekh. Tverd. Tela, No. 1, 155 (1986).
2. S. Gavrilov, J. Sound Vibr. **222**, 345 (1999).
3. L. I. Slepyan, *Transient Elastic Waves* (Sudostroenie, Leningrad, 1972).
4. A. I. Vesnitskiĭ, L. É. Kaplan, and G. A. Utkin, Prikl. Mat. Mekh. **47**, 863 (1983).
5. V. L. Andrianov, Prikl. Mat. Mekh. **57**, 156 (1993).

Translated by E. M. Golyamina

Technical Physics, Vol. 45, No. 4, 2000, pp. 519–520. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 70, No. 4, 2000, pp. 141–142.
Original Russian Text Copyright © 2000 by Solonin, Stepanov, Chirkin.

BRIEF COMMUNICATIONS

Unstable Regime of the Excitation of Kinetic Striations in a Gas-Discharge Plasma

V. V. Solonin, V. A. Stepanov, and M. V. Chirkin
Ryazan State Pedagogical University, Ryazan, 390000 Russia

Received March 30, 1999

Abstract—Chaotic striations in a low-pressure discharge in the regime corresponding to an unstable periodic orbit singled out of a chaotic attractor are stabilized. A specific feature of the wave process in question is the change in the type of some striations in the course of their motion in the positive column. *© 2000 MAIK “Nauka/Interperiodica”.*

The results of applying the methods of nonlinear dynamics to the investigation of chaotic striations (ionization waves) in the positive column of a low-pressure discharge gives grounds to consider the striations as self-oscillations whose geometrical prototype is a low-dimensional chaotic attractor [1, 2] containing a countable set of unstable periodic orbits. Stabilization of the ionization waves near the orbits whose period is severalfold the striation period (controlling chaos) [3] allows investigation of unstable regimes of the striation excitation.

Experiments were carried out with a dc discharge in a 1-cm-diameter tube filled with neon at a pressure of 2.7 torr. The signals in question are the oscillations of the discharge-gap voltage and the intensity of optical radiation from the stratified plasma; 24 photodetectors are located along the tube to record this radiation. When the length of the positive column and the discharge current change, the regular striations go over into the chaotic ones and back [1]. In the case in question, the length (17.3 cm) and the current (16.5 mA) correspond to the chaotic striations.

The experiment is based on the reconstruction of a chaotic attractor in an m-dimensional pseudo-phase space (embedding space) in which the points are given by the vectors $\mathbf{Y}_j = [y(t_j), y(t_j + \tau), y(t_j + 2\tau), \ldots, y(t_j + (m-1)\tau)]$, $j = 1$–N; $y(t_j)$, where $y(t)$ is the discrete temporal set of values of the optical-radiation intensity from the chosen region of the positive column. The delay time τ is determined by the position of the first minimum of the mutual-information function [4] calculated for the recorded temporal series. As the embedding dimension, we take the minimum value of m, which corresponds to the saturation of the logarithmic derivative $d(\ln C)/d(\ln\delta)$ [5], where

$$C(\delta) = \lim_{N\to\infty}\frac{1}{N^2}\sum_{i}^{N}\sum_{j=1,\, j\neq i}^{N}\sigma(\delta - [\mathbf{Y}_i - \mathbf{Y}_j]), \qquad (1)$$

$\sigma(\ldots)$ is the Heaviside function, and δ is the radius of the sphere in the embedding space.

The striations are stabilized by modulating the anode potential with a signal produced in accordance with one of the periodic orbits preliminary singled out of the reconstructed attractor with the use of the Lathrop–Kostelich algorithm [6]. Amplitudes A_n of components with striation frequencies and with the difference frequency in the spectrum of the controlling signal are presented in the table for the orbits with the period T equal to 1.05 ms.

Peculiarities of the wave process in a stabilized complex periodic regime are shown in the x–t diagram in Fig. 1. The solid lines present the motion of the maximums of the optical-radiation intensity. In the region of the positive column near the anode, the striations move to the cathode with a velocity approximately

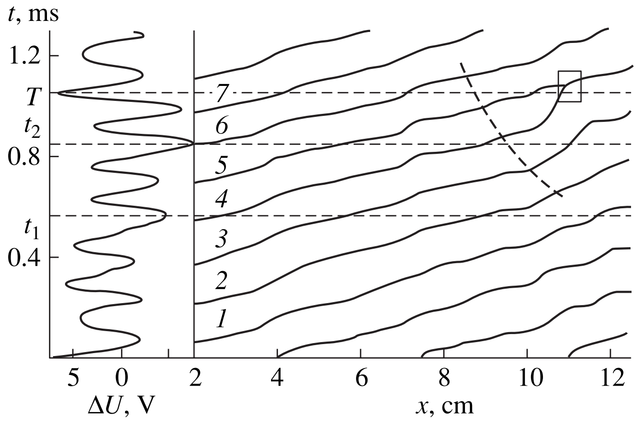


Fig. 1. Alternating component of the voltage across the discharge tube and a fragment of the space–time diagram showing the motion of maximums of optical radiation from a stratified positive column under conditions when the unstable orbit with the period T is stabilized; x is the distance from the anode. Dashed curve shows the propagation of perturbed ionization waves. Numerals near the curves are the striation numbers.

1063-7842/00/4504-0519$20.00 © 2000 MAIK “Nauka/Interperiodica”

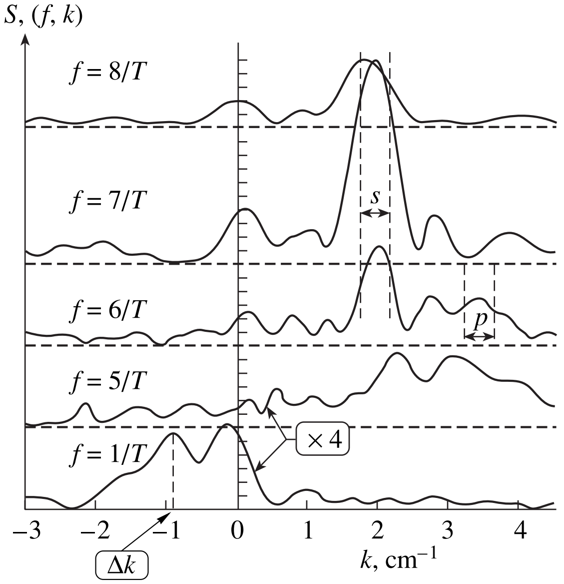


Fig. 2. Spatial spectra of oscillations of the optical-radiation intensity at the striation frequencies and the difference frequency. Dashed lines show the scales corresponding to the *s*-type and *p*-type striations.

equal to 2.2×10^4 cm/s. The space–time spectrum contains peaks, whose frequencies and wavenumbers k are characteristic of the *s*-type kinetic striations [7] with a potential drop of ~19 V (see Fig. 2; positive wavenumbers correspond to the phase velocity directed to the cathode). For $k = 0$, the spectral maxima are related to in-phase oscillations; the sidetones show the variations in the amplitude and the ion velocity.

The discharge-tube voltage is modulated with a frequency close to $7/T$. Twice in the period T (at the instants t_1 and t_2), the voltage decreases (Fig. 1). At the instant t_1, the velocity of the striation *2* decreases to 1.3×10^4 cm/s, which manifests itself in the change in the inclination of the corresponding crest of wave in the x–t diagram. After 0.1 ms, the value of the maximum of optical radiation from this striation drops one-and-a-half times; then, a similar change in the amplitude occurs subsequently for striations *3*–*6*. The perturbation of the ionization waves plotted by the dashed line propagates toward the anode with a velocity of 6.5×10^3 cm/s; this velocity coincides with the ratio $2\pi/(T\Delta k)$ in value and direction (here, Δk is the wavenumber corresponding to the maximum of the spatial spectrum at the difference frequency $f = 1/T$).

f_n	$A_n/\sqrt{\Sigma A_n^2}$
$1/T$	0.060
$5/T$	0.075
$6/T$	0.344
$7/T$	0.811
$8/T$	0.370

The reason for the observed variations is that, for $t = t_1$, the potential drop in the striation *2* does not satisfy the spatial resonance condition for the electron energy distribution function in the *s*-type ionization waves [7]; this affects the ionization rate and the electric field in the following striations. As a result, the length and velocity of the striation *3* approach the values characteristic of the *p*-type ionization waves with a potential drop of ~10 V: 1.8 cm and 10^4 cm/s, respectively. The periodic transformation of the type of some striations corresponds to the spectral maximum typical of the *p*-type striations (Fig. 2).

After the decrease in the voltage across the discharge tube at the instant t_2, the velocity of striation *4* decreases so that striations *4* and *5* merge to produce an object similar to wave dislocations typical of a chaotic regime [8]. The wave merging is accompanied by the increase in the voltage; after that, the intensity of optical radiation from striations *5* and *6* increases.

This effect shows that the distinguishing feature of the unstable periodic regime of the ionization-wave generation is the existence of mechanisms that compensate for the substantial deviation of the electric field in one of the striations from that satisfying the conditions of the spatial resonance for the electron energy distribution function. The observed manifestation of these mechanisms is the change in the type of some striations and merging of the following waves.

REFERENCES

1. V. S. Anishchenko, G. V. Melekhin, *et al.*, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **29**, 951 (1986).
2. K.-D. Weltmann, H. Deutsch, *et al.*, Contrib. Plasma Phys. **33** (2), 73 (1993).
3. K.-D. Weltmann, T. Klinger, and C. Wilke, Phys. Rev. E **52**, 2106 (1995).
4. A. M. Fraser and H. L. Swinney, Phys. Rev. A **33**, 1134 (1986).
5. A. A. Kipchatov, Pis'ma Zh. Tekh. Fiz. **21** (15), 90 (1995) [Tech. Phys. Lett. **21**, 627 (1995)].
6. D. P. Lathrop and E. J. Kostelich, Phys. Rev. A **40**, 4028 (1989).
7. Yu. B. Golubovskiĭ, V. O. Nekuchaev, *et al.*, Zh. Tekh. Fiz. **67** (9), 14 (1997) [Tech. Phys. **42**, 997 (1997)].
8. J. Krasa, J. Phys. D **14**, 1241 (1981).

Translated by A. D. Smirnova