Technical Physics, Vol. 48, No. 1, 2003, pp. 1–6. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 3–9.

THEORETICAL AND MATHEMATICAL PHYSICS

Numerical Simulation of Nanoparticle Images in Scanning Near-Field Optical Microscopy

V. I. Belotelov[1, 2], A. P. Pyatakov[1, 2], A. K. Zvezdin[1, 2], V. A. Kotov[1], and A. S. Logginov[2]

[1] *General Physics Institute, Russian Academy of Sciences, ul. Vavilova 38, Moscow, 117942 Russia*

[2] *Moscow State University, Vorob'evy gory, Moscow, 119899 Russia*

e-mail: bvi@osc.phys.msu.ru

Received March 19, 2002

Abstract—The images of magnetic and nonmagnetic nanoparticles obtained by scanning near-field microscopy in the photon collection mode are numerically simulated. A theoretical approach that uses tensor electrodynamic Green's functions to find the optical near field in a given observation scheme is considered. Typical images of nanoparticles with various shapes are obtained by numerical simulation. Subject to boundary conditions, the plane of polarization is shown to change at topographic features (edges and angles) of objects studied. This makes the observation of the magnetic structure of a nanoparticle with a magnetooptic method difficult. The near-field study of the magnetization distribution in homogeneous thin films appears to be more effective, since the rotation of plane of polarization is associated primarily with the magnetic properties of the sample in this case. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Nowadays, there is great interest in the study of the magnetic and magnetooptic properties of low-dimension systems and nanostructures. Magnetic nanostructures are viewed as a promising material for ultra-high-density (up to 1 Tbit/cm^2) data recording. Note that such a giant record density approaches the superparamagnetic limit (10 Tbit/cm^2) [1], which seems to be the ultimate value of the surface record density (today's value is 30 Gbit/cm^2 [2]). The typical size of a one-domain magnetic particle in structured nanomedia is estimated at 5–100 nm.

Therefore, a number of magnetooptic techniques for studying nanostructures with a subwave resolution that are lumped together as scanning near-field optical microscopy (SNOM) are attracting considerable attention [3]. SNOM combines the rich capabilities of optical methods with the high resolution of scanning probe microscopy. In SNOM, a probe scans the specimen, the distance between them being smaller than the wavelength from its surface. Two SNOM operating modes are distinguished: the illumination mode and the photon collection mode [4, 5]. In both modes, the resolution depends on the aperture size and probe–specimen spacing rather than on the wavelength. In the illumination mode, the probe acts as an optical near-field generator, which illuminates the specimen. Changes in the far-field radiation are related to those in the near field due to optical inhomogeneities. Among SNOM modifications operating in the collection mode, photon scanning tunnel microscopy, where light is incident at the angle of total internal reflection (Fig. 1), has found the widest application. The optical near field localized in the neighborhood of the specimen surface is detected with a near-field probe. This scheme is used in this work.

While near-field microscopy has been extensively applied in recent years for studying the domain structure of magnetic materials both experimentally and theoretically [6–12], its outlook for examining magnetic microparticles applied on a nonmagnetic surface still remains unclear.

In addition, the interpretation of SNOM images is difficult both under magnetooptic observation and in the case of nonmagnetic particles. A natural approach to this problem is the computer simulation of near-field images with some *a priori* model of the object with the subsequent comparison of simulated and observed images.

This work is concerned with the physical problems of magnetooptic SNOM and is aimed at developing a method of numerical simulation as applied to SNOM images of nanoparticles.

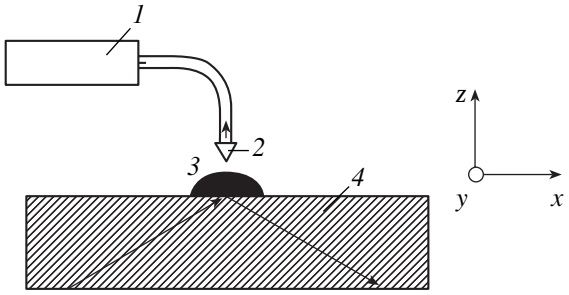

Fig. 1. Scheme of the near-field observation of a nanoparticle: (*1*) detector, (*2*) near-field probe, (*3*) nanoobject, and (*4*) transparent plate.

1063-7842/03/4801-0001$24.00 © 2003 MAIK "Nauka/Interperiodica"

CALCULATION TECHNIQUE

We consider the passive probe model [12], which ignores the effect of the probe on the SNOM image and assumes that the signal detected is proportional to the near-field intensity at the nanostructure surface in the absence of the probe. In the observation geometry depicted in Fig. 1, such a model is valid and allows one to obtain images similar to those observed experimentally [12, 13]. This is because the inducing radiation decays exponentially near the probe.

The near field is calculated with tensor electrodynamic Green's functions. This approach proved to be valid in solving problems of light reflection and transmission in rough-surface media [14, 15], as well as in investigating nonlinear surface magnetooptic effects [16, 17].

Let us consider a semi-infinite medium with a nanoparticle on the surface (Fig. 1). The medium is illuminated by an optical wave $\mathbf{E}^{(i)}(\mathbf{r}, t) = \mathbf{E}^{(i)}\exp(-i(\mathbf{k}_0\mathbf{r} - \omega t))$ at an angle exceeding the critical angle of internal reflection.

From the Maxwell equations, we have

$$\mathrm{curl}\,\mathrm{curl}\mathbf{E}(\mathbf{r}, t) = -\frac{1}{\varepsilon_0 c^2}\frac{\partial^2}{\partial t^2}\mathbf{D}(\mathbf{r}, t).$$

Passing to the field complex amplitudes

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r})\exp(i\omega t)$$
$$\text{and}\quad \mathbf{D}(\mathbf{r}, t) = \mathbf{D}(\mathbf{r})\exp(i\omega t)$$

in view of the material equation

$$\mathbf{D}(\mathbf{r}) = \varepsilon_0\mathbf{E}(\mathbf{r}) + \mathbf{P}(\mathbf{r}),$$

one obtains the equation for $\mathbf{E}(\mathbf{r})$

$$\mathrm{curl}\,\mathrm{curl}\mathbf{E}(\mathbf{r}) - k_0^2\hat{\varepsilon}(\mathbf{r})\mathbf{E}(\mathbf{r}) = \frac{k_0^2}{\varepsilon_0}\mathbf{P}(\mathbf{r}), \qquad (1)$$

where $\mathbf{P}(\mathbf{r})$ is the polarization induced in the particle by the external field and $\varepsilon(\mathbf{r}) = \varepsilon$ (if $z < 0$) or 1 (if $z > 0$).

In the case considered, one can consider the right-hand side of Eq. (1) as a small perturbation ($P \ll \varepsilon_0 E^{(i)}$) and solve this equation in the Born first approximation, where the field is analytically representable as the sum of two terms:

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^{(0)}(\mathbf{r}) + \mathbf{E}^{(1)}(\mathbf{r}).$$

Here, $\mathbf{E}^{(0)}(\mathbf{r})$ is the field without the particle, which satisfies the homogeneous equation

$$\mathrm{curl}\,\mathrm{curl}\mathbf{E}^{(0)}(\mathbf{r}) - k_0^2\varepsilon(\mathbf{r})\mathbf{E}^{(0)}(\mathbf{r}) = 0, \qquad (2)$$

and $\mathbf{E}^{(1)}(\mathbf{r})$ is the field scattered by the particle, which satisfies the inhomogeneous equation

$$\mathrm{curl}\,\mathrm{curl}\mathbf{E}^{(1)}(\mathbf{r}) - k_0^2\varepsilon(\mathbf{r})\mathbf{E}^{(1)}(\mathbf{r}) = \frac{k_0^2}{\varepsilon_0}\mathbf{P}^{(0)}(\mathbf{r}), \qquad (3)$$

where $\mathbf{P}^{(0)}(\mathbf{r})$ is the $\mathbf{E}^{(0)}(\mathbf{r})$-induced part of the particle polarization.

Solutions to Eq. (2) are known as Fresnel formulas, which yield the decaying wave $\mathbf{E}^{(0)}(\mathbf{r}) = \mathbf{E}^{(i)}(\mathbf{r})\exp(-\sqrt{(k_x)^2 - k_0^2}z)$ at $z > 0$, where $k_x = \sqrt{\varepsilon}\,k_0\sin\Theta$ and Θ is an angle exceeding the angle of total internal reflection.

Equation (3) is inhomogeneous and can be solved by using the Green's functions $D_{\mu\nu}(\mathbf{r}, \mathbf{r}')$, which are introduced by means of the equation [14]

$$\left(\frac{\partial^2}{\partial x_\lambda \partial x_\mu} - \delta_{\lambda\mu}\frac{\partial^2}{\partial x_\mu^2} - \varepsilon(\mathbf{r})\frac{\omega^2}{c^2}\delta_{\lambda\mu}\right)D_{\mu\nu}(\mathbf{r}, \mathbf{r}') = -\delta_{\lambda\nu}\delta(\mathbf{r} - \mathbf{r}'), \qquad (4)$$

where $\delta_{\lambda\nu}$ is the Kronecker symbol.

The solution to Eq. (3) can be written as

$$E_\mu^{(1)}(\mathbf{r}) = -\frac{k_0^2}{\varepsilon_0}\int D_{\mu\nu}(\mathbf{r}, \mathbf{r}')P_\nu^{(0)}(\mathbf{r}')\cdot d\mathbf{r}', \qquad (5)$$

where the subscripts μ and ν take the values of x, y, and z. Repeating subscripts imply summation.

Space integration requires a large body of computation. The computation time can be cut in the Fourier space. By introducing the Fourier representations for the Green's functions

$$D_{\mu\nu}(\mathbf{r}, \mathbf{r}') = \frac{1}{4\pi^2}\int d_{\mu\nu}(\mathbf{k}_\parallel, z, z')\exp(i\mathbf{k}_\parallel(\mathbf{r}_\parallel - \mathbf{r}_\parallel'))d\mathbf{k}_\parallel,$$

where $\mathbf{r}_\parallel = (x, y, 0)$, $\mathbf{r}_\parallel' = (x', y', 0)$, and $\mathbf{k}_\parallel = (k_x, k_y, 0)$, we can rewrite formula (5) in terms of the Fourier transforms:

$$\hat{E}_\mu^{(1)}(\mathbf{k}_\parallel, z) = -\frac{k_0^2}{\varepsilon_0}\int d_{\mu\nu}(\mathbf{k}_\parallel, z, z')\hat{\mathbf{P}}_\nu^{(0)}(\mathbf{k}_\parallel, z')dz'. \qquad (6)$$

Note that the Fourier transforms $d_{\mu\nu}(\mathbf{k}_\parallel, z, z')$ of the Green's functions are found by solving ordinary differential equations. The associated procedure is described in [14].

The scattered radiation field can be found with inverse Fourier transformation

$$E_\mu^{(1)}(\mathbf{r}) = \frac{1}{4\pi^2}\int\limits_{\{\mathbf{k}_\parallel\}}\hat{E}_\mu^{(1)}(\mathbf{k}_\parallel, \mathbf{r})e^{i\mathbf{k}\mathbf{r}}d\mathbf{k}_\parallel. \qquad (7)$$

The domain of integration $\{\mathbf{k}_\parallel\}$ specifies the spatial resolution of simulated image. For example, a resolution of 10 nm ($\lambda = 0.5$ μm), which is typical of SNOM, can be achieved if all spatial harmonics with $k_{x,y} \in (-25k_0, 25k_0)$ are integrated.

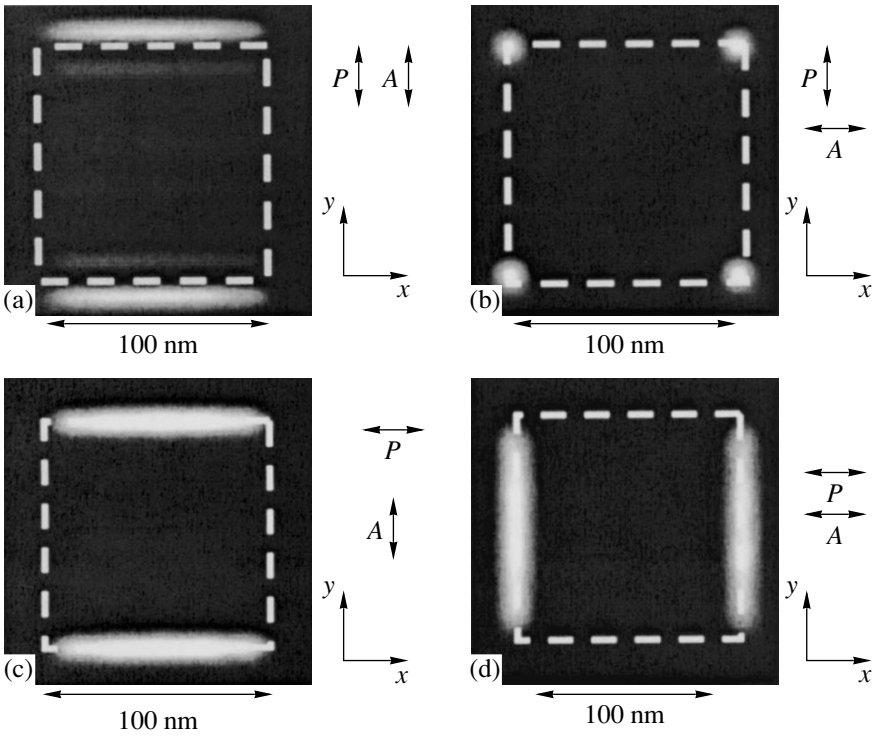

Fig. 2. Simulated images of a nonmagnetic square particle under various illumination and observation conditions. The spatial resolution is 10 nm. *A*, analyzer; *P*, polarizer.

RESULTS OF NUMERICAL SIMULATION FOR NONMAGNETIC PARTICLES

The simulation was carried out for a semi-infinite medium with a refractive index $n = 1.5$ and an angle of incidence $\Theta = 70°$ (at $n = 1.5$, the angle of total internal reflection is 42°). The polarization of a particle on the right of (3) is represented as $P_i^{(0)} = \varepsilon_0 \chi E_i^{(0)}$, where the susceptibility $\chi = -15.9 + i1.1$; in other words, we are dealing with an isotropic Ag particle with a complex refractive index $n_1 = \sqrt{\varepsilon_1} = \sqrt{1 + \chi} = 0.14 + i4.0$ [18]. Two basic illumination geometries were considered: *s*-polarized incident radiation (the field is normal to the plane of incidence) and *p*-polarized incident radiation (polarization in the plane of incidence). The field was calculated at a distance from the specimen surface (the so-called scan height).

(1) Images of a square particle. Figure 2 shows the simulated images of a 100-nm-square particle of height $h = 10$ nm (the scan height $z = 11$ nm) that are obtained for four arrangements of the incident radiation and analyzer. Dashed lines are the boundaries of the particle.

For *s*-polarized incident light with a parallel polarizer and analyzer (Fig. 2a), the intensity markedly differs from the background (the decaying field of illumination) at those edges of the particle normal to the inducing field. With a crossed polarizer and analyzer and *s*-polarized incident light, the intensity exceeds the background only near the corners where the plane of polarization rotates (Fig. 2b). Figure 2c demonstrates the image obtained in *p*-polarized incident light with a crossed polarizer and analyzer. The intensity is other than zero only at the edges parallel to the plane of incidence, the *z* component of the incident field (which is absent in Fig. 2b) making a major contribution to the formation of the image. At these edges, the *z* component transforms into the *y* component, which passes through the analyzer. Because of the *z* component in the *p*-polarized incident wave, the images in Figs. 2b and 2c differ. Figure 2d shows the image in *p*-polarized incident light for a parallel polarizer and analyzer. As in Fig. 2a, the contrast is enhanced at the edges of the particle.

The simulation of images for the same geometries but with larger scan heights demonstrates that the images become more asymmetric and the spatial resolution drops; i.e., the boundaries diffuse. This is because the number of harmonics being detected decreases with increasing scan height, since harmonics with high spatial frequencies decay before they reach the probe.

(2) Images of a circular particle. Figure 3 shows the simulated images of a circular particle of diameter 100 nm ($h = 10$ nm, $z = 11$ nm) for four arrangements of incident light polarization and analyzer. Dashed lines are the boundaries of the particle.

As in the case of a square particle (Fig. 2), with a crossed polarizer and analyzer (Figs. 3a, 3d), the con-

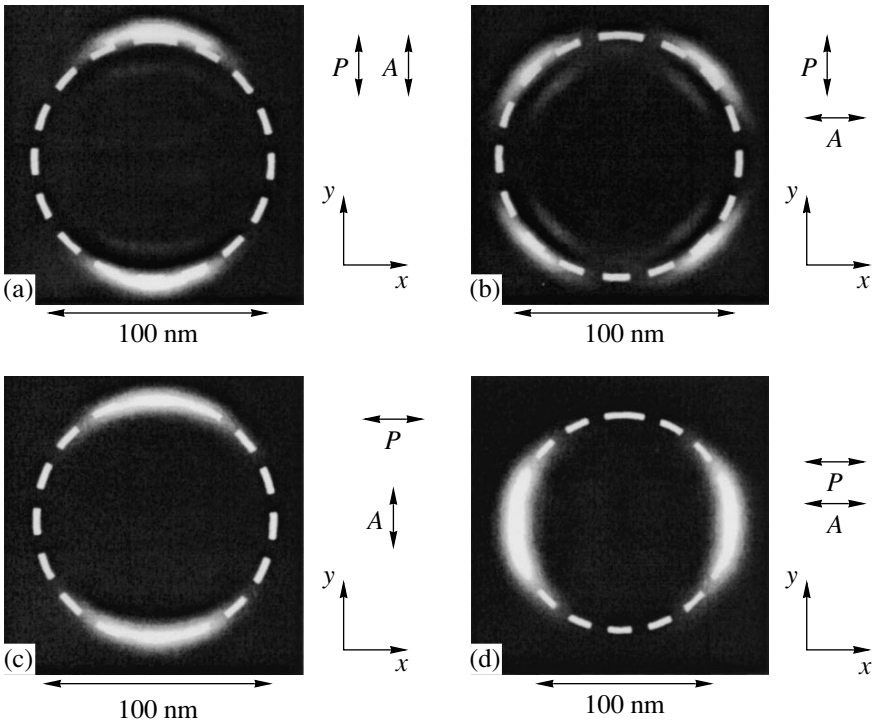

Fig. 3. The same as in Fig. 2 for a circular particle.

trast is enhanced at the edges normal to the incident field, since the boundary conditions favor the occurrence of the fields near arcs perpendicular to the incident field.

With a crossed polarizer and analyzer, the plane of polarization rotates at the edges of the particle (Fig. 3b, 3c). Here, arcs making an angle $\varphi = 45°$ with the plane of polarization of the incident radiation (Fig. 3b) play the role of the corners of a square particle (Fig. 2b). The explanation for this fact is simple. Indeed, if an incident wave is *s*-polarized, the field strength at the side surface of the particle is proportional to $E^{(i)}\cos\varphi$ (Fig. 4) because of the boundary conditions; then, in view of Malus's law, the transmitted intensity varies as $(E^{(i)}\sin\varphi\cos\varphi)^2$ (Fig. 4). This function is maximal at $\varphi = 45°$, 135°, etc.

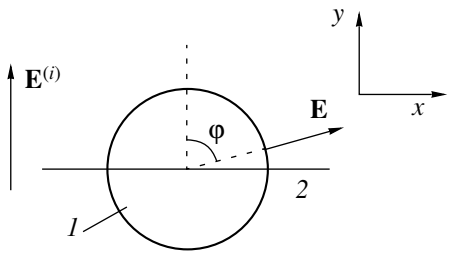

Fig. 4. Explanation of a simulated image of a circular nanoparticle in crossed polarizers for *s*-polarized incident radiation (Fig. 3b). (*1*) Circular nanoparticle (top view) and (*2*) plane of the analyzer.

The presence of the field at the edges parallel to the radiation (Fig. 3c) is explained, as for the case in Fig. 2c, by the rotation of the *z* component of the electric field at the edges of the particle.

It should be noted that the simulated images in Figs. 3a and 3b are in good agreement with the images of submicron metal rings observed under similar conditions [6].

DISCUSSION OF MAGNETOOPTIC IMAGES

As follows from the results of simulation, the plane of polarization rotates at the particle edges. This fact should be taken into account in the interpretation of magnetooptic images, since magnetooptic observation exploits the property of a magnetized material to rotate the plane of polarization of a wave. This property follows from the form of the permittivity tensor for a magnetic medium, which contains off-diagonal elements proportional to the magnetization [19, 20]:

$$\hat{\varepsilon}_1 = n_1^2 \begin{pmatrix} 1 & -im_zQ & im_yQ \\ im_zQ & 1 & -im_xQ \\ -im_yQ & im_xQ & 1 \end{pmatrix},$$

where $(m_x, m_y, m_z) = \mathbf{M}/M$ is the magnetization unit vector and Q is the magnetooptic parameter linearly dependent on the magnetization $\mathbf{M}$.

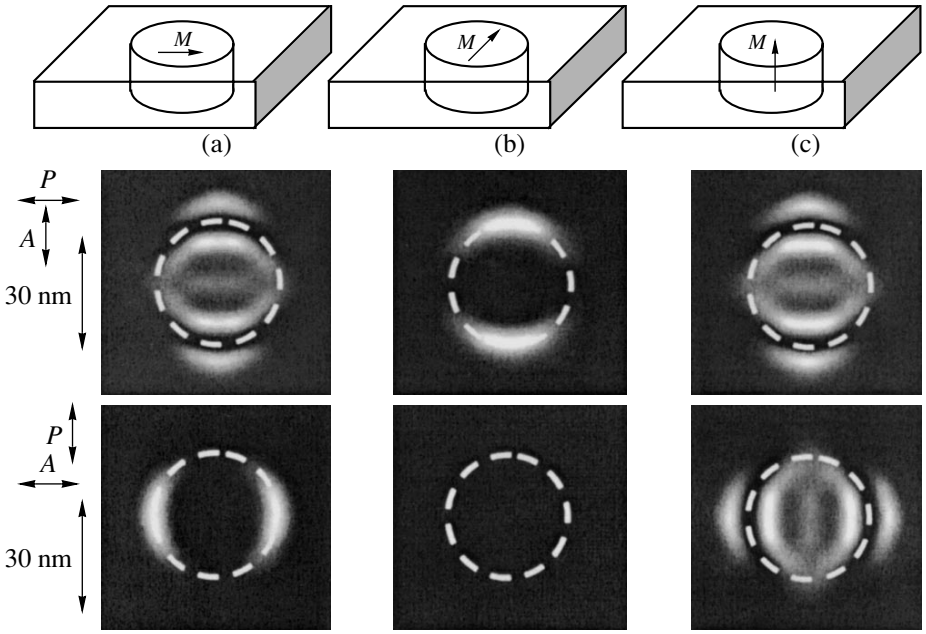

Fig. 5. Simulated images of a 30-nm magnetic region embedded in a 10-nm-thick film under various illumination and observation conditions. The images of the nanoregion magnetized along the (a) *x*, (b) *y*, and (c) *z* axis are shown. The scan height is 11 nm, and the spatial resolution is 5 nm. *A*, analyzer; *P*, polarizer.

Taking into account that Q is on the order of several hundredths ($Q = -0.034 + 0.003i$ for iron [18]), one can infer that a magnetic particle is difficult to observe because of the unavoidable nongyrotropic rotation of the plane of polarization at its edges.

The magnetooptic contribution can be directly detected only at sites where the nonmagnetic contribution to the image approaches zero. Specifically, for a circular particle magnetized along the surface parallel to the plane of light incidence ($\mathbf{m} = (m_x, 0, 0)$), the magnetooptic contribution becomes noticeable in crossed polarizers (the exponentially decaying background field is eliminated) at the edges parallel to the incident light. Note the nonmagnetic contribution is several hundreds of times higher than the magnetooptic component.

From the aforesaid, it follows that nonmagnetic inhomogeneities of nanoobjects (nanoparticle edges, lattice defects, etc.) make magnetooptic observations with SNOM difficult.

Therefore, the near-field observation of magnetic structures may prove to be much more efficient in investigating objects where the gyrotropic part of the permittivity tensor is coordinate dependent. Such are, for example, domain walls, vertical Bloch lines [21], and magnetic nanoregions. These regions have recently been fabricated by laser interference lithography [22, 23]. In particular, nanoregions 20 to 200 nm in size that offer ferromagnetic properties were fabricated in thin (several tens of nanometers thick) homogeneous nonmagnetic Co–C films [22]. Such nanoregions have a fairly smooth surface and are virtually free of any nonmagnetic inclusions; therefore, they can be successfully observed by SNOM.

Figure 5 shows the simulated images of 30-nm circular magnetic regions embedded in a 10-nm-thick nonmagnetic homogeneous layer.

Upon simulation, it was assumed that the additional susceptibility $\chi_{i,j}$ due to the magnetic region is given by the tensor

$$\chi_{i,j} = n_1^2 \begin{pmatrix} 0 & -im_z Q & im_y Q \\ im_z Q & 0 & -im_x Q \\ -im_y Q & im_x Q & 0 \end{pmatrix},$$

where $n_1 = 2.4 + i3.5$ and $Q = -0.034 + 0.003i$ [18].

The simulation was carried out for different directions of the object magnetization. It follows from Fig. 5 that one can judge the magnetization direction in the magnetic nanoregion, varying the orientation of the incident light polarization.

CONCLUSION

Using tensor electrodynamic Green's functions to solve the Maxwell equation, we simulated electromagnetic fields near nonmagnetic and magnetic nanometer particles. This can be viewed as the first approximation to the problem of the numerical simulation of images in scanning near-field optical microscopy. Various images simulated under different illumination and observation conditions (*s*- and *p*-polarized incident light in crossed and parallel polarizers) are demonstrated.

In the near field, the rotation of the plane of polarization at topographical features of the specimen is shown to have a significant effect. This may make the interpretation of SNOM images difficult. However, these difficulties are greatly reduced when near-field microscopy is applied to imaging magnetic nanostructures prepared by modifying homogeneous ultrathin films by laser radiation. This is because, in this case, nonmagnetic inclusions, which are obstacles to magnetooptic investigation, are virtually absent.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 01-02-16595) and INTAS (project no. 99-01839).

REFERENCES

1. K. O'Grady and H. Laidler, J. Magn. Magn. Mater. **200**, 616 (1999).
2. D. Weller, *Plenary Lecture at the Joint European Symposium on Magnetism, JEMS-2001, Grenoble, 2001.*
3. D. W. Pohl, W. Denk, and M. Lanz, Appl. Phys. Lett. **44**, 651 (1984).
4. *Near-Field Nano/Atom Optics and Technology*, Ed. by M. Ohtsu (Springer-Verlag, Berlin, 1998).
5. G. S. Zhdanov, M. N. Libenson, and G. Martsionovskiĭ, Usp. Fiz. Nauk **47**, 801 (1998) [Phys. Usp. **41**, 719 (1998)].
6. E. Betzig, J. K. Trautman, J. S. Weiner, *et al.*, Appl. Opt. **31**, 4563 (1992).
7. E. B. McDaniel, S. C. McClain, and J. P. W. Hsu, Appl. Opt. **37**, 84 (1998).
8. H. Wioland, O. Bergossi, S. Hudlet, *et al.*, Eur. Phys. J.: Appl. Phys. **5**, 289 (1999).
9. A. A. Ezhov, A. S. Logginov, D. A. Muzychenko, *et al.*, Poverkhnost, No. 11, 56 (2000).
10. K. Saito, A. Nakaoki, and M. Kaneko, Jpn. J. Appl. Phys., Part 1 **38**, 6743 (1999).
11. V. A. Kosobukin, Zh. Tekh. Fiz. **68** (7), 86 (1998) [Tech. Phys. **43**, 824 (1998)].
12. *Optics at the Nanometer Scale: Imaging and Storing with Photonic near Fields*, Ed. by M. Nieto-Vesperinas and N. Garcia (Kluwer, Dordrecht, 1996).
13. J. R. Krenn, A. Dereux, J. C. Weeber, *et al.*, Phys. Rev. Lett. **82**, 2590 (1999).
14. A. A. Maradudin and D. L. Mills, Phys. Rev. B **11**, 1392 (1975).
15. *Scattering in Volumes and Surfaces*, Ed. by M. Nieto-Vesperinas and J. C. Dainty (North-Holland, Amsterdam, 1990).
16. V. I. Belotelov, A. P. Pyatakov, S. A. Eremin, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **42**, 1826 (2000) [Phys. Solid State **42**, 1873 (2000)].
17. V. I. Belotelov, A. P. Pyatakov, S. A. Eremin, *et al.*, Opt. Spektrosk. **91**, 663 (2001) [Opt. Spectrosc. **91**, 626 (2001)].
18. E. D. Palik, *Handbook of Optical Constants of Solids* (Academic, Orlando, 1985).
19. A. K. Zvezdin and V. A. Kotov, *Modern Magnetooptics and Magnetooptical Materials* (Inst. of Physics Publ., Bristol, 1997).
20. A. K. Zvezdin and V. A. Kotov, *Magnetooptics of Thin Films* (Nauka, Moscow, 1988).
21. A. P. Malozemoff and J. C. Slonczewski, *Magnetic Domain Walls in Bubble Materials* (Academic, New York, 1979; Mir, Moscow, 1982).
22. R. Skomski, S. H. Liou, D. J. Sellmyer, *et al.*, IEEE Trans. Magn. **37**, 2070 (2001).
23. D. Babonneau, J. Briatico, F. Petroff, *et al.*, J. Appl. Phys. **87**, 3432 (2000).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 1, 2003, pp. 100–103. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 106–109.

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Electrical Conduction in Binary Macrosystems: General Rules

V. A. Sotskov* and S. V. Karpenko**

** Kabardino-Balkar State University, ul. Chernyshevskogo 173, Nalchik, 360004 Russia*

*** Research Institute of Applied Mathematics and Automation, Russian Academy of Sciences, Nalchik, Russia*

e-mail: sv_karpenko@mailru.com

Received April 26, 2002

Abstract—The conductivity of a macroscopically disordered graphite–paraffin system as a function of the graphite content and temperature is studied. It is shown that the properties of the heterosystem differ essentially from those of its components. Experimental data obtained are explained in terms of the percolation theory and concepts of fractal-structure objects. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Electrical conduction in binary systems has been studied for a variety of constituents [1–4]. After certain requirements following from the percolation theory had been realized in experiments, there appeared the possibility of invoking the percolation theory and fractal conceptions to explain the phenomena observed. Most of the results within the percolation theory have been obtained with numerical methods, and only in rare cases were model experiments carried out [1–4].

In this paper, we study the resistivity (ρ) of a binary conductor–insulator macrosystem as a function of the volume content of the conductor (x) and temperature (T). Our model experiment sheds new light on electrical conductivity in binary macroscopically disordered systems.

EXPERIMENTAL

The system components were chosen from the following considerations.

(i) The resistivity of the conducting and insulating phases must be minimal and maximal, respectively.

(ii) The contact resistance of conducting particles must be minimal. For instance, the particles must be free of oxide film.

(iii) The number of particles must be large enough to provide the desired accuracy of the experiment, since the rms deviation of the percolation threshold (x_c) is inversely proportional to the number of particles in a system [3].

(iv) Chemical interactions causing new components to appear are absent at all stages of the experiment.

(v) The technique for preparing the binary system and measuring cell must be as simple as possible and provide the adequate mixing of the components and sample consolidation.

As for contact resistance minimization, the use of graphite as a conducting component has several advantages over metals [4]. Carbon oxides are gases; therefore, the problem of an oxide film present on the particles is eliminated irrespective of the graphite chemical origin. Moreover, carbon does not react with most insulators in a wide temperature range [5]. On using, e.g., aluminum oxide as an insulator, one should actually consider a three-component conductor–insulator–air system, which has several features making it difficult to study (air humidity, etc.). The application of a low-melting insulator seems to be more reasonable, since it provides the consolidation of the conductor–insulator sample. In our opinion, the use of ceresin and paraffin is the most appropriate. Ceresin and paraffin are nonpolar thermoplastic crystalline saturated hydrocarbons of the general formula C_nH_{2n+2} (n = 39–53 for ceresin and 10–36 for paraffin). Ceresin molecules have heavily branched chains of carbon atoms unlike nearly straight weakly branched chains in paraffin molecules. The melting points of ceresin and paraffin are 75 and 52°C, respectively [5]. Based on the above considerations and assuming that the simplicity of sample preparation has priority, we selected S-1 pyrolytic graphite (dry colloidal preparation, specs. 113-08-48-63-90) with a main particle size of 4×10^{-6} m and P2 solid oil paraffin (State Standard 23683-89).

Since paraffin is a saturated hydrocarbon and the sample preparation temperature does not exceed 55°C, intermediates in the system are absent. It was found that lower paraffins are oxidized with difficulty. The oxidation of higher paraffins is carried out at a temperature of about 150°C in the presence of catalysts (e.g., manganese compounds). Carbon is chemically inactive under

the conditions considered in this study [5]. The examination of the sample surface with an optical microscope showed the presence of homogeneous carbon inclusions in the paraffin. Thus, we assumed that the carbon–paraffin system is a macroscopically inhomogeneous system.

The resistivity was measured by the standard two-probe technique. Electrolytic copper was used as an electrode material. The electrodes were covered by a liquid mixture of the system components and were placed inside the sample. Only a small contact area was left outside.

RESULTS

The dependence of the resistivity of the system on the graphite volume content is shown in Fig. 1.

The plot of the function $\log\rho = f(x)$, where x is the graphite volume content in the sample, has a kink at $x_c = 0.16$, as follows from the procedure presented in Fig. 1. In terms of the percolation theory, this value is the percolation threshold [1–3]. At $x_c = 16\%$, a conducting cluster appears, which has been reported in numerous theoretical studies [1, 3]. Our resistivity values for the graphite and paraffin are close to those obtained by other authors [6].

Figure 1 also presents the curve $\log\rho = f(x)$ calculated with the well-known formula for the electrical conductivity [2, 3]

$$\sigma = \sigma_3(x - x_c)^{2\nu},$$

where x_c is the percolation threshold (the best agreement between the theoretical and experimental curves was observed for $\nu = 0.7$), σ_3 is the conductivity of the carbon network at $x = 1$, and $\nu = 0.7$.

Considering a particle as a cube, one can estimate the number of particles in the system, which turns out to be $N \geq 1.4 \times 10^9$. It is known [2, 3] that the notion of a well-defined percolation threshold is valid only for an infinite system. In real finite systems, the percolation threshold cannot be clearly defined: here, there is only the so-called critical region $\delta(N)$, where $\delta(N)$ is the rms deviation. The value of x_c obtained in most of the experiments falls into this region.

Using the value $N = 1.4 \times 10^9$ and the computational procedure suggested in [3], one can estimate the rms deviation of x_c as $\delta(N) \leq 7.64 \times 10^{-4}$ and the width of the percolation threshold region as $\Delta_N \leq 18 \times 10^{-4}$ [4].

The fact that the correlation radius index is equal to 0.7 deserves attention. In many of the papers [1–3], the electrical conductivity indices for a three-dimensional system were reported to be $t = 1.5$–2.0. Assuming $t = 2\nu$ for the three-dimensional case [3], we can argue that the obtained values of the correlation radius index and conductivity index are near the lower limit of the above interval.

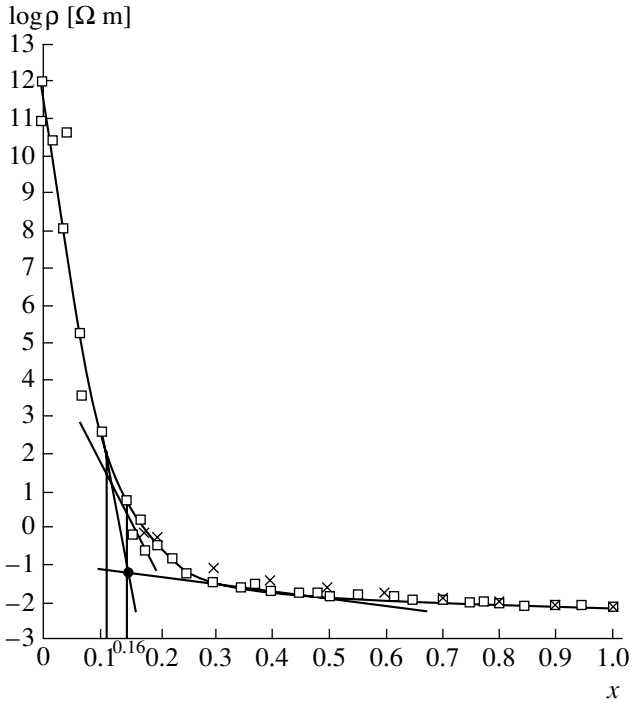

Fig. 1. Resistivity of the graphite–paraffin system as a function of the graphite volume content (the semilogarithmic scale): □, experiment; ×, theory.

In most problems treated in terms of the percolation theory, the resistivity of a system is assumed to be infinite at $x < x_c$. However, this is not the case in real materials. The portion of the experimental curve from $x = 0$ to $x = 0.11$ is well approximated by the relationship $\rho = 10^{\alpha(x)}$, where $\alpha(x) = 12$–$87.5x$, while the portion from $x = 0.11$ to $x = 0.16$ can be approximated by the function $\rho = 10^{\alpha(x)}$, where $\alpha(x) = 5 - 14.3x$. That is, in both portions of the curve, the resistivity depends exponentially on the graphite content in the sample, and the kink is observed at $x = 0.11$. Similar results were obtained in [2], where the tentative theoretical dependence of the conductivity on the maximal values of the resistance exponents was studied.

The temperature dependence of the resistivity of the graphite–paraffin composition was investigated in a thermostat with the current leads outside the temperature zone. The maximal heating temperature was determined from the onset of plastic deformation in the sample and hot spots appearing near the electrodes. The associated results for a number of x are presented in Fig. 2. The temperature resistance coefficient (TRC) α_R was defined as

$$\alpha_R = \frac{1}{\rho}\frac{d\rho}{dT}.$$

Three types of the $\log\rho = f(T)$ curves can be distinguished depending on the x range.

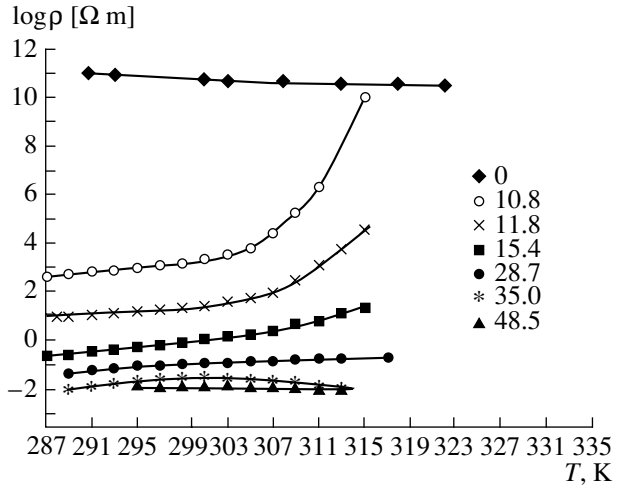

Fig. 2. Resistivity of the graphite–paraffin system as a function of temperature. The figures by the symbols indicate the values of x (%).

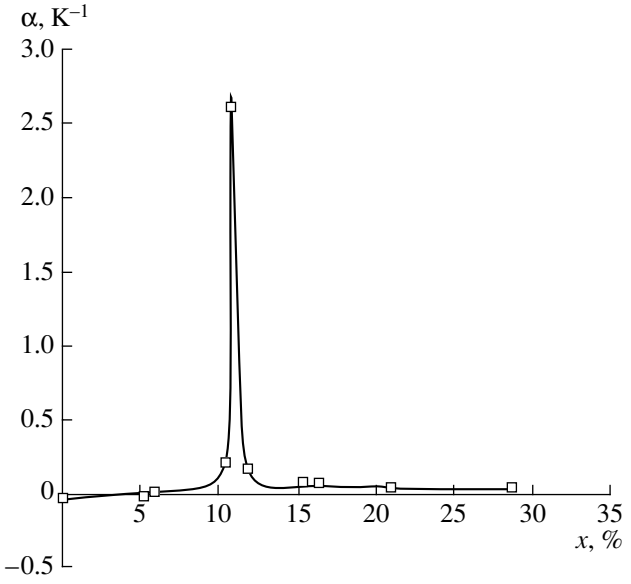

Fig. 3. Temperature resistance coefficient of the graphite–paraffin system as a function of the graphite volume content at $T = 313$ K.

(1) $0 \le x \le 0.06$. The curves decrease monotonically and almost linearly. In molecular-lattice substances, including paraffin, the conductivity is low and depends mainly on the impurity content [5]. The decrease in the resistivity at low temperature can be explained by the presence of ionized impurities. The curves $\log\rho = f(T)$ do not have kinks up to the melting point, which suggests that the conductivity in the temperature range considered is not intrinsic.

At $0.42 \le x \le 1$, the curves are similar to those considered above ($0 \le x \le 0.06$) with the TRC close to that of pyrolytic graphite.

(2) $0.06 \le x < 0.29$. The curves $\log\rho = f(T)$ increase monotonically. Such a behavior cannot be explained by the properties of the system constituents, paraffin and graphite. As was indicated above, the curve $\log\rho = f(T)$ for pure paraffin is decreasing. The TRC of graphite is equal to -2×10^{-4} K^{-1} [7], whereas $\alpha_R \approx +3 \times 10^{-2}$ K^{-1} at a graphite content in the mixture of $x = 0.287$, i.e., at $x > x_c$, when the conducting carbon framework has already formed. Such a great difference can be explained only by the formation dynamics of conducting graphite regions, as well as by taking into account the difference in the linear expansion coefficients of graphite (7×10^{-6} K^{-1}) and paraffin (130×10^{-6} K^{-1}) [7]. For $0.06 \le x \le 0.16$, the curves have a kink at $T = 304$–307 K, whereas in the region $0.06 \le x < 0.29$, they are almost linear. It might be expected that at $x \approx x_c$ the temperature dependence of the resistivity of the composite is the most pronounced, since in this case the first conducting chain of graphite particles forms [1–3] and the difference in the linear expansion coefficients of graphite and paraffin is the most essential. However, the dependence $\log\rho = f(T)$ is most clearly defined at $x \approx 0.108$, i.e., near the second kink of the curve in the range $0.06 \le x < 0.16$ (Fig. 1).

(3) $0.29 \le x \le 0.42$. The curves have two characteristic segments where they increase and decrease; i.e., they have a kink at $T = 297$–328 K.

The x dependence of the TRC (excluding the range $0.3 < x < 1$, where the TRC changes sign) is shown in Fig. 3. The anomalously high values of α_R (up to 2.6 K^{-1}) at $T = 313$ K, as well as the wide range of α_R variation (from -4×10^{-2} to 2.6 K^{-1}) at relatively low temperature variations ($\Delta T = 25$ K), are worth noting.

DISCUSSION

The results obtained can be considered within the theory of fractal-structure objects [8–10]. In terms of the concept of fractal clusters, the particles in the system conglomerate as the number of carbon particles in the melt grows. We do not know the laws which the formation of DLA, RLA, or intermediate clusters obeys [9]. However, it can be assumed that the fastest processes, which are responsible for the formation of clusters with a minimal binding energy, i.e., those processes corresponding to the overcoming of barriers that have a minimal height $Q_{\alpha\beta}$, occur first [8]. Then, clusters with a higher formation energy arise. The latter contain more particles than those formed at the previous energy stage. As the graphite content increases, this process may go up to $x = 1$. We can separate the clusters by their structure (number of particles and their arrangement) and by their binding energy or assign them a particular level. Such levels can be classified by both cluster configuration and formation energy. This classification may cover not only individual clusters but the paraffin–graphite system as a whole.

Thus, we arrive at a multilevel system that is similar, for example, to a Cayley tree [8]. It is a hierarchical subordinate system. The subordination in the system is in the fact that clusters of a following level cannot arise until those with a lower binding energy form.

Clearly, clusters of some of the levels will possess physical properties that differ from the properties of clusters belonging to previous or following levels if only because of their different sizes [9]. This distinction may show up in electrical, thermodynamic, and other properties of the system. Specifically, when the clusters geometrically overlap, a conducting chain, which changes the conductivity type in the whole graphite–paraffin system, forms. If the levels take into account the electrophysical properties of our composition, at least six levels can be predicted: $0 \le x < 0.06$, $0.06 \le x < 0.11$, $0.11 \le x < 0.16$, $0.16 \le x < 0.29$, $0.29 \le x < 0.42$, and $0.42 \le x \le 1$.

For each of the levels, the physical properties depend on the cluster type (or types) that dominates, since the hierarchical level is unknown. In the range $0.29 \le x < 0.42$, the dependence $\alpha_R = f(T)$ is governed by the interaction between the clusters as the temperature rises. Individual fractals are connected cool ("blue") end to cool end, producing a monofractal largely with hot ("red") bonds, and the properties of the entire system are defined only by the carbon framework, whose properties approach those of pyrolytic graphite.

CONCLUSIONS

(1) The dependence of the resistivity on the graphite volume content and temperature in the graphite–paraffin system is considered.

(2) The numerical values of the percolation threshold $x_c = 0.16$, correlation radius index $\nu = 0.7$, and conductivity index $t \approx 1.4$ are obtained.

(3) Within the theory of fractal-structure objects, a model of a hierarchical subordinate system is proposed. The levels in this system correspond to the following ranges of graphite volume content: $0 \le x < 0.06$, $0.06 \le x < 0.11$, $0.11 \le x < 0.16$, $0.16 \le x < 0.287$, $0.287 \le x < 0.417$, $0.417 \le x \le 1$.

(4) The possibility of preparing a binary system whose electrophysical properties differ radically from those of its constituents is demonstrated experimentally. In particular, it is shown that one can obtain samples with an anomalously high temperature resistance coefficient varying over a wide range.

ACKNOWLEDGMENTS

The authors are grateful to G.V. Dedkov for valuable discussion.

REFERENCES

1. J. M. Ziman, *Models of Disorder: The Theoretical Physics of Homogeneously Disordered Systems* (Cambridge Univ. Press, Cambridge, 1979; Mir, Moscow, 1982).
2. B. Sh. Shklovskii and A. M. Éfros, *Electronic Properties of Doped Semiconductors* (Nauka, Moscow, 1979; Springer-Verlag, New York, 1984).
3. A. M. Éfros, *Physics and Geometry of Disorder* (Nauka, Moscow, 1982).
4. S. V. Karpenko, V. A. Sotskov, and M. V. Sotskov, in *Proceedings of Northern Caucasia Regional Conference "Prospects-2001," Nalchik, 2001*, Vol. 2, p. 121.
5. A. M. Petrov, Kh. V. Bal'yan, and A. T. Troshchenko, *Organic Chemistry* (Vysshaya Shkola, Moscow, 1963).
6. V. V. Pasynkov, *Materials for Electronics* (Vysshaya Shkola, Moscow, 1980).
7. *Tables of Physical Quantities: A Handbook*, Ed. by I. K. Kikoin (Atomizdat, Moscow, 1976).
8. A. I. Olemskoĭ and A. Ya. Fiat, Usp. Fiz. Nauk **163** (12), 3 (1993) [Phys. Usp. **36**, 1087 (1993)].
9. V. I. Roldurin and D. V. Voronin, Kolloidn. Zh. **63**, 394 (2001).
10. J. Feder, *Fractals* (Plenum, New York, 1988; Mir, Moscow, 1991).

Translated by M. Lebedev

Technical Physics, Vol. 48, No. 1, 2003, pp. 104–109. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 110–116.

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Stability of Subatomic Carbon Films on Metal Surfaces against Low-Energy Gas Ion Bombardment

A. L. Suvorov, V. P. Babaev, A. G. Zaluzhnyi, Yu. N. Devyatko, N. E. Lazarev, and A. A. Zaluzhnyi

State Scientific Center of the Russian Federation, Institute of Theoretical and Experimental Physics, Bol'shaya Cheremushkinskaya ul. 25, Moscow, 117218 Russia

e-mail: alex.suvorov@itep.ru

Received May 29, 2002

Abstract—The earlier developed original experimental technique for measuring and analyzing the parameters of low-frequency fluctuations of the field-emission current in metal film systems is used to measure the sputtering yield Y_f of carbon films (with a coverage Θ ranging from 1 to 4) applied on Fe, Nb, Ta, and U substrates. The value of Y_f is calculated by an expression derived within a theoretical model developed. The sputtering ratios were measured for the case when the carbon films are sputtered by H^+ and He^+ ions with an energy E_i between 2 and 10 keV. With Θ fixed, the energy dependences of Y_f are obtained for each of the ions. In addition, for each of the ions, the Θ dependences of Y_f are found for several values of E_i. In all the cases, the measured values of Y_f far exceed those for pure carbon. With another original technique that combines field-ion microscopy (FIM) and precise measurement of current and/or luminous properties of local regions in FIM images, the energy thresholds E_{th} of sputtering carbon films applied on the metal surfaces are found. The energy distributions of Y_f in the near-threshold energy range for various Θ are obtained. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

It is known that the sensitivity and accuracy of conventional techniques for measuring the yield Y of sputtering surface atoms by various particles [1, 2] decrease noticeably in going to the low-energy range. Available methods for theoretically estimating Y also give contradictory and unreliable values [3, 4]. Therefore, there is a great need to elaborate precise methods for measuring Y when particles bombarding the material surface have a subthreshold or near-threshold energy. In this work, we performed precise measurements of the sputtering yield Y for subatomic carbon films (with the coverage ranging from 1 to 4) covering the Fe, Nb, Ta, and U single-crystal surface. The sputtering particles were H^+ and He^+ ions with energies between 2 and 10 keV. Another aim was to find the energy threshold E_{th} of sputtering these films as a function of Θ. When choosing a film material, we proceeded from its importance as a structural material for various devices and instruments. The choice of the substrate material was associated largely with the specific feature of our original technique for calculating Y_f from experimentally measured parameters of low-frequency field-emission current fluctuations (LFFECFs) in metal film systems [5, 6].

Along with the our method of analyzing the LFFECF parameters, we also used another method developed by the authors, which combines FIM analysis with precise measurement of current and/or luminous properties of FIM images [7, 8]. With the latter, we determined the energy thresholds E_{th} of sputtering carbon films on the surfaces of the substrates selected and also measured the Y_f vs. energy for various Θ in the subthreshold energy range.

SPUTTERING OF CARBON FILMS BY H^+ AND He^+ IONS WITH ENERGIES OF 2–10 keV

As was already noted in the Introduction, the values of Y_f in this energy range were measured with an original technique for measuring the LFFECF parameters in metal film systems where the emitting surface is subjected to ion bombardment. In this work, this technique was implemented with an original device, a field-emission analyzing microscope [9]. The essence of the method is as follows. The surface of a metal tip was smoothed and cleaned (thermally or by the action of a field) and then covered by a thin film of a material for which the condition $\varphi_f > \varphi_s$ is met (φ_f is the work function of the film, φ_s is that of the substrate). Next, a negative potential V_0 that is sufficient for emitting electrons was applied to the tip (the electric field strength at the tip surface is on the order of 10^{10} V/m; V_0 equals several kilovolts). By placing a fluorescent screen (indicator) into the tip–electron detector gap opposite the detector, we set the desired crystallographic face (*hkl*) of the tip surface. Then, the screen is removed, a beam of bombarding ions is directed to the sample at a certain angle, and the continuous recording of the LFFECF parameters (a set of amplitudes $\Delta I_n(hkl)$) is accomplished.

1063-7842/03/4801-0104$24.00 © 2003 MAIK "Nauka/Interperiodica"

To explain: if a single impact of the ion cleans the selected area of the crystallographic face being analyzed, the field-emission current from this area grows sharply by a value of ΔI_n, which is the greater, the larger the difference is between φ_f and φ_s. Such an approach is by no means unique; others are discussed in [10].

In order for impact-cleaned tip areas to be covered again (to provide multiple measurements of ΔI_n, i.e., to improve statistics), complementary operations are used. First, the tip is maintained at a temperature that is sufficient for the effective surface migration of the film material; second, if possible, the additional evaporation of the film material from an external source is carried out ensuring that the coverage Θ remains invariable.

In the final calculation of Y_f, it is assumed that the areas of the metal surface that are cleaned by a single striking ion are circular with a radius r_i. Then, the volume removed is $\pi r_i^2\Theta$ and the number of atoms removed (the "intermediate" value of the sputtering yield) is $Y_f' = c' r_i^2\Theta/r_{af}^3$, where r_{af} is the atomic radius of the film material and c' is a dimensional factor. Using the Fowler–Nordheim equation for field electron emission [11], we get

$$\Delta I_n = \frac{c_1 V_0 r_i^2}{\beta R_{0,\,loc}}\left\{\frac{1}{\varphi_s(hkl)}\exp\left(-\frac{c_2\beta R_{0,\,loc}}{V_0}[\varphi_s(hkl)]^{3/2}\right)\right. \quad (1)$$
$$\left. -\frac{1}{\varphi_f(hkl)}\exp\left(-\frac{c_2\beta R_{0,\,loc}}{V_0}[\varphi_f(hkl)]^{3/2}\right)\right\}^{-1},$$

where $R_{0,\,loc}$ is the local radius of tip curvature in the area of the crystallographic face (hkl) being analyzed, $\varphi_s(hkl)$ is the work function of the face (hkl) of the pure metal substrate, $\varphi_f(hkl)$ is the work function of the same face covered by the film, c_1 and c_2 are dimensional constants, and β is a geometrical factor (β = 2–5).

Then, summing over all n measurements, we eventually obtain

$$Y_f = \frac{\beta\Theta}{c_1 n V_0 r}\sum_{(hkl)}^{n}\left\{\Delta I_n R_{0,\,loc}\left[\left[\frac{1}{\varphi_s(hkl)}\right.\right.\right.$$
$$\times\exp\left(-\frac{c_2\beta R_{0,\,loc}}{V_0}[\varphi_s(hkl)]^{3/2}\right) \quad (2)$$
$$\left.\left.\left. -\frac{1}{\varphi_f(nkl)}\exp\left(-\frac{c_2\beta R_{0,\,loc}}{V_0}[\varphi_f(hkl)]^{3/2}\right)\right]\right]\right\}.$$

Note that such an approach to measuring Y_f offers not only a much higher sensitivity but also opens up a qualitatively new possibility of finding Y_f in those cases when the film thickness is no more than several atomic layers (subatomic films).

It should be emphasized that the sputtering yield Y_f thus determined is somewhat overestimated: $Y_f = (Y_f)_{max}$. This is because Y_f, in the above procedure, is calculated based on those ion–substrate interaction events causing the cleaning of local substrate areas and generating anomalously high field-emission current fluctuations. If the number of such events per unit time is designated as ξ_n^* (i.e., $\xi_n^* = n$) and the total number of ion impacts on the same area is ξ_n^{**}, the experimental value of Y_f calculated by (2) is related to the actual value of Y_f^* as

$$Y_f^* = Y_f(\xi_n^*\xi_n^{**}) = a' Y_f, \quad (3)$$

where the factor α' allows for substrate sputtering without cleaning ($\alpha' < 1$).

In this work, the value of ξ_n^{**} was calculated from the irradiation parameters measured experimentally according to the ideas put forward in [12].

Table 1 lists the values of φ_s (taken from [13] and measured in this work) for four substrate materials and a carbon film. Note that the data for uranium were obtained by the authors in both cases [14]. The LFFECF parameters were measured in the central parts of the tips. Therefore, when estimating Y_f with (2), we employed data for ΔI_n found in this work for single crystals oriented in accordance with the (known) texture of wires used for tip preparation (see Table 1).

Carbon films were deposited on tips made of the four substrate materials by the *in vacuo* evaporation of a carbon target under the action of laser pulses. After each pulse, the total field-emission current I_{fe} from the tip was measured at a fixed tip potential. As the film surface coverage grows ($\Theta < 1$), the value of I_{fe} decreases. The saturation of I_{fe} was taken as an indication of the deposition of one monoatomic carbon layer (Θ = 1). The doubling of the number of laser pulses corresponded to Θ = 2, etc.

The irradiation of the tips by H^+ and He^+ ions was carried out directly in the field-emission analyzing microscope with a special source. The ions were incident on the tip surface at nearly a right angle.

Some of the experimental data obtained in this work are illustrated in Figs. 1–6. Figures 1 and 2 demonstrate the energy dependences of the sputtering yield Y_f for H^+ bombarding ions; Figs. 3–6 show the same for He^+ ions. The values of Y_f were calculated by Eq. (2) with regard for α' (i.e., the final calculation was performed with Eq. (3)).

Prior to discussing the results obtained, several points need to be made. First, we call attention once again to the fact that the values of Y_f found with the above procedure are to some extent conventional. Second, it should be emphasized that such a procedure for

Table 1. Work functions φ of the substrate and film materials for polycrystals and single crystals (published data and our measurements)

Substrate material	Published values of φ for single crystals, eV [13]	Published values of φ for polycrystals measured by the field-emission method, eV [13]	Published values of φ for variously oriented single crystals (*hkl*), eV [13]	Values of φ for polycrystals determined in this work, eV	Values of φ for single crystals determined in this work, eV
Fe	4.31	4.85	4.31–4.81 (111)	4.20	4.50 (111)
Nb	3.99	No	4.09–4.80 (110)	3.90	4.10 (110)
Ta	4.12	4.12	4.95 (110)	4.20	4.50 (110)
U	3.30	3.50 [14]	3.40 [14] (100)	3.50	3.40 (100)
C	4.70	No	4.70–4.80 (100)	4.80	4.80 (100)

calculating Y_f is valid only if a bombarding ion penetrates the film, i.e., cleans some substrate area. In other words, starting from a certain value of $\Theta \approx \Theta^* + 1$, LFFECFs disappear and Y_f cannot be measured. Finally, the shape of a crater made by a bombarding ion (at $\Theta \geq 2$) may be other than cylindrical (the numbers of atoms knocked out of the upper and lower layers of the film differ). Therefore, the value of Y_f determined from Eq. (2) turns out to be underestimated. Nevertheless, Y_f obtained with this technique reflect major trends in sputtering subatomic films by atoms of various energy.

Let us discuss our results. As follows from Figs. 1–6, with the ion energies chosen, both H^+ and He^+ ions clean certain areas of the substrates at $\Theta = 1$. However, even at $\Theta = 2$, H^+ ions with an energy of 2 keV almost do not clean the Nb, Ta, and U substrates; for $\Theta = 3$, this is true for ion energies of 2 and 4 keV. For H^+ ions, the coverage $\Theta = \Theta^* = 3$ is limiting for substrate cleaning in this energy range; for He^+ ions, the limiting value is $\Theta^* = 4$.

From Figs. 1–6, it also follows that when the carbon films are sputtered by H^+ ions, Y_f is roughly one order of magnitude higher than Y for pure carbon; for He^+ ions, this difference is less significant [2]. Our data also indicate that the difference in Y_f when the films are sputtered by H^+ and He^+ ions of the same energy is smaller than the difference observed upon sputtering pure carbon. Finally, there is a distinct tendency toward increasing the values of Y_f from Nb and Ta to U and Fe (under the same experimental conditions). Presumably, the conditions for carbon film deposition on the different substrates differ. However, the formation of carbides seems to be unlikely in any of the cases.

CARBON FILM SPUTTERING BY H^+ AND He^+ IONS IN THE NEAR-THRESHOLD ENERGY RANGE

As was mentioned in the Introduction, our original technique allows for the determination of energy thresholds E_{th} for sputtering subatomic carbon films of various thickness applied on metal surfaces. Also, it

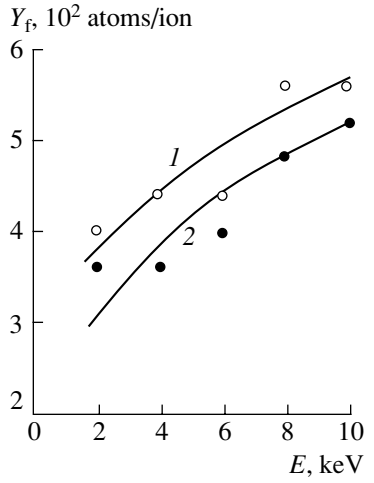

Fig. 1. Energy dependence of the sputtering yield Y_f for carbon films on the (o) iron and (●) uranium surfaces in the case of H^+ bombardment. $\Theta = 1$ in both cases.

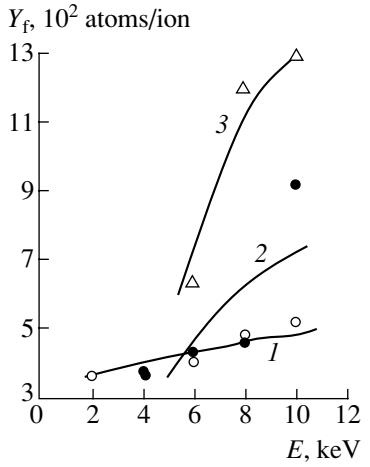

Fig. 2. Energy dependence of the sputtering yield Y_f for carbon films on the uranium surface in the case of H^+ bombardment. Figures by the curves indicate the values of Θ.

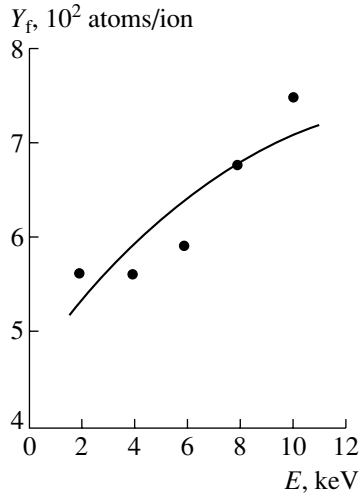

Fig. 3. Energy dependence of the sputtering yield Y_f for carbon films on the iron surface in the case of He^+ bombardment. $\Theta = 1$.

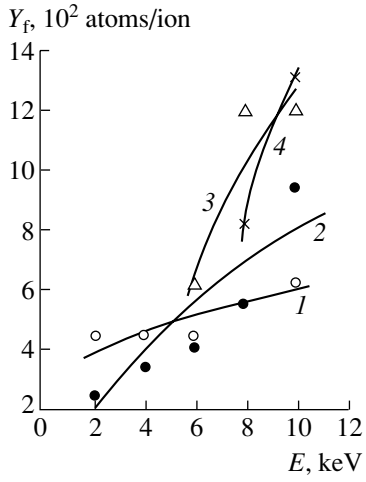

Fig. 4. Energy dependence of the sputtering yield Y_f for carbon films on the niobium surface in the case of He^+ bombardment. Figures by the curves indicate the values of Θ.

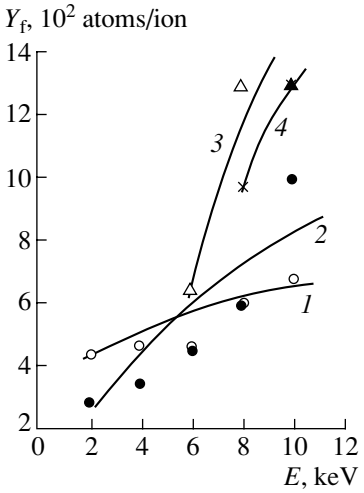

Fig. 5. Energy dependence of the sputtering yield Y_f for carbon films on the tantalum surface in the case of He^+ bombardment. Figures by the curves indicate the values of Θ.

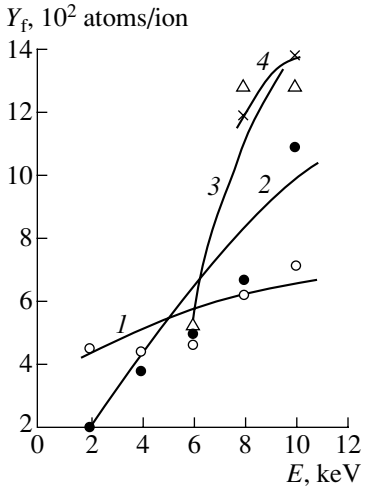

Fig. 6. Energy dependence of the sputtering yield Y_f for carbon films on the uranium surface in the case of He^+ bombardment. Figures by the curves indicate the values of Θ.

makes it possible to take the energy dependences of Y_f at near-threshold energies. The essence of the technique and its implementation are detailed in [8, 15]. Briefly, the idea is as follows. In an ion-field microscope, the variation of the imaging ion current to the microchannel plate and the variation of the FIM image contrast on the screen are measured. The microchannel plate, as is usually done in FIM, enhances the luminance of the images. The ion current and luminance (more exactly, the luminous flux from either the entire fluorescent screen of the microscope or a local area) of the images were measured each time after the surface had been exposed to imaging ion (H^+ or He^+) pulses. The number of atoms removed from the surface (i.e., Y_f) was calculated by estimating the imaging ion current per atom and luminous flux from the associated image. As a result, we succeeded in obtaining the sputtering yield Y_f even for materials that do not give atomic-resolution FIM images. Carboniferous materials are known to be among them [16, 17]. It should be noted that the technique was tested and refined with tungsten. The values of Y thus obtained were corrected by comparing with the values found by the direct calculation of the number of surface vacancies on the tungsten images resolved on an atomic level [7, 8].

We used an all-metal field-emission microscope described in [18]. Pulses between the tip (cathode) and microchannel plate (anode) were generated using the grid of a metal-ceramic cathodic unit. The unit was

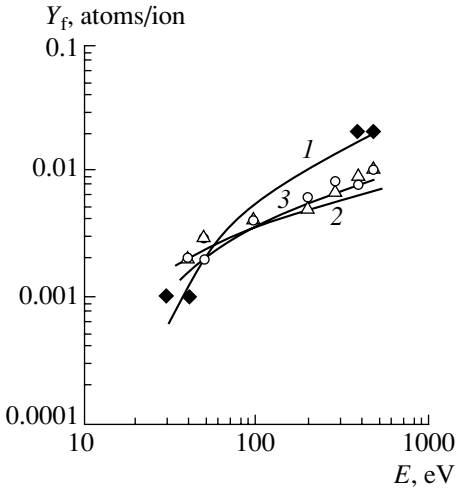

Fig. 7. Energy dependence of the sputtering yield Y_f for carbon films on the uranium surface in the case of H^+ bombardment with near-threshold energies. Figures by the curves indicate the values of Θ.

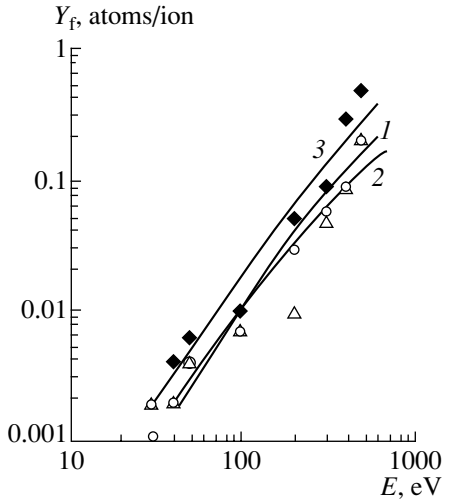

Fig. 8. The same as in Fig. 7 in the case of He^+ bombardment.

placed between the sample and screen only during the generation of individual irradiation pulses.

The samples were irradiated by an original technique of pulsed two-step change of the high voltage polarity. The bombarding (and simultaneously imaging) gas was ionized by electronic impact at the instant a short high-voltage pulse of reverse (negative) polarity was applied and electrons were generated by the tip. The duration of such ion-producing pulses was no longer than 0.1 μs, and their amplitude was equal to 5 kV. Along with the high-voltage pulse, a low-voltage negative pulse was applied to the sample. Its duration was $\Delta\tau \geq 10$ μs, and its amplitude ΔV, which provided the desired energy of bombarding ions, was varied from 10 to 500 V in 10 V intervals. In the course of this work, many control experiments using the diode and triode schemes with various sequences and parameters of pulses were performed (in the latter scheme, the grid of the cathodic unit mentioned above was employed).

An important stage in evaluating the sputtering yield Y_f is the determination of the irradiation fluence. It should be stressed that this problem is by no means simple. It is this problem (and also the estimation of the bombarding ion energy and the number of irradiation-induced single surface vacancies) that makes the values of Y_f obtained by the technique described somewhat unreliable. As was indicated above, the irradiation fluence in this work was estimated from the irradiation parameters measured experimentally [12].

Some of the experimental results obtained in this work with the technique described above are illustrated in Figs. 7 and 8. Table 2 summarizes experimentally found threshold energies E_{th} of sputtering carbon atoms by H^+ and He^+ ions for all the substrates and various coverages Θ.

As follows from the data tabulated, the values of E_{th} roughly correlate with Y_f determined from the LFFECF parameters: they decline from Nb and Ta to Fe and U. If these values of E_{th} are compared with those obtained by the same technique for pure carbon [15], one can conclude that the film values of E_{th} are lower. Moreover, it is seen that the values of Y_f at near-threshold energies differ markedly for Θ = 1 and Θ = 2 but are very close for Θ = 2 and Θ = 3; that is, the effect of the substrates is substantially reduced as Θ increases.

Table 2. Experimental threshold energies E_{th} (eV) of sputtering carbon films applied on various substrates

Bombarding ions	Θ	Substrate material			
		Fe	Nb	Ta	U
H^+	1	30	50	40	30
	2	30	50	40	40
	3	30	40	40	40
He^+	1	20	40	30	30
	2	30	40	40	30
	3	30	40	40	30

CONCLUSIONS

The conclusions drawn from the results of this work are the following.

The original techniques for measuring the sputtering yield in metal film systems and pure conducting materials make it possible to study the sputtering process on subatomic films ($\Theta \geq 1$) at low (including subthreshold and near-threshold) energies of bombarding particles. These techniques offer a high sensitivity but

sometimes lack quantitative reliability. However, they are undoubtedly promising, and call for further refinement.

In the energy range 2–10 keV of bombarding ions, the film thickness $\Theta = \Theta^* = 3$ is limiting for surface cleaning by H^+ ions; for He^+ ions, the limiting value is $\Theta^* = 4$.

The sputtering yield measured upon H^+ ion bombardment of the carbon films is roughly one order of magnitude higher than for pure carbon. For He^+ ions, this difference is smaller. The difference in Y_f when the carbon films are exposed to H^+ and He^+ ions of the same energy is smaller than the corresponding difference for pure carbon.

With all other things being equal, the sputtering yield Y_f grows from Nb and Ta to U and Fe.

At near-threshold energies of the ions, the energy dependences of Y_f for the carbon films differ noticeably at $\Theta \approx 1$ and $\Theta \geq 2$ for both H^+ and He^+ ions.

The energy thresholds E_{th} of sputtering the carbon films by H^+ and He^+ ions decrease from Nb and Ta to Fe and U. For metal film systems, they are lower than for pure carbon in all cases.

The energy thresholds at $\Theta \approx 1$ and $\Theta \geq 2$ differ, the difference being greater in the case of He^+ ions.

ACKNOWLEDGMENTS

This work was partially supported by the Russian Foundation for Basic Research (grant no. 01-02-17934).

REFERENCES

1. *Sputtering by Particle Bombardment*, Ed. by R. Behrisch (Springer-Verlag, New York, 1981, 1983; Mir, Moscow, 1984, 1986), Vols. I and II.
2. W. Ekstein, C. Garcia-Rosales, J. Roth, *et al.*, *Sputtering Data*, Report IPP 9/82 (Max-Planck Institut fur Plasmaphysik, 1993).
3. O. S. Oen and M. T. Robinson, Nucl. Instrum. Methods **132**, 647 (1976).
4. J. P. Biersack and W. Eckstein, Appl. Phys. A **34**, 73 (1984).
5. A. L. Suvorov, E. P. Sheshin, D. E. Dolin, *et al.*, Appl. Surf. Sci. **76/77**, 26 (1994).
6. A. L. Suvorov, V. I. Kvintradze, A. F. Bobkov, *et al.*, Zh. Tekh. Fiz. **58**, 1134 (1988) [Sov. Phys. Tech. Phys. **33**, 665 (1988)].
7. M. I. Guseva, A. L. Suvorov, S. N. Korshunov, *et al.*, J. Nucl. Mater. **266–269**, 222 (1999).
8. M. I. Guseva, A. L. Suvorov, S. N. Korshunov, *et al.*, Zh. Tekh. Fiz. **69** (9), 137 (1999) [Tech. Phys. **44**, 1123 (1999)].
9. V. A. Kasatkin, V. I. Kvintradze, and A. L. Suvorov, Prib. Tekh. Éksp., No. 1, 191 (1987).
10. A. L. Suvorov, *Structure and Properties of Surface Atomic Layers of Metals* (Énergoatomizdat, Moscow, 1990).
11. M. I. Elinson and G. F. Vasil'ev, *Field Emission* (Fizmatgiz, Moscow, 1958).
12. P. A. Bereznyak and V. V. Slezov, Radiotekh. Élektron. (Moscow) **17**, 354 (1972).
13. V. S. Fomenko, *Emission Properties of Materials: A Handbook* (Naukova Dumka, Kiev, 1981).
14. A. L. Suvorov, At. Énerg. **42**, 280 (1977).
15. A. L. Suvorov, M. I. Guseva, S. N. Korshunov, *et al.*, in *Proceedings of the Conference "Physics of Radiation Effects and Radiation Material Science"* (NNTs KhFTI, Kharkov, 2001), Vol. 1, p. 78.
16. D. E. Dolin, A. A. Sosunov, A. L. Suvorov, *et al.*, Zh. Tekh. Fiz. **60** (12), 115 (1990) [Sov. Phys. Tech. Phys. **35**, 1430 (1990)].
17. A. L. Suvorov, Yu. N. Cheblukov, N. E. Lazarev, *et al.*, Adv. Sci. Technol. (Faenza, Italy) **19**, 191 (1999).
18. A. F. Bobkov, S. V. Zaĭtsev, V. A. Kasatkin, *et al.*, Prib. Tekh. Éksp., No. 6, 188 (1985).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 1, 2003, pp. 110–114. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 117–122.

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Effect of Various Components of Slightly Diluted Tungsten-Based Alloys on the Formation of Defects in Atomic Displacement Cascade Regions: Field-Ion Microscopic Analysis

A. L. Suvorov*, A. G. Zaluzhnyi*, A. F. Bobkov*, S. V. Zaitsev*, V. P. Babaev*, M. I. Guseva, S. N. Korshunov**, I. N. Nikolaeva*, and A. A. Zaluzhnyi***

** State Scientific Center of the Russian Federation, Institute of Theoretical and Experimental Physics, Bol'shaya Cheremushkinskaya ul. 25, Moscow, 117218 Russia*

e-mail: alex.suvorov@itep.ru

*** Russian Research Centre Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia*

Received May 29, 2002

Abstract—New complex field-ion microscopic data for the formation of radiation-induced defects in VChV ultrapure tungsten, VA-3 commercial-grade tungsten, and four slightly diluted tungsten-based alloys (W–Hf–C, P39A; W–1.5% ThO_2, VT-15; W–5% Re, VR-5; and W–2% Fe, VZh-2) are reported. Samples were irradiated in an external unit by Ar^+ and Ni^+ ions of energy 35 keV. In the experiments, the ion current is kept at j = 2.0 μA and the irradiation fluence equals $\Phi t = 5 \times 10^{14}$ ions/cm^2. The clustering of single vacancies in samples irradiated is studied in relation to the impurity concentration and type. The distribution of vacancy clusters over the number of aggregated single vacancies is studied. These distributions are found to differ noticeably inside and outside the depletion zones. The mean lengths of focused substitutional atomic collision chains in samples with different impurity concentrations and types are measured indirectly. From these data, the efficiency of trapping intrinsic interstitials by various impurities in tungsten is estimated. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The potentialities of field-ion microscopy (FIM) as a tool for studying various radiation-induced defects in conducting materials have by no means been exhausted in spite of the huge body of available data [1, 2]. This method provides a unique possibility of directly observing individual point defects; identifying the atomic structure of other lattice defects; analyzing the distribution of the defects in the volume; and, finally, performing (using an atomic probe [3]) elemental mass-spectrometric analysis of single particles (atoms and molecules) and their complexes both observed directly and selected on micrographs.

The purpose of this work is to elucidate how a number of impurities influence the radiation resistance of tungsten upon cascade-forming irradiation. This work is a logical elaboration of early studies carried out by the authors [4–6].

EXPERIMENTAL

The objects studied were VChV ultrapure single-crystal (100)W (its properties were described in [7]), VA-3 commercial-grade tungsten, and four slightly diluted tungsten based alloys (W–1.5% ThO_2, VT-15; W–5% Re, VR-5; W–2% Fe, VZh-2; and W–HfC–W_2C, P39A). The last-listed alloy was prepared by the powder metallurgy method where carbon served as a deoxidizer: it was added to remove oxygen and form tungsten carbide. The chemical composition of all the samples was analyzed with a SuperProbe-733 electron probe instrument (JEOL, Japan). In all the samples, the basic impurities (other than those given in specifications) were Ni (<0.02%), Al (0.02–0.06%), Mo (<0.05%), Fe (<0.02%), Si (0.02–0.03%), Nb (<0.05%; absent in VA-3 commercial-grade tungsten), Mn (<0.02%), Ti (<0.02%), Cu (<0.02%), Mg (<0.02%), and V (<0.05%).

All the samples were tips prepared by the conventional electrochemical etching of wires (except for the samples of ultrapure tungsten; they were made of small rectangular bars). They were preexamined in a field-ion microscope and then transferred in air to the irradiator—an ILU-4 ion-beam setup designed in the Institute of Nuclear Fusion at the Russian Scientific Centre Kurchatov Institute. The samples were irradiated by 35-keV Ar^+ and Ni^+ ions. The ion current and the irradiation fluence were $j \approx 2.0$ μA and $\Phi t \approx 5 \times 10^{14}$ ions/cm^2 in all the cases. The ion beam was directed normally to the tip axis. The experiments were carried out at room temperature.

The field-ion microscopic analysis of the samples involved two standard steps: (1) cleaning and polishing the tip surface by field desorption and evaporation and (2) imaging and digitization of the images from several successive surface atomic layers. Then, the samples

were removed from the microscope, mounted in a special cooled support, and placed into the irradiator. The irradiated samples were placed into the microscope again, and no less than one hundred atomic layers were imaged sequentially under the condition of controllable field evaporation. In all the cases, the image gas was helium at a pressure of 10^{-2} Pa. The sample temperature was 77 K.

RESULTS AND DISCUSSION

In all the samples irradiated, the primary defects were single vacancies and their clusters of low multiplicity, as well as depletion zones, which arise because of atomic displacement cascades in the crystal lattice. The defects were observed on the side of ion beam incidence within a depth of ≈20 nm. In many cases, the high quality of the FIM images (see, e.g., Fig. 1) made it possible to determine the depletion zone parameters (the number of constituent vacancies, vacancy concentration, and volume occupied by vacancies) with a fairly good accuracy (±5–10%). We sometimes succeeded in determining these parameters even when the samples failed under the action of ponderomotive forces of the electric field during the layer evaporation (Fig. 2). Note that the fraction of failed samples was about 10% and the decrease in the mechanical strength of irradiated samples in comparison with nonirradiated ones was estimated at 15–20%. The mean values of the depletion zone parameters for all the materials studied are listed in Table 1. It should be emphasized that the quantity ν^* can be called the cascade function only conventionally; strictly speaking, it corresponds to the total number of vacancies that "survived" at the site of cascade development. In other words, the true value ν of the cascade function is always less than ν^*. Note also that the values listed in Table 1 are averaged for the two types of bombarding ions. This is largely due to the fact that, for the ions used in the experiments, we did not reveal an appreciable spread in the depletion zone parameters measured. Moreover, the maximal transferred energies were rather close (20.53 and 25.68 keV for Ar^+ and Ni^+ ions, respectively). Finally, it should be noted that the field evaporation sometimes removed the material from the depletion zones, turning them into artefact pores. Such cases were rejected from the analysis of the depletion zone parameters.

As follows from Table 1, the depletion zone parameters (hence, the associated parameters of atomic displacement cascades) are close for VChV ultrapure tungsten, VA-3 commercial-grade tungsten, and VR-5 alloy. For the other alloys, they differ substantially. These differences can be explained in large measure within the concept according to which focused substitutional displacement cascade chains appear in the crystal lattice [8]. In this work, following the method first suggested in [9], we measured the mean lengths $\bar{l}_{ch}$ of the chains in two basic crystallographic directions, ⟨110⟩ and ⟨111⟩. This length was also determined from the spacing between an interstitial and the nearest vacancy irrespective of the crystallographic direction. The associated results are summarized in Table 2. Because of the specific features of the measurements (the distance from each of the identified vacancies to the nearest interstitial atom–impurity atom complex was determined from many FIM images both along the two crystallographic directions and in any direction), the chain lengths in the ultrapure tungsten were not found. In other words, the lengths were estimated for chains interrupted by impurities, because only dumb-

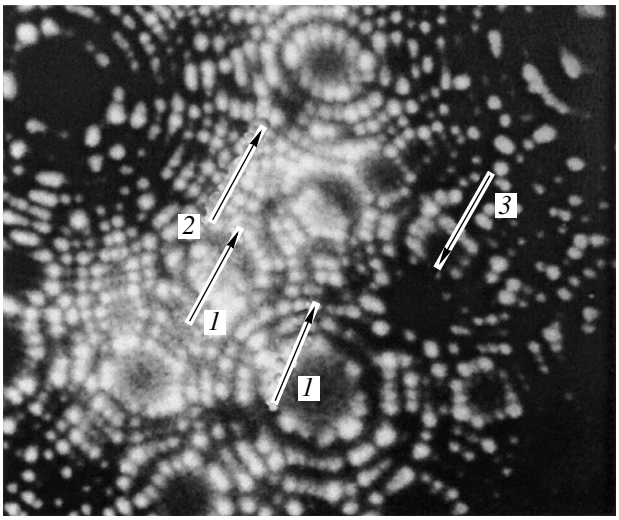

Fig. 1. FIM image from the surface of VA-3 commercial-grade tungsten irradiated by Ni^+ ions of energy 35 keV. (*1*) Single vacancies, (*2*) trivacancy, and (*3*) depleted zone etched by field evaporation.

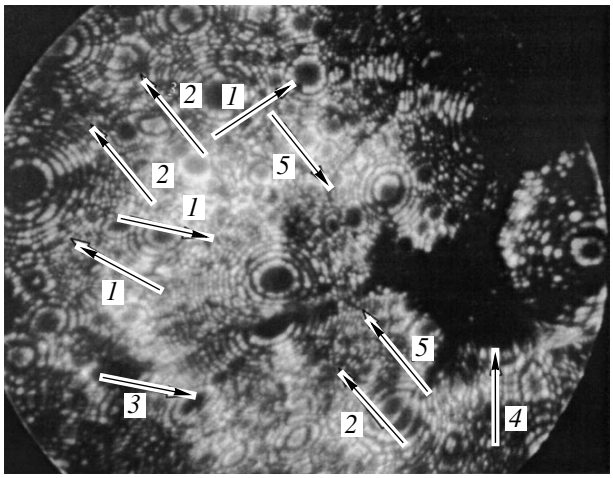

Fig. 2. FIM image from the surface of W–5% Re (VR-5) alloy irradiated by Ni^+ ions of energy 35 keV and deformed (partial failure) by the ponderomotive forces of the electric field. (*1*) Single vacancies, (*2*) vacancy clusters of small multiplicity, (*3*) dislocations, (*4*) large craters, and (*5*) triple boundary between grains.

Table 1

Material	VChV	VA-3	VR-5	VT-15	VZh-2	P39A
Cascade function ν*	180	160	150	130	120	140
Vacancy concentration C_v, %	>20	20	20	<15	<15	20
Zone volume, Ω_a	900	800	750	870	800	700

Table 2

Material	VA-3	VR-5	VT-15	VZh-2	P39A
Mean chain length $\dot{l}_{ch}$ along the ⟨110⟩ direction, nm	4.0	4.0	2.5	2.3	3.6
Mean chain length $\dot{l}_{ch}$ along the ⟨111⟩ direction, nm	3.8	3.7	2.2	2.0	3.4
Mean chain length $\dot{l}_{ch}$ calculated from measured interstitial atom–nearest vacancy spacing, nm	3.0	3.2	1.9	1.9	2.5

bell-like complexes consisting of an intrinsic interstitial and impurity atom could be observed in FIM images [10].

As follows from Table 2, the values of l_{ch} for VA-3 and VR-5 alloys are close, exceed noticeably that for P39A alloy, and are much greater than for VT-15 and VZh-2 alloys. It is quite natural to suppose that, if the chains are short, the probability of recombination between vacancies and nearest complexes greatly increases both inside and outside cascade regions. The first effect may reduce the vacancy concentration inside a depletion zone (near its edges); the second may shrink the depletion zone volume (size). Since these considerations are to a great extent speculative, it seems to be reasonable to make a quantitative estimation by means of computer simulation [11]. Of great interest also is bringing to end the measurement of the length l_{ch} of focused substitutional displacement chains in tungsten and its alloys, i.e., the construction of the statistical chain length distributions along the basic crystallographic directions. Note that much of such a task has already been solved in the course of this work. Figures 3 and 4 illustrate the statistical distributions of l_{ch} for P39A and VZh-2 alloy, respectively. It is also clear that direct measurements of spacings between single vacancies and interstitials in the case when the irradiation and study of the samples are carried out at temperatures below the migration temperature of far interstitials (≈28 K for tungsten) would be of much higher value [12].

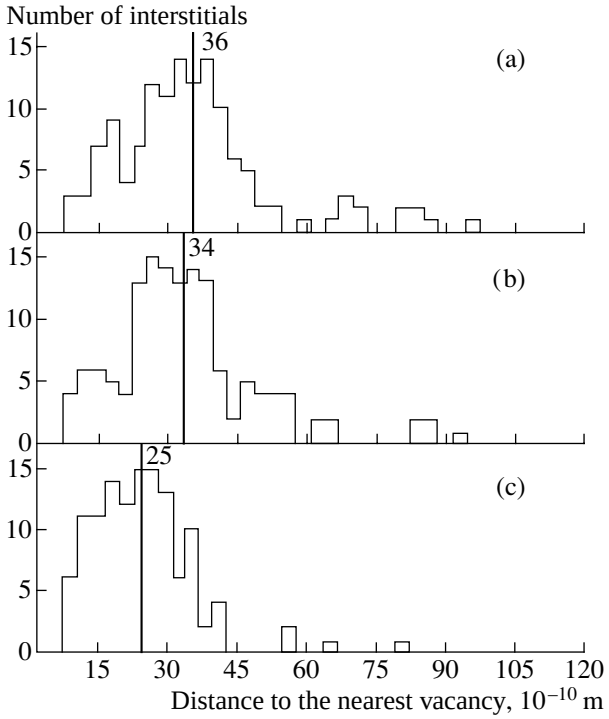

Fig. 3. Distributions of focused substitutional atomic collisions over chain length l_{ch} in W–HfC–C (P39A) alloy irradiated by 35-keV Ar^+ or Ni^+ ions along the (a) ⟨110⟩ and (b) ⟨111⟩ directions (FIM data). (c) Calculation from the measured distances between an interstitial and the nearest vacancy.

In addition, Table 2 suggests that, among the impurities, Fe atoms (the shortest chains) have the highest efficiency of trapping (terminating) chains of focused substitutional collisions in the tungsten alloys under study. Next in order of efficiency are Th, Hf, and Re atoms (or complexes). Alkali metal atoms in commercial-grade tungsten appear to have the lowest efficiency. All the interstitial atom–impurity atom complexes are stable at room temperature and start to disintegrate (or move remaining intact [13]) at temperatures of several hundreds of degrees centigrade.

In [14, 15], the idea of measuring the binding energy E_{io}^{b} between intrinsic interstitial atoms and impurity atoms, as well as the migration energy E_{io}^{m} of such complexes, was put forward. The procedure consists in constructing and theoretically processing the tempera-

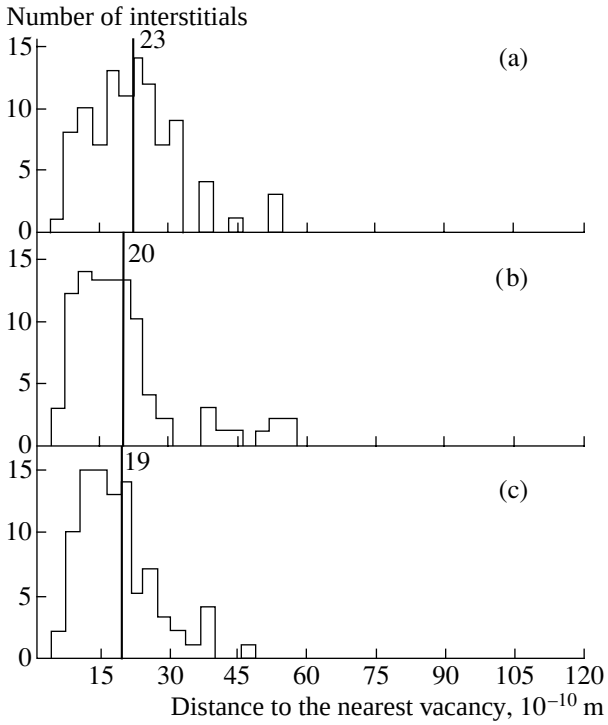

Fig. 4. The same as in Fig. 3 for W–2% Fe (VZh-2) alloy.

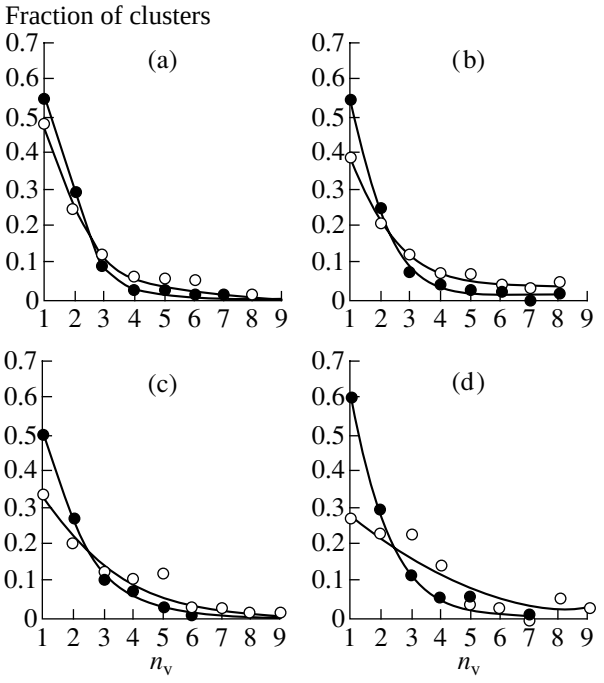

Fig. 5. Distributions of vacancy clusters over the multiplicity n_v in (a) VChV ultrapure tungsten, (b) VA-3 commercial-grade tungsten, (c) W–HfC–C (P39A) alloy, and (d) W–2% Fe (VZh-2) alloy. (●) Distribution in the entire sample volume and (o) distribution in the depletion zone volume.

ture dependence of the number of the complexes observed on the surface of an FIM specimen during isochronous step annealing. Such measurements of E_{io}^{b} and E_{io}^{m} as applied to the objects studied in this work appear to be a very fruitful approach, which will be used in further investigation.

Finally, it should be stressed that the distribution of vacancy clusters over their multiplicity n_v in relation to the impurity type and concentration and to the depletion zone parameters might help to illuminate the formation and evolution of defects both inside and outside atomic displacement cascade regions [16]. The initial attempt to gain such information (for irradiated specimens of commercial-grade tungsten) was made in [5]. In this work, we obtained the cluster multiplicity distribution in irradiated samples of VChV ultrapure tungsten, VA-3 commercial-grade tungsten (fresh data), and P39A and VT-15 alloys (Fig. 5). The difference between the distributions inside and outside the depletion zones stands out. It follows from these distributions that they correlate distinctly with the lengths of focused atomic collision chains in the depletion zone (Table 2). As $\dot{l}_{ch}$ shortens, the fraction of vacancy clusters in the depletion zone markedly increases, while that of single vacancies and divacancies substantially drops. At the same time, away from the depletion zones, the distribution of vacancy clusters over the multiplicity n_v are very similar for all the materials.

As was mentioned above, the termination of focused substitutional collision chains by impurities (or, more generally, the very existence of the chains and their finite length) markedly affects the parameters of the depletion zones, which are formed at the sites where atomic displacement cascades develop. The influence is twofold. On the one hand, the farther an interstitial is from a vacancy when a Frenkel pair forms (i.e., the larger $\dot{l}_{ch}$ is), the lower the probability of its returning to this vacancy (i.e., the lower the probability of point defect annihilation in the depletion zone). Note however that this in no way means an increase in the vacancy concentration in the depletion zone when interstitial atom–impurity atom stable complexes form, because these complexes may coexist with vacancies in the depletion zones. In pure metals, this is impossible: because of the high mobility of interstitials even at cryogenic temperatures, single vacancies and their clusters of various multiplicity are only present in the depletion zone. On the other hand, the larger $\dot{l}_{ch}$ is, the farther from the depletion zone edges forms a "coat" of impurity-stabilized intrinsic interstitials, i.e., the lower the probability that this coat will absorb the depletion zone edges (the probability of this absorption is responsible for a decrease in the cascade function ν^* and depletion zone volume V_z in the case of short chains).

It is seen that Tables 1 and 2 support the above considerations indirectly. Moreover, analyzing the distributions in Fig. 5 from these standpoints, one can conclude that, as the chains shorten ($\bar{l}_{ch}$ decreases), largely single vacancies annihilate in the depletion zone: their fraction in the cluster multiplicity distributions diminishes. Therefore, the construction of the vacancy distributions over clusters as a function of the irradiation temperature and post-irradiation annealing temperature seems to be a very important problem.

CONCLUSION

This work reports the second part of results obtained in the course of our complex investigations. Data presented here suggest a correlation between the parameters of depletion zones (including vacancy distributions over clusters in them), which form at the sites where atomic displacement cascades develop, and the length of focused substitutional collision chains (which is a somewhat conventional parameter in the context of this paper). In addition, our data illustrate the efficiency of terminating the chains by specific impurities present in tungsten, i.e., allow one to estimate the effect of impurities on the depletion zone parameters and, in the general case, on the parameters of defect structures in slightly diluted metal alloys.

REFERENCES

1. D. N. Seidman, Surf. Sci. **70**, 532 (1978).
2. A. L. Suvorov, *Field-Ion Microscopy of Radiation Defects in Metals* (Énergoatomizdat, Moscow, 1982).
3. M. K. Miller and G. D. W. Smith, *Atom Probe Microanalysis: Principles and Applications to Material Problems* (Materials Research Society, Pittsburgh, 1989; Mir, Moscow, 1993).
4. A. F. Bobkov, M. I. Guseva, S. V. Zaitsev, *et al.*, in *Abstracts of the Third International Ural Seminar "Radiation Damage Physics of Metals and Alloys," Snezhinsk, Russia, 1999*, p. 21.
5. A. L. Suvorov, V. P. Babaev, A. F. Bobkov, *et al.*, in *Proceedings of 11th Conference on Radiation Physics and Chemistry of Condensed Matter, Tomsk, Russia, 2000*, Vol. 1, p. 273.
6. A. L. Suvorov, A. G. Zaluzhnyi, A. F. Bobkov, *et al.*, in *Proceedings of 47th International Field Emission Symposium, Berlin, 2001*, IP-13.
7. V. V. Marchenko, V. E. Arkhipov, A. L. Suvorov, *et al.*, Fiz. Met. Metalloved. **88** (1), 41 (1999).
8. V. V. Kirsanov, A. L. Suvorov, and Yu. V. Trushin, in *Processes of Radiation-Induced Defect Formation in Metals* (Énergoatomizdat, Moscow, 1985).
9. A. L. Suvorov, Vopr. At. Nauki Tekh., Ser.: Fiz. Radiats. Povrezhdeniĭ Radiats. Materialoved., No. 1 (15), 39 (1981).
10. A. L. Suvorov and D. E. Dolin, Appl. Surf. Sci. **94/95**, 384 (1996).
11. V. V. Kirsanov, *Computer Experiment in Atomic Material Science* (Énergoatomizdat, Moscow, 1990).
12. R. M. Scanlan, D. L. Styris, and D. N. Seidman, Philos. Mag. **23**, 1439 (1971).
13. A. L. Suvorov, *Structure and Properties of Surface Atomic Layers of Metals* (Énergoatomizdat, Moscow, 1990).
14. Yu. V. Trushin, A. L. Suvorov, D. E. Dolin, *et al.*, Pis'ma Zh. Tekh. Fiz. **16** (17), 81 (1990) [Sov. Tech. Phys. Lett. **16**, 676 (1990)].
15. A. L. Suvorov, Yu. V. Trushin, D. E. Dolin, *et al.*, Mater. Sci. Forum **97–99**, 217 (1992).
16. A. I. Melker, AIP Conf. Proc. **303**, 156 (1994).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 1, 2003, pp. 115–117. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 123–126.

BRIEF COMMUNICATIONS

Fabrication of Holographic Diffraction Gratings Based on As_2S_3 Layers

S. A. Kostyukevych*, P. E. Shepeliavyi*, P. F. Romanenko*, and I. V. Tverdochleb**

** Institute of Semiconductor Physics, National Academy of Sciences of Ukraine, pr. Nauki 45, Kiev, 03028 Ukraine*

*** Shevchenko National University, Kiev, Vladimirskaya ul. 64, Kiev, 03125 Ukraine*

e-mail: spe@isp.kiev.ua

Received January 28, 2002

Abstract—The formation of holographic diffraction gratings based on As_2S_3 layers is investigated. The variation of the groove profiles with exposure is studied by atomic force microscopy. The spectral curves of the diffraction efficiency are taken, and a relationship between these curves and grating surface relief is analyzed. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Inorganic photoresists (e.g., amorphous layers of a chalcogenide glass (ChG)) are widely used in the production of holographic diffraction gratings (HDGs) [1]. Typical ChGs are arsenic sulfide (As_2S_3) and selenide (As_2Se_3). The use of ChGs as high-resolution inorganic resists is based on the photostimulated variation of the solubility of ChG films prepared by vacuum evaporation in selective etchants [2]. In [3–6], HDGs were made of As_2Se_3 layers, which offer a reasonably high sensitivity in a wide spectral interval including the UV, visible, and IR ranges. The HDG formation in As_2S_3 layers is as yet little understood [7]. However, these layers continue to be of interest because of their advantages, for example, in HDG recording by a helium–cadmium (He–Cd) laser. The quality of the postexposure selective etching of a ChG substantially influences the formation of the HDG relief. Earlier, the authors developed a selective nonaqueous organic etchant for the precision etching of negative inorganic As_2S_3-based resists [8]. The purpose of this work was to investigate the HDG formation in As_2S_3 layers using the etchant developed. In particular, we studied the exposure dependence of the diffraction efficiency (DE) and HDG groove profile. The results obtained enable the optimization of the exposure regime for As_2S_3 layers when manufacturing high-efficiency sinusoidal-groove HDGs.

EXPERIMENTAL

A test specimen for HDG recording was prepared by the sequential vacuum evaporation (at a residual pressure of 2×10^{-3} Pa) of Cr and As_2S_3 layers with thicknesses of 80 and 1000 nm, respectively, onto an optically finished glass substrate. The As_2S_3 layer was exposed in a holographic setup employing the method of light-wave amplitude division. An interference pattern with a spatial frequency of 1200 mm^{-1} was obtained by using a He–Cd laser with a wavelength $\lambda = 0.441$ μm. For various areas of the photoresist layer, the exposure H was 20, 40, 80, 150, and 300 mJ/cm^2. After the exposure, we performed chemical processing in the nonaqueous organic etchant [8], providing the selective dissolution of unexposed As_2S_3 areas (negative etching). Such a processing forms a phase grating whose surface profile depends on the exposure distribution in the photoresist. By varying the exposure, one can make symmetric gratings with various modulation depths h/d (here, h is the relief depth and d is the grating period). The groove profile can be sinusoidal, near-sinusoidal, or cycloidal. The reflectivity of the gratings was increased by applying a 100-nm-thick aluminum coating.

The spectral dependences of the HDG diffraction efficiency $\eta(\lambda)$ were taken with a scheme similar to the Littrow scheme: the angle between the incident and diffracted beams was about 10°. The measurements were performed in the spectral interval between 400 and 800 nm for S-polarized light (the electric field vector of the incident wave is orthogonal to the grooves).

A Dimension 3000 Scanning Probe Microscope (Digital Instruments) was used to measure the surface topography and groove profiles.

RESULTS AND DISCUSSION

Figure 1 shows the spectral dependences of the diffraction efficiency $\eta(\lambda)$ for holographic gratings recorded on the As_2S_3 layer at exposures H = 20, 40, 80, 150, and 300 mJ/cm^2 (curves *1–5*, respectively). The anomalous behavior of the curves in the spectral interval 625–650 nm is related to the energy redistribution

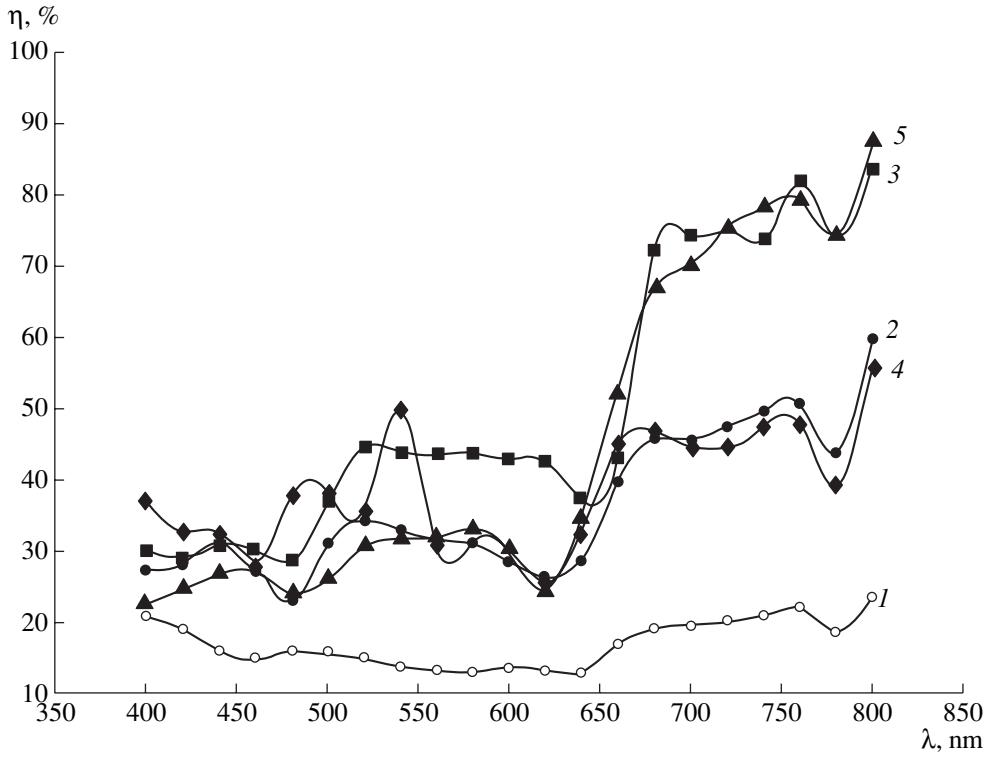

Fig. 1. Spectral curves of the diffraction efficiency η(λ) for holographic gratings recorded on the As_2S_3 layer.

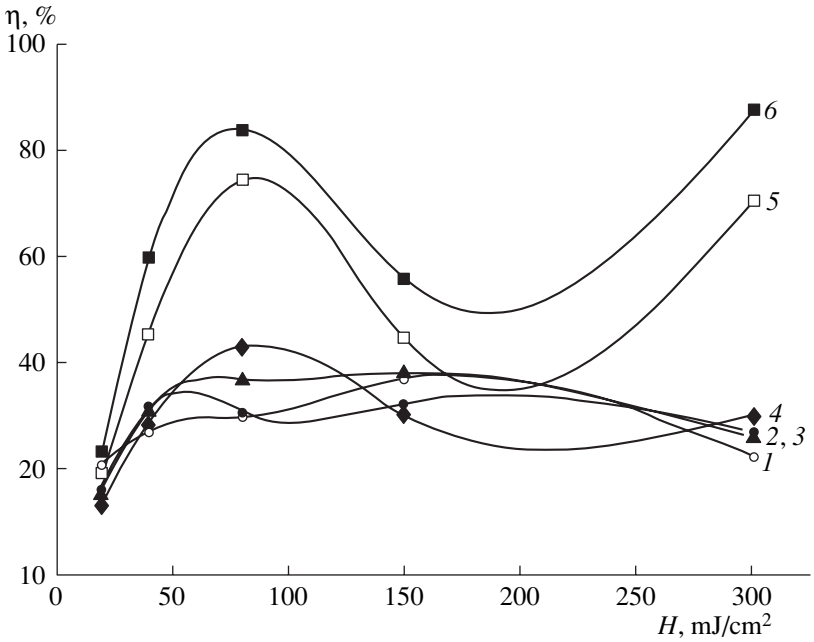

Fig. 2. Exposure dependence of the HDG diffraction efficiency (data are taken from Fig. 1).

between diffraction orders. As the wavelength decreases, η increases and reaches a maximum at λ = 520–530 nm. Another anomaly is observed at λ = 475 nm. Then, η grows slightly to 30–35%. Note curve *4* (H = 150 mJ/cm^2), which is characterized by the achromatism of the diffraction efficiency (η = 40–45%) in a relatively wide spectral range (λ = 520–760 nm). It is seen that the highest efficiencies (η ≈ 80%)

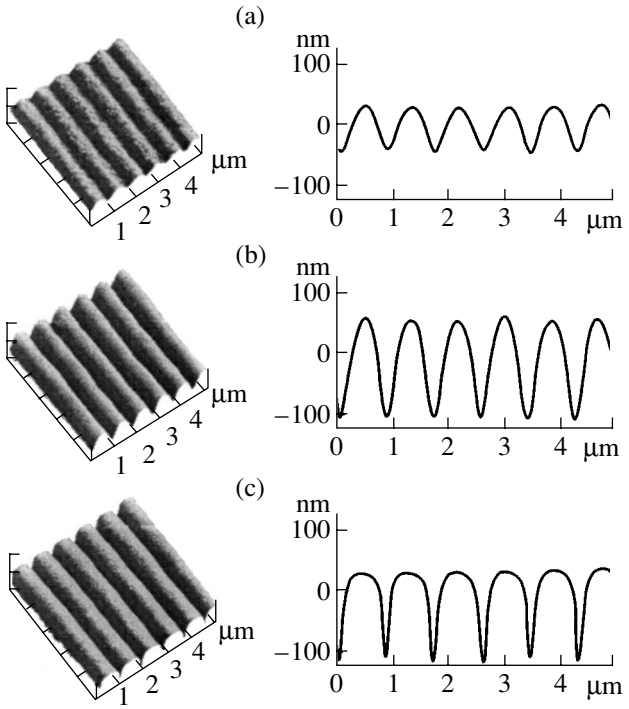

Fig. 3. HDG topography and groove profiles at H = (a) 20, (b) 80, and (c) 300 mJ/cm^2.

are reached at wavelengths between 675 and 800 nm for exposures of 80 and 300 mJ/cm^2 (curves *3* and *5*, respectively).

The data from Fig. 1 were used in Fig. 2, which plots the exposure dependences of the HDG diffraction efficiency at 400, 440, 500, 600, 700, and 800 nm (curves *1*–*6*, respectively). For wavelengths of 600, 700, and 800 nm, the diffraction efficiency is the highest at H = 80–85 mJ/cm^2. The further increase in the exposure minimized η at H = 175–220 mJ/cm^2. Then, the efficiency starts growing again. Of interest is the first maximum of the diffraction efficiency, where η amounts to 70–80%. It is seen from Fig. 2 that the optimal exposures belong to the interval H = 75–100 mJ/cm^2, which also follows from the data shown in Fig. 1 (see curve *3*). Figure 3 shows the SPM images of the groove topography and profiles, enabling one to contrast the optical characteristics of the gratings with various groove profiles. Let us compare the extreme cases corresponding to exposures of 20 and 300 mJ/cm^2 (Figs. 3a and 3c, respectively). At the lower exposure, the grooves are relatively shallow, the widths of peaks and valleys differ insignificantly, and the profile is close to sinusoidal. Because of this, the grating recorded at H = 20 mJ/cm^2 exhibits a low efficiency throughout the spectral range studied. At the higher exposure (H = 300 mJ/cm^2), the groove profile changes to cycloidal. The peaks become flat and wide; the valleys, deep and narrow. At intermediate exposures, the tendency is the same with an increase in the exposure: the peaks flatten and the valleys become narrower and deeper. Therefore, the optimal exposure for recording high-efficiency sinusoidal-groove HDGs on an As_2S_3 layer is H = 80 mJ/cm^2. At 300 mJ/cm^2 (Fig. 1, curve *5*), diffraction efficiencies are high but such exposure conditions are too power-consuming. In addition, cycloidal-groove HDGs are unsuitable for copying, since master grating and its replica are impossible to be identical in this case.

CONCLUSION

The basic results of this work can be summarized as follows.

(i) High-efficiency HDGs can be formed on As_2Se_3-based inorganic photoresists with a He–Cd laser at exposures of 50–100 mJ/cm^2. The diffraction efficiency of such gratings in the wavelength interval λ = 675–800 nm is $\eta > 70\%$.

(ii) The groove profile varies with exposure from sinusoidal (H = 20 mJ/cm^2) or near-sinusoidal (H = 50–100 mJ/cm^2) to cycloidal (H = 150–300 mJ/cm^2).

REFERENCES

1. M. C. Hatley, *Diffraction Grating* (Academic, London, 1982).
2. B. T. Kolomiets, V. M. Lyubin, and V. P. Shilo, Fiz. Khim. Stekla **4**, 351 (1978).
3. I. Z. Indutnyi, I. I. Robur, P. F. Romanenko, and A. V. Stronskii, Proc. SPIE **1555**, 248 (1991).
4. I. Z. Indutnyi, A. V. Stronski, S. A. Kostioukevitch, *et al.*, Opt. Eng. **34**, 1030 (1995).
5. P. F. Romanenko, I. I. Robur, and A. V. Stronskiĭ, Optoélektron. Poluprovodn. Tekh. **27**, 47 (1994).
6. P. F. Romanenko, N. V. Sopinskiĭ, I. Z. Indutnyĭ, *et al.*, Zh. Prikl. Spektrosk. **66**, 587 (1999).
7. R. R. Gerke, M. D. Mikhaĭlov, I. Yu. Yusupov, and O. A. Yakovuk, in *Holographic Optical Elements and Systems*, Ed. by Yu. N. Denisyuk (Nauka, St. Petersburg, 1994), pp. 84–90.
8. E. F. Venger, S. A. Kostyukevich, P. E. Shepeliavyĭ, *et al.*, RF Patent No. 2,165,902, Byull. Izobret., No. 12 (2001).

Translated by A. Chikishev

Technical Physics, Vol. 48, No. 1, 2003, pp. 118–121. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 127–130.

BRIEF COMMUNICATIONS

Simulation of Temperature Fields in a Superconducting Microstrip Bridge

K. A. Titkov, E. I. Vernoslova, M. F. Sitnikova, and I. B. Vendik

St. Petersburg State Electrotechnical University, St.Petersburg, 197376 Russia

e-mail: MWLab@eltech.ru

Received April 17, 2002

Abstract—The temperature distribution in a microstrip bridge made of a high-temperature superconductor (HTSC) as a result of nonuniform heating by a microwave current is analyzed. Differential equations of heat conduction are solved in the one- and two-dimensional approximations. The size of a localized normal region is estimated. Comparative analysis of the results is performed. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

In the development and use of microwave devices based on high-temperature superconducting films, especially those operating at elevated powers, one encounters an important problem associated with local overheating of a superconducting film by a microwave current with subsequent destruction of a superconducting state and spontaneous transition of the film in to a normal state (*S*/*N*-transition) [1–3]. The microwave current in an HTSC film causes Joule heating that may give rise to the formation of a region with normal conductivity (normal zone or domain) in it, which is, as a whole, in the superconducting state. The cause of this phenomenon consists in nonuniform heating of the film due to the presence of structural inhomogeneities. Under certain conditions, the normal region can expand and even occupy the entire film, thereby transferring it into a normal state. The existence of the normal zone significantly affects the value of microwave losses and results in undesirable nonlinear effects that are responsible for changing the operational characteristics and the abnormal operation of devices. Thus, a microwave power providing a possibility for the coexistence of normal and superconducting regions can be considered as a threshold for arising thermal nonlinearity. In this connection, the modeling of temperature distributions in a superconducting film during the flow of a microwave current under given conditions of heat removal is a topical problem. The microwave power and the corresponding microwave current density that determine the threshold for thermal nonlinearity in devices based on HTSC films, as well as the simulation of thermal fields in HTSC films deposited on dielectric substrates, can be found by solving the differential equation of heat conduction.

In this paper, we perform one- and two-dimensional simulations of thermal fields in a film and a substrate, making it possible to obtain the corresponding temperature distribution in a superconducting film. The results under different conditions are compared.

CHARACTERISTIC CURRENT DENSITY

Consider a microwave microstrip bridge [2] in the geometry shown in Fig. 1. As a rule, the width of a microbridge is much smaller than both its length and the substrate thickness.

To describe the thermal processes in the one-dimensional approximation, we use the heat conduction equation for an infinitely long sample in the form

$$K_f\frac{\partial^2 T}{\partial x^2} + q(T) - w(T) = c_f(T)\frac{\partial T}{\partial t},$$
$$T(\pm\infty) = T_0, \quad \left.\frac{\partial T}{\partial x}\right|_{\pm\infty} = 0, \tag{1}$$

where $q(T)$ and $w(T)$ are the power densities of the Joule heat release and the heat removal, respectively; K_f is the thermal conductivity of the film; T_0 is the temperature of the thermostat; and $c_f(T)$ is the thermal capacity of the film.

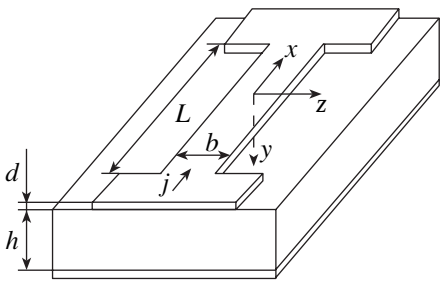

Fig. 1. Model of an HTSC microstrip bridge.

1063-7842/03/4801-0118$24.00 © 2003 MAIK "Nauka/Interperiodica"

The power density $q(T)$ is determined by the following expression:

$$q(T) = \frac{R_{sur}(T) j_{sur}^2}{2d}, \quad (2)$$

where j_{sur} is the surface current density and $R_{sur}(T)$ is the surface resistance of the film.

The surface resistance is determined by the phenomenological model [4] as a function of the operating frequency f and current temperature T. This model provides an adequate quantitative description of $R_{sur}(T)$ in the temperature regions both below the critical temperature T_c and at $T > T_c$. The critical temperature is a key parameter of the given model of surface resistance.

The critical current density j_c depends on the temperature and can be described by the following approximation [5]:

$$j_c(T) = j_c(0)\left(1 - \left(\frac{T}{T_c}\right)^{\gamma}\right)^{1.5}, \quad (3)$$

where $j_c(0)$ is the critical current density at $T = 0$ and γ is a phenomenological parameter.

The power density of the heat removal is

$$w(T) = \frac{K_s}{d_{eff} d}(T - T_0), \quad (4)$$

where K_s is the heat conduction of the substrate material, which is assumed to be isotropic and temperature-independent in the problem under consideration.

In view of the small contributions of the other kinds of heat removal, we suppose that the heat dissipates only into the substrate. If the film width b is much smaller than the substrate thickness h, a rod with diameter b can be considered as a heat source, provided that the heat flow into the semi-infinite substrate is cylindrically symmetric and the temperature varies linearly in a layer with a thickness of d_{eff} [3]:

$$d_{eff} \cong \frac{b}{\pi} \ln \frac{2h}{b}. \quad (5)$$

In general, both an HTSC film and a substrate can contain structural inhomogeneities. In the problem under consideration, this results in the appearance of a coordinate dependence of the power densities of heat release $q(T, x)$ and/or heat removal $w(T, x)$ in expressions (1)–(4). An inhomogeneity may represent both an inclusion of a nonsuperconducting defect [1] and a local variation in the parameters of a superconducting film.

A typical value of the surface density j_p of a microwave current, which determines the possibility of coexistence of normal and superconducting regions, can be found by the procedure described in [6]. The value of j_p corresponds to a minimum surface current density of the S–N-interface propagation. At $j < j_p$, the stable existence of the finite-size regions of the normal phase in the superconducting sample becomes impossible. At $j > j_p$, finite-size superconducting regions cannot exist in a normal phase. The values of j_p and corresponding maximum temperature T_m within the normal region are defined by the system of equations

$$\begin{aligned} w(T_m, x) - q(T_m, j_p, x) &= 0, \\ \int_{T_0}^{T_m} K_f [w(T, x) - q(T, j_p, x)]\, dT &= 0. \end{aligned} \quad (6)$$

The dependence of $j_p(x)$ thus obtained makes it possible to estimate [6] the size of a normal region for a given surface current density j. In what follows, we assume that the inhomogeneity is symmetrically located with respect to the origin of coordinates at the center of a microstrip bridge and is caused by a variation in the film parameters.

Let us represent the inhomogeneity in our model as follows. The critical temperature is assumed to vary smoothly along the longitudinal coordinate. The model representation of distribution along the microstrip bridge length that is used in this study (Fig. 2) meets the aforementioned requirement. The distribution $j_p(x)$, which is the solution of the system of equations (6) for the inhomogeneity, is also shown in Fig. 2. The corresponding sizes of the normal region D arising at the inhomogeneity at different current densities are shown in Fig. 3.

Simulations were performed for the following characteristic parameters of the model and the geometry of the microstrip bridge: $h = 0.5$ mm, $d = 0.2$ μm, $b = 25$ μm, $L = 100$ μm, $T_0 = 77$ K, $K_s = 18.5$ W/(m K) (lanthanum aluminate), $K_f = 30$ W/(m K) (YBaCuO), $f = 10$ GHz, and $j_c(0) = 1.5 \times 10^7$ A/cm^2 (at $T_c = 90$ K).

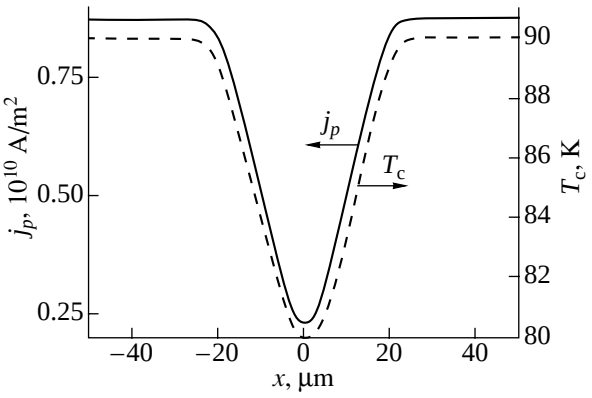

Fig. 2. Distributions of the critical temperature T_c (dashed line) and the typical current density (solid line) in the microstrip bridge along the longitudinal coordinate x.

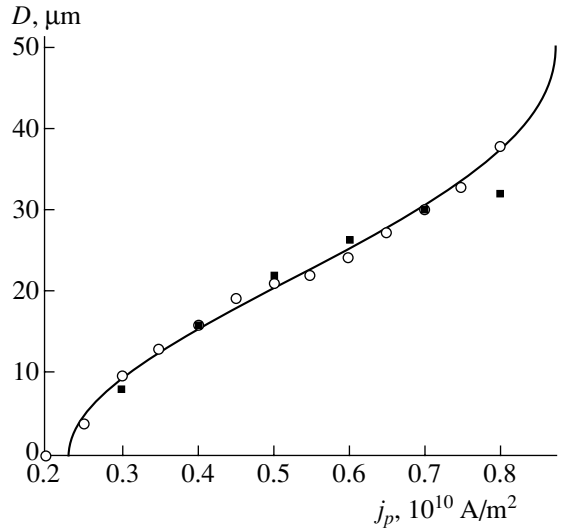

Fig. 3. The size of a normal region (thermal domain) for different current densities: the solid curve is the theoretical curve calculated by the system of equations (6); open and filled circles correspond to the one- and two-dimensional approximations of the thermal fields, respectively.

MODELING OF THE THERMAL FIELDS

The modeling of a stationary temperature field in the microstrip bridge is based on solving the stationary equation of heat conduction. For a linear variation in temperature across the substrate thickness, the one-dimensional approximation is valid. In this case, the stationary equation of heat conduction has the form

$$K_f \frac{\partial^2 T}{\partial x^2} + q(T, x) - w(T) = 0. \tag{7}$$

To find the temperature field in the microstrip bridge, this equation is solved by the finite difference method using the following difference approximation of the initial differential equation:

$$\begin{gathered} \frac{T_{i-1} - 2T_i + T_{i+1}}{\Delta x^2} = K_f^{-1}[w(T_i) - q(T_i, x_i)], \\ i = 1, 2, \ldots, N-1; \quad N\Delta x = l_x; \\ T(x_0) = T(x_N) = T_0, \end{gathered} \tag{8}$$

where l_x is the region involving the microstrip bridge, as well as the parts of the leads with a length of d_{eff}.

The heat release in the leads can be neglected, since the current density in them is smaller compared to that in the bridge. The solution to (8) can be found in different ways, for example, by the iteration method using the scheme

$$\begin{gathered} T_i^{(k+1)} = b^{-1}\left[\frac{T_{i-1}^{(k)} + T_{i+1}^{(k)}}{\Delta x^2}\right] + b^{-1} f_i^{(k)}; \\ i = 1, 2, \ldots, N-1, \\ T_0^{(k+1)} = T_N^{(k+1)} = T_0, \\ b = \frac{2}{\Delta x^2}, \quad f_i^k = K_f^{-1}[q(T_i, x_i) - w(T_i)]. \end{gathered} \tag{9}$$

For a more accurate description of the temperature field, the two-dimensional character of the problem should be taken into account. In this case, the temperature distribution is characterized by the two-dimensional equation

$$\begin{gathered} K(y)\frac{\partial^2 T}{\partial x^2} + K(y)\frac{\partial^2 T}{\partial y^2} + q(T, x, y) = 0, \\ K(y) = \begin{cases} K_f & y \le d \\ K_s & y > d, \end{cases} \end{gathered} \tag{10}$$

with the boundary conditions

$$\left.\frac{\partial T}{\partial y}\right|_{y=0} = 0, \quad T(\pm 0.5 l_x, y) = T(x, d+h) = T_0. \tag{11}$$

As well as in the one-dimensional case, the solution is found by the finite difference method with a difference scheme similar to that of (8) and (9). The simulation results are shown in Fig. 4. The curves obtained make it possible to estimate the size of the normal region D localized at the inhomogeneity. The desired sizes are determined at a level of $0.5(T_{\max} - T(0.5L))$, where $T_{\max}$ is the maximum temperature in the distribution. A comparison of these estimates and those obtained in the one-dimensional approximation for the infinitely long sample is shown in Fig. 4.

DISCUSSION

In this study, after solving the differential equation of heat conduction, we found the values of the microwave current density that determine the possibility of coexisting normal and superconducting regions. In addition, we simulated the stationary thermal fields in HTSC nonhomogeneous films.

These thermal field distributions make it possible to estimate the size of a normal region localized at an inhomogeneity for various values of the current density. A comparison of the results derived in the one- and two-dimensional approximations indicates acceptable agreement of the calculation results with each other and with those obtained for the infinite sample. The maximum deviation does not exceed 10%, which is tolerable for our discretization step of ±1. Thus, the model of an infinite sample can be used to determine the threshold

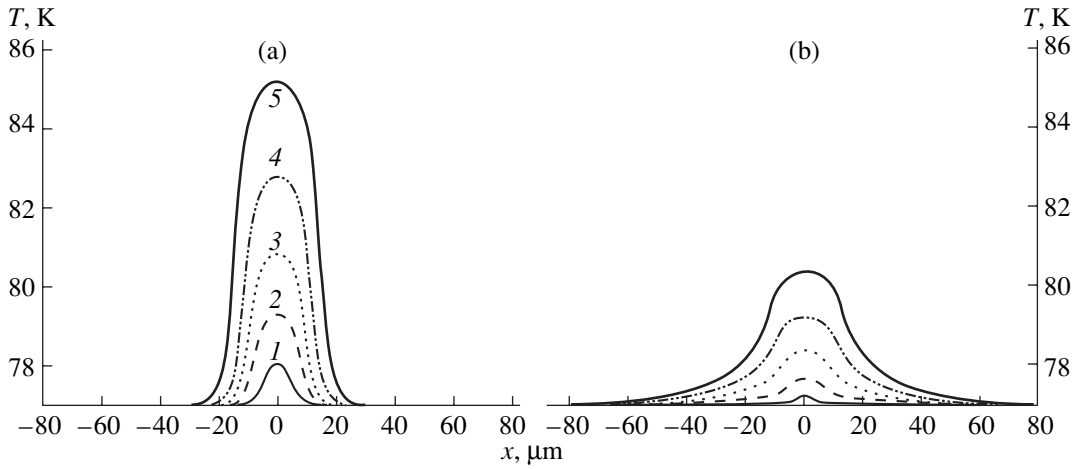

Fig. 4. Temperature distributions in the microstrip bridge for various current densities; (a) and (b) correspond to the one- and two-dimensional simulations of the thermal fields, respectively; $j = 3.0 \times 10^9$ (*1*), 4.0×10^9 (*2*), 5.0×10^9 (*3*), 6.0×10^9 (*4*), and 7.0×10^9 A/m^2 (*5*).

of thermal nonlinearity and estimate the size of a normal region.

The difference in the temperature distributions for the microstrip bridge that were obtained in the one- and two-dimensional approximations is explained by the nonuniform heating of the substrate. This results in deviation of the model of heat removal from the linear one and, consequently, in a decrease in the power density of the heat removal. Thus, despite coincidence in the estimates for the sizes of a normal region, which were obtained in the one- and two-dimensional approximations, the latter case is preferable, since it allows for the influence of a substrate more correctly. The temperature distributions obtained can also be used in calculating operational characteristics of devices made of HTSC films in view of the effect of thermal nonlinearity.

REFERENCES

1. A. Zharov and A. Reznik, Zh. Tekh. Fiz. **68** (1), 131 (1998) [Tech. Phys. **43**, 117 (1998)].
2. E. Loskot, M. Sitnikova, and V. Kondratiev, Inst. Phys. Conf. Ser. **167**, 367 (1999).
3. E. Vernoslova, K. Titkov, M. Sitnikova, *et al.*, Proc. IEEE, No. 1, 46 (2000).
4. I. Vendik, Supercond. Sci. Technol. **13**, 974 (2000).
5. I. Vendik and O. Vendik, *High Temperature Superconductor Devices for Microwave Signal Processing* (Skladen, St. Petersburg, 1997).
6. A. V. Gurevich and R. G. Mints, Usp. Fiz. Nauk **142** (1), 61 (1984) [Sov. Phys. Usp. **27**, 19 (1984)].

Translated by Yu. Vishnyakov

Technical Physics, Vol. 48, No. 1, 2003, pp. 122–124. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 131–133.

BRIEF COMMUNICATIONS

On Internal Strain-Induced Changes in Massive Bodies Due to Their Own Gravitational Field

S. O. Gladkov

Moscow Pedagogical University, ul. Radio 10a, Moscow, 107005 Russia

Received April 17, 2002

Abstract—The behavior of the deformation vector **u** in massive structures under the action of their own gravitational field is analyzed. The variation of the radial component of **u** with physical properties of a massive body is demonstrated with the Earth. A set of differential equations is derived based on variational calculus methods. *© 2003 MAIK "Nauka/Interperiodica".*

Despite the seemingly obvious fact that gravity influences the deformation properties of massive bodies (such as planets), the variation of the components of the object displacement vector with distance in the gravitational field has not yet been explored. Anyhow, in publications hitherto known to us (see, for example, [1–16] and Refs. therein), this problem has not been dealt with, although it is not only of purely scientific (satisfaction of our inquisitiveness) but also of practical interest.

It is also important that the elucidation of the relation between the gravitational potential φ and deformation parameters of the planet would make it possible to put forward one more concept of, and shed additional light upon, natural cataclysms such as earthquakes and volcanic eruptions.

Our approach is phenomenological; yet, using a single adjustable parameter (see below), one can predict and describe a number of intriguing features of the behavior of intraplanetary deformation displacements that are unnoticeable on the Earth's surface but are undoubtedly very important in considering fine gravitational effects like the experimental detection of gravity waves [17, 18].

When considering the effect of gravitational potential φ on the deformation displacement **u** of the medium, we will use the well-proven and thus reliable variational method [19, 20]. Within the framework of our problem, it is reduced merely to finding an extremum of the system's Hamiltonian (because of the absence of the time dependence of the parameters **u** and φ).

It should be noted that such a method of describing the dynamic development of linear and nonlinear Hamiltonian systems can be successfully applied to various physical phenomena (such as the evolution of the combustion front [21] or the crystallization dynamics of composites [22] in some phase space of angular momenta and angles) typical of synergistic systems [23].

In our case, the Hamiltonian of the system can be represented in the invariant form

$$H\{\mathbf{u}, \varphi\} = \int\{[0.5(\nabla\varphi)^2 + 4\pi G\rho\varphi](\rho/\omega_0^2) + [0.5K(\operatorname{div}\mathbf{u})^2 + \mu(u_{ik} - 1/3\delta_{ik}\operatorname{div}\mathbf{u})^2] + (\omega_0^2/G)\mathbf{u}\nabla\varphi\}d^3x, \quad (1)$$

where G is the gravitational constant; ρ is the planet density; and K and μ are the compression and shear moduli, respectively. They are related to the Poisson's ratio σ and Young's modulus E as

$$K = E/3(1-2\sigma), \quad \mu = E/2(1+\sigma). \quad (2)$$

In (1), the constant ω_0 has the dimension of frequency and represents a phenomenological parameter of interaction between the Newtonian (nonrelativistic) potential φ and displacement vector **u** of the medium (the last term in (1)), $u_{ik} = (1/2)(\partial u_i/\partial x_k + \partial u_k/\partial x_i)$ is the symmetric strain tensor, and δ_k is the Kronecker symbol.

Consider Hamiltonian (1). The first two terms lead to the well-known Poisson equation for the distribution of the nonrelativistic gravitational potential φ. The next pair of terms represents the purely elastic deformation part of the Hamiltonian. They lead to the static equations of the elasticity theory [24]. Finally, the last term characterizes the desired interaction between the gravitational potential and displacement vector. The dimensional factor ρ/ω_0^2 is chosen so as to minimize the number of uncertain constants. It is shown below that even one constant that has the dimension of frequency will suffice.

By varying the parameters **u** and φ of the functional $H\{\varphi, \mathbf{u}\}$, we arrive at the system of linear differential equations

$$\Delta\varphi = 4\pi\rho G - \beta'\operatorname{div}\mathbf{u},$$
$$\Delta\mathbf{u} + [\sigma/(1-2\sigma)]\operatorname{grad}\operatorname{div}\mathbf{u} = \beta[(1+\sigma)/E]\nabla\varphi, \quad (3)$$

1063-7842/03/4801-0122$24.00 © 2003 MAIK "Nauka/Interperiodica"

where $\beta' = (\omega_0^2/\rho)\beta$, $\beta = \omega_0^2/G$, and Δ is Laplacian.

From the first equation, we immediately obtain

$$\varphi(\mathbf{r}) = -G\int\frac{\rho(\mathbf{r}')d^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|} + \frac{\beta'}{4\pi}\int\frac{\mathrm{div}\,\mathbf{u}(\mathbf{r}')d^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}. \quad (4)$$

It is easy to check that for a spherical body,

$$\varphi(\mathbf{r}) = (2\pi\rho G/3)(r^2 + R^2) + \beta'\int_r^R (1/r')(\partial/r')(r'^2 u_r(r'))dr'. \quad (5)$$

Substituting this solution into the second equation of system (3) written in the spherical coordinate system yields

$$(1/r)(ru_i)'' + [\sigma/(1-2\sigma)][(1/r^2)(r^2u_r)']' = 4\pi\rho\beta G(1+\sigma)r/E - \beta'\beta[(1+\sigma)/E\rho][1/r(r^2u_r)']'.$$

After simple rearrangements, we find

$$u_r'' + u_r'(r\gamma + 2/r) + 2u_r[\gamma - \sigma/(1+\sigma)r^2] = 4\pi\rho\beta G(1-2\sigma)r/3E(1+\sigma), \quad (6)$$

where $\gamma = \beta\beta'(1-2\sigma)/(1+\sigma)\rho$.

Making Eq. (6) dimensionless with the substitutions $x = r/r_0$ and $u = u_0 y$ (where $r_0 = (2/\gamma)^{1/2}$ and $u_0 = 4\pi\rho\beta G(1-2\sigma)r_0^3/3E(1+\sigma)$) and making two other substitutions, $\xi = x^2$ and $y = W(\xi)\xi^k$, where k satisfies the equation $k^2 - \sigma/2(1+\sigma) = 0$ (here, only the positive root has a physical meaning, because at $x = 0$ the strain must be finite), we come to equation

$$\xi W'' + (2k + 1 + 0.5\xi)W' + (1 + 0.5k)W = 0. \quad (7)$$

This equation is nothing but an equation of a degenerate hypergeometric function. Then, its solution is $W = F(k+2, 1+2k, -\xi/2)$. Eventually, the desired deformation displacement can be written as

$$u_r = u_0 x^{2k} F(k+2, 1+2k, -x^2/2). \quad (8)$$

It is seen that at small x ($x \longrightarrow 0$), the solution $u \longrightarrow u_0 x^{2k}$. At large x ($x \gg 1$), we find, using the asymptotics of a hypergeometric function, that

$$u_k \approx u_0\left\{\frac{2^{k12}\Gamma(k+2)}{\Gamma(k-1)x^4} + \frac{\Gamma(k+2)\exp\{-x^2/2\}x^2}{\Gamma(1+2k)}(-0.5)^{1-k}\right\}. \quad (9)$$

Since $k < 1$, the first term in (9) vanishes and finally the correct solution takes the form

$$u_k \approx u_0\frac{\Gamma(k+2)\exp\{-x^2/2\}x^2}{\Gamma(1+2k)}(-0.5)^{1-k}.$$

Thus, at a distance of r_0 or larger, the deformation displacement decays exponentially towards the Earth's surface, which physically is a reasonable result in our opinion.

This critical radial distance (measured from the planet center) can be estimated by the formula

$$r_0 = (1/\gamma)^{1/2} = (G/\omega_0^3)[E(1+\sigma)\rho/(1-2\sigma)]^{1/2}. \quad (10)$$

To estimate r_0, we take the Young's modulus $E = 10^{11}$ J/m^3, Earth's density $\rho = 10$ kg/m^3, Poisson's ratio $\sigma = 1/3$, and gravitational constant $G = 6.672 \times 10^{-11}$ m^3/(kg s^2). The interaction constant measured in units of inverse time is taken as equal to $\omega_0 = 10^{-3}$ s^{-1}. Note in passing that if ω_0 is other than 10^{-3}, the result will be somewhat unrealistic. It turns out that $r_0 = 134$ km.

This estimate indicates the decay of strains due to the interaction between the gravitational field of the planet and deformation displacements occurring in any massive body under its own gravity force. Characteristic distances are estimated as several hundreds of kilometers from the center. Basically, this is quite obvious, because the relationship between the displacement **u** and potential φ was shown to be very strong and the neglect of this interaction would lead to a meaningless solution. Indeed, it is easy to verify that solving the equation $\mathbf{g} = -g\mathbf{r}/R$ with $\Delta\varphi = 4\pi\rho G$ (which follows from the Poisson equation $\Delta\mathbf{u} + [\sigma/(1-2\sigma)]\mathrm{grad}\,\mathrm{div}\,\mathbf{u} = -\rho\mathbf{g}$ at $r < R$, where R is the Earth's radius) would yield a displacement on the order of 160 km at a distance of about 2000 km from the center. Such a value could still be recognized as more or less adequate; however, one encounters the greatest problems when calculating the displacement on the Earth's surface, which turns out to be about 1600 km.

The above value merely indicates that it is necessary to take into account the nonlinear terms describing the interaction between the gravitational and deformation potentials. To stress this need was the aim of the present study. The results obtained are more than realistic. They suggest the absence of displacements on the Earth's surface despite considerable displacements near the core.

REFERENCES

1. M. S. Molodenskii, Comm. Obs. R. Belgique **25**, 288 (1961).
2. *Natural Oscillations of the Earth: Collection of Scientific Works* (Nauka, Moscow, 1964).
3. V. A. Magnitskiĭ, *Internal Structure and Physics of the Earth* (Nauka, Moscow, 1965).

4. P. Melchior, *The Earth Tides* (Pergamon, Oxford, 1966; Nauka, Moscow, 1968).
5. J. Weber, Phys. Rev. Lett. **20**, 1307 (1968).
6. E. G. Markaryan and V. P. Myasnikov, *Hydrodynamic Model of the Earth's Evolution* (Institut Kosmicheskikh Issledovaniĭ Akad. Nauk SSSR, Moscow, 1977).
7. J. Won and J. Kuo, Geophys. Res. **78**, 905 (1973).
8. V. P. Myasnikov and V. E. Fadeev, Itogi Nauki Tekh., Ser.: Fiz. Zemli **5**, 3 (1980).
9. M. A. Sadovskiĭ, V. F. Pisarenko, *et al.*, Izv. Akad. Nauk SSSR, Fiz. Zemli, No. 12, 3 (1983).
10. M. A. Sadovskiĭ, V. F. Pisarenko, *et al.*, Izv. Akad. Nauk SSSR, Fiz. Zemli, No. 1, 12 (1984).
11. O. A. Akimov, V. A. Malugin, and A. B. Manukin, Izv. Akad. Nauk SSSR, Fiz. Zemli, No. 10, 97 (1985).
12. D. Agnew, Rev. Geophys., No. 3, 579 (1986).
13. P. Melchior and B. Ducarme, Phys. Earth Planet. Inter. **129**, 42 (1986).
14. A. V. Kobzev and V. P. Myasnikov, Dokl. Akad. Nauk SSSR **296**, 381 (1987).
15. *Dynamic Processes in the Earth's Crust* (Nauka, Moscow, 1994).
16. V. B. Braginskiĭ, Ya. B. Zel'dovich, and V. N. Rudenko, *Determination of the Gravitational Constant and Measurement of Fine Gravitational Effects* (Moscow, 1973), pp. 8–19.
17. I. Bichak and V. N. Rudenko, *Gravity Waves and Their Detection* (Nauka, Moscow, 1987).
18. V. N. Rudenko, Phys. Lett. A **223**, 421 (1996).
19. L. El'sgol'ts, *Differential Equations and Variational Calculus* (Nauka, Moscow, 1969; Mir, Moscow, 1970).
20. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 1: *Mechanics* (Nauka, Moscow, 1973; Pergamon, New York, 1988).
21. S. O. Gladkov and A. M. Tokarev, Fiz. Goreniya Vzryva **25**, 30 (1990).
22. S. O. Gladkov, Perspekt. Mater., No. 1, 50 (2000).
23. A. Yu. Loskutov and A. S. Mikhaĭlov, *Introduction to Synergism* (Nauka, Moscow, 1990).
24. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 7: *Theory of Elasticity* (Nauka, Moscow, 1987; Pergamon, New York, 1986).

Translated by B. Kalinin

Technical Physics, Vol. 48, No. 1, 2003, pp. 125–127. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 134–136.

BRIEF COMMUNICATIONS

Acoustooptic Spectroscopy of Metal Structure Modifications under Plastic Deformation Due to Submicrosecond Impulsive Shock Loading

Yu. V. Sud'enkov and Z. A. Sazhko

St. Petersburg Sate University, Universitetskaya nab. 7/9, St. Petersburg, 198504 Russia

e-mail: spm@unicorn.math.spbu.ru

Received May 31, 2002

Abstract—Restructuring in metals due to elastoplastic deformation produced by submicrosecond impulsive shock loading is studied experimentally. Tests are performed on 0.3ZhR iron and N1 nickel specimens with various structures. Structure modifications after shock tests are studied with acoustooptic spectroscopy. The frequency dependences of the longitudinal acoustic velocity and attenuation coefficient make it possible to estimate both the characteristic scales of restructuring in relation to the initial structure and the sensitivity of the acoustooptic method to restructuring in various materials. © *2003 MAIK "Nauka/Interperiodica".*

The interrelation between the plastic deformation of materials and modifications in their structure is well known [1–4]. However, information on characteristic structure levels responsible for deformation when loading conditions and initial material structures vary is very scarce. This is because the techniques for studying restructuring under deformation are not sufficiently advanced and few complex experimental investigations of this phenomenon (especially after shock tests) have been carried out.

Upon studying elastoplastic deformation in metals due to submicrosecond shock impulses, it was found [5] that the grain size dependence of hardening differs much from the Hall–Petch relationship. It was assumed that this is associated with a change in the scale levels of restructuring that are responsible for plastic deformation in materials with various initial structures.

To reveal the scales of modifications, we studied 0.3ZhR iron and N1 nickel specimens with various initial structures before and after shock loading. The loading was accomplished by pressure impulses of amplitude ≈1.0 GPa and duration $\approx 8 \times 10^{-8}$ s [5]. The initial structure of the specimens was modified by vacuum annealing for 1 h at a given temperature. After the annealing, the specimens were cooled in the furnace.

To obtain the frequency dependence of the attenuation and acoustic velocity over a wide spectral range, we applied the pulsed acoustooptic method [6–8] using a 1.06-μm laser with a pulse width of 25 ns and a pulse energy of ≤30 mJ.

Acoustic pulses were recorded with a piezoelectric detector made of a 30-μm-thick polyvinyldiphenyl film placed on the surface of an acoustooptic cell made of fused quartz. The cell served as an optical path and

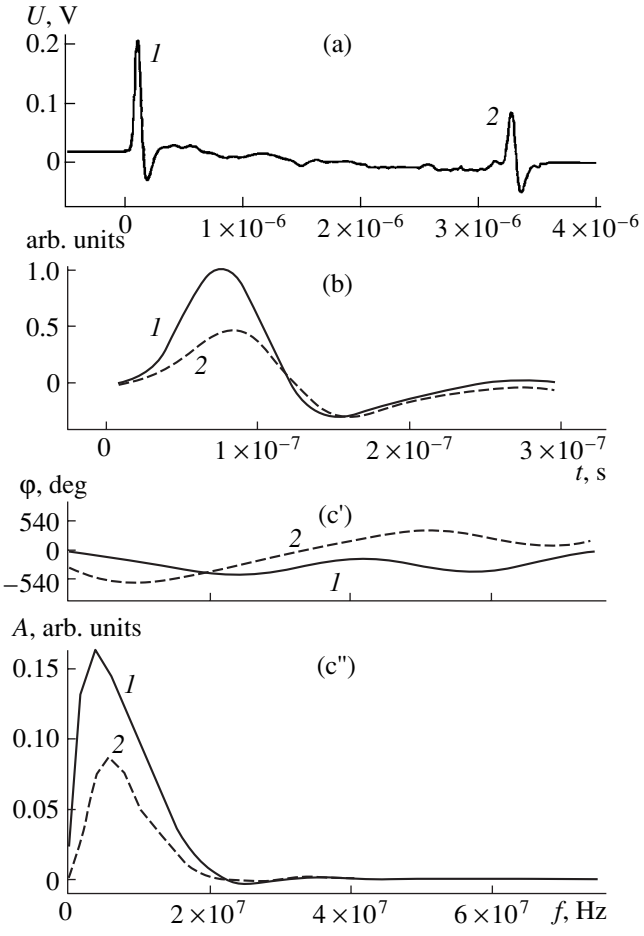

Fig. 1. (a) Waveforms of the (*1*) first and (*2*) second acoustic pulses in iron (annealed at 100°C), (b) pulses are imposed on each other, and (c') phase and (c") magnitude of the Fourier spectra of the first and second acoustic pulses.

1063-7842/03/4801-0125$24.00 © 2003 MAIK "Nauka/Interperiodica"

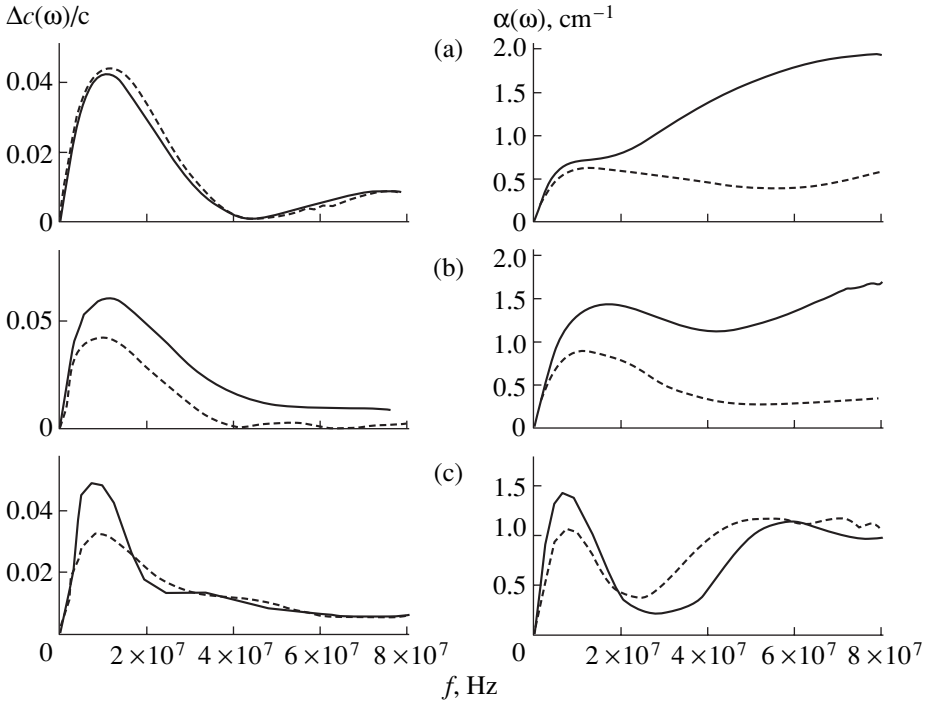

Fig. 2. Frequency dependences of the acoustic velocity $\Delta c(\omega)/c_0$ and attenuation $\alpha(\omega)$. Annealing at (a) 1000°C ($d \approx 20$ μm), (b) 700°C ($d \approx 20$ μm), and (c) 100°C ($d \approx 220$ μm). Dashed line, before the shock; solid line, after the shock.

acoustic waveguide [9]. The passband of the detector was about 150 MHz.

Figure 1 shows the typical waveforms of the (*1*) first and (*2*) second acoustic echo pulses in the iron specimen (annealed at 100°C) and the first and second pulses imposed on each other, as well as the magnitudes and phases of their Fourier transforms.

In the experiments, we measured the velocity c_0 of acoustic pulse propagation, as well as determined (by analyzing the spectra of the pulses) the frequency (ω) dependence of the velocity ($c(\omega)$) and attenuation ($\alpha(\omega)$) from the relationships [8]

$$\Delta c(\omega) = \frac{c_0}{\omega \Delta x}\left[\arctan\frac{B_1}{A_1} - \arctan\frac{B_2}{A_2}\right],$$

$$\alpha(\omega) = \frac{1}{2\Delta x}\ln\frac{A_1^2 + B_1^2}{A_2^2 + B_2^2},$$

where $c(\omega) \approx c_0 + \Delta c(\omega)$, Δx is the specimen thickness, and $A_{1,2}$ and $B_{1,2}$ are the coefficients before the real and imaginary parts of the Fourier spectrum for the first and second acoustic pulses.

Figure 2 demonstrates the thus-obtained frequency dependences of the acoustic velocity and attenuation before and after shock loading the iron specimens annealed at 1000, 700, and 100°C. From the variation of these curves, one can estimate the scales of restructuring due to shock loading.

For the specimens with a grain size $d \approx 220$ μm and low dislocation density that were annealed at 1000°C, the curves $\Delta c(\omega)/c_0$ before and after the shock are almost the same throughout the frequency range. The acoustic velocity also remains unchanged, $c_0 = 5948 \pm 5$ m/s. The attenuation coefficients do not differ up to a frequency of ≤25 MHz. However, in the high-frequency part of the spectrum, the attenuation after the shock increases significantly because of the high concentration of linear defects (dislocations and twins) generated by plastic deformation. The characteristic size (scale) of such structure imperfections (modifications) is on the order of $\approx 10^{-7}$–10^{-6} m [1–4]. It appears that the strain hardening of the specimens with such an initial structure was the highest for precisely this reason [5].

In the specimens with a structure modified by annealing at 700°C ($d \approx 20$ μm), the shock loading is followed by an appreciable increase in the spread in the velocity $\Delta c(\omega)/c_0$ and attenuation $\alpha(\omega)$ throughout the frequency range. In the shock-loaded specimens, the acoustic velocity grows from 5956 to 5974 m/s. This difference exceeds the measurement accuracy threefold and is associated with the change in the crystallographic orientation of entire grains or their fragments that have a characteristic size of 10^{-6}–10^{-5} m. This change is due to the rotational mechanism of plastic deformation [2–4]. Accordingly, the attenuation grows at frequencies ≤25 MHz because acoustic waves scatter by fragment or grain boundaries. The increase in the attenuation after the shock in the high-frequency part of

the spectrum is due to a higher concentration of the linear defects, as before. Thus, both the rotational mechanism and defect structure modifications are responsible for the plastic deformation of the iron specimen whose initial structure was obtained by annealing at 700°C.

In the specimens annealed at 100°C in which the grain size was the same as before ($d \approx 20$ μm) but the starting defect density was higher, $\Delta c(\omega)/c_0$ and $\alpha(\omega)$ grow markedly after the loading only at frequencies ≤25 MHz. In the shock-loaded specimens, the acoustic velocity increases by 24 m/s from $c_0 = 5966 \pm 5$ to 5990 ± 5 m/s. In the high-frequency part, the velocity spreads and attenuation coefficients before and after the tests differ only slightly. A possible reason for such behavior of $\Delta c(\omega)/c_0$ and $\alpha(\omega)$ is that the rotational mechanism of plastic yield causing restructuring on a scale comparable to the grain size becomes prevailing in the specimens with the initial structure described above.

The data for the N1 nickel specimens qualitatively agree with those for the iron specimens shown in Fig. 2.

Thus, in the case of a submicrosecond shock, the rotational mechanism governs plastic deformation in polycrystals with a grain size on the order of the shock loading front ($\approx 10^{-5}$ m). According to [5], the strain hardening in such specimens is the lowest.

Our results combined with those in [5] also indicate that the violation of the Hall–Petch relationship (a rise in the hardening with grain size) is explained by the decisive role of the intragranular defect structure modification (an increase in the concentration of dislocations and twins) in the strain hardening. For a submicrosecond shock loading, this effect is most pronounced in coarse-grained specimens and apparently in single crystals.

To conclude, we note that the method of acoustooptic spectroscopy shows promise for studying restructuring in metals. With this method, one can gain information on restructuring on a scale of 10^{-7}–10^{-6} m.

REFERENCES

1. *Shock Waves and Effects of Fast Deformation of Metals* (Metallurgiya, Moscow, 1984).
2. *Problems in Theory of Defects in Crystals* (Nauka, Leningrad, 1987).
3. *Strain Hardening and Destruction of Polycrystalline Metals* (Naukova Dumka, Kiev, 1989).
4. *Disclination and Rotational Deformation of Solids* (FTI, Leningrad, 1990).
5. Yu. V. Sud'enkov and Yu. B. Nikitin, Pis'ma Zh. Tekh. Fiz. **19** (12), 66 (1993) [Tech. Phys. Lett. **19**, 383 (1993)].
6. V. É. Gusev and A. A. Karabutov, *Laser Opto-Acoustics* (Nauka, Moscow, 1991).
7. A. G. Merkulov and V. A. Tokarev, Defektoskopiya, No. 4, 1 (1970).
8. A. N. Bondarenko, *Laser Methods of Excitation and Recording of Acoustic Signals* (Izd. Standartov, Moscow, 1989).
9. Yu. V. Sud'enkov and D. Yu. Nikitin, in *Proceedings of X Session of the Russian Acoustic Society* (GEOS, Moscow, 2000), Vol. 2, p. 345.

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 1, 2003, pp. 128–130. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 137–139.

BRIEF COMMUNICATIONS

Fast Electrostatic Sweep of a Mass Spectrum in Permanent-Magnet Mass Spectrometers

O. N. Peregudov, V. F. Shkurdoda, and L. F. Sukhodub

Institute of Applied Physics, National Academy of Sciences of Ukraine, ul. Petropavlovskaya 58, Sumy, 40030 Ukraine

e-mail: op@pochtamt.ru

Received June 21, 2002

Abstract—As was shown earlier [1], in a permanent-magnet mass spectrometer, one can record mass spectra in a considerably wide range of mass numbers by using several specially arranged collectors. This scheme is easy to implement in magnetic mass analyzers with an angle of rotation other than 180°. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

In permanent-magnet mass analyzers, the fast recording of mass spectra in a wide range of masses can be accomplished in two ways: with the use of position-sensitive detectors (PSDs) and by sweeping the accelerating voltage in a system with several collectors. PSDs offer a high rate of recording and a high sensitivity. However, they have a complex design and are also expensive and difficult to maintain. The latter variant is structurally simpler and offers a sufficiently high stability and reliability. The sensitivity of a multicollector system is somewhat lower, but it records a wider mass spectrum than the PSD of the same mass analyzer. Moreover, it will be shown that the recording time in a multicollector system is small.

TRAJECTORY EQUATION

The motion of an ion in a magnetic field can be described in terms of its deviation from the central trajectory [2]:

$$\frac{\delta}{r_0} = f(r_0, x_1(r_0), \ldots, x_n(r_0)), \quad (1)$$

where δ/r_0 is the deviation of the ion moving along the paraxial trajectory; r_0 is the central trajectory radius; and $x_1(r_0)$, …, $x_n(r_0)$ are parameters describing the analyzer geometry and deviation from the initial conditions for ions moving along the paraxial trajectory.

In a uniform magnetic field, the radius of ion rotation is given by [2]

$$r = \frac{1}{B}\sqrt{\frac{2MU}{q}}, \quad (2)$$

where B is the magnetic field induction, U is the ion-accelerating voltage, and q is the charge of an ion.

Let us introduce the dimensionless parameter (trajectory index)

$$\gamma = \frac{M - M_0}{M_0}, \quad (3)$$

where M is the mass of ions moving along an arbitrary trajectory and M_0 is the mass of ions moving along the central trajectory.

Using the concept of the trajectory index, we express the radius of an arbitrary trajectory through the radius of the central trajectory,

$$r = \frac{1}{B}\sqrt{\frac{2M_0(1+\gamma)U}{q}} = r_0\sqrt{1+\gamma}, \quad (4)$$

and rewrite relationship (1) as

$$\frac{\delta(\gamma)}{r_0} = \sqrt{1+\gamma}\,\tilde{f}(r_0, \gamma, x_1(r_0, \gamma), \ldots, x_n(r_0, \gamma)). \quad (5)$$

Thus, we arrived at the equation for an arbitrary trajectory in a given mass range. This approach was used by the authors in [3].

SIMULATION OF SWEEPING

Using the trajectory index, we will simulate the process of electrostatic sweeping. According to formula (2), the radius of the central trajectory is given by

$$r_0 = \frac{1}{B}\sqrt{\frac{2M_0U_0}{q}}, \quad (6)$$

where U_0 is the initial accelerating voltage.

During sweeping, ions passing along the central trajectory have various masses. At the end of sweeping, ions passing along the central trajectory will have an

1063-7842/03/4801-0128$24.00 © 2003 MAIK "Nauka/Interperiodica"

index $\tilde{\gamma}$. The radius of their trajectory is

$$r = \frac{1}{B}\sqrt{\frac{2M_0(1+\tilde{\gamma})U_0(1+\beta)}{q}} = r_0\sqrt{(1+\tilde{\gamma})(1+\beta)}, \quad (7)$$

where $\beta = (\Delta U)/U_0$ is the relative change in the accelerating voltage during sweeping.

In view of the fact that radius (7) at the end of sweeping coincides with the central trajectory radius ($r = r_0$),

$$(1+\tilde{\gamma})(1+\beta) = 1, \quad (8)$$

we determine the relative shift of the spectrum due to the change in the accelerating voltage:

$$\tilde{\gamma} = -\frac{\beta}{1+\beta}. \quad (9)$$

Since the change in the accelerating voltage is the same for all ions irrespective of their masses, the relative shift of the spectrum is the same for any trajectory in a given mass range. From this condition, it is easy to draw up the rule for collector arrangement. We have

$$\begin{aligned} M_1 &= M_0(1+\tilde{\gamma}), \\ M_2 &= M_1(1+\tilde{\gamma}) = M_0(1+\tilde{\gamma})^2, \\ &\cdots \\ M_n &= M_0(1+\tilde{\gamma})^n. \end{aligned} \quad (10)$$

At the same time, according to (3),

$$M_n = M_0(1+\gamma_n). \quad (11)$$

Equating the right-hand sides of the last equation in (10) and of Eq. (11), we come to the rule for collector arrangement:

$$\gamma_n = (1+\beta)^{-n} - 1. \quad (12)$$

From the last equation in (10), we can find the desired number of collectors:

$$\frac{M_n}{M_0} = (1+\tilde{\gamma})^n. \quad (13)$$

Taking the logarithm of both sides yields

$$n = \frac{\ln M_0 - \ln M_n}{\ln(1+\beta)}. \quad (14)$$

Then, with the limits of a specified mass range ($M_0 = M_{\min}$ and $M_n = M_{\max}$), we have

$$N = \frac{\ln M_{\min} - \ln M_{\max}}{\ln(1+\beta)} \quad (15)$$

and take the integer part of N.

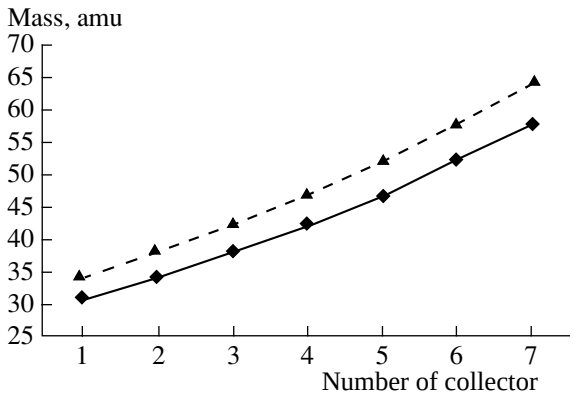

Fig. 1. Mass spectra at the (rhombus) beginning and (triangles) end of electrostatic sweeping. The curves were obtained for a symmetric mass analyzer. $\varphi_m = 130°$, $r_m = 60$ mm, entrance angle 52°, $U_0 = 2240$ V, and $\Delta U = -0.1U_0$.

Figure 1 shows mass spectra at the beginning and end of electrostatic sweeping. When simulating, we used the program described in [3].

For a mass spectrum to be reliable, each line must be within the range of an appropriate collector for a time τ. The time τ corresponds to the time constant of an electrometer and/or integration time of an analog-to-digital converter. Using our model, we choose $\Delta\gamma$ such that peaks with the indices γ and $\gamma + \Delta\gamma$ do not fall into one collector simultaneously (Fig. 2); then, the equality

$$\Delta x - \Delta w = s_2 \quad (16)$$

is valid, where Δx is the center spacing of the peaks with the indices γ and $\gamma + \Delta\gamma$, Δw is the width of the peak, and s_2 is the slit width of the collector.

In formula (16), all the parameters are expressed through the central trajectory radius.

Thus, $\Delta\gamma$ is the maximal shift of a mass spectrum with which reliable recording is provided. For the set of collectors at hand, we take the shift $\Delta\gamma$ with the least magnitude. According to (8), the value of $\Delta\gamma_{\min}$ can be found by measuring the maximal change in the accelerating voltage $\beta_{\max}$.

Electrostatic sweeping is most easily accomplished by a discharge of an additional capacitor inserted in the lens voltage divider of the source (Fig. 3). According to the similarity principle for ion–optical systems, the voltages of the source lenses will vary in proportion and beam focusing will be completely retained. In this case, the voltage across a resistor is known to vary by the law [4]

$$U(t) = U_0\exp\left(-\frac{1}{RC}t\right), \quad (17)$$

where U_0 is the voltage across the capacitor before the discharge, C is the capacitor value, R is the resistance

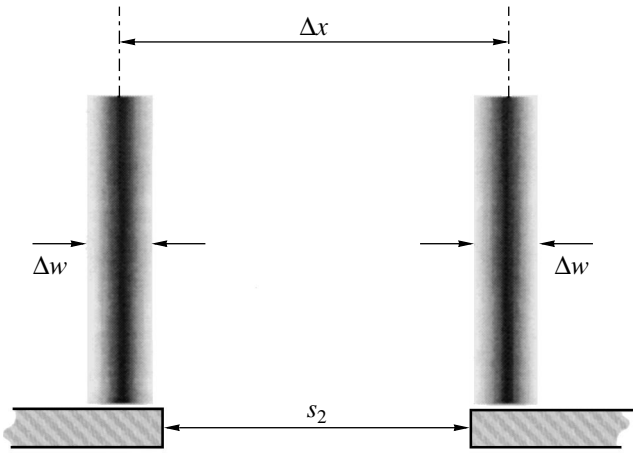

Fig. 2.

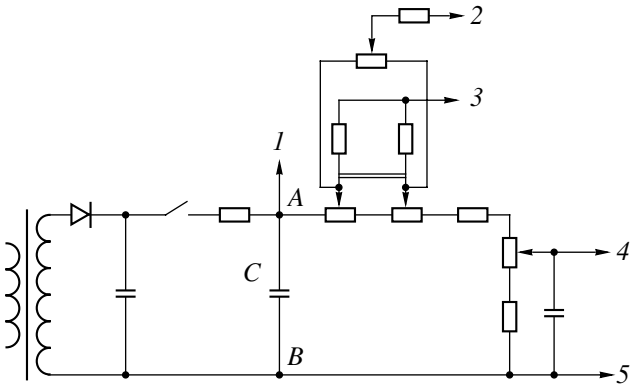

Fig. 3. Typical lens voltage divider in an electron-impact source. (*1*) Accelerating voltage, (*2*) focusing voltage, (*3*) extraction voltage, (*4*) ionization voltage, and (*5*) applied voltage.

of the lens voltage divider between points *A* and *B*, and *t* is the time from the beginning of the discharge.

As follows from the reliable recording rule,

$$\Delta U(\tau) = U(\tau) - U_0 = \beta_{\max} U_0. \qquad (18)$$

Then, the value of the parameter *RC* is found from the formula

$$RC = -\frac{\tau}{\ln(1+\beta_{\max})}. \qquad (19)$$

Knowing the parameter *RC* and total change in the accelerating voltage during sweeping, we find the time taken to record the entire spectrum:

$$T = -RC\ln\left(1+\frac{\Delta U}{U_0}\right). \qquad (20)$$

In our example, for an integration time of the analog-to-digital converter $\tau = 25$ ms and a collector slit width $s_2 = 400$ μm, the time required to record the entire spectrum is $T \approx 1.9$ s.

CONCLUSION

The sweep rate is a very important parameter for portable mass spectrometers with a chromatographic interface. With this parameter known, requirements for the gas-delivery and pumping systems of a mass spectrometer can be worked out. Moreover, using the approach suggested in this paper, along with the program used in [3], one can find all the parameters needed to design multicollector recording systems.

Our approach makes it possible to determine the maximum allowable sweep rate for a mass spectrum. Reliable recording is provided if the *RC* value is taken 5–10% greater than the calculated one.

REFERENCES

1. Yu. V. Brazhnik and V. F. Shkurdoda, Ukr. Fiz. Zh. **45**, 1015 (2000).
2. A. S. Kuzema, O. R. Savin, and I. Ya. Chertkov, *Analyzing Systems of Magnetic Mass Spectrometers* (Naukova Dumka, Kiev, 1987).
3. O. N. Peregudov, V. F. Shkurdoda, and L. F. Sukhodub, Zh. Tekh. Fiz. **72** (6), 141 (2002) [Tech. Phys. **47**, 792 (2002)].
4. V. I. Kalinin and G. M. Gershteĭn, *Introduction to Radiophysics* (GITTL, 1957).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 1, 2003, pp. 131–132. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 140–142.

BRIEF COMMUNICATIONS

Long-Term Stability of Long-Wavelength (>1.25 μm) Quantum-Dot Lasers Fabricated on GaAs Substrates

E. Yu. Lundina*, Yu. M. Shernyakov*, M. V. Maksimov*, I. N. Kayander*, A. F. Tsatsul'nikov*, N. N. Ledentsov*, A. E. Zhukov*, N. A. Maleev*, S. S. Mikhrin*, V. M. Ustinov*, Zh. I. Alferov*, and D. Bimberg**

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

*** Institut für Feskörphysik, Technische Universität Berlin, D-10623 Berlin, Germany*

e-mail: maximov@beam.ioffe.ru

Received June 24, 2002

Abstract—Accelerated degradation testing of long-wavelength (>1.25 μm) quantum-dot lasers made on GaAs substrates is carried out at a fixed current of 1.7 A, initial optical output of about 0.3 W, and a heat sink temperature of 60°C. No signs of degradation are revealed after testing for 450 h. The test bed is not sealed, inert gas purging is not performed, and the laser faces are not passivated. *© 2003 MAIK "Nauka/Interperiodica".*

Since the first demonstration of lasing through the ground state of quantum dots produced by self-organization [1], quantum-dot lasers have attracted much attention throughout the world [2, 3]. An advantage of quantum-dot lasers is the suppression of carrier transport in the lateral direction towards defects, dislocations, and the laser faces owing to the strong localization of electrons and holes within the dots. This may improve the operating stability and radiation resistance of quantum-dot lasers in comparison with those of quantum-well lasers. Recently, Ribbat *et al.* [4] demonstrated that quantum-dot lasers are much less sensitive to irradiation by a proton beam than quantum-well lasers. Note that quantum dots make it possible to lase at a wavelength of 1.3 μm in structures grown on a GaAs substrate [2, 3, 5, 6]. The maximum cw output of such lasers is 2.7 W [7, 8]. Surface-emitting lasers built around a vertical microcavity with InAs/GaAs quantum dots grown on a GaAs substrate have been shown to lase in the wavelength interval 1.295–1.305 μm with an internal differential efficiency of 64% and external efficiency of 20% [6].

In this work, we study the stability of a long-wavelength quantum-dot laser and demonstrate that this laser is more stable than a quantum-well laser fabricated in the same MBE equipment and tested under the same conditions.

The object studied was an MBE-grown double-heterostructure laser with separate confinement of the electrons and the light wave. The lasing area was sandwiched in 1.5-μm-thick $Al_xGa_{1-x}As$ emitting layers with an effective aluminum content of 75%. The emitters were doped by Si and Be to reach the concentrations of electrons and holes at a level of 5×10^{17} cm^{-3}.

The lasing area consisted of three quantum-dot layers separated by GaAs spacers with a thickness of 33 nm. Quantum dots in the lasing area were made by the activated decomposition of the solid solution. Initial islands formed by the deposition of 2.7 InAs monolayers were covered by an $In_{0.15}Ga_{0.85}As$ layer with a mean thickness of 5 nm. Indium atoms of the overlayer diffuse toward the islands, whose lattice parameter is close to that of bulk InAs, whereas Ga atoms diffuse toward the areas between the islands, where the lattice parameter is close to that of bulk GaAs. This process leads to the activated decomposition of InGaAs, increases the effective volume of the islands, and causes the red shift of the photoluminescence band. The quantum dots covered become embedded in an InGaAs quantum well, which also contributes to the red shift of the photoluminescence band because of the stress field redistribution in the dots and the narrowing of the matrix band gap. The density of dots measured by transmission electron microscopy was found to be 4×10^{10} cm^{-2}.

The lasers had a stripe geometry with a stripe width of 100 μm (shallow mesa) and a stripe length of 2000 μm. The cleaved faces were neither passivated nor covered by an insulating coating. The lasers were soldered onto copper heat sinks with an indium solder. We did not use inert gas purging; that is, the degradation characteristics were studied under room conditions. In degradation tests, the laser current in the cw mode was stabilized.

The differential efficiency of the quantum-dot lasers near the lasing threshold is typically about 42% in a temperature range of 15–60°C. The quantum efficiency and the internal losses determined from the dependence of the inverse differential efficiency on the stripe length were 95 ± 5% and 7 cm^{-1}, respectively. The threshold

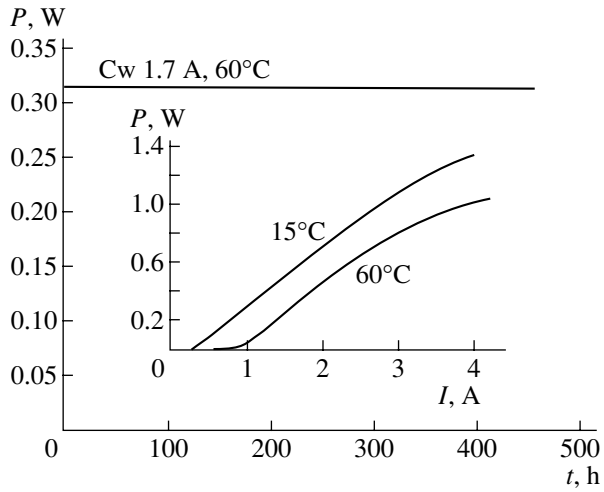

Fig. 1. Cw quantum-dot laser output versus time at 60°C. During the test, the laser current was stabilized. The inset shows the watt–ampere characteristics at heat sink temperatures of 15 and 60°C. Stripe length and width are 2000 and 100 μm, respectively.

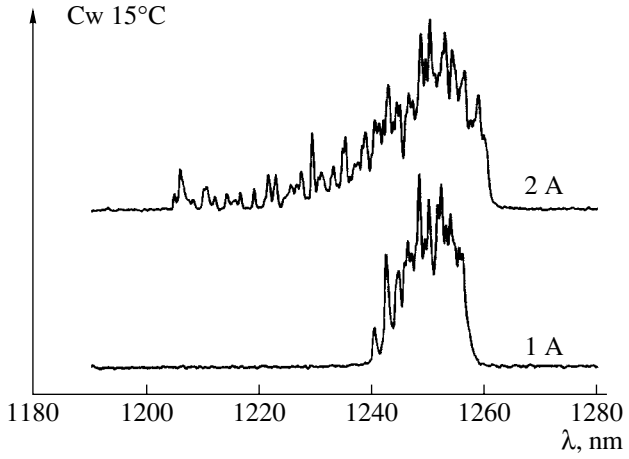

Fig. 2. Lasing spectra of the quantum-dot laser for two currents. The heat sink temperature is 15°C.

current density equaled 150 and 340 A/cm^2 at heat sink temperatures of 15 and 60°C, respectively. The accelerated degradation test was performed at a temperature of 60°C and current of 1.7 A, which corresponds to a laser output of 0.28 W at the beginning of testing. Figure 1 shows a typical time dependence of the quantum-dot laser output for the current 1.7 A. After testing for 450 h, we took the spectral characteristics of the lasers. Figure 2 demonstrates the electroluminescence spectra in the cw lasing mode at two values of the current for 15°C. Any variations in the laser wavelength after the degradation tests were not observed.

We also applied degradation tests to conventional GaAs/AlGaAs quantum-well lasers fabricated in the same MBE equipment. The output power decreases by more than 10% after the 100-h tests in this case presumably because of the degradation of the unprotected cleaved faces of the laser diodes. The higher reliability of the quantum-dot lasers is explained by the weak diffusion of the carriers toward the cleaved faces. Indeed, the lateral diffusion of electrons and holes in quantum dots is strongly suppressed owing to their severe confinement in three dimensions. This decreases sharply the overheat of the laser faces and inhibits chemical reactions on them stimulated by the recombination of free carriers.

At present, we are planning to apply degradation tests to quantum-dot lasers with mirror coatings on their faces and use inert gas purging of the test bed. Note, however, that even the early results of the accelerated degradation tests presented suggest a high stability of quantum-dot lasers.

CONCLUSION

Long-wavelength (>1.25 μm) quantum-dot lasers produced by the activated decomposition of the solid solution are shown to be highly stable. Our results are in agreement with the high radiation resistance of quantum-dot lasers exposed to high-energy protons.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, Program “Physics of Solid-State Nanostructures,” INTAS, and NanOp.

M.V. Maksimov acknowledges the financial support of the Alexander von Humboldt-Stiftung.

A.E. Zhukov and V.M. Ustinov thank the Foundation for Support of National Science.

REFERENCES

1. N. N. Ledentsov, V. M. Ustinov, A. Yu. Egorov, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **28**, 1484 (1994) [Semiconductors **28**, 832 (1994)].
2. D. L. Huffaker, G. Park, Z. Zou, *et al.*, IEEE J. Sel. Top. Quantum Electron. **6**, 452 (2000).
3. A. Stintz, G. T. Lui, H. Li, *et al.*, IEEE Photonics Technol. Lett. **12**, 591 (2000).
4. C. Ribbat, R. Sellin, M. Grundmann, *et al.*, Electron. Lett. **37**, 174 (2001).
5. N. N. Ledentsov, M. Grundmann, F. Heinrichsdorff, *et al.*, IEEE J. Sel. Top. Quantum Electron. **6**, 439 (2000).
6. N. N. Ledentsov, D. Bimberg, V. M. Ustinov, *et al.*, in *Proceedings of International Conference on Indium Phosphide and Related Materials, Nara, Japan, 2001*, p. 5.
7. N. N. Ledentsov, Springer Tracts Mod. Phys. **156**, 3 (1999).
8. Yu. M. Shernyakov, D. A. Bedarev, E. Yu. Kondrat’eva, *et al.*, Electron. Lett. **35**, 898 (1999).

Translated by A. Chikishev

Technical Physics, Vol. 48, No. 1, 2003, pp. 15–25. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 19–29.

THEORETICAL AND MATHEMATICAL PHYSICS

On Unsteady-State Heat-and-Mass Transfer in Polyatomic Gases

S. A. Savkov, A. A. Yushkanov, and Yu. I. Yalamov

Orel State University, Orel, 302015 Russia

Received March 19, 2002; in final form, July 5, 2002

Abstract—The problem of a nonstationary source of heat or particles in a polyatomic gas is solved analytically. The temperature and concentration distributions vs. distance to the source are constructed. It is shown that a nonstationary heat source may generate acoustic vibration. The solution obtained can be helpful in studying acoustic wave propagation, temperature and concentration fields produced by heat or particle sources of various configuration (in particular, by a laser beam in an absorbing media), etc. © *2003 MAIK "Nauka/Interperiodica".*

Interest in transfer phenomena in molecular gases has increased over the past few years [1]. However, analytical solutions have been derived only for stationary problems [2–6] and the nonstationary kinetic equation has been solved with regard only for the translational motion of gas molecules [7–9].

In this work, a nonstationary kinetic equation for a polyatomic gas is considered for the first time.

The contribution of internal degrees of freedom is known (see, e.g., [10]) to depend on their energy spectrum. The energy spacing between rotational degrees of freedom is defined by the ratio $\hbar^2/2J$ (J is the moment of inertia of a molecule) and is comparable to the energy of thermal motion kT only for the lightest gases. Specifically, for hydrogen molecules, $\hbar^2/2Jk = 85.4$ K. For polyatomic gases, this value is much lower. Therefore, one can ignore the discreteness of the rotational motion energy and consider rotational degrees of freedom in the classical approximation. As for vibrational degrees of freedom, they are excited at temperatures near 1000 K and so can be considered to be completely frozen.

Following [11], we start with the solution of the one-dimensional equation

$$\frac{\partial f}{\partial t} + \mathbf{V}_x \frac{\partial f}{\partial x} = J_{\mathrm{col}}[f].$$

Here, f is the distribution function, $\mathbf{V}$ is the intrinsic translational velocity of gas molecules, and J_{col} is the integral collision operator.

Leaving only linear effects, we represent the solution to this equation in the form

$$f = f_0(1 + \varphi),$$

where φ is a correction to the equilibrium (Maxwell) distribution function f_0. For a polyatomic gas, this function is given by

$$f_0 = n_0\left(\frac{m}{2\pi k T_0}\right)^{3/2}\prod_{i=1}^{3}\left(\frac{J_i}{2\pi k T_0}\right)^{1/2}\exp(-C^2-\gamma^2),$$

where

$$\mathbf{C} = \mathbf{V}\left(\frac{m}{2kT_0}\right)^{1/2}, \quad \gamma = \left(\sum_{i=1}^{3}\frac{J_i\omega_i^2}{2kT_0}\right)^{1/2},$$

m, J_i, and ω_i are the mass, principal moments of inertia, and related components of the angular velocity of molecules; k is the Boltzmann constant; and T_0 and n_0 are the undisturbed temperature and concentration values.

Since reliable data for the nature of intermolecular interaction in molecular gases are lacking, we will restrict our analysis to an analogue of the Bhatnagar–Gross–Krook model of collision integral [3]. Since the relaxation times for translational and rotational motions differ insignificantly for most gases under normal conditions, we will assume them to be the same.

Let x and t be measured in terms of

$$l = \chi\sqrt{\frac{2m}{kT}} \quad \text{and} \quad \tau = \frac{\chi m}{kT},$$

where l and τ are the mean free path and time of gas molecules. Then,

$$\frac{\partial \varphi}{\partial t} + C_x\frac{\partial \varphi}{\partial x} = \sum_{i=1}^{3} P_i M_i - \varphi. \qquad (1)$$

Here,

$$M_i = \pi^{-3}\int P_i \varphi \exp(-C^2-\gamma^2)\,d^3\gamma\,d^3C,$$

1063-7842/03/4801-0015$24.00

$$P_1 = 1, \quad P_2 = \frac{1}{\sqrt{3}}(C^2 + \gamma^2 - 3), \quad P_3 = \sqrt{2}C_x,$$

and χ is the thermal diffusivity of the gas.

Putting $C_x = \mu$, we represent φ in the form

$$\varphi = e_1 Y_1(t, x, \mu) + e_2 Y_2(t, x, \mu),$$

where

$$e_1 = 1, \quad e_2 = \frac{1}{\nu}(C^2 - \mu^2 + \gamma^2 - \nu^2), \quad \nu = \sqrt{5/2}. \quad (2)$$

As a result, (1) is reduced to an integro-differential equation for the vector $\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$:

$$\left(\frac{\partial}{\partial t} + \mu\frac{\partial}{\partial x} + 1\right)\mathbf{Y}(t, x, \mu)$$

$$= \pi^{-1/2}\int_{-\infty}^{+\infty}\mathbf{K}(\mu, \mu_1)\mathbf{Y}(t, x, \mu_1)\exp(-\mu_1^2)d\mu_1,$$

$$\mathbf{K}(\mu, \mu_1) = \begin{bmatrix} 1 + 2\mu\mu_1 + \frac{1}{3}\left(\mu^2 - \frac{1}{2}\right)\left(\mu_1^2 - \frac{1}{2}\right) & \frac{\nu}{3}\left(\mu^2 - \frac{1}{2}\right) \\ \frac{\nu}{3}\left(\mu_1^2 - \frac{1}{2}\right) & \frac{\nu^2}{3} \end{bmatrix}.$$

Separating variables, we represent a solution to this equation in the from

$$\mathbf{Y}(t, x, \mu) = \exp(\sigma t - (\sigma + 1)x/\eta)\mathbf{F}(\sigma, \eta, \mu).$$

The components of the vector $\mathbf{F}$ are found from a set of characteristic equations

$$\pi^{1/2}\left(1 - \frac{\mu}{\eta}\right)(\sigma + 1)F_1 = N_1^0 + 2\mu N_1^1 + \frac{2\mu^2 - 1}{12}(2N_1^2 - N_1^0 + 2\nu N_2^0), \quad (3)$$

$$\pi^{1/2}\left(1 - \frac{\mu}{\eta}\right)(\sigma + 1)F_2 = \frac{\nu}{6}(2N_1^2 - N_1^0 + 2\nu N_2^0),$$

$$N_i^\alpha = \int_{-\infty}^{\infty} F_i\mu^\alpha\exp(-\mu^2)d\mu. \quad (4)$$

According to [12], we express the higher order moments of the function F_1 that enter into (3) and (4) through N_1^0. To this end, we multiply (3) by $\exp(-\mu^2)$ and (4) by $\mu\exp(-\mu^2)$ and integrate over the range of μ. Solving the resulting set of equation yields

$$N_1^\alpha = N_1^0\left(\frac{\eta\sigma}{\sigma + 1}\right)^\alpha.$$

Thus, Eqs. (3) and (4) can also be represented in vector form:

$$\pi^{1/2}\eta\mathbf{F} = (\sigma + 1)(\eta - \mu)\Delta \cdot \mathbf{N}, \quad (5)$$

$$\Delta = \begin{bmatrix} 1 + 2\frac{\mu\sigma\eta}{\sigma + 1} + \frac{2\mu^2 - 1}{6}\left(\left(\frac{\sigma\eta}{\sigma + 1}\right)^2 - \frac{1}{2}\right) & \frac{\nu}{6}(2\mu^2 - 1) \\ \frac{\nu}{3}\left(\left(\frac{\sigma\eta}{\sigma + 1}\right)^2 - \frac{1}{2}\right) & \frac{\nu^2}{3} \end{bmatrix}, \quad (6)$$

$$\mathbf{N} = \begin{bmatrix} N_1^0 \\ N_2^0 \end{bmatrix} = \int_{-\infty}^{+\infty}\exp(-\mu^2)\mathbf{F}d\mu. \quad (7)$$

Equality (7) can be viewed as the normalization condition for the function $\mathbf{F}$.

If η is not a real number, we find from (5)

$$\mathbf{F} = \frac{\pi^{1/2}\eta}{(\eta - \mu)(\sigma + 1)}\Delta \cdot \mathbf{N}. \quad (8)$$

The values of η corresponding to this solution are found from condition (7). Then,

$$\mathbf{N} = \frac{\pi^{-1/2}\eta}{\sigma + 1}\int_{-\infty}^{+\infty}\frac{\Delta}{\eta - \mu}\exp(-\mu^2)d\mu \cdot \mathbf{N}$$

or (in the equivalent form)

$$\det\Lambda(\sigma, \eta) = 0. \quad (9)$$

Here,

$$\Lambda(\sigma, \eta) = (\sigma + 1)\mathbf{E} - \pi^{-1/2}\eta\int_{-\infty}^{+\infty}\frac{\Delta}{\eta - \mu}\exp(-\mu^2)d\mu$$

$$= \begin{bmatrix} \sigma + 1 + \lambda_c(\eta) + \frac{2\eta^2\sigma}{\sigma + 1}(\lambda_c(\eta) + 1) + \lambda_1 l & \nu\lambda_1 \\ \frac{\nu}{3}\lambda_c(\eta)l & \sigma + 1 + \frac{5}{6}\lambda_c(\eta) \end{bmatrix}, \quad (10)$$

$$\lambda_1 = \frac{\lambda_c(\eta)(2\eta^2 - 1) + 2\eta^2}{6}, \quad l = \left(\frac{\sigma\eta}{\sigma + 1}\right)^2 - \frac{1}{2},$$

$$\lambda_c(z) = \pi^{-1/2} z \int_{-\infty}^{+\infty} \exp(-\mu^2) \frac{d\mu}{\mu - z},$$

and **E** is the unit matrix.

The normalizing vector **N** can be determined accurate to an arbitrary constant:

$$\mathbf{N} = \text{const} \begin{bmatrix} \Lambda_{22} \\ -\Lambda_{21} \end{bmatrix}. \tag{11}$$

To solve dispersion relation (9), we turn to the theory of boundary-value problems for functions of complex variable (see, e.g., [13]). Note that the function $D(z) = \det\Lambda(\sigma, z)$ is an even piecewise analytic function in the complex plane with a cut along the real axis. Let us designate the restrictions of this function on the upper and lower half-planes as D^+ and D^-, respectively, and consider the Riemann homogeneous boundary-value problem

$$X^+(x) = G(x)X^-(x), \quad x \in \mathbb{R} \tag{12}$$

with the coefficient $G(x) = D^+(x)/D^-(x)$, where $D^\pm(x) = \lim_{y \to \pm 0} D(x + iy)$.

By virtue of the Liouville generalized theorem, the general solution to problem (12) is given by

$$D^\pm(z) = A(z + i)^{-2\kappa} X^\pm(z) \prod_{\alpha = 1}^{\kappa} (\eta_\alpha^2 - z^2),$$

where

$$X^\pm(z) = \left(\frac{z + i}{z \pm i}\right)^{2\kappa} \exp(\Gamma(z)),$$

$$\Gamma(z) = \frac{1}{2\pi i} \int_{-\infty}^{+\infty} \ln\left(\left(\frac{\mu + i}{\mu - i}\right)^{2\kappa} G(\mu)\right) \frac{d\mu}{\mu - z},$$

$$\kappa = \frac{1}{2}\text{ind}G(x) = \frac{1}{2\pi}[\arg G(x)]_{\mathbb{R}_+}, \quad \mathbb{R}_+ \in (0, \infty).$$

Thus, to find η_α, it is sufficient to calculate $D^\pm$ and $X^\pm$ for $(\kappa + 1)$ arbitrary values of z and solve the resulting set of equations for these quantities and constant A. The values of $\pm\eta_\alpha$ thus found are the desired roots of Eq. (9).

In view of (11), a discrete solution spectrum can be represented in the form

$$\mathbf{F}(\sigma, \eta_\alpha, \mu) = \frac{\pi^{-1/2}\eta_\alpha}{(\eta_\alpha - \mu)(\sigma + 1)} \begin{bmatrix} \Lambda_{22}\Delta_{11} - \Lambda_{21}\Delta_{12} \\ \Lambda_{22}\Delta_{21} - \Lambda_{21}\Delta_{22} \end{bmatrix}, \tag{13}$$

where Λ_{ij} and Δ_{ij} are the components of matrices (10) and (6) taken at $\eta = \eta_\alpha$.

If the values of η are real, solutions to Eqs. (5) and (7) are the functions

$$\Phi(\sigma, \eta, \mu) = \left(\pi^{-1/2} \frac{\eta}{\eta - \mu} \Delta + \exp(\eta^2)\Lambda\delta(\eta - \mu)\right) \frac{\mathbf{N}}{\sigma + 1}, \tag{14}$$

which constitute a continuous solution spectrum. All integrals of function (14) should be taken in the sense of the principal value of the Cauchy integral.

Because of the arbitrariness of the normalizing vector **N**, solutions (14) can be represented as the superposition of two independent functions

$$\Phi_1 = \pi^{-1/2} \frac{\eta}{\eta - \mu} \begin{bmatrix} \mu^2 - \frac{1}{2} \\ \nu \end{bmatrix} + \exp(\eta^2)\delta(\eta - \mu) \begin{bmatrix} \lambda_p(\eta)\left(\eta^2 - \frac{1}{2}\right) + \eta^2 \\ \frac{3}{\nu}(\sigma + 1) + \nu\lambda(\eta) \end{bmatrix},$$

$$\Phi_2 = \pi^{-1/2} \frac{\eta}{\eta - \mu} \begin{bmatrix} 1 + \frac{2\mu\eta\sigma}{\sigma + 1} \\ 0 \end{bmatrix} + \exp(\eta^2)\delta(\eta - \mu) \times \begin{bmatrix} \sigma + 1 + \lambda_p(\eta) + 2(\lambda_p(\eta) + 1)\frac{\eta^2\sigma}{\sigma + 1} \\ \frac{\sigma + 1}{2\nu} - \frac{\sigma^2\eta^2}{(\sigma + 1)\nu} \end{bmatrix},$$

where

$$\lambda_p(\eta) = -2\eta \exp(-\eta^2) \int_0^\eta \exp(\mu^2) d\mu.$$

One can prove (see, e.g., [11]) that the set of solutions obtained is a complete set of orthogonal functions satisfying the following conditions:

$$\begin{aligned} &\int_{-\infty}^{+\infty} \mathbf{F}(\sigma, \eta_\alpha, \mu)\mathbf{F}(\sigma, \eta_\beta, \mu)\exp(-\mu^2)\mu d\mu = \delta_{\alpha\beta}N_\alpha, \\ &\int_{-\infty}^{+\infty} \mathbf{F}(\sigma, \eta_\alpha, \mu)\Phi_\beta(\sigma, \eta, \mu)\exp(-\mu^2)\mu d\mu = 0, \\ &\int_{-\infty}^{+\infty} \mathbf{X}_\alpha(\sigma, \eta', \mu)\Phi_\beta(\sigma, \eta, \mu)\exp(-\mu^2)\mu d\mu \\ &\qquad = \delta_{\alpha\beta}\delta(\eta - \eta')N_0, \end{aligned} \tag{15}$$

where

$$\mathbf{X}_1 = N_{11}\Phi_1 - N_{12}\Phi_2, \quad \mathbf{X}_2 = N_{22}\Phi_2 - N_{12}\Phi_1,$$

$$N_{11} = \left(\sigma + 1 + \lambda_p(\eta) + 2(\lambda_p(\eta) + 1)\frac{\sigma\eta^2}{\sigma + 1}\right)^2$$

$$+ \frac{2}{5}\left(\frac{\sigma + 1}{2} - \frac{\eta^2\sigma^2}{\sigma + 1}\right)^2 + \pi\eta^2\exp(-2\eta^2)\left(1 + \frac{2\sigma\eta^2}{\sigma + 1}\right),$$

$$N_{12} = \left(\frac{\sigma + 1}{2} - \frac{\eta^2\sigma^2}{\sigma + 1}\right)\left(\frac{6}{5}(\sigma + 1) + \lambda_p(\eta)\right)$$

$$+ \left(\sigma + 1 + \lambda_p(\eta) + 2(\lambda_p(\eta) + 1)\frac{\sigma\eta^2}{\sigma + 1}\right)$$

$$\times\left(\lambda_p(\eta)\left(\eta^2 - \frac{1}{2}\right) + \eta^2\right)$$

$$+ \pi\eta^2\exp(-2\eta^2)\left(\eta^2 - \frac{1}{2}\right)\left(1 + \frac{2\sigma\eta^2}{\sigma + 1}\right),$$

$$N_{22} = \left(\lambda_p(\eta)\left(\eta^2 - \frac{1}{2}\right) + \eta^2\right)^2 + \frac{5}{2}\left(\frac{6}{5}(\sigma + 1) + \lambda_p(\eta)\right)^2$$

$$+ \pi\eta^2\exp(-2\eta^2)\left(\left(\eta^2 - \frac{1}{2}\right)^2 + \frac{5}{2}\right),$$

$$N_0 = \eta\exp(\eta^2)(N_{11}N_{22} - N_{12}^2).$$

The values of N_α are more convenient to calculate by numerical integration because the associated expressions are extremely awkward.

As an application of the solution found, let us consider an infinite planar heat source of power $W(t) = \exp(\sigma t)$ placed in the plane $x = 0$.

In this case, the distribution function can be represented as

$$\mathbf{Y}(t, x, \mu) = \mathbf{Y}^{\pm}(\sigma, x, \mu)\exp(\sigma t), \quad \text{for} \quad \pm x > 0,$$

where

$$\mathbf{Y}^{\pm}(\sigma, x, \mu) = \pm\sum_\alpha A_\alpha^{\pm}\mathbf{F}_\alpha(\sigma, \pm\eta_\alpha, \mu)\exp(\mp(\sigma + 1)x/\eta_\alpha)$$

$$+ \sum_{\beta = 0}^{2}\int_0^{\pm\infty} B_\beta\Phi_\beta(\sigma, \eta, \mu)\exp(-(\sigma + 1)x/\eta)d\eta. \tag{16}$$

Here, summation in the first term should be performed only over those α for which $\mathrm{Re}((\sigma + 1)/\eta_\alpha) > 0$.

The coefficients A and B are found from the step condition

$$\mu(\mathbf{Y}^+ - \mathbf{Y}^-) = \mathbf{S} \quad \text{at} \quad x = 0.$$

Hence, by virtue of condition (15), we find

$$A_\alpha^{\pm} = \int_{-\infty}^{+\infty}\mathbf{F}(\sigma, \pm\eta_\alpha, \mu)\mathbf{S}\exp(-\mu^2)d\mu,$$

$$B_\beta = \int_{-\infty}^{+\infty}\mathbf{X}_\beta\mathbf{S}\exp(-\mu^2)d\mu. \tag{17}$$

In our case,

$$\mathbf{S} = \mathbf{S}_h = \frac{1}{3}\begin{bmatrix}\mu^2 - \frac{1}{2}\\ \nu\end{bmatrix}.$$

Accordingly,

$$A_\alpha^{\pm} = a_\alpha^{\mathrm{h}}/N_\alpha, \quad B_1 = \frac{\sigma + 1}{N_0}(N_{11}b_1^{\mathrm{h}} - N_{12}b_2^{\mathrm{h}}),$$

$$B_2 = \frac{\sigma + 1}{N_0}(N_{22}b_2^{\mathrm{h}} - N_{12}b_1^{\mathrm{h}}),$$

$$a_\alpha^{\mathrm{h}} = \frac{1}{3}\frac{\eta_\alpha^2\sigma^2}{\sigma + 1} - \frac{\sigma + 1}{6}, \quad b_1^{\mathrm{h}} = 1, \quad b_2^{\mathrm{h}} = 0, \tag{18}$$

$$\mathbf{Y}_{\mathrm{h}}^{\pm}(\sigma, x, \mu) = \pm\sum_\alpha\frac{a_\alpha^{\mathrm{h}}}{N_\alpha}\mathbf{F}_\alpha(\sigma, \pm\eta_\alpha, \mu)\exp(\mp(\sigma + 1)x/\eta_\alpha)$$

$$+ \int_0^{\pm\infty}\frac{\sigma + 1}{N_0}\mathbf{X}_1\exp(-(\sigma + 1)x/\eta)d\eta.$$

For relative drops in temperature

$$\Delta T = \frac{T - T_0}{T_0}$$

$$= \frac{1}{3\pi^3}\int(C^2 + \gamma^2 - 3)\varphi\exp(-C^2 - \gamma^2)d^3\gamma d^3C$$

and gas molecule concentration

$$\Delta N = \frac{n - n_0}{n_0} = \pi^{-3}\int\varphi\exp(-C^2 - \gamma^2)d^3\gamma d^3C,$$

we have

$$\Delta T_{\mathrm{h}} = \pi^{-1/2}\exp(\sigma t)\left(\pm\sum_\alpha\frac{(a_\alpha^{\mathrm{h}})^2}{N_\alpha}\exp(\mp(\sigma + 1)x/\eta_\alpha)\right.$$

$$\left. + (\sigma + 1)^2\int_0^{\pm\infty}\frac{N_{11}}{N_0}\exp(-(\sigma + 1)x/\eta)d\eta\right),$$

$$\Delta N_h = \pi^{-1/2}\exp(\sigma t)\left(\pm\sum_\alpha \frac{a_\alpha^h a_\alpha^p}{N_\alpha}\exp(\mp(\sigma+1)x/\eta_\alpha)\right.$$
$$\left.-(\sigma+1)^2\int_0^{\pm\infty}\frac{N_{12}}{N_0}\exp(-(\sigma+1)x/\eta)d\eta\right).$$

If the translational and vibrational degrees of freedom,

$$S_v = \frac{2}{3}C^2 - 1 \quad \text{and} \quad S_\omega = \frac{2}{3}\gamma^2 - 1$$

are excited separately, the vectors $\mathbf{F}$ and $\mathbf{X}_\beta$ should be written in the expanded form

$$\mathbf{F} = e_1F_1 + e_2F_2, \quad \mathbf{X}_\beta = e_1X_{1\beta} + e_2X_{2\beta}.$$

Also, relationships (17) should be supplemented by integration over the remaining velocity projections:

$$A_\alpha^\pm = \pi^{-5/2}\int(e_1F_1(\sigma,\pm\eta_\alpha,\mu) + e_2F_2(\sigma,\pm\eta_\alpha,\mu))$$
$$\times S\exp(-C^2-\gamma^2)d^3\gamma d^3C,$$
$$B_\beta = \pi^{-5/2}\int(e_1X_{1\beta} + e_2X_{2\beta})S\exp(-C^2-\gamma^2)d^3\gamma d^3C.$$

Of certain interest are the temperature distributions over translational and vibrational degrees of freedom:

$$\Delta T^v = \frac{2}{3}\pi^{-3}\int\left(C^2-\frac{3}{2}\right)\varphi\exp(-C^2-\gamma^2)d^3\gamma d^3C,$$
$$\Delta T^\omega = \frac{2}{3}\pi^{-3}\int\left(\gamma^2-\frac{3}{2}\right)\varphi\exp(-C^2-\gamma^2)d^3\gamma d^3.$$

For completeness, we should also consider a particle source of type

$$\mathbf{S}_p = \begin{bmatrix}1\\0\end{bmatrix}.$$

For relative temperature and concentration drops, we have in this case

$$M_s^m(t,\sigma,x) = \pm\pi^{-1/2}\sum_\alpha\frac{a_\alpha^s a_\alpha^m}{N_\alpha}$$
$$\times\exp(\sigma t\mp(\sigma+1)x/\eta_\alpha) + \pi^{-1/2}(\sigma+1)^2$$
$$\times\int_0^{\pm\infty}\frac{N_{11}b_1^sb_1^m + N_{22}b_2^sb_2^m - N_{12}(b_1^sb_2^m + b_2^sb_1^m)}{N_0}\exp(\sigma t-(\sigma+1)x/\eta)d\eta. \quad (19)$$

Hereafter, superscript m specifies the distribution function moment as follows: $m = 1$, ΔT; $m = 2$, ΔT^v; $m = 3$, ΔT^ω; and $m = 4$, ΔN. Superscript s specifies the source type: $s = 1$, S_h; $s = 2$, S_v; $s = 3$, S_ω; and $s = 4$, S_p. Also,

$$b_1^2 = \frac{4}{5}, \quad b_1^3 = \frac{6}{5}, \quad b_2^4 = 1, \quad b_1^4 = 0,$$
$$b_2^2 = -b_2^3 = \frac{2}{5}\left(\frac{\sigma\eta}{\sigma+1}\right)^2 - \frac{1}{5},$$
$$a_\alpha^2 = \frac{8(\sigma+1)+\lambda_c(\eta_\alpha)}{24}\left(2\left(\frac{\eta_\alpha\sigma}{\sigma+1}\right)^2-1\right),$$
$$a_\alpha^3 = \frac{\lambda_c(\eta_\alpha)}{24}\left(1-2\left(\frac{\eta_\alpha\sigma}{\sigma+1}\right)^2\right), \quad a_\alpha^4 = \sigma+1+\frac{4}{5}\lambda_c(\eta_\alpha).$$

The coefficients a_α^1, b_1^1, and b_2^1 are found from relationship (18).

Note in passing that the moments listed above are symmetric with respect to index permutation. In other words, the relative drop, e.g., of concentration in the vicinity of the heat source, $\Delta N_h = M_1^4$, coincides with the temperature distribution near the particle source, $\Delta T_p = M_4^1$.

Clearly, a planar source can be considered as a set of isotropic point sources. Hence, the distribution of any scalar quantity ρ_{pl} from a planar source can be expressed through the distribution of this quantity from a point source, ρ_{pt}:

$$\rho_{pl}(x) = \int\rho_{pt}(r)d\sigma = 2\pi\int_x^\infty r\rho_{pt}(r)dr,$$

where r is the distance between a surface element $d\sigma$ and a selected point in space.

Hence, it follows that

$$\rho_{pt}(r) = -\frac{1}{2\pi r}\frac{d\rho_{pl}(r)}{dr}.$$

Thus, for an isotropic point source,

$$M_s^m(t,\sigma,r) = \frac{\sigma+1}{2\pi^{3/2}r}\sum_\alpha\frac{a_\alpha^s a_\alpha^m}{\eta_\alpha N_\alpha}$$
$$\times\exp(\sigma t-(\sigma+1)r/\eta_\alpha) + \frac{(\sigma+1)^3}{2\pi^{3/2}r}$$
$$\times\int_0^{\pm\infty}\frac{N_{11}b_1^sb_1^m + N_{22}b_2^sb_2^m - N_{12}(b_1^sb_2^m + b_2^sb_1^m)}{\eta N_0}$$
$$\times\exp(\sigma t-(\sigma+1)r/\eta)d\eta. \quad (20)$$

Let us analyze the solution obtained.

Figure 1 shows the regions C_1, C_2, and C_3 of the parameter σ where the dispersion relation has, respectively, two, four, and six roots. For the negative half-

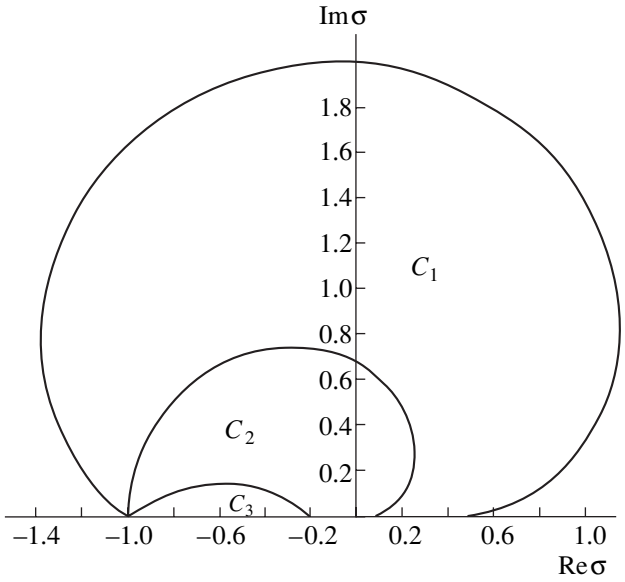

Fig. 1. Ranges of σ where the dispersion relation has two (C_1), four (C_2), and six (C_3) roots of the discrete spectrum.

space Imσ, the pattern is symmetric about the real axis. It was shown numerically that as σ approaches the boundary of these regions from the inside, the imaginary part of one pair of the roots tends to zero and the associated solutions turn to the solutions of the continuous spectrum, with the general solution (the sum of the discrete and continuous spectra) remaining a continuous function of σ.

If σ is small, the solutions of the continuous spectrum pass to the related solutions of the stationary equation and the eigenvalues of the discrete spectrum η_α can be found in explicit form. Indeed, substituting the obvious asymptotic representation

$$\lambda_c(z) = -\sum_{n=0}^{\infty} \frac{(2n-1)!!}{2^n z^{2n}}$$

into dispersion relation (9) yields (up to first nonvanishing terms in z)

$$\frac{1}{3\eta^4} - \frac{2\sigma}{3\eta^4} + \sigma^3 = 0.$$

Hence,

$$\eta_{\pm 1} = \pm\sqrt{2/3}\sigma^{-1}, \quad \eta_{\pm 2} = \pm(2\sigma)^{-1/2}.$$

Accordingly,

$$N_{\pm 1}\sqrt{\pi}\eta_{\pm 1} = \frac{4}{81\sigma}, \quad N_{\pm 2}\sqrt{\pi}\eta_{\pm 2} = \frac{1}{9},$$

$$a_{\pm 1}^1 = 1/18, \quad a_{\pm 1}^4 = a_{\pm 2}^4 = -a_{\pm 2}^1 = 1/6,$$

$$a_{\pm 1}^2 = 2/72, \quad a_{\pm 2}^2 = -7/24,$$

$$a_{\pm 1}^3 = 1/72, \quad a_{\pm 2}^3 = -1/24.$$

Thus, in the limit $\sigma \longrightarrow 0$, we have

$$\Delta T_h = \Delta T_h^{as} + I_{11},$$

$$\Delta T_h^v = \Delta T_v = \Delta T_h^{as} + \frac{4I_{11} + I_{12}}{5},$$

$$\Delta T_h^\omega = \Delta T_\omega = \Delta T_h^{as} + \frac{6I_{11} - I_{12}}{5},$$

$$\Delta T_p = \Delta T_p^{as} - I_{12},$$

$$\Delta T_v^v = \Delta T_h^{as} + \frac{16I_{11} + 8I_{12} + I_{22}}{25},$$

$$\Delta T_v^\omega = \Delta T_\omega^v = \Delta T_h^{as} + \frac{24I_{11} + 2I_{12} - I_{22}}{25},$$

$$\Delta T_\omega^\omega = \Delta T_h^{as} + \frac{25I_{11} - 10I_{12} + I_{22}}{16}, \qquad (21)$$

$$\Delta T_p^v = \Delta T_p^{as} - \frac{4I_{12} + I_{22}}{5}, \quad \Delta T_p^\omega = \Delta T_p^{as} - \frac{4I_{12} - I_{22}}{5},$$

$$\Delta N_h = \Delta N_h^{as} - I_{12}, \quad \Delta N_v = \Delta N_h^{as} - \frac{4I_{12} + I_{22}}{5},$$

$$\Delta N_\omega = \Delta N_h^{as} - \frac{6I_{12} - I_{22}}{5}, \quad \Delta N_p = \Delta N_p^{as} + I_{22},$$

where

$$I_{ij} = \frac{1}{2\pi^{3/2} r}\int_0^\infty \frac{N_{ij}}{\eta N_0}\exp(-r/\eta)d\eta.$$

The first terms in expressions (21) are defined by that solution of the discrete spectrum to which the eigenvalue η_2 corresponds. Here, the functions

$$\Delta T_h^{as} = -\Delta N_h^{as} = \frac{1}{8\pi r}$$

describe the temperature and concentration distributions of gas molecules that are produced by the stationary point heat source at a sufficiently large distance from it and are independent of the way the energy is excited, and

$$\Delta N_p^{as} = -\Delta T_p^{as} = \frac{1}{8\pi r}$$

are the asymptotic temperature and concentration distributions of gas molecules that are produced by the stationary heat source.

For an arbitrary time dependence of the source power, the power can be represented by the Fourier

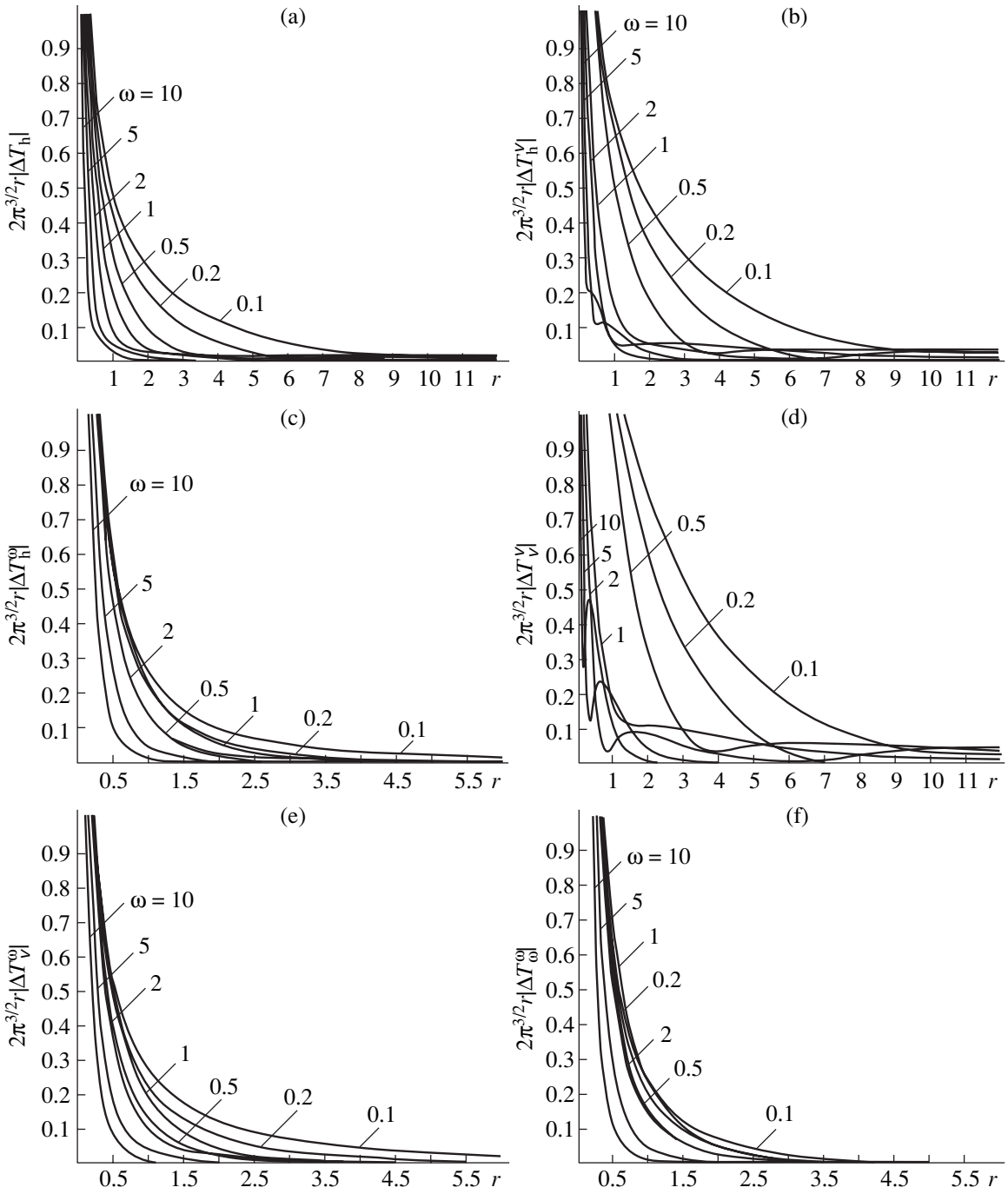

Fig. 2. Relative temperature and density distributions in the gas for fixed values of $\sigma = i\omega$.

integral

$$W(t) = \int_{-\infty}^{+\infty} W_\omega \exp(i\omega t)\, d\omega,$$

$$W_\omega = \frac{1}{2\pi}\int_{-\infty}^{+\infty} W(t)\exp(-i\omega t)\, dt.$$

Then, owing to the linearity of the problem, one can consider the distribution M_s^m as a distribution of the related quantities that are produced by individual harmonics:

$$M_s^m(t, r) = \int_{-\infty}^{+\infty} W_\omega M_s^m(t, i\omega, r)\exp(i\omega t)\, dt.$$

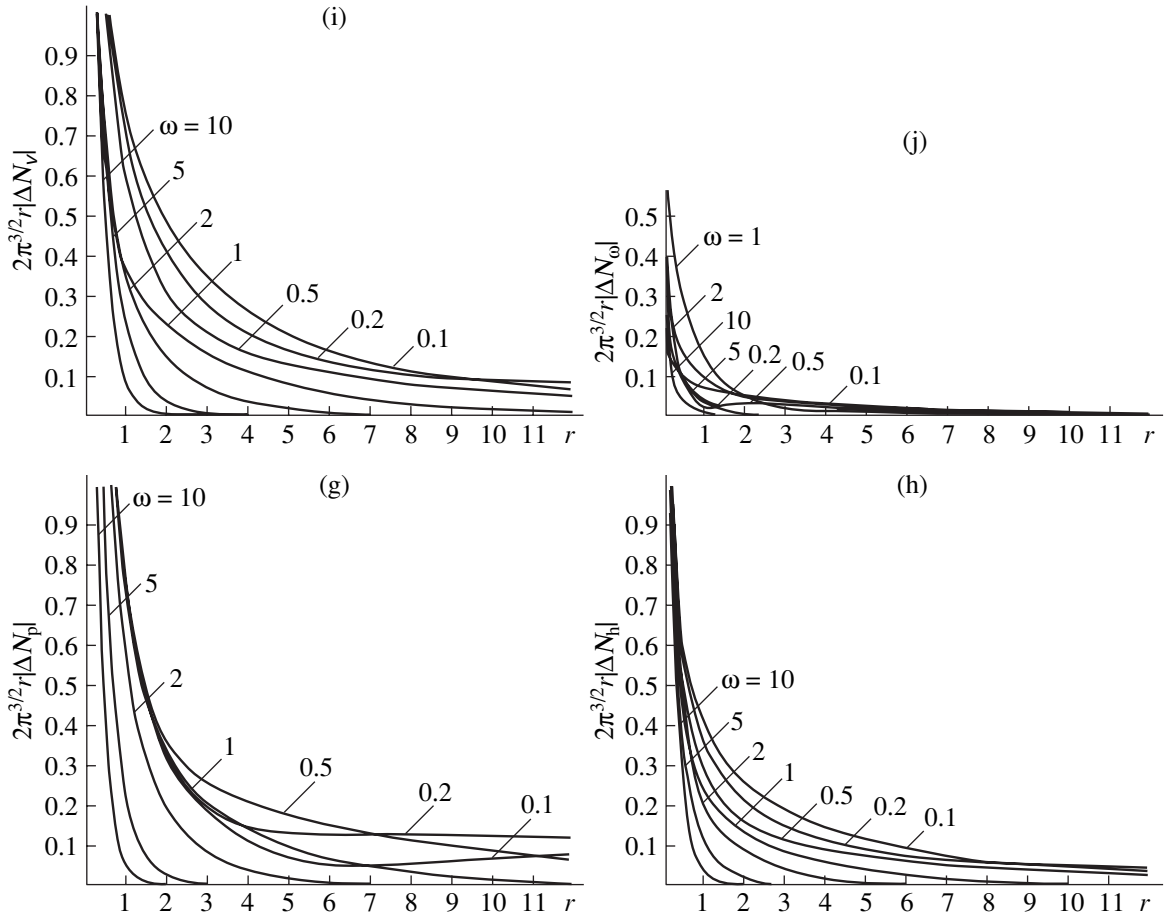

Fig. 2. (Contd.)

This fact emphasizes the special importance of considering the distribution M_s^m for pure imaginary values $\sigma = i\omega$ represented in Fig. 2.

As follows from the plots in Fig. 2, the solutions of the continuous spectrum make a major contribution to expression (20) in the neighborhood of the source ($r \ll |\sigma + 1|$). In this case, the integrals depend on small parameters η, for which

$$\lambda_p = 0, \quad N_{11} = \frac{11}{10}(\sigma + 1)^2,$$

$$N_{12} = \frac{3}{5}(\sigma + 1)^2, \quad N_{22} = \frac{18}{5}(\sigma + 1)^2,$$

$$N_0 = \frac{18}{5}(\sigma + 1)^4\eta, \quad \int_0^\infty \exp(-(\sigma + 1)r/\eta)\frac{d\eta}{\eta^2} = \frac{1}{r(\sigma + 1)}.$$

Accordingly,

$$\Delta T_h = \frac{11}{72\pi^{3/2}r^2}, \quad \Delta T_h^\omega = \Delta T_\omega = \frac{1}{6\pi^{3/2}r^2},$$

$$\Delta N_h = \Delta T_p = -\frac{1}{12\pi^{3/2}r^2},$$

$$\Delta T_v = \Delta T_v^v = \Delta T_h^v = \Delta T_v^\omega$$

$$= \Delta T_\omega^v = -\Delta N_v = -\Delta T_p^v = \frac{5}{36\pi^{3/2}r^2}.$$

Thus, in the neighborhood of the source, the distribution of most of the moments is independent of σ and depends only on its instantaneous power. The only exception is $\Delta T_p^\omega = \Delta N_\omega$. In the limit considered, their values are defined by the next terms of the expansion in η, whose contribution to the distribution of these moments varies as $1/r$.

Away from the source, the second term in expression (20) decays faster than the first one. Therefore, in the limit $r \longrightarrow \infty$, the distribution M_s^m depends on the solutions (if any) of the discrete spectrum. The rate of decay of these solutions depends on η_α. Then, the solution making a greater contribution to the distribution of macroscopic parameters in the neighborhood of the source decays faster. Depending on a specific value of

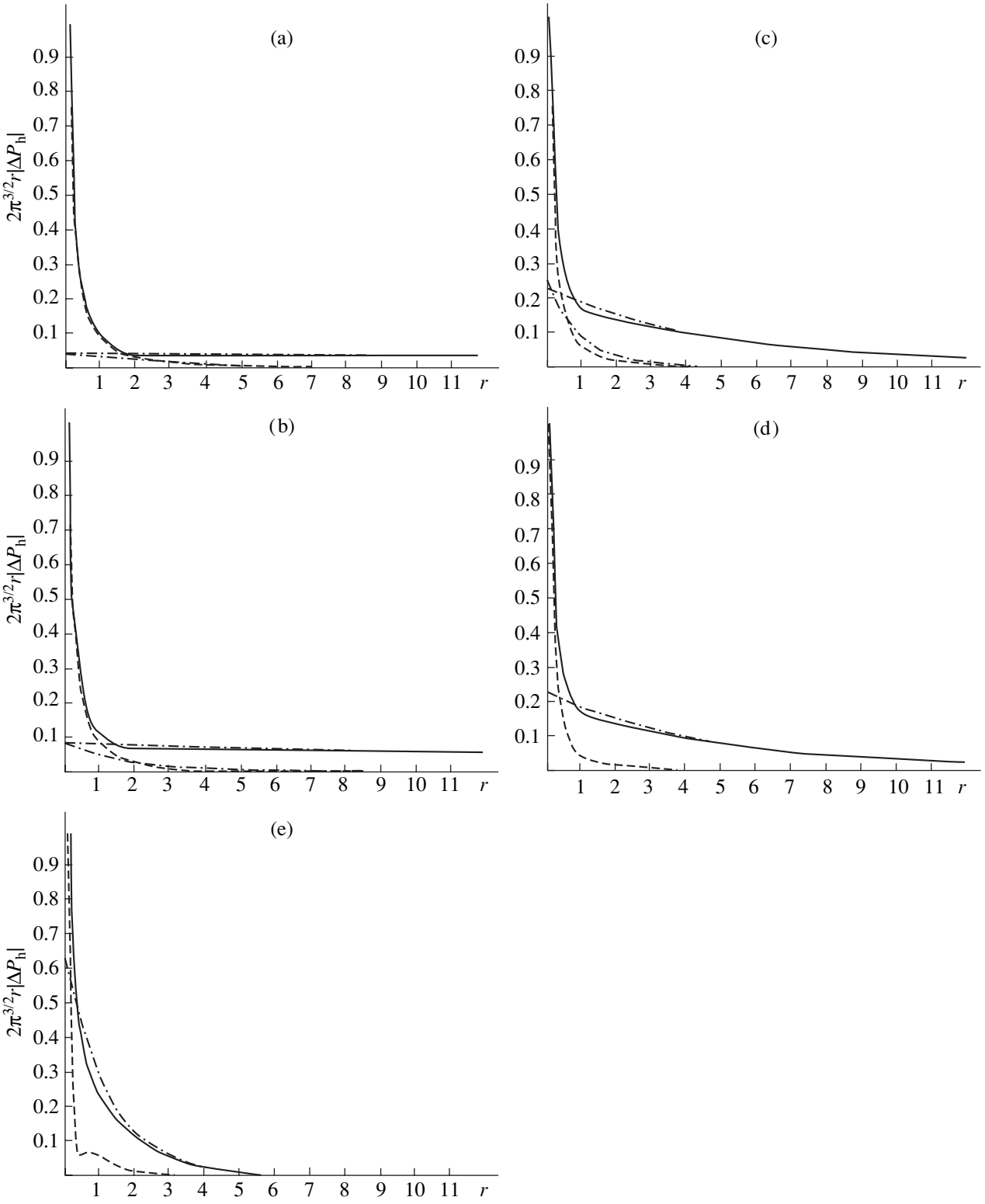

Fig. 3. Relative pressure distribution (solid curve) for ω = (a) 0.1, (b) 0.2, (c) 0.671, (d) 0.672, and (e) 1.991.

ω, the solutions indicated make inphase or opposite-phase contributions to the distribution M_s^m, which specifies the presence or absence of an inflection in the plots.

Of independent interest is the relative pressure drop of the gas

$$\Delta P = \frac{P - P_0}{P_0} = \Delta T + \Delta N.$$

Table 1. Parameters of the wave corresponding to η_1

ω	η_1	A	k	d	δ_0
0.001	1.42886 – 816.497*i*	0.00079	0.00122	0.00000	0.49936π
0.01	1.42875 – 81.6499*i*	0.00785	0.01225	0.00009	0.49364π
0.1	1.41851 – 8.16438*i*	0.07627	0.12096	0.00877	0.43976π
0.2	1.39647 – 4.06906*i*	0.14367	0.23495	0.03148	0.39308π
0.3	1.37388 – 2.68515*i*	0.20329	0.34046	0.06247	0.35883π
0.4	1.35441 – 1.97676*i*	0.25846	0.43861	0.09817	0.33254π
0.5	1.33832 – 1.53905*i*	0.31135	0.53085	0.13674	0.31128π
0.6	1.32505 – 1.23764*i*	0.36320	0.61830	0.17718	0.29340π
0.671	1.31699 – 1.07345*i*	0.39978	0.67797	0.20670	0.28221π
0.672	1.31689 – 1.07136*i*	0.40029	0.67880	0.20712	0.28205π
0.7	1.31398 – 1.01497*i*	0.41471	0.70184	0.21892	0.27793π
0.8	1.30461 – 0.84221*i*	0.46628	0.78210	0.26162	0.26427π
0.9	1.29658 – 0.70324*i*	0.51818	0.85958	0.30504	0.25204π
1.0	1.28961 – 0.58834*i*	0.57056	0.93466	0.34902	0.24097π
1.1	1.28348 – 0.49125*i*	0.62352	1.00764	0.39346	0.23084π
1.2	1.27803 – 0.40777*i*	0.67714	1.07877	0.43826	0.22151π
1.3	1.27316 – 0.33495*i*	0.73147	0.14825	0.48336	0.21287π
1.4	1.26875 – 0.27067*i*	0.78653	0.21624	0.52871	0.20483π
1.5	1.26475 – 0.21334*i*	0.84234	1.28288	0.57427	0.19730π
1.6	1.26108 – 0.16177*i*	0.89893	1.34829	0.62001	0.19023π
1.7	1.25771 – 0.11503*i*	0.95630	0.41257	0.66591	0.18357π
1.8	1.25459 – 0.07238*i*	1.01446	1.47581	0.71194	0.17728π
1.9	1.25168 – 0.03324*i*	1.07340	1.53809	0.75808	0.17131π
1.991	1.24921 – 0.00028*i*	1.12772	1.59398	0.80015	0.16615π

Table 2. Parameters of the wave corresponding to η_2

ω	η_2	A	k	d	δ_0
0.001	15.8399 – 15.7837*i*	0.00079	0.03160	0.03165	−0.49992π
0.01	5.08528 – 4.91060*i*	0.00786	0.09928	0.10078	−0.49918π
0.1	1.78961 – 1.28682*i*	0.07783	0.30169	0.34185	−0.48830π
0.2	1.35997 – 0.71992*i*	0.15012	0.41892	0.51355	−0.48294π
0.3	1.17165 – 0.44780*i*	0.21788	0.50804	0.65933	−0.48483π
0.4	1.06031 – 0.27551*i*	0.28358	0.58295	0.79165	−0.48990π
0.5	0.98479 – 0.15189*i*	0.34862	0.64890	0.91536	−0.49622π
0.6	0.92928 – 0.05661*i*	0.41375	0.70858	1.03294	−0.50291π
0.671	0.89773 – 0.00048*i*	0.46029	0.74803	1.11352	−0.50767π

By virtue of (21), the pressure distribution in the limit $\omega \longrightarrow 0$ depends on the solutions of the continuous spectrum. As the frequency increases, the contribution from the discrete spectrum grows, the solution corresponding to η_1 decaying much more slowly. Therefore, at distances exceeding several free paths of gas molecules, the pressure is given by

$$\Delta P_s = \frac{i\omega + 1}{2\pi^{3/2} r} \frac{a_1^s (a_1^1 + a_1^4)}{\eta_1 N_1} \exp\left(i\omega t - \frac{i\omega + 1}{\eta_1} r \right).$$

This expression can be viewed as the equation for a

spherical acoustic wave,

$$\Delta P = \frac{A}{2\pi r}\exp(i(\omega t - kr) - dr + \delta_0)$$

with an amplitude $A = \pi^{-1/2}|(i\omega + 1)a_1^s(a_1^1 + a_1^4)/(\eta_1 N_1)|$, wavenumber $k = \mathrm{Im}((i\omega + 1)/\eta_1)$, damping constant $d = \mathrm{Re}((i\omega + 1)/\eta_1)$, and initial phase $\delta_0 = \arg((i\omega + 1)a_1^s(a_1^1 + a_1^4)/(\eta_1 N_1))$.

For a source $S = S_h(s = 1)$, the above values are listed in Table 1. For comparison, Table 2 shows the parameters of the wave corresponding to $\eta = \eta_2$. Figure 3 demonstrates the pressure distribution curves. Continuous ones correspond to the values $2\pi^{3/2}r|\Delta P_h|$. The contributions from the continuous and discrete spectra are depicted by dashed and dash-and-dot curves, respectively.

A similar pattern is observed for sources of other types.

As follows from the tables and figures, the amplitude and the rate of damping of each of the waves increase with frequency. As ω approaches the boundary of the region C_1, the damping of the acoustic wave tends to that of the solutions of the continuous spectrum. Thus, pure acoustic vibrations can be separated in the vicinity of ω = 2.

REFERENCES

1. V. M. Zhdanov and M. Ya. Alievskiĭ, *Transport and Relaxation Processes in Molecular Gases* (Nauka, Moscow, 1989).
2. A. V. Latyshev and A. A. Yushkanov, Teor. Mat. Fiz. **95**, 530 (1993).
3. A. V. Latyshev and A. A. Yushkanov, Zh. Éksp. Teor. Fiz. **114**, 956 (1998) [JETP **87**, 518 (1998)].
4. A. B. Poddoskin and A. A. Yushkanov, Izv. Akad. Nauk, Mekh. Zhidk. Gaza, No. 5, 182 (1998).
5. A. B. Poddoskin, A. A. Yushkanov, and Yu. I. Yalamov, Izv. Akad. Nauk, Mekh. Zhidk. Gaza, No. 4, 163 (1999).
6. A. V. Latyshev and A. A. Yushkanov, Prikl. Mekh. Tekh. Fiz. **42** (3), 91 (2001).
7. C. Cercignani, *Theory and Application of the Boltzmann Equation* (American Elsevier, New York, 1975; Mir, Moscow, 1978).
8. A. V. Latyshev and A. A. Yushkanov, Teor. Mat. Fiz. **92**, 127 (1992).
9. S. A. Savkov and A. A. Yushkanov, Teor. Mat. Fiz. **113**, 139 (1997).
10. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics* (Nauka, Moscow, 1976; Pergamon, Oxford, 1980), Part 1.
11. K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, 1967; Mir, Moscow, 1972).
12. A. V. Latyshev, Izv. Akad. Nauk, Mekh. Zhidk. Gaza, No. 2, 151 (1992).
13. F. D. Gakhov, *Boundary Value Problems* (Nauka, Moscow, 1977; Addison-Wesley, Reading, 1966).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 1, 2003, pp. 26–30. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 30–34.

THEORETICAL AND MATHEMATICAL PHYSICS

Nonlinear Magnetic Phenomena in Atomic Bose Gas

E. Orlenko*, I. Mazets, and B. Matisov***

** St. Petersburg State Technical University,*
ul. Politekhnicheskaya 29, St. Petersburg, 195251 Russia
e-mail: quark@citadel.stu.neva.ru
*** Ioffe Physicotechnical Institute, Russian Academy of Sciences,*
Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia

Received May 22, 2002

Abstract—Optically confined ($j = 1$, $j = 2$)-Bose gas is studied. The spin Hamiltonian and the ground state of a set of particles with spins $j = 1$ and $j = 2$ are obtained. Nonlinear magnetic phenomena in the set of particles with $j = 1$ are considered. *© 2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The observation and description of quantum atomic systems with internal degrees of freedom have recently become the mainstream in the study of Bose–Einstein condensate. Optical traps used for the confinement of such a system do not break the degree of freedom that is due to the atomic spin. This enables the observation of macroscopic quantum phenomena related to spin coorientation in a cooled quantum gas experiencing interaction. Basically, only atoms in the lowest multiplet state are confined, while others are rejected from traps due to spin-flip scattering. Hyperfine multiplets ($j = 2$ and $j = 1$) of ^{23}Na and ^{87}Rb are normal in the sense that the larger spin has a higher energy. Therefore, spin-2 Bose gas in ^{23}Na is difficult to obtain. At the same time, much weaker spin-flip effects in ^{87}Rb make it promising in the production of optically confined spin-2 Bose condensate. The lowest multiplet in ^{85}Rb has the spin $j = 2$; therefore, this system of atoms may also serve as the basis for the $j = 2$-condensate [1].

Systems with nonzero atomic spin are of interest in that, when in the state of Bose condensate, they may have several, rather than one, ground states. The so-called fragmentation of the ground state occurs. For $j = 1$-systems, the ground state is singlet, although magnetic fluctuations resulting in the appearance of the so-called magnetic, or cyclic, phase are very strong [2]. The aforesaid was supported experimentally in [3]. As for systems with $j = 2$, their ground state is ferromagnetic [4].

The theoretical description of the ground states in the above systems is based on the point interaction approximation, where the general form of the interaction operator is given by

$$V(|\mathbf{r}_1 - \mathbf{r}_2|) = \delta(\mathbf{r}_1 - \mathbf{r}_2)\sum_{j=0}^{2\gamma} g_j P_j,$$

where $g_j = 4\pi\hbar^2 a_j/M$, M is the mass of an atom, a_j is the amplitude of scattering in the s wave in a channel with the total spin j, and P_j is the projector of a pair of atoms onto a state with a particular spin j.

A system of bosons interacting pairwise is described by a model Hamiltonian one term of which depends on the total spin of the system in the form

$$\sum_n C_n(\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2)^n,$$

where j_1 and j_2 are the atomic spins and the parentheses mean the scalar product.

This term accounts for nonlinear phenomena due to the magnetic fluctuation of the ground state in a system of atoms.

In this study, we are interested mainly in magnetic phenomena that may arise in pure form in exotic spin systems with an integer spin without regard for the specific features of the Bose condensate state. Wave properties of atoms at low temperatures are taken into consideration upon the construction of the symmetric wave function, which is used to calculate the constant of interatomic exchange interaction in the system. Then, from first principles, we derive the spin Hamiltonian for a system of interacting ($j = 1$, $j = 2$)-spin atoms that is similar to the Heisenberg Hamiltonian for particles with $s = 1/2$. The interaction constant is calculated by the perturbation theory with allowance for the nonlocal character of atomic interaction but without using the point approximation. The Hamiltonian thus obtained includes high powers of the product $n(\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2)^n$. The presence of such nonlinear terms may give rise to nonlinear spin waves in atomic systems with the spin $j = 1$ and, in particular, to a soliton. The soliton is stable if the nonpoint potential varies slowly within the soliton length. Atomic systems with the spin $j = 2$ are charac-

1063-7842/03/4801-0026$24.00 © 2003 MAIK "Nauka/Interperiodica"

terized by fourth-order nonlinearity and have a truly ferromagnetic ground state.

EXCHANGE INTERACTION BETWEEN TWO ATOMS WITH $j = 1$

To begin with, let us consider a pair of interacting atoms that are bosons with the spin $j = 1$. The first-order energy correction for atomic interaction calculated within the exchange theory of perturbations (ETP), which takes into consideration the indistinguishability of particles, has the form [5]

$$E^{(1)} = K \pm A, \tag{1}$$

where K and A are the direct and exchange contributions to the interaction energy, respectively.

The coefficients ± 1 in (1) correspond to the symmetric (+1) or antisymmetric (–1) coordinate part of the wave function. The full wave function of a two-boson system, which is the product of the coordinate and spin parts, must be totally symmetric under permutations; i.e., the coordinate and spin parts must be symmetric or antisymmetric functions simultaneously:

$$\Psi(1, 2) = \Psi^{s}_{a}(\mathbf{r}_1, \mathbf{r}_2)\Upsilon^{s}_{a}(\xi_1, \xi_2). \tag{2}$$

Here, the indices "s" and "a" denote symmetric and antisymmetric linear combinations and $\Psi(\mathbf{r}_1, \mathbf{r}_2)$ and $\Upsilon(\xi_1, \xi_2)$ are the coordinate and spin parts of the wave function, respectively.

The coordinate part of the two-particle function written in the center-of-mass system appears as

$$\Psi^{s}_{a}(\mathbf{R}) = \left[\left(e^{ik_z z} + \frac{f(\Theta)e^{ikR}}{R}\right) \pm \left(e^{-ikz} + \frac{f(\pi - \Theta)e^{ikR}}{R}\right)\right]N, \tag{3}$$

where $\mathbf{R} = \mathbf{r}_1 - \mathbf{r}_2$ is the radius vector connecting the centers of the atoms, $\mathbf{k}$ is the wave vector, $f(\Theta)$ is the scattering amplitude, and N is the normalizing factor. At low temperatures, $f(\Theta) = \text{const}(\Theta)$ (s-scattering).

In a more compact form,

$$|\Psi^{s}_{a}(\mathbf{R})\rangle = |\Phi(\mathbf{R})\rangle \pm |\Phi(\mathbf{R})\rangle. \tag{4}$$

Then, taking the interaction potential in the Suserland model in the form $\hat{V} = c/R^6$ if $R > r_0$ or $\hat{V} = \infty$ if $R < r_0$ (c is the van der Waals constant, and r_0 is the hard sphere radius characterizing the maximum possible approach of atoms), we come to

$$K = (\Phi|\hat{V}|\Phi); \quad A = (\Phi^{|}|\hat{V}|\Phi).$$

For the detailed calculation of the interaction constants, see [6].

SPIN HAMILTONIAN

The spin part of the two-particle function of bosons with the spin $j = 1$ is structured as follows: linear combinations that are symmetric under permutations correspond to states with an even total spin of a pair of atoms ($\sigma = 2$ and $\sigma = 0$; see Appendix), and antisymmetric functions describe the state with the total spin $\sigma = 1$. For these three spin numbers, the eigenvalues of the scalar product operator $\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2$ are known:

$$\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2 = \frac{1}{2}(\hat{\sigma}^2 - \hat{j}_1^2 - \hat{j}_2^2),$$

$$\overline{\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2} = \frac{1}{2}(\sigma(\sigma + 1) - 2j(j + 1)),$$

$$\sigma = 2, \quad \overline{\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2} = 1; \quad \sigma = 1, \quad \overline{\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2} = -1;$$

$$\sigma = 0, \quad \overline{\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2} = -2.$$

Taking into consideration the symmetry properties of the above states, we construct the projection operator $\hat{P}$ so that it has eigenvalues of +1 and –1 when acting on the symmetric and antisymmetric spin function of a pair of bosons, respectively. Such an operator has the form

$$\hat{P}_{11} = \hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2 + (\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2)^2 - 1. \tag{5}$$

The substitution of this operator into Eq. (1) for ± 1 yields the energy correction for interaction in the form of an operator acting on the spin variable:

$$\hat{H}_{\text{int}} = K + A(\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2 + (\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2)^2 - 1). \tag{6}$$

To include the effect of an applied magnetic field h, Eq. (6) should be completed with the term

$$\hat{H}_0 = 2\mu_0 h(\hat{j}_{z_1} + \hat{j}_{z_2}),$$

where μ_0 is the Bohr magneton.

Now the spin Hamiltonian for a system of N Bose particles with $j = 1$ interacting in pairs can readily be obtained:

$$\hat{H}_{\text{int}} = (K - A)\frac{N(N-1)}{2} + A\sum_{i<j}[(\hat{\mathbf{j}}_i \cdot \hat{\mathbf{j}}_j)^2 + (\hat{\mathbf{j}}_i \cdot \hat{\mathbf{j}}_j)]. \tag{7}$$

The interaction of atoms with an external magnetic field adds to the energy operator the quantity

$$\hat{H}_0 = -2\mu_0 h\sum_i \hat{j}_{z_i}.$$

This spin Hamiltonian includes the quadratic term, which is responsible for nonlinear phenomena in the spin system.

GROUND STATE OF A SPIN SYSTEM

Let us derive the energy eigenvalues for a many-particle spin system.

The second sum in Eq. (7) can be expressed in terms of the operator $\hat{J}^2$ of the total spin squared:

$$\sum_{i<j}\hat{\mathbf{j}}_i\cdot\hat{\mathbf{j}}_j = \frac{1}{2}\left(\hat{J}^2-\sum_i\hat{j}_i^2\right). \tag{8}$$

The quadratic term can be rewritten through the operator $\hat{\sigma}^2$ of the paired spin of the system squared:

$$\sum_{i<j}(\hat{\mathbf{j}}_i\cdot\hat{\mathbf{j}}_j)^2 = \frac{N(N-1)}{2}\sum_\sigma(\hat{\sigma}^2-2\hat{j}^2)^2 \times\sum_{\sigma_z,\,j_{z_1},\,j_{z_2}}\langle\sigma,\sigma_z|j_{z_1}j_{z_2}j_1j_2\rangle^2, \tag{9}$$

where $\langle\sigma,\sigma_z|j_1j_{z_1}j_2j_{z_2}\rangle$ are the Clebsch–Gordan coefficients.

Acting on the $|J,J_z\rangle$ state with the total spin J and its projection J_z, operator (8) takes the eigenvalue

$$\overline{\sum_{i<j}\hat{\mathbf{j}}_i\cdot\hat{\mathbf{j}}_j} = \frac{1}{2}(J(J+1)-Nj(j+1)). \tag{10}$$

The eigenvalue of operator (9) has the form

$$\overline{\sum_{i<j}(\hat{\mathbf{j}}_i\cdot\hat{\mathbf{j}}_j)^2} = \frac{4}{3}\frac{J(J-1)}{2}. \tag{11}$$

Then, the eigenvalue of energy operator (7) is

$$E_{\text{int}} = (K-A)\frac{N(N-1)}{2} + A\left[\frac{2J(J-1)}{3}+\frac{1}{2}(J(J+1)-Nj(j+1))\right]. \tag{12}$$

For $j=1$, the interaction energy becomes

$$E_{\text{int}} = K\frac{N(N-1)}{2}-A\frac{N(N+1)}{2}+\frac{A}{6}J(7J-1). \tag{13}$$

It is evident that, at $A>0$, the ground state corresponding to the minimal energy is that with $J=0$; i.e.,

$$|0\rangle = \prod_{k<l}\frac{1}{\sqrt{3}}(|j_k=1,j_{zk}=1\rangle_k|j_l=1,j_{zl}=-1\rangle_l + |j_k=1,j_{zk}=-1\rangle_k|j_l=1,j_{zl}=1\rangle_l - |j_k=1,j_{zk}=0\rangle|j_l=1,j_{zl}=0\rangle), \tag{14}$$

which is totally symmetric.

HAMILTONIAN OF BOSONS WITH $j=2$

For a system of two bosons with the spin $j=2$, the symmetry of the terms is as follows: even states with the spins of a pair of atoms $\sigma=4$, 2, and 0 are symmetric and odd states with the spins $\sigma=3$ and 1 are antisymmetric. All eigenvalues of the operator $\hat{\mathbf{j}}_1\cdot\hat{\mathbf{j}}_2$ of the scalar product of the spins are presented below:

$$\begin{aligned}
&\sigma=4,\ \overline{\hat{\mathbf{j}}_1\cdot\hat{\mathbf{j}}_2}=4\ \text{(symmetric)},\\
&\sigma=3,\ \overline{\hat{\mathbf{j}}_1\cdot\hat{\mathbf{j}}_2}=0\ \text{(antisymmetric)},\\
&\sigma=2,\ \overline{\hat{\mathbf{j}}_1\cdot\hat{\mathbf{j}}_2}=-3\ \text{(symmetric)},\\
&\sigma=1,\ \overline{\hat{\mathbf{j}}_1\cdot\hat{\mathbf{j}}_2}=-5\ \text{(antisymmetric)},\\
&\sigma=0,\ \overline{\hat{\mathbf{j}}_1\cdot\hat{\mathbf{j}}_2}=-6\ \text{(symmetric)}.
\end{aligned} \tag{15}$$

Let us also construct an operator $\hat{P}$ taking into account the symmetry properties of the spin function of a pair of atoms in such a way that its action on a symmetric or antisymmetric state yields an eigenvalue of +1 and –1, respectively. Such an operator has the form

$$\hat{P}_{22} = \frac{1}{36}(\hat{\mathbf{j}}_1\cdot\hat{\mathbf{j}}_2)^4+\frac{1}{6}(\hat{\mathbf{j}}_1\cdot\hat{\mathbf{j}}_2)^3 -\frac{13}{36}(\hat{\mathbf{j}}_1\cdot\hat{\mathbf{j}}_2)^2-\frac{5}{2}(\hat{\mathbf{j}}_1\cdot\hat{\mathbf{j}}_2)-1. \tag{16}$$

Replacing ±1 in Eq. (1) by this operator yields the Hamiltonian for a system of N bosons with $j=2$:

$$\hat{H}_{\text{int}} = (K-A)\frac{N(N-1)}{2}+\frac{A}{6}\sum_{i<j}\left\{\frac{1}{6}(\hat{\mathbf{j}}_i\cdot\hat{\mathbf{j}}_j)^4 +(\hat{\mathbf{j}}_i\cdot\hat{\mathbf{j}}_j)^3-\frac{13}{6}(\hat{\mathbf{j}}_i\cdot\hat{\mathbf{j}}_j)^2-15(\hat{\mathbf{j}}_i\cdot\hat{\mathbf{j}}_j)-6\right\}. \tag{17}$$

It is clearly seen that, in a Bose gas with $j=2$, one may expect still more pronounced nonlinear effects due to fluctuations of a macroscopic magnetic field.

GROUND STATE OF BOSONS WITH $j=2$

The calculation of the eigenvalues of the operators involved in (17) in view of the Clebsch–Gordan coefficients gives

$$\begin{aligned}
&\overline{\sum_{i<j}(\hat{\mathbf{j}}_i\cdot\hat{\mathbf{j}}_j)^2}=6J(J-1),\quad \overline{\sum_{i<j}\hat{\mathbf{j}}_i\cdot\hat{\mathbf{j}}_j}=\frac{1}{2}(J(J+1)-6N),\\
&\overline{\sum_{i<j}(\hat{\mathbf{j}}_i\cdot\hat{\mathbf{j}}_j)^3}=-3J(J-1),\quad \overline{\sum_{i<j}(\hat{\mathbf{j}}_i\cdot\hat{\mathbf{j}}_j)^4}=\frac{588}{5}J(J-1).
\end{aligned} \tag{18}$$

Then, the interaction energy can be written as

$$E_{\text{int}} = (K - A)\frac{N(N-1)}{2} - \frac{A}{2}\left\{\frac{7J}{10}(13J + 37) - 15N\right\}. \quad (19)$$

Equation (19) suggests that if $A > 0$, the ground state of the system is ferromagnetic with the maximal total spin $J = 2N$.

The wave function of the ground state is evidently expressed as

$$|0\rangle = \frac{1}{\sqrt{N!}}\prod_i |j = 2, j_z = 2\rangle_i.$$

It seems that there is the following regularity in the realization of the ground state: odd-spin particles are condensed mostly into the antiferromagnetic state, whereas the ground state of a Bose gas consisting of even-spin particles is ferromagnetic.

SPIN DARK–BRIGHT SOLITON IN A HOMOGENEOUS BOSE GAS

Let us consider a system of cooled spin-1 atoms in an optical trap that preserves antiferromagnetic ground state (14) of this system. The Hamiltonian of the system that takes into account only exchange interaction is taken in the form of (7):

$$\hat{H} = A\left\{\sum_n [\hat{\mathbf{j}}_n \cdot \hat{\mathbf{j}}_{n+1} + (\hat{\mathbf{j}}_n \cdot \hat{\mathbf{j}}_{n+1})^2] - \frac{N(N-1)}{2}\right\}. \quad (20)$$

The ground state of this system is antiferromagnetic. Then, assuming that the interaction takes place with the nearest neighbors, we can set the spin of a pair of atoms to be equal to $\sigma = 0$ and the eigenvalue of the operator $\hat{\mathbf{j}}_n \cdot \hat{\mathbf{j}}_{n+1}$, to $\overline{\hat{\mathbf{j}}_n \cdot \hat{\mathbf{j}}_{n+1}} = -2$.

If a spin k in a chain of atoms is flipped, the operator of the excitation energy can be written in the form

$$\hat{V} = \hat{H} - \hat{E}_0 = A\{\hat{\mathbf{j}}_{k-1} \cdot \hat{\mathbf{j}}_k + \hat{\mathbf{j}}_k \cdot \hat{\mathbf{j}}_{k+1} + (\hat{\mathbf{j}}_{k-1} \cdot \hat{\mathbf{j}}_k)^2 + (\hat{\mathbf{j}}_k \cdot \hat{\mathbf{j}}_{k+1})^2 - (-2)\hat{j}_k^2 - 2\hat{j}_k^4\}, \quad (21)$$

where (considering the flip of a single spin) $\hat{E}_0 = -A(2\hat{j}_k^4 - 2\hat{j}_k^2)$.

In the semiclassical continuous approximation of magnetic moment eigenvalues, the magnetic moment of an atom may be represented by a function that smoothly varies with distance. Then, the spin $k \pm 1$ can be expanded as

$$j_{k\pm1} = j_k \pm \frac{\partial j_k}{\partial x} + \frac{a^2}{2}\frac{\partial^2 j_k}{\partial x^2},$$

where a is the mean atomic spacing. After the substitution of the above expansion into (21), the excitation energy becomes

$$V = A\left\{a^2 j_k^2 \frac{\partial^2 j_k}{\partial x^2}\left(1 + 4j_k^2 + a^2 j_k \frac{\partial^2 j_k}{\partial x^2}\right) + 2j_k^2 a^2 \left(\frac{\partial j_k}{\partial x}\right)^2\right\}. \quad (22)$$

On the other hand, in the approximation of an effective field H^* produced by all spins of the system, the excitation energy of the system can be defined as the energy of interaction of each of the spins with this field:

$$V = H^* \mu_0 j_k. \quad (23)$$

Then, each of the spins precesses (in accordance with the Bloch theorem) in the effective magnetic field

$$\frac{\partial \mathbf{m}}{\partial t} = \frac{e}{Mc}\mathbf{H}^* \times \mathbf{m}, \quad (24)$$

$$H^* = \frac{A}{\mu_0}\left[a^2 \frac{\partial^2 j_k}{\partial x^2}(1 + 4j_k^2) + 2j_k a^2 \left(\frac{\partial j_k}{\partial x}\right)^2\right]. \quad (25)$$

Applying (24) to the components of the magnetic moment vector $\mathbf{m} = \mathbf{j}\mu_0$ and assuming $m_z \gg m_x, m_y$, one can rewrite Eq. (24) in the cyclic coordinate system:

$$\dot{m}_+ = -\frac{Aa^2}{\hbar} i\left[\frac{\partial^2 m_+}{\partial x^2} + \left(\frac{(m_+)^2}{a^2\mu_0^2} - \frac{4(m_-)^2}{a^2\mu_0^2}\right)m_+\right],$$

$$\dot{m}_- = -\frac{Aa^2}{\hbar} i\left[\frac{\partial^2 m_-}{\partial x^2} + \left(\frac{(m_-)^2}{a^2\mu_0^2} - \frac{4(m_+)^2}{a^2\mu_0^2} - \Delta\right)m_-\right], \quad (26)$$

$$m_\pm = m_x \pm i m_y,$$

where

$$\Delta = \frac{A}{\mu_0^2}a.$$

The equation obtained is similar to that for a dark–bright soliton [7, 8]. Following [7], we represent a solution in the form

$$m_- = \sqrt{\frac{m_0^2 ka}{2}} e^{i\Omega t} e^{ixk\tan\alpha} \operatorname{sech}\{k(x - q(t))\},$$

$$m_+ = \sqrt{Aa^3}(i\sin\alpha + \cos\alpha \tanh\{k[x - q(t)]\}).$$

Here, $1/k$ is the soliton length, $m_z = m_0$,

$$k = \frac{1}{a}\left\{\sqrt{\frac{4m_0^2}{\mu_0^2}\cos^2\alpha + \left(\frac{m_0^2}{4\mu_0^2}\right)^2} - \frac{m_0^2}{4\mu_0^2}\right\},$$

$$\Omega = \frac{A}{\hbar}\left(1 - \frac{\tan^2\alpha}{2}\right)(ka)^2 - \Delta$$

is the precession frequency with regard to the shift, and

$$q(t) = q(0) + kt\frac{Aa^2}{\hbar}\tan\alpha$$

is the soliton coordinate.

Such integrable systems in which the free energy is conserved are known as Manakov equations [9]. The higher the velocity of a soliton, the lower the free energy of system [9]; therefore, the soliton is formally unstable (it is accelerated). However, the effect of other excitations (solitary or ordinary spin waves), which is disregarded here, does not cause dissipation [9]. With an additional inhomogeneous potential added to (26), the system becomes nonintegrable, allowing for the nontrivial interaction of a soliton with the environment. Nevertheless, if the interaction potential varies slowly within the scale of the soliton length k^{-1}, the variations (propagations) in the soliton and potential correlate and the free energy can be treated as an adiabatic invariant in this case [8].

Thus, for the spin system considered above, we will arrive at the pattern of nonlinear magnetic vortices that transform into macroscopic vortices of a magnetic field in a Bose condensate system. It is important to note that we analyzed the case when only one spin was flipped. Considering the flip of two, three, etc. spins, we will generate a set of solitons with different frequencies Ω_1, Ω_2, …, Ω_n that differ in the exchange interaction constants $\Omega_1 \sim A$, $\Omega_2 \sim 2A$, …, $\Omega_n = nA$.

ACKNOWLEDGMENTS

We are grateful to Dr. Christoph Ehricht, as well as to Peter and Monika Lobers, for support and valuable discussions.

This study was supported by the Russian Foundation for Basic Research (grant no. 02-02-17686), the program "Universities of Russia" (project no. UR.01.01.040), and the Ministry of Education of the Russian Federation (project no. E00-3.2-12).

APPENDIX

For particles with the spin $j = 1$, two-particle spin functions are as follows (the one-particle states of a kth particle are denoted as $|j, j_z\rangle_k$.

Symmetric spin states are given by

$$|J = 2, J_z = 2\rangle = |1, 1\rangle_1|1, 1\rangle_2,$$

$$|J = 2, J_z = 1\rangle = \frac{1}{\sqrt{2}}(|1, 1\rangle_1|1, 0\rangle_2 + |1, 0\rangle_1|1, 1\rangle_2),$$

$$|J = 2, J_z = 0\rangle = |1, 0\rangle_1|1, 0\rangle_2,$$

$$|J = 2, J_z = -1\rangle = \frac{1}{\sqrt{2}}(|1, -1\rangle_1|1, 0\rangle_2 + |1, 0\rangle_1|1, -1\rangle_2),$$

$$|J = 2, J_z = -2\rangle = |1, -1\rangle|1, -1\rangle,$$

$$|J = 0, J_z = 0\rangle = \frac{1}{\sqrt{3}}(|1, 1\rangle_1|1, -1\rangle_2$$

$$+ |1, -1\rangle_1|1, 1\rangle_2 - |1, 0\rangle_1|1, 0\rangle_2).$$

Antisymmetric spin states have the form

$$|J = 1, J_z = 1\rangle = \frac{1}{\sqrt{2}}(|1, 1\rangle_1|1, 0\rangle_2 - |1, 0\rangle_1|1, 1\rangle_2),$$

$$|J = 1, J_z = 0\rangle = \frac{1}{\sqrt{2}}(|1, 1\rangle_1|1, -1\rangle_2 - |1, -1\rangle_1|1, 1\rangle_2),$$

$$|J = 1, J_z = -1\rangle = \frac{1}{\sqrt{2}}(|1, -1\rangle_1|1, 0\rangle_2 - |1, 0\rangle_1|1, -1\rangle_2).$$

REFERENCES

1. K. B. Davis, M.-O. Mewes, M. R. Andrews, *et al.*, Phys. Rev. Lett. **75**, 3969 (1995).
2. Po Nozieres and D. Saint James, J. Phys. (Paris) **43**, 1133 (1982); Tin Lun Ho and Sung Kit Yip, Phys. Rev. Lett. **84**, 4031 (2000).
3. J. Stenger, Nature (London) **396**, 345 (1998).
4. T.-L. Ho, Phys. Rev. Lett. **81**, 742 (1998).
5. E. V. Orlenko and T. Yu. Latyshevskaya, Zh. Éksp. Teor. Fiz. **113**, 2139 (1998) [JETP **86**, 1167 (1998)].
6. E. V. Orlenko and A. A. Rumyantsev, Fiz. Nizk. Temp. **15**, 485 (1989) [Sov. J. Low Temp. Phys. **15**, 272 (1989)].
7. Th. Busch and J. R. Anglin, Phys. Rev. Lett. **87**, 010401 (2001).
8. S. Trillo, S. Wabnitz, E. M. Wright, and G. I. Stegeman, Opt. Lett. **13**, 871 (1988); D. N. Christodoulides, Phys. Lett. A **132**, 451 (1988); M. Shalaky and A. J. Barthelemy, IEEE J. Quantum Electron. **28**, 2736 (1992).
9. S. V. Manakov, Zh. Éksp. Teor. Fiz. **65**, 505 (1974) [Sov. Phys. JETP **38**, 248 (1974)].

Translated by A. Sidorova

Technical Physics, Vol. 48, No. 1, 2003, pp. 31–37. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 35–41.

GASES AND LIQUIDS

Damping Effect of an External Uniform Magnetic Field on Gas Bubble Vibration in a Viscous Conductive Liquid

A. P. Vasilyev

Orenburg State University, Orenburg, 460352 Russia

e-mail: post@mail.osu.ru

Received March 6, 2002; in final form, June 26, 2002

Abstract—Dynamic and heat-transfer processes accompanying the free vibrations of a gas bubble immersed in a viscous conductive liquid exposed to a uniform magnetic field are considered. Solutions to a set of equations describing bubble relaxation are obtained by numerical methods. It is shown that the magnetic field causes the fast damping of the vibration due to Joule dissipation. At the stage of vibration, the energy dissipates mainly through the Joule mechanism. At the final stage, thermal dissipation prevails. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The dynamics of gas and vapor bubbles and heat transfer between them are of interest in many branches of technology [1]. In particular, in a liquid-metal magnetohydrodynamic (MHD) compressor [2, 3], gas compression is accompanied by the dissipation of the energy of vibrating bubbles in a viscous conductive liquid exposed to a magnetic field.

Along with the well-studied mechanisms of viscous and thermal dissipations accompanying gas bubble vibration [4], the long-range mechanism of Joule dissipation is also observed in the channel of an MHD compressor. It may influence dynamic and thermal processes in a vibrating bubble, causing the fast damping of vibration. The study of these processes is aimed at improving the MHD compressor design.

Let a spherical bubble of perfect gas be placed at the center of the spherical coordinate system (r, Θ, φ). The bubble is in mechanical and thermal equilibrium with an ambient viscous incompressible conductive liquid. An external magnetic system produces a uniform magnetic field $\mathbf{B} = \{0, 0, -B\}$ (B = const) in the Cartesian system.

Under the action of a step pressure pulse excited at infinity, the bubble is disturbed from equilibrium and the relaxation process brings it into a new equilibrium state.

To describe this transient process, we will use the continuity, momentum, and energy (or heat transfer) equations; the Ohm and Joule laws; and the thermal equation of state of gas, assuming that the damping of gas bubble vibration takes place at small magnetic Reynolds numbers. The equations have the form

$$\mathrm{div}\mathbf{v}_1 = 0, \quad (1)$$

$$\rho_1\left(\frac{\partial \mathbf{v}_1}{\partial t} + (\mathbf{v}_1 \cdot \nabla)\mathbf{v}_1\right) = -\nabla p_1 + \mathbf{j} \times \mathbf{B} + \mu\nabla^2\mathbf{v}_1, \quad (2)$$

$$\rho_1 c_1\left(\frac{\partial T_1}{\partial t} + (\mathbf{v}_1 \cdot \nabla)T_1\right) = -\nabla \cdot \mathbf{q}_1 + \Phi + q_v, \quad (3)$$

$$\frac{\partial \rho_2^0}{\partial t} + \nabla \cdot (\rho_2^0 \mathbf{v}_2) = 0, \quad (4)$$

$$\rho_2^0\left(\frac{\partial \mathbf{v}_2}{\partial t} + (\mathbf{v}_2 \cdot \nabla)\mathbf{v}_2\right) = -\nabla p_2, \quad (5)$$

$$\rho_2^0\frac{d_2}{dt}\left(u_2 + \frac{\mathbf{v}_2^2}{2}\right) = -\nabla \cdot \mathbf{q}_2 + \nabla \cdot (\mathbf{p}_{2n} \cdot \mathbf{v}_2), \quad (6)$$

$$p_2 = \rho_2^0 R_2 T_2, \quad (7)$$

$$\mathbf{j} = \sigma\mathbf{v}_1 \times \mathbf{B}, \quad q_\mathrm{v} = j^2/\sigma. \quad (8)$$

Hereafter, parameters with the subscript 1 refer to the liquid phase and those with the subscript 2, to the gaseous phase. In Eqs. (1)–(8), $\mathbf{v}_i$, p_i, T_i, $\mathbf{q}_i$, c_i, and λ_i are, respectively, the velocity, pressure, temperature, heat flux density vector, mass specific heat at constant volume, and thermal conductivity of an ith phase; $\mathbf{p}_t$ is the surface tension tensor; ρ_1 and ρ_2^0 are the true densities of the liquid and gas, respectively; u_2 is the mass specific internal energy; R_2 is the specific gas constant; Φ is the dissipative function; q_v is the volume density of heat release due to Joule dissipation; μ and ν are the dynamic and kinematic viscosity coefficients of the liquid, respectively; and σ is the conductivity of the liquid.

The presence of the uniform magnetic field disturbs the spherically symmetric pattern of dynamic and thermal processes around the bubble. In fact, in the case of

1063-7842/03/4801-0031$24.00

a radial liquid flow with a velocity $\mathbf{v}_1 = \{w_1, 0, 0\}$ in a magnetic field $\mathbf{B} = -B\cos\Theta\mathbf{i}_r + B\sin\Theta\mathbf{i}_\Theta$, an electric current $\mathbf{j} = \sigma w_1 B\sin\Theta\mathbf{i}_\varphi$ is induced around the bubble. The interaction of this current with the magnetic field causes an electromagnetic force

$$\mathbf{f} = \mathbf{j}\times\mathbf{B} = -\sigma w_1 B^2\sin^2\Theta\mathbf{i}_r - \sigma w_1 B^2\cos\Theta\sin\Theta\mathbf{i}_\Theta = f_r\mathbf{i}_r + f_\Theta\mathbf{i}_\Theta,$$

where $\mathbf{i}_r$, $\mathbf{i}_\Theta$, and $\mathbf{i}_\varphi$ are the unit vectors of the physical basis in the spherical coordinate system.

The nonpotential component f_Θ of the electromagnetic force results in a meridional flow of the liquid around the bubble with a velocity w_Θ, which disturbs the purely radial flow.

When averaged over a sphere of radius r, the meridional component of the electromagnetic force is equal to zero in the case of a radial flow, and the average value of the radial component is

$$\langle f_r\rangle = \frac{1}{4\pi r^2}\int_0^\pi\int_0^{2\pi} f_r r^2\sin\Theta d\Theta d\varphi = -\frac{2}{3}\sigma w_1 B^2.$$

In the frame of the one-dimensional description of the velocity field in a viscous liquid, we neglect the meridional flow and replace the component f_r by its average value $\langle f_r\rangle$. We also average the Joule heat release: $\langle q_v\rangle = (2/3)\sigma w_1^2 B^2$.

Under these assumptions, from Eq. (1) follows an expression for the liquid velocity: $w_1(r, t) = w_{int}(t)a^2/r^2$, where w_{int} is the interface velocity and $a(t)$ is the bubble radius. Using this law, momentum equation (2), and the equilibrium condition for the interface, one can derive the equation for bubble dynamics

$$a\frac{d^2a}{dt^2} + \frac{3}{2}\left(\frac{da}{dt}\right)^2 + \frac{2}{3}\frac{\sigma B^2}{\rho_1}a\frac{da}{dt} + 4\nu\frac{1}{a}\frac{da}{dt} + 2\frac{\Sigma}{a} = \frac{p_{2int}(t) - p_\infty}{\rho_1},$$

where Σ is the surface tension and p_∞ is the liquid pressure at infinity.

This equation contains two unknown functions: the bubble radius $a(t)$ and the gas pressure $p_{2int}(t)$ at the interface.

Equation (3), which describes the heat release in the liquid phase for a given velocity $w_1(r, t)$, in view of Fourier's law for heat conduction $\mathbf{q}_i = -\lambda_i\nabla T_i$, averaged value of q_v, and calculated dissipative function Φ can be written in the form

$$\rho_1 c_1\left(\frac{\partial T_1}{\partial t} + w_1\frac{\partial T_1}{\partial r}\right) = \frac{1}{r^2}\frac{\partial}{\partial r}\left(\lambda_1 r^2\frac{\partial T_1}{\partial r}\right) + 12\mu\frac{w_1^2}{r^2} + \frac{2}{3}\sigma w_1^2 B^2.$$

Assuming that the gas pressure p_2 in the bubble is uniform, one can reduce Eqs. (4)–(6) subject to (7) to the equation for the gas velocity in the bubble [4]

$$w_2(r, t) = \frac{\gamma-1}{\gamma}\frac{1}{p_2(t)}\left(\lambda_2\frac{\partial T_2}{\partial r}\right) - \frac{r}{3\gamma p_2(t)}\frac{dp_2}{dt},$$

where γ is the adiabatic exponent; the equation for the gas pressure in the bubble

$$\frac{dp_2}{dt} = -3\gamma\frac{p_2 w_{int}}{a} - 3(\gamma-1)\frac{q_{int}}{a},$$

where

$$q_{int} = -\lambda_2\left.\frac{\partial T_2}{\partial r}\right|_{r=a}$$

is the heat flux density at the interface; and the equation for the heat release in the gas phase

$$\frac{\partial T_2}{\partial t} + w_2(r, t)\frac{\partial T_2}{\partial r} = \frac{\gamma-1}{\gamma}\frac{T_2}{p_2}\frac{1}{r^2}\frac{\partial}{\partial r}\left(\lambda_2 r^2\frac{\partial T_2}{\partial r}\right) + \frac{\gamma-1}{\gamma}\frac{T_2}{p_2}\frac{dp_2}{dt}.$$

To numerically investigate the above set of equations, we will represent it in a more convenient form by introducing the following characteristic quantities: the initial bubble radius a_0 as a length scale; the characteristic velocity $v_0 = \sqrt{p_0/\rho_1}$, where p_0 is the initial gas pressure in the bubble; the characteristic time $t_0 = a_0/v_0$; and the initial temperature T_0, K as a characteristic temperature.

Let us pass from the Euler coordinates (r, t) to the movable system of dimensionless coordinates $\eta = r/a(t)$ and $t_* = t/t_0$.

Using the introduced scales of the physical parameters and the transformation formulas [4]

$$\frac{\partial}{\partial r} = \frac{1}{a}\frac{\partial}{\partial\eta}, \quad \left(\frac{\partial}{\partial t}\right)_r = \left(\frac{\partial}{\partial t}\right)_\eta - \frac{\dot a}{a}\eta\frac{\partial}{\partial\eta},$$

we rearrange the above equations to the following set: the equation of bubble dynamics (Rayleigh equation)

$$a_*\frac{d^2a_*}{dt_*^2} + \frac{3}{2}\left(\frac{da_*}{dt_*}\right)^2 + \frac{2}{3}\frac{\mathrm{Ha}^2}{\mathrm{Re}_1}a_*\frac{da_*}{dt_*} + \frac{4}{\mathrm{Re}_1}\frac{1}{a_*}\frac{da_*}{dt_*} + \frac{2}{\mathrm{We}_1}\frac{1}{a_*} = p_* - (k+1); \tag{9}$$

the equation of heat release in the liquid

$$\frac{\partial\Theta_1}{\partial t_*} + \frac{\dot a_*}{a_*}\left(\frac{1}{\eta^2} - \eta\right)\frac{\partial\Theta_1}{\partial\eta} = \frac{1}{\mathrm{Pe}_1}\frac{1}{a_*^2}\left(\frac{\partial^2\Theta_1}{\partial\eta^2} + \frac{2}{\eta}\frac{\partial\Theta_1}{\partial\eta}\right) + 12\frac{\mathrm{Ec}_1}{\mathrm{Re}_1}\left(\frac{\dot a_*}{a_*}\right)^2\frac{1}{\eta^6} + \frac{2}{3}\frac{\mathrm{Ha}^2\mathrm{Ec}_1}{\mathrm{Re}_1}\frac{\dot a_*^2}{\eta^4}; \quad (10)$$

the equation for gas pressure in the bubble

$$\frac{dp_*}{dt_*} = 3(\gamma-1)\frac{\rho_*}{\mathrm{Pe}_2\mathrm{Ec}_2}\frac{1}{a_*^2}\frac{\partial\Theta_2}{\partial\eta}\bigg|_{\eta=1} - 3\gamma p_*\frac{\dot a_*}{a_*}; \quad (11)$$

the equation for gas velocity in the bubble

$$w_{2*}(\eta, t_*) = \frac{\rho_*}{\mathrm{Pe}_2\mathrm{Ec}_2}\frac{\gamma-1}{\gamma}\frac{1}{p_* a_*}\frac{\partial\Theta_2}{\partial\eta} - \frac{1}{3\gamma}\frac{dp_*}{dt_*}\eta; \quad (12)$$

and the equation of heat release in the gas phase

$$\frac{\partial\Theta_2}{\partial t_*} + \frac{w_{2*} - \eta\dot a_*}{a_*}\frac{\partial\Theta_2}{\partial\eta} = \frac{\gamma-1}{\gamma}\frac{\rho_*}{\mathrm{Pe}_2\mathrm{Ec}_2} \times\left(\frac{\partial^2\Theta_2}{\partial\eta^2} + \frac{2}{\eta}\frac{\partial\Theta_2}{\partial\eta}\right) + \frac{\gamma-1}{\gamma}\frac{\Theta_2}{p_*}\frac{dp_*}{dt_*}. \quad (13)$$

In these equations, $a_* = a(t)/a_0$ is the reduced bubble radius; $p_* = p(t)/p_0$ is the reduced pressure; $w_{2*} = w_2/v_0$ is the reduced gas velocity in the bubble, $\Theta_1 = T_1/T_0$ and $\Theta_2 = T_2/T_0$ are the reduced temperatures in the liquid and gas, respectively; and k is the parameter of gas disturbance, which is from the equality $\delta p_\infty = kp_0$.

Boundary conditions for the set of Eqs. (9)–(13) are defined as follows:

$$t_* = 0: a_* = 1, \quad \dot a_* = w_{a_*} = 0,$$
$$\Theta_1(\eta, 0) = 1, \quad \Theta_2(\eta, 0) = 1;$$
$$\eta = 0: \frac{\partial\Theta_2}{\partial\eta} = 0, \quad \Theta_2 < \infty, \quad w_{2_*} = 0; \quad (14)$$
$$\eta = 1: \lambda_2\frac{\partial\Theta_2}{\partial\eta} = \lambda_1\frac{\partial\Theta_1}{\partial\eta}, \quad \Theta_1(1, t_*) = \Theta_2(1, t_*);$$
$$\eta \longrightarrow \infty: w_{1_*} \longrightarrow 0, \quad \Theta_1(\infty, t_*) \longrightarrow 1.$$

Similarity criteria and numbers for this set of equations are as follows: Hartmann number $\mathrm{Ha} = a_0 B\sqrt{\sigma/\mu}$, Reynolds number $\mathrm{Re}_1 = a_0 v_0/\nu$, Weber number $\mathrm{We}_1 = \rho_1 v_0^2/(\Sigma/a_0)$, Eckert numbers $\mathrm{Ec}_1 = v_0^2/(c_1(T_0 - 273))$ and $\mathrm{Ec}_2 = v_0^2/(c_{2p}(T_0 - 273))$, Peclet numbers $\mathrm{Pe}_1 = a_0 v_0/(\lambda_1/(\rho_1 c_1))$ and $\mathrm{Pe}_2 = a_0 v_0/(\lambda_2/(\rho_{20}^0 c_{2p}))$, and reduced density $\rho_* = \rho_{20}^0/\rho_1$ (where ρ_{20}^0 is the initial density of the gas).

A solution to the set of Eqs. (9)–(13) that satisfies boundary conditions (14) was sought by numerical methods. The integration of the ordinary differential equations of the set was performed with the fourth-order Runge–Kutta method. When integrating the equation for heat release, we used an explicit finite-difference scheme. Over the spatial closed interval $0 \le \eta \le L$, where L is an infinitely remote point, we introduced an integration step $\delta\eta = L/M$, so that the current coordinate was $\eta_j = j\delta\eta$, where $j = 0, 1, \ldots, M$. Over the temporal interval $0 \le t_* < \tau$, we introduced a time step $\delta t = \tau/N$; thus, $t_{*i} = i\delta t$, where $i = 0, 1, \ldots, N$. The infinitely remote point corresponded to $L = 2$, and the time interval τ was taken to be coincident with the duration of the oscillatory stage of the process. The grid parameters M and N were selected so as to provide reliability of the computational results and still minimize the computer time. Sometimes, the calculations were compared with those obtained on a grid with the doubled parameter M. Most of the calculations were performed on a grid with $N = 5000$ and $M = 50$.

During the integration of the dynamic equation and heat release equation, the program generated data for the heat flux density on the bubble surface and the solutions at the interface were joined together in accordance with the boundary conditions of the fourth kind.

Below, we present the results of the numerical investigation into the dynamic and thermal processes accompanying gas bubble vibrations under various conditions.

The calculations were carried out for liquid gallium as a liquid phase and air as a disperse phase. The thermophysical properties were taken at 100°C. In all the cases, unless otherwise stated, the bubble radius was equal to $a_0 = 1$ mm and the initial pressure inside the bubble, $p_0 = 100$ kPa. The pressure disturbance parameter was $k = 0.5$, which corresponded to a final pressure of 150 kPa in the system. For these parameters, the characteristic time was $t_0 = 2.5 \times 10^{-4}$ s and the characteristic velocity, $v_0 = 4.05$ m/s. The similarity numbers were $\mathrm{Re}_1 = 13\,000$, $\mathrm{Pe}_1 = 3787$ and $\mathrm{Pe}_2 = 131$, $\mathrm{We}_1 = 205$, and $\mathrm{Ec}_1 = 1.2 \times 10^{-4}$ and $\mathrm{Ec}_2 = 1.54 \times 10^{-4}$. The magnetic field induction was varied from 0 to 1.5 T.

Figure 1 shows the variation of the bubble radius a_* with time t_* for Hartmann numbers Ha (*1*) 22.5, (*2*) 45.1, and (*3*) 67.7. The curves demonstrate the strong effect of the magnetic field on the duration of the vibration damping, that is, on the vibratory stage of the process. The same is confirmed by the time dependences of the reduced gas pressure p_* inside the bubble that are shown in Fig. 2 for different magnetic fields.

The heat exchange between the bubble and ambient liquid is characterized by the derivative $\partial\Theta_2/\partial\eta$ at $\eta = 1$. Its time dependences at different magnetic fields are presented in Fig. 3, which shows that the magnetic field affects the heat exchange between the gas bubble and

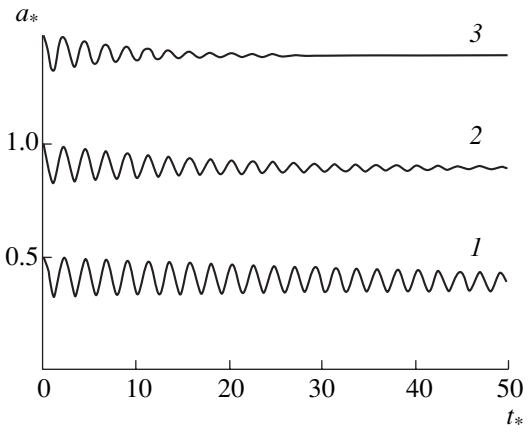

Fig. 1. Reduced radius a_* of the bubble vs. reduced time t_* for damped bubble oscillations in liquid gallium. Magnetic induction B = (*1*) 0.5, (*2*) 1.0, and (*3*) 1.5 T. Curve *1* is shifted down by –0.5, and curve *3* is shifted up by 0.5.

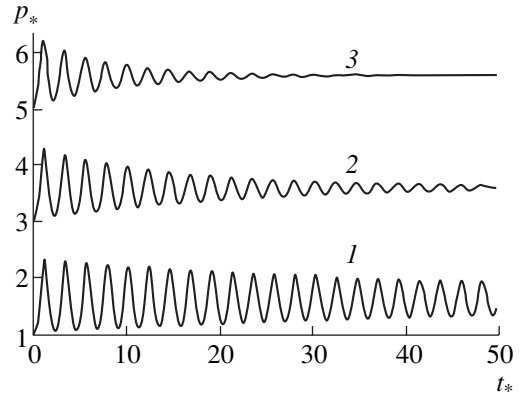

Fig. 2. Reduced pressure p_* as a function of the reduced time t_* at B = (*1*) 0.5, (*2*) 1.0, and (*3*) 1.5 T. Curve *2* is shifted up by 2 and curve *3*, by 4.

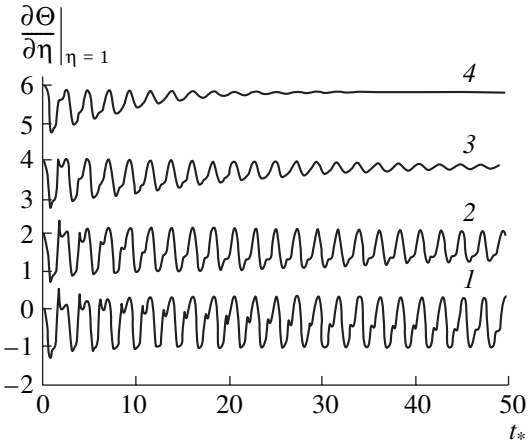

Fig. 3. Derivative of the temperature with respect to the radius on the bubble surface at different time instants t_*. B = (*1*) 0, (*2*) 0.5, (*3*) 1.0, and (*4*) 1.5 T. Curve *2* is shifted up by 2; curve *3*, by 4; and curve *4*, by 6.

the ambient liquid. Without the magnetic field (curve *1*), the derivative of the temperature with respect to the radius on the bubble surface (conventional heat flux density) is an alternating-sign function; however, its average value is negative. With time, this derivative tends to zero, remaining negative on average. A negative value of the derivative means that the heat is removed from the bubble to the ambient liquid, while a positive value of the derivative indicates heat absorption by the bubble.

Curve *1* in Fig. 3 shows that the bubble cools down, periodically giving up the heat to the ambient liquid (in the compression phase) and absorbing the heat from the liquid (in the expansion phase). As the magnetic field grows (curves *3*, *4*), the derivative of the temperature with respect to the radius becomes negative everywhere and asymptotically tends to zero; that is, the heat flux is constantly directed from the bubble to the ambient liquid. Thus, the magnetic field shortens the vibratory phase of the process and thus prevents the inverse heat transfer from the liquid to the bubble, increasing the external pressure.

Figure 4 displays the temperature at the center and on the surface of the bubble vs. time at different magnetic fields. The curves suggest that the temperatures in the different parts of the gas bubble vary synchronously. At the initial stage of the process, the temperature at the center of the bubble (curves *a*) oscillates with a larger amplitude than on its surface (curves *b*), remaining greater than unity; in other words, the gas temperature inside the bubble is higher than before the perturbation. Under weak fields, the temperature at the surface may be both higher and lower than unity. As the magnetic induction increases, the overcooling of the bubble surface becomes weaker and disappears.

The overcooling of the near-interface gas in the bubble under weak magnetic fields (curves *1*, *2*) is due to the bubble expansion, causing the temperature to decrease. As the bubble expands, the heat-releasing surface grows, the temperature gradient diminishes, and the total amount of the heat acquired by the bubble from the liquid during the expansion phase is less than the heat given up by the bubble in the compression phase. The magnetic field, on the one hand, decreases the number of oscillations, thus cutting the time for heat transfer from the bubble to the ambient liquid, and, on the other hand, decreases the spread of the interface surface area, making it difficult to transfer the heat to the bubble in the expansion phase (curves *3*, *4*).

Another effect of the magnetic field during the gas bubble relaxation is the separation of mechanical and thermophysical processes. In the absence of the field, the dynamic and thermal processes are interrelated throughout the relaxation time (Fig. 4, curve *1*). The magnetic field suppresses the oscillation of the dynamic parameters, and the vibratory phase of the process shortens. Under such conditions, the system has no time to come to equilibrium and the second

phase, the phase of regular heat exchange, starts, after which the bubble passes into a new state of thermodynamic equilibrium (curve *4* in Fig. 4).

In Fig. 5, the liquid temperature at various distances from the bubble surface is shown. Here, the magnetic induction is 1.5 T.

The curves in Fig. 5 demonstrate that liquid layers adjacent to the interface experience temperature oscillations and the farther the layer from the bubble surface, the weaker the effect of the gas temperature.

Generally, the liquid around the bubble weakly heats up, accumulating the heat of the gas.

The temperature distributions along the bubble radius for various time instants at $B = 1.5$ T are presented in Fig. 6. Here, $\eta = 0$ corresponds to the bubble center; $\eta = 1$, to its surface; and $\eta = L$, to an infinitely remote point. Curves *1* and *5* correspond to two cases when the bubble is the most compressed in the first cycle, that is, to (*1*) $t = 0$ and (*5*) $T = 2\pi/\omega$ (ω is the natural angular frequency of the damped oscillations). Curve *2* corresponds to $t = T/4$; curve *3*, to $t = T/2$; and curve *4*, to $t = 3T/2$. In the central part of the bubble, the gas temperature field remains uniform almost up to the surface. Temperature gradients appear near the bubble surface, and their sign depends on the process phase (compression or expansion). The expansion of the bubble is the greatest when the temperature of the gas is the lowest.

Of interest are partial contributions of viscous, thermal, and Joule mechanisms to the energy dissipation.

The kinetic energy of a vibrating bubble converted to the heat by viscous forces by a time t from the onset of vibration is given by

$$Q_\mu(t) = \int_0^t \int_{V_0}^\infty 12\frac{\mu w_1^2}{r^2} dt dT = 16\pi\mu a_0 t_0 \nu_0^2 J_\mu(t),$$

where

$$J_\mu(t) = \frac{Q_\mu(t)}{16\pi\mu a_0 t_0 \nu_0^2} = \int_0^{t_*} a_* w_*^2 dt_*$$

is the dissipative integral of viscous forces.

The kinetic energy of a vibrating bubble converted to the heat through the mechanism of Joule dissipation by a time t is

$$Q_j(t) = \int_0^t \int_{V_0}^\infty \frac{j^2}{\sigma} dt dV = \frac{8}{3}\pi\sigma B^2 a_0^3 \nu_0^2 t_0 J_j(t),$$

where

$$J_j(t_*) = \frac{Q_j(T)}{\frac{8}{3}\pi\sigma B^2 a_0^3 \nu_0^2 t_0} = \int_0^{t_*} a_*^3 w_*^2 dt_*.$$

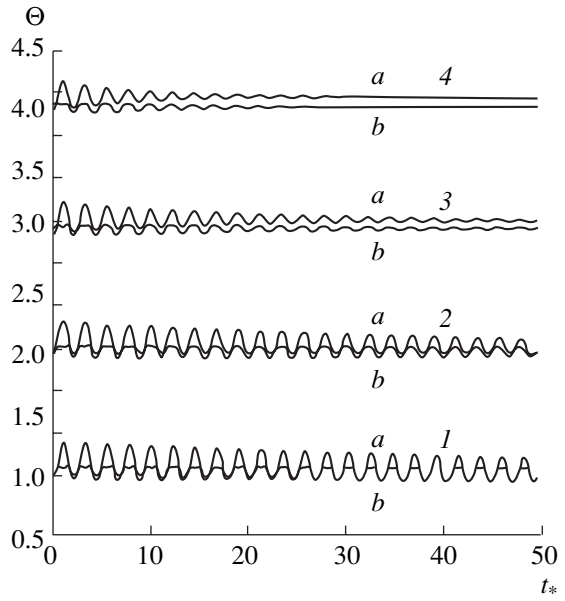

Fig. 4. Reduced temperatures Θ (*a*) at the center and (*b*) on the surface of the bubble as a function of the reduced time t_*. B = (*1*) 0, (*2*) 0.5, (*3*) 1.0, and (*4*) 1.5 T. Curve *2* is shifted up by 1; curve *3*, by 2; and curve *4*, by 3.

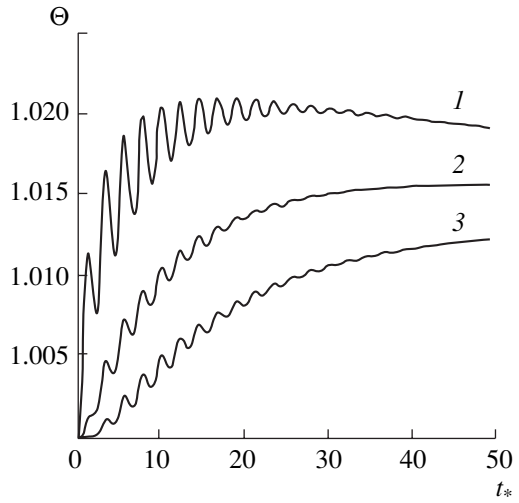

Fig. 5. Distribution of the reduced temperature Θ in liquid gallium with reduced time t_* at different distances from the bubble. Curve (*1*) corresponds to a distance of one, (*2*) two, and (*3*) three spatial steps $\delta\eta = L/M$.

The energy dissipated due to nonequilibrium heat transfer to (and heat removal from) the bubble through its surface is given by the integral

$$Q_\lambda(t) = \int_0^t \oint_\Sigma q_{\text{int}} dt d\Sigma = 4\pi a_0 T_0 \lambda_2 t_0 J_\lambda(t),$$

where

$$J_\lambda(t_*) = \frac{Q_\lambda(t)}{4\pi a_0 T_0 \lambda_2 t_0} = -\int_0^{t_*} a_* \left.\frac{\partial\Theta}{\partial\eta}\right|_{\eta=1} dt_*$$

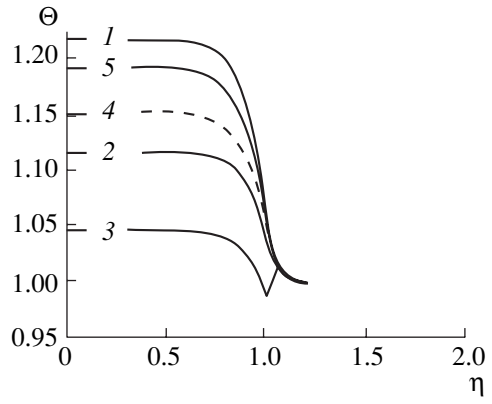

Fig. 6. Reduced temperature Θ as a function of the reduced distance η at different phases of vibration. The figures at the left and right of the vertical axis refer to the gas and liquid, respectively.

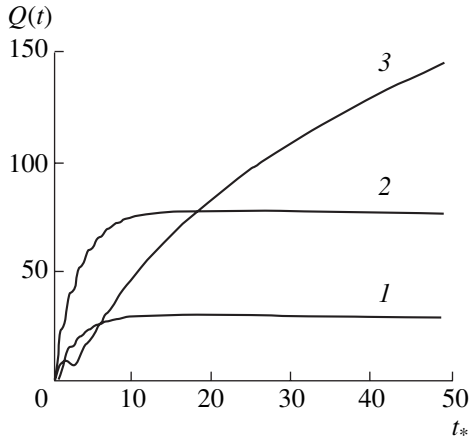

Fig. 7. Energy dissipation vs. time t_*. (*1*) $Q_\mu \times 10^8$, (*2*) $Q_j \times 10^5$, and (*3*) $Q_\lambda \times 10^5$ J. $B = 1$ T, and the initial bubble radius $a_0 = 3$ mm.

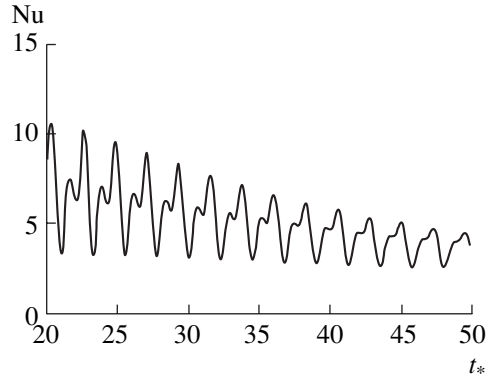

Fig. 8. Dependence of the Nusselt number Nu on the reduced time t_* at the stage of bubble damped vibrations. $B = 0.5$ T, $a_0 = 3$ mm.

is the thermal dissipative integral.

Figure 7 shows the dependences of the energies Q_μ, Q_j, and Q_λ on the process time. The saturation of the curve (plateau) indicates that a given mechanism of energy dissipation is terminated. At the vibratory stage of the process, all the three curves rise, the dependence $Q_\lambda(t)$ oscillating because of the alternating direction of the heat flux early in the process. After this stage has been completed, the curves Q_μ and Q_j saturate. The Joule dissipation of the energy is three orders of magnitude higher than the viscous one. The latter depends on the viscosity, which is low for liquid gallium. The thermal dissipation of the energy is comparable to the Joule dissipation by an order of magnitude at the vibratory stage but exceeds the latter at the final stage of the process. However, at the first stage, the energy dissipated by the Joule mechanism is several times greater than by the thermal one. The rise of the curve Q_λ throughout the time interval is an indication of the incompleteness of the thermal processes.

Thermal processes in a gas bubble surrounded by a liquid are also characterized by the Nusselt number

$$\mathrm{Nu} = \frac{2a\alpha}{\lambda_2} = \frac{2a}{T_{\mathrm{int}} - \langle T_2\rangle}\left.\frac{\partial T_2}{\partial r}\right|_{r=a} = \frac{2}{\Theta_{\eta=1} - \langle\Theta\rangle}\left.\frac{\partial\Theta}{\partial\eta}\right|_{\eta=1},$$

where α is the heat transfer coefficient and $\langle T_2\rangle$ is the bulk temperature of the gas in the bubble.

Figure 8 shows the variation of the Nusselt number with the reduced time t_* at the vibratory stage of the process for the induction $B = 0.5$ T.

At the initial stage of the process, the Nusselt number is an alternating function because of the heating and cooling of the bubble. At the instants of the greatest compression of the bubble, the Nusselt number reaches a maximum. At the instants of the greatest expansion, it is minimal. Positive Nusselt numbers correspond to heat removal from the bubble, which takes place under compression, when the bubble temperature is greater than that of the liquid. Negative Nusselt numbers correspond to heat absorption by the bubble, which is observed when the bubble expands and its temperature lowers. As the process comes to an end, the Nusselt number asymptotically tends to two.

CONCLUSIONS

Let us note the basic features of the damping of free vibrations in the volume of a gas bubble immersed in a conductive liquid and exposed to a magnetic field.

(1) To the well-known mechanisms of viscous and thermal dissipation [4], the highly effective mechanism of Joule dissipation is added.

(2) The magnetic field “subdivides” the relaxation process into vibratory and regular (inertialess) stages. The former is characterized by the oscillation of all the dynamic and thermodynamic parameters. At this stage,

the kinetic energy of the liquid is dissipated mainly by the Joule mechanism. At the end of this stage, mechanical equilibrium is established, the oscillation of the parameters disappears, but thermodynamic equilibrium is absent. At the latter stage, the mechanisms of viscous and Joule dissipation "switch off" and energy dissipation and the transition of the bubble to the equilibrium state take place through heat exchange with the ambient liquid. At the end of this stage, complete thermodynamic equilibrium in terms of all bubble parameters is established.

(3) At the vibratory stage, the dominant mechanism of energy dissipation is Joule heat release; at the regular stage, thermal dissipation. In low-viscous liquids like water, the kinetic energy converted to heat by viscous forces is three orders of magnitude less as compared with the ohmic dissipation.

(4) As the magnetic induction increases, the vibratory stage of the bubble relaxation shortens. At high magnetic fields (large Hartmann numbers), damped harmonic oscillations become aperiodic.

(5) Eddy currents due to the radial vibration of the bubble heat liquid layers near the bubble, affecting the bubble–liquid heat exchange by preventing heat transfer from the liquid to the bubble at the expansion phase.

REFERENCES

1. A. N. Zharov and A. I. Grigor'ev, Zh. Tekh. Fiz. **71** (11), 12 (2001) [Tech. Phys. **46**, 1358 (2001)].
2. A. P. Vasilyev, I. M. Kirko, *et al.*, USSR Inventor's Certificate No. 1549445, MKI H02K 44/00.
3. A. P. Vasilyev, V. A. Bondarenko, D. A. Tarakov, and A. D. Prokopenko, Kholod. Tekh., No. 12, 22 (1991).
4. R. I. Nigmatulin, *Dynamics of Multiphase Media* (Nauka, Moscow, 1987), Vols. 1, 2.

Translated by N. Mende

Technical Physics, Vol. 48, No. 1, 2003, pp. 38–42. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 42–45.

GAS DISCHARGES, PLASMA

Electron Energy Distribution Function in a Hollow-Cathode Glow Discharge in Mixtures of Nitrogen with Electronegative Gases

A. G. Kalyuzhnaya, A. V. Ryabtsev, and A. I. Shchedrin

Institute of Physics, National Academy of Sciences of Ukraine, pr. Nauki 46, Kiev, 03028 Ukraine

e-mail: ashched@iop.kiev.ua

Received April 2, 2002

Abstract—The electron energy distribution function (EEDF) in a hollow-cathode glow discharge in mixtures of nitrogen with electronegative gases is investigated. It is shown that small admixtures of SF_6 or CCl_4 to nitrogen significantly increase the number of electrons in the energy range 2–6 eV, which corresponds to the inverse part of the EEDF. In nitrogen with small admixtures of F_2 or NF_3, the EEDF differ slightly from that calculated for pure nitrogen. The EEDF in these mixtures substantially depends on electron attachment to electronegative gas molecules. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The possibility of the formation of an inverse electron energy distribution function (EEDF) in a low-temperature plasma has been repeatedly studied both experimentally and theoretically (see, e.g., [1–12] and the literature cited therein). Media with inverse distribution functions are of great interest because they can be used to create the inverse population of the electronic levels of the admixture atoms and molecules.

In [1–3], the feasibility of an inverse EEDF at low energies in the relaxing plasma of heavy noble gases was predicted theoretically and confirmed experimentally. A necessary condition for the formation of an inverse EEDF is the presence of a Ramsauer minimum in the cross section for the elastic scattering of electrons by the working gas atoms. In a weakly ionized noble gas, elastic scattering is the main mechanism for electron energy losses at low energies. Hence, the electrons, which were initially distributed uniformly in energy space, rapidly leave the regions of intense scattering and are accumulated near the Ramsauer minimum. The arising inversion exists for a time on the order of the plasma thermalization time.

The conditions for the formation of a steady-state inverse EEDF were considered in [4–7]. A mixture of a noble gas with an electronegative gas was ionized with a fast electron beam or X-rays. It was observed that the thermal electrons were generated rather evenly over a broad energy range. The EEDF inversion is related to the loss of low-energy electrons due to their attachment to the electronegative gas molecules and accumulation near the Ramsauer minimum.

In both cases, the EEDF inversion occurs in the energy range 0–2 eV, i.e., at energies that are much lower than the characteristic threshold energies for electronic excitation. For this reason, such media cannot be used to create the inverse population of the electronic levels of atoms and molecules.

Low EEDF inversion near an energy of 5 eV in an nitrogen afterglow plasma was obtained in [8–11]. A characteristic feature of the EEDF formation under those conditions is that the electron gas is heated due to superelastic electron collisions with vibrationally and electronically excited molecules. A local maximum in the EEDF, which occurs for a time interval of ~10^{-4} s, is related to superelastic electron collisions with nitrogen molecules in a metastable electronic state.

In [12], the feasibility of the formation of a steady-state inverse EEDF in a hollow-cathode glow discharge in nitrogen in the energy range 2–6 eV was demonstrated both experimentally and theoretically. In this case, the inversion is related to the following factors. On the one hand, the electrons with energies of 4–7 eV do not undergo any inelastic collisions with nitrogen molecules. In elastic scattering, the energy is transferred to molecules rather slowly and electrons are accumulated in the above energy range. On the other hand, in the energy range 2–4 eV, the electrons rapidly lose their energy when exciting the vibrational levels of nitrogen molecules; as a result, a dip in the EEDF is formed. An important condition for the inversion formation is that the electric field be sufficiently low, which is the case of a hollow-cathode glow discharge. The EEDF inversion occurs in the energy range 2–6 eV. However, the number of electrons in this range is low. Most of the electrons have energies lower than 2 eV, which decreases the efficiency of producing the inverse population of the electronic levels.

The aim of this study is to find conditions under which the relative number of the electrons with ener-

1063-7842/03/4801-0038$24.00

gies corresponding to the inverse part of the EEDF increases. It is proposed to use nitrogen with a small admixture of an electronegative gas. The admixture, which has a large attachment cross section in the low-energy range, increases the loss of low-energy electrons from the discharge; as a result, the EEDF in the inversion region increases.

The EEDF was numerically simulated for mixtures of nitrogen with SF_6, F_2, NF_3, and CCl_4. It is shown that small admixtures of SF_6 or CCl_4 to nitrogen increases the EEDF in the inversion region. In mixtures of nitrogen with F_2 or NF_3, this effect is absent.

NUMERICAL SIMULATIONS

To find the EEDF, we solved the Boltzmann equation in the two-term approximation [13]. Numerical simulations were carried out with allowance for elastic and inelastic collisions between electrons and neutrals, electron–electron scattering, and the external ionization of the gas mixture with a fast electron beam. The form of the equation and the solution technique were described in detail in [5].

The EEDF in mixtures of nitrogen with an electronegative gas at a total pressure of 0.1 torr was calculated for different mixture compositions. The interaction processes between electrons and nitrogen molecules that were taken into account when calculating EEDF are listed in the table; the corresponding cross sections are presented in [14]. The cross sections for the interaction of electrons with electronegative gas molecules are taken from [15–21].

Calculations were performed for the parameters of the experimental facility used in [12] to measure the EEDF in a hollow-cathode glow discharge in pure nitrogen. In this discharge, the applied voltage almost entirely drops across a narrow cathode sheath. The electric field in the main part of the discharge is low (~0.1 V/cm). The gas was ionized and the electrons were heated by the beam of 400-eV electrons escaping from the cathode sheath. In calculations, the electron density was set at $\sim 10^{10}$ cm^{-3}, which corresponded to that measured in [12]. To evaluate the effect of the electric field and the electron density on the form of the EEDF, we calculated EEDFs in the same mixtures but at $E = 1$ V/cm and $n_e = 10^{11}$ cm^{-3}.

(i) $N_2 : SF_6$ mixture. When calculating the EEDF in an $N_2 : SF_6$ mixture, the cross sections for the interaction between electrons and SF_6 molecules were taken from [15] (see inset in Fig. 1). It is seen from Fig. 1 that, at low energies, the most important process is electron attachment.

When SF_6 is added to nitrogen (in the proportion 1 : 100), the EEDF in the inversion region increases significantly (Fig. 1). This effect is caused by the intense attachment of low-energy electrons to SF_6 molecules, i.e., the loss of these electrons from the discharge. As a result, the fraction of low-energy electrons decreases, whereas the fraction of electrons in the inversion region increases. Thus, the change in the EEDF reflects its rearrangement when the number of low-energy electrons changes.

Table

Reaction	Threshold energy, eV
$N_2 + e \longrightarrow N_2(v) + e$, $v = 1$–8	1.5
$N_2 + e \longrightarrow N_2(A^3\Sigma_u^+) + e$	7.63
$N_2 + e \longrightarrow N_2(B^3\Pi_g) + e$	8.54
$N_2 + e \longrightarrow N_2(W^3\Delta_u) + e$	9.11
$N_2 + e \longrightarrow N_2(B'^3\Sigma_u^-) + e$	9.83
$N_2 + e \longrightarrow N_2(a^1\Pi_g) + e$	9.89
$N_2 + e \longrightarrow N_2^+ + e + e$	15.6
$N_2 + e \longrightarrow N + N + e$	9.76

The increase in the electronegative gas concentration in the mixture leads to a competition between two processes. On the one hand, the electron attachment in the low-energy range promotes a further increase in the EEDF in the inversion region. On the other hand, the increasing vibrational excitation of SF_6 molecules gradually decreases the inversion itself: involving the electrons with energies 4–7 eV in this process impedes their accumulation in this energy range, i.e., formation of an EEDF peak.

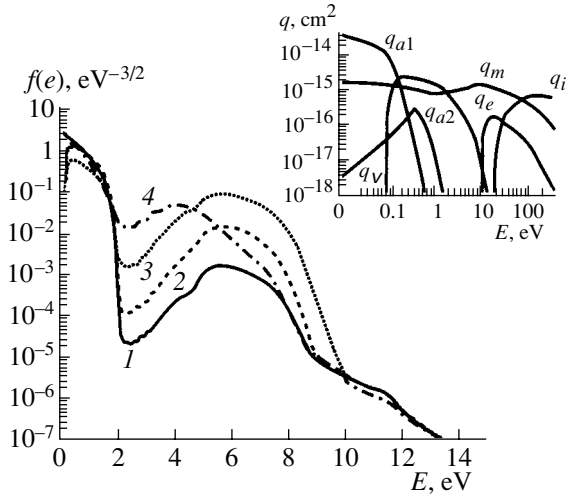

Fig. 1. EEDF in nitrogen and $N_2 : SF_6$ mixtures at a total pressure of $p = 0.1$ torr: $N_2 : SF_6 =$ (*1*) 1 : 0, (*2*) 1 : 0.01, (*3*) 1 : 0.1, and (*4*) 1 : 1. The inset shows the cross sections for electron interaction with SF_6 [15]: q_m is the transport cross section; q_v is the vibrational excitation cross section; q_e is the electronic excitation cross section; q_i is the ionization cross section; and q_{a1} and q_{a2} are the cross sections for electron attachment with the production of SF_6^- and SF_5^-, respectively.

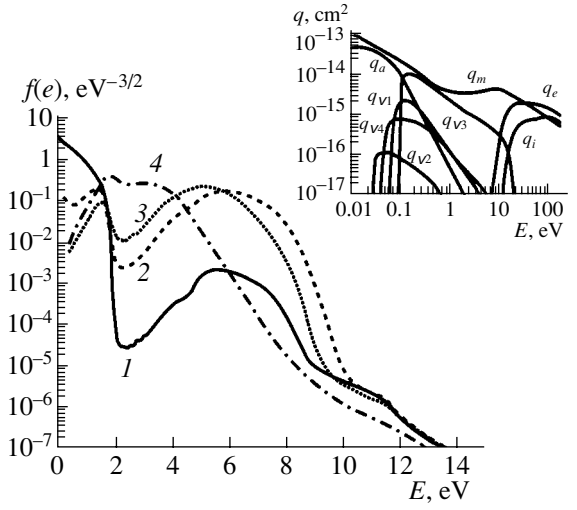

Fig. 2. EEDF in nitrogen and N_2 : CCl_4 mixtures at a total pressure of p = 0.1 torr: N_2 : CCl_4 = (*1*) 1 : 0, (*2*) 1 : 0.01, (*3*) 1 : 0.1, and (*4*) 1 : 1. The inset shows the cross sections for electron interaction with CCl_4 [17, 18]: q_m is the transport cross section; q_e is the electronic excitation cross section; q_i is the ionization cross section; q_a is the electron attachment cross section; and q_{V1}, q_{V2}, q_{V3}, and q_{V4} are the vibrational excitation cross sections.

The calculations show that, in the N_2 : SF_6 = 10 : 1 mixture, the dominant process is electron attachment, which increases the EEDF in the inversion region even higher. However, at SF_6 concentrations higher than 0.5, vibrational excitation becomes predominant and the distribution function becomes almost monotonic (Fig. 1).

The electronic excitation and ionization of SF_6 molecules only slightly affect the form of the EEDF because the threshold energies for these processes are fairly high (~10 eV). When the set of the cross sections for the electron interaction with SF_6 from [16] is used in calculations, the EEDF also does not change significantly.

(ii) N_2 : CCl_4 mixture. The electron attachment to CCl_4 molecules is very intense [17] (see inset in Fig. 2). Thus, in this mixture, the effect under study is very pronounced. A small admixture of CCl_4 to nitrogen results in a significant decrease in the number of low-energy electrons; as a result, the maximum of the EEDF in the inversion region is nearly the same in magnitude as the EEDF maximum in the low-energy range (Fig. 2).

The increase in the electronegative gas concentration leads to a smoothing of the inversion region of the EEDF. As with the previous mixture, this effect is related to the increasing influence of the vibrational excitation of the admixture molecules. However, in the N_2 : CCl_4 = 10 : 1 mixture, the dominant process is electron attachment and the number of electrons in the inversion region is higher than at low energies. As the CCl_4 concentration increases further, the inversion in the discussed energy range vanishes (Fig. 2).

(iii) N_2 : F_2 mixture. The calculated EEDF in nitrogen with a small fluorine admixture differs only slightly from that in pure nitrogen (Fig. 3). Hence, the attachment rate of low-energy electrons to fluorine molecules [19] is insufficiently high to significantly affect the number of these electrons. Calculations performed with

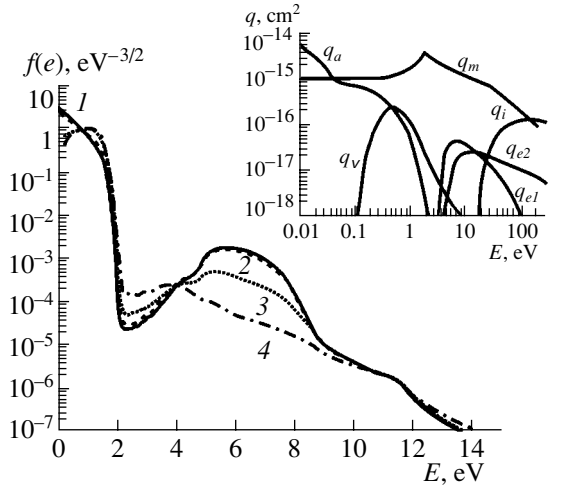

Fig. 3. EEDF in nitrogen and N_2 : F_2 mixtures at a total pressure of p = 0.1 torr: N_2 : F_2 = (*1*) 1 : 0, (*2*) 1 : 0.01, (*3*) 1 : 0.1, and (*4*) 1 : 1. The inset shows the cross sections for electron interaction with F_2 [19]: q_m is the transport cross section; q_v is the vibrational excitation cross section; q_i is the ionization cross section; q_a is the electron attachment cross section; and q_{e1} and q_{e2} are the cross sections for electron excitation of the $a^3\Pi_u$ and $A^1\Pi_u$ levels, respectively.

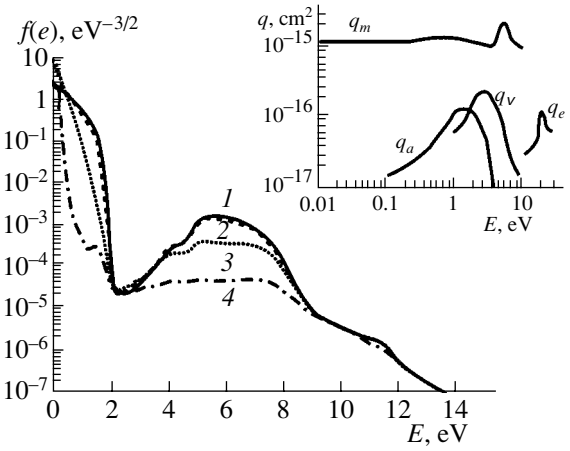

Fig. 4. EEDF in nitrogen and N_2 : NF_3 mixtures at a total pressure of p = 0.1 torr: N_2 : NF_3 = (*1*) 1 : 0, (*2*) 1 : 0.01, (*3*) 1 : 0.1, and (*4*) 1 : 1. The inset shows the cross sections for electron interaction with NF_3 [20, 21]: q_m is the transport cross section, q_v is the vibrational excitation cross section, q_e is the electronic excitation cross section, q_i is the ionization cross section, and q_a is the electron attachment cross section.

larger electron attachment cross sections result in a significant increase in the EEDF in the inversion region.

A further increase in the fluorine content reduces the EEDF inversion. This is mainly due to the electronic excitation of the admixture gas. The excitation thresholds for the low-lying electronic levels of fluorine are 3.16 and 4 eV (see inset in Fig. 3). Therefore, electrons with energies higher than the threshold energies intensely interact with F_2 molecules. This process impedes the accumulation of electrons in the energy range 4–7 eV and, accordingly, the formation of an EEDF peak. Moreover, having transferred their energy to the molecules, the electrons most probably occur in the range 2–4 eV, which corresponds to the dip in the EEDF in pure nitrogen. At equal amounts of nitrogen and fluorine in the mixture, the inversion region vanishes almost completely (Fig. 3).

(iv) $N_2 : NF_3$ mixture. When calculating the EEDF in an $N_2 : NF_3$ mixture, we used the cross sections for elastic scattering and the electronic excitation of NF_3 from [20] and the cross sections for electron attachment and vibrational excitation from [21] (see inset in Fig. 4).

The cross section for electron attachment to NF_3 molecules is relatively small. Hence, as with the previous mixture, a small admixture of this gas to nitrogen only slightly affects the EEDF (Fig. 4).

The higher the NF_3 concentration, the higher the role of inelastic electron collisions with NF_3 molecules. Since the threshold energy for electronic excitation is fairly high, this process only slightly affects the EEDF. In contrast, the excitation of the vibrational levels significantly affects the EEDF. This is because the threshold energy for the excitation of the vibrational levels is low and the quantum energy of the NF_3 vibrational excitation is fairly large (1 eV as compared with 0.1 eV in F_2 and SF_6). Consequently, the electrons lose their energy much more rapidly than in the above mixtures and are efficiently transferred in the low-energy region. As a result, the increase in the NF_3 concentration leads to the vanishing of the EEDF inversion, because almost all the electrons have energies lower than the threshold energy for the vibrational excitation of NF_3 (Fig. 4).

The EEDF was also calculated at an elevated electric field and elevated electron density. In all the mixtures, there was a tendency toward a smoothing of the EEDF in the inversion region. These effects reflect the Maxwellianization influence of the electric field and electron–electron scattering, as was previously noted in [22].

CONCLUSION

The numerical simulations of the EEDF in a hollow-cathode glow discharge in mixtures of nitrogen with electronegative gases have shown that adding SF_6 or CCl_4 to nitrogen can significantly increase the number of electrons in the inversion region. This effect is related to the intense attachment of low-energy electrons to the electronegative gas molecules. The change in the EEDF is caused by its renormalization as the number of the low-energy electrons decreases. No increase in the inversion region of the EEDF in $N_2 : F_2$ and $N_2 : NF_3$ mixtures has been observed because the rate of electron attachment to these gas molecules is insufficiently high to significantly affect the EEDF.

When a small admixture of an electronegative gas is added to nitrogen, the change in the EEDF is determined by the cross section for electron attachment to the admixture molecules. The other inelastic interactions between the electrons and the admixture molecules only slightly affect the EEDF. The higher the admixture concentration, the higher the role of the electronic and vibrational excitation of the admixture molecules. For each of the discussed mixtures, these processes lead to the smoothing of the EEDF in the inversion region and the formation of a monotonic EEDF.

The largest increase in the EEDF in the inversion region was achieved in the $N_2 : SF_6 = 10 : 1$ and $N_2 : CCl_4 = 10 : 1$ mixtures.

ACKNOWLEDGMENTS

This study was supported by the Foundation for Basic Research of the Ministry of Education and Science of Ukraine, contract no. F7/298-2001.

REFERENCES

1. G. L. Bragila and L. Ferrari, Nuovo Cimento **4**, 245 (1971).
2. J. M. Warman, V. Sowada, and M. P. De Haas, Phys. Rev. A **31**, 1974 (1985).
3. A. V. Rokhlenko, Zh. Éksp. Teor. Fiz. **75**, 1315 (1978) [Sov. Phys. JETP **48**, 663 (1978)].
4. N. A. Dyatko, I. V. Kochetov, and A. P. Napartovich, Pis'ma Zh. Tekh. Fiz. **13**, 1457 (1987) [Sov. Tech. Phys. Lett. **13**, 610 (1987)].
5. P. M. Golovinskiĭ and A. I. Shchedrin, Zh. Tekh. Fiz. **59** (2), 51 (1989) [Sov. Phys. Tech. Phys. **34**, 159 (1989)].
6. Z. Rozenberg, M. Lando, and M. Rokki, J. Phys. D **21**, 1593 (1988).
7. Z. Rozenberg, M. Lando, and M. Rokki, Phys. Rev. A **37**, 2569 (1988).
8. N. A. Gorbunov, N. B. Kolokolov, and A. A. Kudryavtsev, Zh. Tekh. Fiz. **58**, 1817 (1988) [Sov. Phys. Tech. Phys. **33**, 1104 (1988)].
9. N. A. Gorbunov, N. B. Kolokolov, and A. A. Kudryavtsev, Zh. Tekh. Fiz. **61** (6), 52 (1991) [Sov. Phys. Tech. Phys. **36**, 616 (1991)].
10. N. A. Dyatko, I. V. Kochetov, and A. P. Napartovich, J. Phys. D **26**, 418 (1993).
11. N. A. Gorbunov, N. B. Kolokolov, and F. E. Latyshev, Fiz. Plazmy **27**, 1143 (2001) [Plasma Phys. Rep. **27**, 1079 (2001)].

12. V. Yu. Bazhenov, A. V. Ryabtsev, I. A. Soloshenko, *et al.*, Fiz. Plazmy **27**, 859 (2001) [Plasma Phys. Rep. **27**, 813 (2001)].
13. J. P. Shkarofsky, T. W. Johnston, and M. P. Bachynski, *The Particle Kinetics of Plasmas* (Addison-Wesley, Reading, 1966; Atomizdat, Moscow, 1969).
14. C. J. Gillant, J. Tennyson, B. M. McLaughlin, *et al.*, J. Phys. B **29**, 1531 (1996).
15. H. Itoh, Y. Miura, N. Ikuta, *et al.*, J. Phys. D **21**, 922 (1988).
16. A. V. Phelps and R. J. Brunt, J. Appl. Phys. **64**, 4269 (1988).
17. A. Chutjian and S. H. Alajajian, Phys. Rev. A **31**, 2885 (1985).
18. L. C. Pitchford, B. V. McKoy, A. Chutjian, *et al.*, *Electron Collision Cross Sections for Molecules Determined from Beam and Swarm Data* (Springer-Verlag, New York, 1987).
19. M. Hayashi and T. Nimura, J. Appl. Phys. **54**, 4879 (1983).
20. T. N. Rescigno, Phys. Rev. A **52**, 329 (1995).
21. N. A. Dyatko, M. Capitelli, S. Longo, *et al.*, Fiz. Plazmy **24**, 745 (1998) [Plasma Phys. Rep. **24**, 691 (1998)].
22. P. M. Golovinskiĭ and A. I. Shchedrin, Pis'ma Zh. Tekh. Fiz. **12**, 1162 (1986) [Sov. Tech. Phys. Lett. **12**, 480 (1986)].

Translated by N. Ustinovskiĭ

Technical Physics, Vol. 48, No. 1, 2003, pp. 43–47. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 46–50.

GAS DISCHARGES, PLASMA

Pulsed Microwave Discharge in Atmospheric Air in the Focus of a Two-Mirror Resonator

K. V. Aleksandrov, L. P. Grachev, I. I. Esakov, S. M. Pokras, and K. V. Khodataev

State Unitary Enterprise Moscow Radiotechnical Institute, Russian Academy of Sciences, Moscow, 117519 Russia

e-mail: esakov@dataforce.net

Received May 22, 2002

Abstract—Free-localized pulsed microwave discharge in atmospheric air in the focus of an open two-mirror high-Q resonator excited by linearly polarized electromagnetic radiation with a wavelength of 4.3 cm is described. This discharge is analogous to the previously studied streamer resonance microwave discharge ignited under similar conditions but with an electromagnetic radiation wavelength of 8.9 cm. Starting from a certain overcritical electric field, the discharge plasma channel has a high-temperature core. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

In 1967, a scientific team (in which one of the authors participated) headed by R.F. Avramenko and G.M. Batanov observed pulsed microwave discharges in atmospheric air in the standing wave of a linearly polarized electric field [1]. The discharge was ignited in the central (focal) region of an open high-Q resonator formed by two coaxial spherically concave mirrors and operating at a wavelength of $\lambda \approx 10$ cm. The discharge stayed a few tens of centimeters apart from the nearest construction elements. It had a shape of a plasma channel with a diameter of $2a \approx 0.1$ cm and length of $2l \approx \lambda/2$. The channel was stretched along the vector of the electric component E of the electromagnetic (EM) field. In the center of the channel, there was a core, whose glow intensity was significantly higher than that of the main discharge channel. A significant fraction of the energy stored in the resonator by the instant of breakdown was absorbed in the core plasma.

In 1975 [2], freely localized microwave discharges were observed in a similar experimental facility but with $\lambda = 0.81$ cm. The electrodeless breakdown of air in the focus of a two-mirror high-Q resonator was accomplished at a pressure of $p \leq 400$ torr. The discharge ignited at the maximum p values also had a shape of a plasma channel stretched along **E** but without a central core.

Since 1990, further detailed studies of this type of discharge have been carried out in the experimental facility with $\lambda = 8.9$ cm [3–5]. It was shown that the discharge develops from a single "free" electron. Initially, a small spherically symmetric plasmoid appears, which then stretches along **E** in both sides from the point of origin in the form of a microwave streamer channel. When the channel becomes as long as about $\lambda/2$, it begins to manifest its resonance properties. As a result, the current amplitude in the central region of the streamer significantly increases. It was supposed that, in the final stage of the discharge development, the discharge is pinched, which results in the formation of a high-temperature core. In experiments, this stage has a threshold with respect to the gas pressure. Thus, in air, it occurs only at $p \geq 540$ torr.

The fact that there is a small region in the discharge in which the electromagnetic (EM) energy is efficiently and rapidly accumulated opens up possibilities for practical applications of these discharges (see, e.g., [6]) and stimulates their further study. On the other hand, these studies are limited by the fact that the experimental facility used in [3–5] is basically the only one in which such studies have been carried out. This facility allows one to vary the resonator configuration, gas species, and gas pressure but operates at a fixed EM field wavelength.

This paper presents the experimental results obtained in the facility described in [7], with a microwave source operating at $\lambda = 4.3$ cm. Its energy characteristics and the possibility of controlling the radiation frequency allow one to study the electrodeless microwave breakdown in atmospheric air in the focus of a two-mirror open resonator at this λ value and obtain a freely localized streamer resonance microwave discharge with a cumulative core.

EXPERIMENTAL SETUP

A schematic of the experimental facility is shown in Fig. 1. The facility incorporates a vertical transmitting antenna array with a 2.7×2.7-m rectangular aperture. The antenna array consists of rectangular horns placed in six rows with six horns in a row, all of them located on a spherical surface with a radius of curvature of 2.7 m. Each horn is fed by an amplifying microwave klystron oscillator. A signal from a single master oscil-

1063-7842/03/4801-0043$24.00

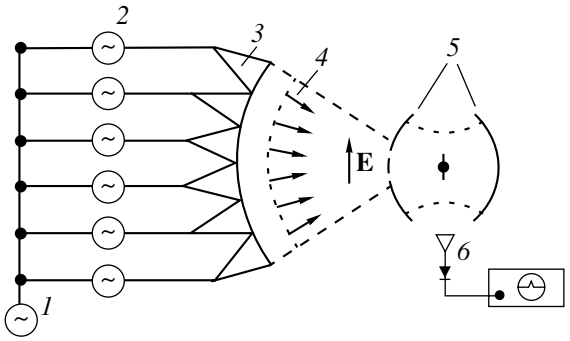

Fig. 1. Schematic of the experimental facility: (*1*) master oscillator, (*2*) amplifying klystrons, (*3*) emitting horns, (*4*) microwave beam, (*5*) open resonator, and (*6*) measuring circuit.

lator is applied to the inputs of all the klystrons. The phase shifters in the klystron input circuits and the metal-plate lenses at the horn outlets form an even phase front over the aperture of the spherical antenna array. As a result, a linearly polarized TEM microwave beam focused in the near-field zone is formed. The beam axis is horizontal and is perpendicular to the plane of the antenna mouth. In the focal region, the field vector $\mathbf{E}_{foc}$ is vertical. The beam focusing angle is $\alpha \approx 60°$, and the FWHM beam diameter in the focal plane is $2r_{foc} \approx \lambda$. The total diameter (the distance between the field zeros) of the main lobe, which contains nearly 88% of the total power emitted by the antenna, is approximately twice as high [8].

The central frequency of the microwave field is $f_0 = 7005$ MHz ($\lambda = 4.283$ cm). At this frequency, each array element can emit microwave pulses with a duration of up to $t_{pul} = 800$ μs and an average power of $P_1 = 30$ kW; i.e., the maximum total microwave power is $P_{gen} \approx 1$ MW. The radiation power can be decreased by nearly one order of magnitude. When the field frequency is shifted from the central frequency f_0 by ±2.5 MHz, the amplitude of the emitted wave decreases by a factor of 2. During a microwave pulse, the master oscillator allows the facility to operate in the frequency modulation regime with a constant sweep rate v_{sw}. The pulse repetition rate can be varied from a single-pulse mode to 1 Hz.

The facility also incorporates a quasi-optical open resonator formed by two round coaxial spherical mirrors with a radius of curvature $R_{mir} = 17.5$ cm and diameter of $2r_{mir} = 34$ cm. The maximum distance between the mirrors along the axis is $2h = 29.6$ cm. This distance can be smoothly varied within 29.6 ± 0.5 cm. The mirrors are made from a 0.2-cm-thick copper sheet. The resonator is oriented horizontally and is coaxial with the axis of the microwave beam emitted by the antenna array. It is placed apart from the antenna at a distance that ensures approximate coincidence between the curvature of the beam phase front and the curvature of the outer spherically convex surface of the nearest mirror. The central part of the mirror with a diameter of $2r_{con} \approx 15$ cm is perforated with round 0.63-cm-diameter holes, which are positioned at the nodes of a quadratic mesh with a 2×2-cm cell size. One cell side is parallel to the vector **E** of the microwave radiation incident onto the mirror, whereas the other side is perpendicular to this vector. These holes serve to excite the resonator with EM radiation; i.e., it provides the coupling with the microwave oscillator. The $2r_{con}$ value is about the characteristic transverse dimension of the microwave beam (at the mirror position) exciting the resonator.

According to [9], for the above resonator dimensions and $\lambda = 4.3$ cm, the simplest azimuthally symmetric mode with a transverse Gaussian profile and a standing-wave field distribution along the resonator axis is excited in the resonator. The distance from the resonator axis at which the field decreases *e*-fold is $r = 3$ cm in the central focal plane and $r_{car} = 7.3$ cm on the mirror surface. There are $q = 13$ variations in the field amplitude along the resonator axis; i.e., the field is maximum in the center of the resonator.

In estimates, the copper conductivity σ is usually assumed to be $5.8 \times 10^7\ \Omega^{-1}\ m^{-1}$. Hence, at $f_0 = 7$ GHz, the ratio of the microwave power absorbed by a copper surface to the microwave power incident on it is [10] $\alpha_\sigma = \sqrt{2\varepsilon_0\omega/\sigma} = 2.3 \times 10^{-4}$, where $\varepsilon_0 = 10^{-8}/(36\pi)$ F/m and $\omega = 2\pi f$. Under the assumption that the microwave diffraction loss from the resonator can be neglected, this α_σ value results in the intrinsic resonator Q-factor $Q_0 = \pi q/\alpha_\sigma = 1.8 \times 10^5$. Indeed, the diffraction loss coefficient is $\alpha_{dif} = \exp[-2(r_{mir}/r_{car})^2] = 2.6 \times 10^{-5} \ll \alpha_\sigma$. The Q-factor of a loaded resonator Q also depends on the coupling factor α_{con} between the resonator and oscillator. In our case, α_{con} is the ratio of the microwave power passed through the perforated mirror hole system to the microwave power incident onto the mirror. In the steady-state operation mode and for the optimum coupling, when $\alpha_{con} = 2\alpha_\sigma$, all the energy of the exciting wave enters the resonator and is absorbed by the mirrors; in this case, $Q = Q_0/2 = 9 \times 10^4$. At this Q value, the full width of the resonance curve at the level of $1/\sqrt{2}$ with respect to the mechanical detuning of the resonator length $2\Delta h = 2h/Q$ amounts to only 3 μm, whereas that with respect to the frequency detuning is $2\Delta f = f_0/Q = 70$ kHz.

In the exact resonance, the maximum field in the resonator focus can be estimated by the formula [11]

$$E_{max} = [4/(\pi r)]\sqrt{Q\eta P_{gen} Z_0/q}, \qquad (1)$$

where $Z_0 = 120\pi\ \Omega$ and η is the power utilization factor (the fraction of P_{gen} that is actually spent on the excitation).

In the experiments, the field in the resonator focus was monitored with a waveguide horn that received the

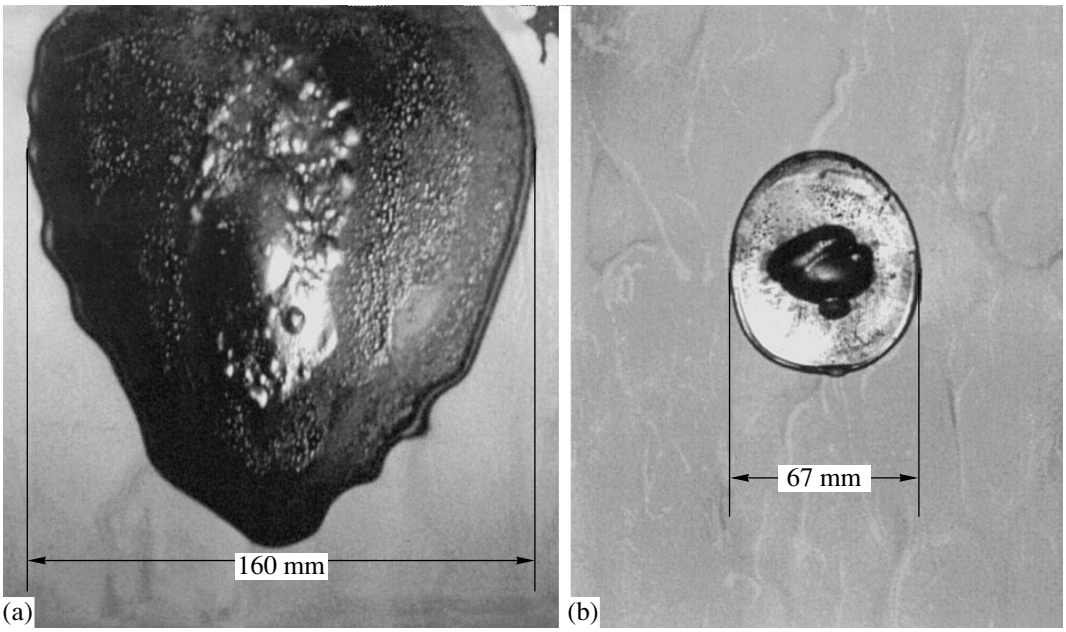

Fig. 2. Configuration of the focal region of a microwave beam exciting the resonator.

microwave signal emitted from the resonator due to diffraction. The microwave signal from this horn was fed to a linear detector and a storage oscilloscope. The absolute calibration of this measuring circuit will be described below.

EXPERIMENT

In our experiments, the facility operated not at the maximum power P_{gen}, but mainly in the regime with single microwave pulses with t_{pul} = 350 μs and ν_{sw} = 2 kHz/μs.

In the preliminary experiments, the configuration of the microwave beam in the focal region was refined and the P_{gen} value was estimated. In so doing, the resonator was removed.

To refine the beam configuration, a 0.3-cm-thick textolite plate 40×40 cm in size was placed in the focal region. One side of the plate was covered with a paint whose color changes with temperature. After a certain number of the microwave pulses, the plate, having absorbed microwave energy, was substantially heated; thus, the changed paint color visualized the beam geometry. Figure 2a shows a photo of the plate oriented along the beam axis and perpendicular to the field vector **E**. The beam propagates from top to bottom. The beam focusing angle measured by this photo is $\alpha \approx 60°$. Figure 2b shows the plate located in the beam focal plane perpendicular to the beam propagation direction. As was expected, the focal spot diameter measured by the photo is between λ and 2λ.

To measure P_{gen}, cylindrical metal microwave vibrators with a resonance length of $2l_{vib}$ = 2.15 cm, spherically rounded ends, and different diameters $2a_{vib}$ were set one after another in the beam focus parallel to the field vector. In the experiments, the initial (unperturbed) field amplitude E_{foc} in the beam focus was significantly lower than the critical breakdown amplitude for atmospheric air, E_{cr} = 32 kV/cm [12]. However, at the vibrator ends, the field was amplified, and, for the given λ, its value [13]

$$E_{pol} = 3.6E_{foc}/(2a_{vib}) \quad (2)$$

(where $2a_{vib}$ is in cm) could be higher than E_{cr}. The measured maximum diameter of the vibrator at which air breakdown still occurred (i.e., $E_{pol} \approx E_{cr}$) was $2a_{vib}$ = 0.17 cm. This corresponds to $E_{foc} \approx 1.5$ kV/cm; i.e., the microwave intensity is $\Pi_{foc} = E_{foc}^2/(2Z_0) = 3 \times 10^3$ W/cm^2 and $P_{gen} \approx \Pi_{foc} \times 2\lambda^2 = 120$ kW.

According to Eq. (1), at this beam power and for $Q = 9 \times 10^4$ and η = 1, the maximum field in the focus of the tuned resonator is expected to be E_{max} = 240 kV/cm > E_{cr}; hence, air breakdown in the resonator can occur. Indeed, in the main series of experiments, nearly every microwave pulse was accompanied by air breakdown in the resonator focus.

Figure 3 presents the oscillogram obtained in the absence of a discharge. The time scale is 50 μs/div. Actually, the measuring circuit monitors the interference signal from the microwave beam exciting the resonator and the scattered signal, which reflects the $E_0(t)$ dependence. At the input of the measuring circuit, these signals are summed with certain phases. As a result, in Fig. 3, the interference signal with a duration of t_{pul} =

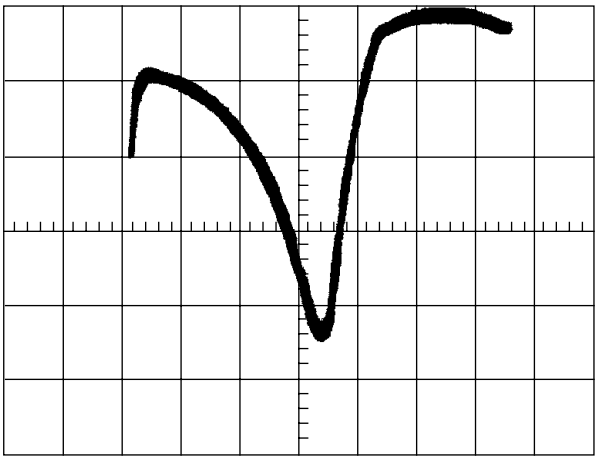

Fig. 3. Oscillogram of the control signal; time scale is 50 μs/div.

350 μs is drawn upward from the zero line, whereas the superimposed resonance curve is drawn downward. The width of the resonance curve $2\Delta f$ at a level of $1/\sqrt{2}$ is approximately 70 μs. In this case, the amplitude of the resonance curve and the $1/\sqrt{2}$ level are counted from the upper right end of the oscilloscope trace. The measured $2\Delta f$ value corresponds to $Q = 10^5$ and $\alpha_{con} = 2\alpha_\sigma$.

After breakdown, in full accordance with the experiments of [3], the time behavior of the field in the resonator changes significantly. In this case, the oscillogram reproduces only the initial portion of the resonance curve up to the breakdown level E_{br}; then, for a time shorter than 1 μs, the oscilloscope trace sharply ascends to the level corresponding to the interference signal from the microwave beam exciting the resonator. This upper right horizontal part of the oscilloscope trace in Fig. 3 will be considered a zero level of the resonator field, and the field E_0 in the focus will be counted downward from this level.

In the experiments, the field $E_0 = E_{br}$ in the successive pulses differed by several times. In most pulses, the field E_{br} was much higher than E_{cr}, and breakdown occurred on both the descending and ascending segments of the resonance curve. This statistical scatter is related to the random character of the free-electron origin (which can cause an electron avalanche [4]) in air under natural conditions. Moreover, in some pulses, over a time interval in which $E > E_{cr}$ in the focal region of the resonator, such electrons may not even occur at all. In this case, there is no discharge during the microwave pulse, and the oscillogram shows the complete resonance curve.

It is natural to associate the minimum increase in the signal in the oscillogram before breakdown (averaged over a large series of pulses) with $E_0 = E_{cr}$ [4]. This allows us to relate the vertical scale in the oscillogram to the absolute value of the field in the resonator focus. In our experiments, the minimum increase in the signal was five to six times smaller than the maximum of the resonance curve, which gives $E_{max} = (5–6)E_{cr} = 150–190$ kV/cm; hence, the power utilization factor η is in the range 0.4–0.6.

Fig. 4. Freely localized streamer microwave discharge in atmospheric air.

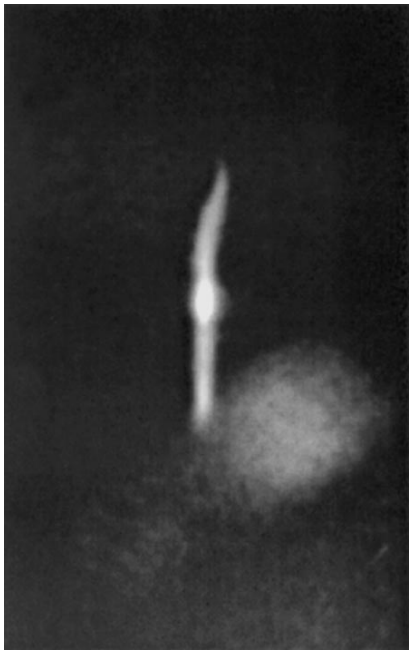

Fig. 5. Freely localized streamer microwave discharge with a high-temperature core in atmospheric air.

In the experiments, different E_{br} values corresponded to different types of discharges. As an example, Fig. 4 presents the integral photo of a discharge obtained at an exposure time that is significantly longer than the duration of the discharge glow and at E_{br} only slightly exceeding E_{cr}. It is seen that the discharge has the form of a plasma channel directed along **E**. In different pulses, the channel length $2l$ varied from 1.3 to 1.65 cm, whereas the diameter was $2a \approx 0.1$ cm. No cumulative core is seen in this discharge. Figure 5 shows a discharge at E_{br} close to E_{max}. It is seen that the discharge has a central core, although the discharge length and diameter remain almost the same.

In the experiments, the high-temperature core was observed only at $E_0 \geq 2E_{br}$.

DISCUSSION

The results obtained agree with the concept of a streamer resonance cumulative microwave discharge described in [5]. The possibility of discharge pinching, which, in our opinion, causes the appearance of a cumulative core in the central region of the streamer, substantially depends on the ratio of the current through this streamer region to its radius, I_0/a. It was shown experimentally that the decrease in λ from 8.9 to 4.3 cm only slightly affects the streamer radius a. At the same time, the current $I_0 = E_{br}h_{eff}/R_\Sigma$ depends on the effective streamer length h_{eff}, which is proportional to λ. Here, R_Σ is the wave impedance of the streamer, which is considered to be a microwave vibrator. The R_Σ value depends on the $2l/(\lambda/2)$ ratio, which changes only slightly with decreasing λ. In contrast, as the wavelength decreases to $\lambda = 4.3$ cm, the h_{eff} value decreases by a factor of 2 and, at $E_{br} = E_{cr}$, the current I_0 also decreases about two times. As a result, at $E_{br} \approx E_{cr}$, the conditions necessary for discharge pinching may not be satisfied. However, in some pulses, this decrease in h_{eff} can be balanced by an increase in E_{br}. That is why, at $E_{br} \geq 2E_{cr}$, the discharge has a cumulative core.

CONCLUSION

Thus, the electrodeless breakdown was ignited in atmospheric air in the standing wave of a linearly polarized microwave field with a wavelength of $\lambda = 4.3$ cm. In the experiments, we used a quasi-optical beam technique to feed an open two-mirror high-Q resonator and the dynamic regime for tuning it to the resonance by sweeping the frequency of the exciting field.

The experiments showed that the streamer microwave discharge at a wavelength $\lambda = 4.3$ cm develops qualitatively in the same manner as that at $\lambda = 8.9$ cm, but with a natural quantitative difference at the resonance stage of its development. Thus, at a breakdown field E_{br} approximately equal to the critical breakdown field E_{cr}, which corresponds to air at atmospheric pressure, the discharge streamer does not have a cumulative core. At $\lambda = 4.3$ cm, a freely localized streamer resonance microwave discharge with a high-temperature core can occur only at $E_{br} \geq 2E_{cr}$. This discharge is similar to the previously studied discharge of this type ignited in atmospheric air at $\lambda = 8.9$ cm and $E_{br} = E_{cr}$.

REFERENCES

1. R. F. Avramenko, L. P. Grachev, and V. I. Nikolaeva, in *Electropuncture and the Problems of Information–Energy Regulation of Human Vital Functions*, Ed. by V. G. Nikiforov and V. N. Pushkin (Tsentr. Nauchno-Issled. Inst. Ékonomiki i Nauchno-Tekhnicheskoĭ Informatsii Ugol'noĭ Promyshlennosti Ministerstva Ugol'noĭ Promyshlennosti SSSR, Moscow, 1976), p. 197.
2. A. P. Vikharev and B. G. Eremin, Zh. Éksp. Teor. Fiz. **68**, 452 (1975) [Sov. Phys. JETP **41**, 219 (1975)].
3. L. P. Grachev, I. I. Esakov, G. I. Mishin, and K. V. Khodataev, Zh. Tekh. Fiz. **64** (2), 26 (1994) [Tech. Phys. **39**, 130 (1994)].
4. V. S. Barashenkov, L. P. Grachev, I. I. Esakov, *et al.*, Zh. Tekh. Fiz. **70** (10), 34 (2000) [Tech. Phys. **45**, 1265 (2000)].
5. V. S. Barashenkov, L. P. Grachev, I. I. Esakov, *et al.*, Zh. Tekh. Fiz. **70** (11), 31 (2000) [Tech. Phys. **45**, 1406 (2000)].
6. L. P. Grachev, I. I. Esakov, G. I. Mishin, and K. V. Khodataev, Preprint No. 1577 (Ioffe Physicotechnical Institute, Russian Academy of Sciences, St. Petersburg, 1992).
7. G. I. Batskikh and Yu. I. Khvorostyanoĭ, Radiotekh. Élektron. (Moscow) **37**, 311 (1992).
8. J. W. Sherman, IRE Trans. Antennas Propag. **10**, 399 (1962).
9. L. A. Vainshtein, *Open Resonators and Open Waveguides* (Sov. Radio, Moscow, 1966; Golem, Boulder, 1969).
10. I. V. Lebedev, *Microwave Engineering and Devices* (Vysshaya Shkola, Moscow, 1970).
11. L. P. Grachev, I. I. Esakov, S. G. Malyk, and K. V. Khodataev, Zh. Tekh. Fiz. **71** (6), 66 (2001) [Tech. Phys. **46**, 709 (2001)].
12. A. D. MacDonald, *Microwave Breakdown in Gases* (Wiley, New York, 1966; Mir, Moscow, 1969).
13. L. P. Grachev, I. I. Esakov, G. I. Mishin, and K. V. Khodataev, Zh. Tekh. Fiz. **65** (7), 60 (1995) [Tech. Phys. **40**, 666 (1995)].

Translated by N. Ustinovskiĭ

Technical Physics, Vol. 48, No. 1, 2003, pp. 48–51. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 51–54.

GAS DISCHARGES, PLASMA

Electrical Breakdown in Solid Insulators: Duration of the Discharge Channel Formation

V. M. Rozhkov

Siberian State University of Telecommunication and Informatics, Novosibirsk, 630102 Russia

Received June 11, 2002

Abstract—The duration of the discharge channel formation is calculated for 0.01- to 0.5-mm-thick insulators in terms of the avalanche–streamer theory. At this stage of breakdown development, the statistical delay is assumed to be the time from the instant the testing voltage becomes equal to the breakdown value to the instant of streamer origination and that the formation time is the interval between the instant of streamer origination and the instant the electrodes close through the discharge channel. *© 2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

A new approach to the description of electrical breakdown in solid insulators has been recently proposed in [1]. This approach uses new definitions of the first and second breakdown stages. In earlier theories, the first stage, i.e., the formation of the discharge channel, is most completely described in terms of the avalanche–streamer mechanism, where the duration of this stage is assumed to be the total discharge time

$$t_{dis} = t_{st} + t_f. \quad (1)$$

Here, t_{st} is the statistical discharge delay, which is the interval between the instant the testing voltage $U_{t,\,st}$ becomes equal to the breakdown value U_{br} and the instant the first effective electron appears, and t_f is the discharge formation time defined as the interval between the appearance of the first effective electron and the instant the electrodes close through the conducting channel [2, 3].

By an effective electron is meant an electron that is capable of triggering an electron avalanche whose size is sufficient for creating a streamer. The formation time is defined as

$$t_f = \frac{d}{v_{av}}, \quad (2)$$

where d is the thickness of the solid insulator and v_{av} is the discharge average velocity. The statistical delay time is given by

$$t_{st} = \frac{1}{\nu(n_0)} = \frac{1}{\nu}\exp\frac{n_0}{\bar{n}}, \quad (3)$$

where $\bar{n}$ is the average number of electrons in an avalanche, n_0 is the minimal number of electrons in the avalanche that cause a streamer, ν is the number of initial free electrons generated in the insulator per second, and $\nu(n_0)$ is the number of avalanches of size exceeding n_0 that are generated per second.

STATEMENT OF THE PROBLEM

The procedure of calculating t_{st} by formula (3) [2, p. 759] is rather tedious; therefore, this technique has not found wide application. However, the value of t_{st} contributes significantly to the discharge formation time in specimens 20–30 μm thick and must therefore be evaluated. Below, we propose new formulas for t_{st} and t_f that can be conveniently applied in practice to evaluate, for instance, the parameters of insulators less than 0.01 mm thick.

BASIC REASONING AND EVALUATION OF THE INITIAL DISCHARGE STAGE DURATION

Let a testing voltage be applied to a solid insulator (Fig. 1) with a ramp of 5–10 kV/s, which corresponds to a breakdown at the leading edge of an oblique wave. Let the instant when $U_{t,\,st} = U_{br}$ be $t = 0$. By this instant, the insulator already contains free electrons in the discharge gap. These electrons, which are capable of initiating impact ionization, are emitted mostly from microasperities on the cathode. In the situation illustrated in Fig. 1a, there are three such electrons, but only electron no. 1 will be effective, because only this electron travels a distance d_1, which is sufficient for initiating an avalanche with the number of electrons n_0 to provide a cathode streamer. In Fig. 1b, the thickness d_2 of the insulator is greater than d_1; hence, the number of electrons causing ionization is also greater (say four). All except electron no. 4 are effective, but the avalanche duration is specified by electron no. 3, which is d_1 distant from the anode.

1063-7842/03/4801-0048$24.00

Let the velocity of the electron avalanche moving to the anode and that of the cathode streamer be the same and equal to v_{av} [2, 3]. Then, the formation time of the discharge channel will be $t_{f,a} = 2d_1/v_{av}$ in the case of Fig. 1a and $t_{f,b} = d_1/v_{av} + d_2/v_{av}$ in the case of Fig. 1b.

Both these values are greater than t_f given by (2) and seemingly must increase t_{dis} calculated from (1); this may however be avoided by using other definitions for t_{st} and t_f.

We propose to count the discharge channel formation time from the instant the streamer is initiated rather than from the instant the first effective electron appears, since the latter instant is difficult to determine exactly in solids. As for the duration of the first electron avalanche that produces n_0 electrons over the distance d_1 and initiates the streamer, it is more logical to regard it as a component of the statistical discharge delay. The same conclusion follows, for example, from the experimental time dependences of the discharge channel length in rock salt, where the statistical delay time lasts until the discharge channel appears [3, Fig. 36]. In [4], a similar time dependence of the discharge channel length was observed for propylene and it was also noted that all processes preceding the appearance of the discharge channel in insulators should be viewed as contributing to the statistical delay.

Thus, within the avalanche–streamer theory, it is proposed to consider t_f as the time from the instant the discharge channel appears to the instant the electrodes close up through the conducting channel and to define this time by formula (2).

As for the time t_{st}, note that, when the insulator thickness increases, the number n of electrons in formula (3) grows while n_0 remains constant. Therefore, according to (3), t_{st} goes down [2, 3]. The same conclusion also follows from Fig. 1. The distance d_1 between the first effective electron and anode remains unchanged with increasing d_2; consequently, the time of travel for the first intense avalanche is thickness dependent. Then, if $d \gg d_1$, this time can be neglected in comparison with t_f just as t_{st} in the form of (3) is ignored when the material is thick. With decreasing d_2 and then d_1, $\bar{n}$ decreases and t_{st} goes up, becoming greater than t_f. In this case, it is necessary to apply formula (1). The thickness d_1 in Fig. 1 is the critical thickness d_{crit}, because, with $d < d_{crit}$, none of the free electrons in the insulator can trigger an avalanche that is sufficient to initiate a cathode streamer.

An increase in t_{st} by the time it takes for the avalanche to travel the distance $d_1 = d_{crit}$ is taken into account by an additional term in (3):

$$t'_{st} = \frac{1}{\nu}\exp\frac{n_0}{\bar{n}} + \frac{d_{crit}}{v_{av}}. \quad (4)$$

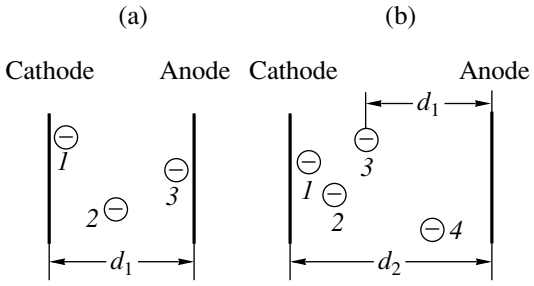

Fig. 1. Discharge gap at the onset of breakdown.

Formula (4) shows that the statistical breakdown delay is the time from the instant $U_{t,st}$ becomes equal to U_{br} to the instant the discharge channel arises, i.e., the streamer is initiated.

The critical thickness of the insulator can be found from a formula for the breakdown field E_{br}. In terms of the avalanche–streamer breakdown mechanism, it is given by [3]

$$E_{br} = \frac{en_{crit}}{k4\pi\varepsilon_0\varepsilon\lambda_{fp,expt}^2}; \quad (5)$$

$e = 1.6 \times 10^{-19}$ C is the electron charge, n_{crit} is the critical number of electrons in the avalanche at the critical thickness of the insulator, $\lambda_{fp,expt}$ is the experimentally found mean free path of electrons between two ionization events, $\varepsilon_0 = 8.86 \times 10^{-12}$ is the permittivity of free space, ε is the relative permittivity of the insulator, and $k = 0.1$–1 is the ratio E_{br}/E at the head of the streamer at the instant it is initiated.

Since the electron avalanche is produced by impact ionization,

$$n_{crit} = 2^{\frac{d_{crit}}{\lambda_{fp,expt}}}.$$

Then,

$$d_{crit} = 1.44\lambda_{fp,expt}\ln n_{crit}, \quad (6)$$

where n_{crit} is calculated from (5) for a particular insulator.

The table summarizes n_{crit} and d_{crit} calculated for several solid insulators whose dielectric strength is higher than 1 MV/cm. It is commonly thought that such fields can support the processes of impact ionization. In our calculations, $k = 0.5$; $\varepsilon = 2.2$ for polyethylene, 2.0 for Teflon, and 3 for poly(ethylene terephthalate) (PETF) films; and $v_{av} = 10^4$ m/s. The experimental mean free paths $\lambda_{fp,expt}$ were evaluated from the curve shown in Fig. 2. This curve averages $\lambda_{fp,expt}$ vs. electric field dependences obtained experimentally by different authors [2, 3, 5]. Since the statistic delay is difficult to calculate from (4) because of the term given by expression (3), formula (4) was replaced by

$$t_{st} = d_{crit}/v_{dis,c}. \quad (7)$$

Characteristics of solid insulators at the stage of discharge channel formation

Material	d, μ	E, MV/cm	$n_{crit} \times 10^{-5}$	d_{crit}, μ	t_{st}, ns	t_f, ns
Polyethylene	500	6.5 [5]	0.8	6.5	0.65	50
″	500	3.5 [6]	1.31	11.9	1.19	50
″	50	6.5	0.8	6.5	0.65	5
″	50	3.5	1.31	11.9	1.19	5
″	10	4.0	1.14	10.2	1.02	1
PETF	50	1.6	6.67	38.7	3.87	5
PETF	10	1.7	5.74	34.4	3.44	1
Teflon	20	1.2 [6]	20.8	105	10.5	2
″	10	1.2	20.8	105	10.5	1
″	10	1.3 [6]	11.1	70	7	1

The error in (7) is much less than in (4) and can be neglected, because the presence of prebreakdown currents in the insulator [3, 5] *a fortiori* means that at least one electron capable of generating an intense electron avalanche at the critical thickness of the insulator is present in the discharge gap. Also, formula (7) uses the so-called cathode discharge velocity $v_{dis,c}$ instead of the avalanche velocity v_{av}. It is known that the discharge velocity under electrical breakdown conditions grows with overvoltage (an excess of the pulsed testing voltage over the dc testing voltage) and increases during the action of the testing voltage [3]. The interval t_{st} is the beginning of the electrical breakdown process; therefore, the velocity within this interval must be minimal. From the physical understanding of electrical breakdown, the variation of the discharge velocity can be attributed to the difference in the electric field structure at different breakdown stages. Up to the instant t_{st}, electrons are injected into the insulator from microasperities on the cathode (in the thickness range considered, the role of these peaks is substantial) and the field in the insulator can therefore be viewed as the field in a negative tip–plane gap. After the instant t_{st}, the positive space charge starts growing near the anode. This charge enhances the electric field intensity in the insulator, and the field transforms into that of the positive tip–plane type. In the latter field, charged particles always move faster than in the former.

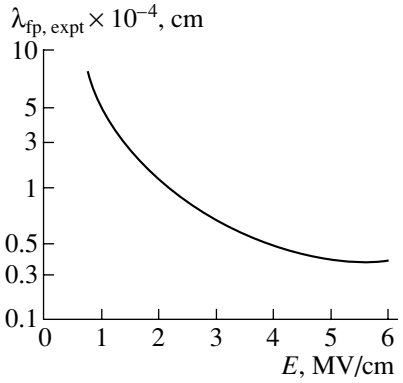

Fig. 2. Mean free path of electrons versus field intensity.

Such a pattern agrees with the difference in the cathode, $v_{dis,c}$ and anode, $v_{dis,a}$ discharge velocities observed experimentally [7]. The motion of electrons from the cathode to the anode for the time t_{st} can be regarded as an anode discharge propagating with a velocity lower than the velocity of sound in a solid insulator [7]. The motion of the cathode streamer can be treated as a discharge propagating from the anode with the velocity $v_{dis,a} = v_{av}$, $v_{dis,a}$ being higher than the velocity of sound. In [7], an exact formula for $v_{dis,c}$ was not derived; therefore, in the calculations by formula (7), we used the maximum possible value of $v_{dis,c}$, which is equal to the velocity of sound in a particular dielectric. According to [8], the velocity of sound is 2.48×10^3 m/s in polyethylene and 1.34×10^3 m/s in Teflon. No data are available for PETF; therefore, we used the value of 2.48×10^3 m/s.

The data summarized in the table show that (i) the general variation of all quantities with d and d_{crit} is consistent with the physical concepts of electrical breakdown in solid insulators; i.e., the validity of the notions of n_{crit} and d_{crit} and of estimating the statistical delay by formula (7) is corroborated; (ii) the theoretical value of d_{crit} subdivides insulators into "thick" ($d > d_{crit}$) and "thin" ($d < d_{crit}$), the boundary thickness falling approximately into the range observed in experiments [3]; (iii) the rapid increase in t_{dis} given by (1) in 10- to 20-μm-thick insulators owing to t_{st} agrees with experimental data reported in [2, 3] and means, at the limit, a microsecond time of breakdown development in insulating films [9]; (iv) t_{dis} decreases with increasing breakdown field strength when $d < d_{crit}$, which is observed in insulating films [9]; and (v) the quantitative relationship between t_{st} and t_f observed is in good agreement with the qualitative relationships considered above.

It should also be noted that the error introduced into d_{crit} by the uncertainty in the coefficient k in (5) is no more than 14% at $k = 0.1$ and 5% at $k = 1$ for d_{crit} given in the table.

The new relationships for t_{st} and t_f provide a novel approach to calculating the breakdown formation time in thin solid insulators, where $d < d_{crit}$. In [3], electrical breakdown in such materials is described in terms of the multiavalanche–streamer mechanism, which relates the initiation of a streamer near the anode to the accumulation of a great amount of positive space charge from many electron avalanches. This approach also uses the notion of the critical thickness of an insulator, which is introduced only qualitatively without deriving an analytical expression. The formula for the discharge time is given instead:

$$t_{dis} = \frac{m}{\nu_0 \lambda_{fp}^2}, \quad (8)$$

where m is the number of electron avalanches necessary for generating the streamer, λ_{fp} is the mean free path of electrons, and ν_0 is the number of electrons picked up by the field from the unit area of the cathode per second.

However, calculations by formula (8) face considerable problems, because it is difficult to determine m and ν_0. It has also been shown [10] that the breakdown current calculated based on the multiavalanche–streamer mechanism decays with time, which is unrealistic. Therefore, the formulas for electrical breakdown in solid insulators presented in [3] prove to be inapplicable when $d < d_{crit}$. At the same time, it is of interest that the technique for calculating t_{dis} from formulas (1), (2), and (7) at $d < d_{crit}$ also implies the multiavalanche–streamer mechanism of developing electrical breakdown. Indeed, if $d < d_{crit}$, this means that the critical number n_{crit} of electrons is generated by several, rather than one, avalanches. There is no need to exactly determine the number of these avalanches. Instead, one can use d_{crit}, which in this case is a hypothetical thickness, and formula (7) to calculate the time over which n_{crit} electrons will arrive at the anode, i.e., to calculate the statistical component of the breakdown development time. By adding t_f given by (2) to this time, we obtain the discharge time. This procedure is much simpler than the calculation of t_{dis} from formula (8).

The novel formulas for t_{st} and t_f can be used in other theories of electrical breakdown in solid insulators (see, e.g., [1]).

CONCLUSIONS

(1) To simplify calculations based on the avalanche–streamer theory of electrical breakdown in solid insulators, it is proposed to assume that the statistical breakdown delay is the interval between the instant the testing voltage becomes equal to the breakdown voltage and the instant a streamer is initiated and that the discharge formation time is the interval from the instant the streamer is initiated to the instant the electrodes close up through the discharge channel.

(2) The definition of the critical thickness d_{crit} of a solid insulator, i.e., the thickness allowing an electron moving to the anode to trigger an avalanche with the number of electrons sufficient for initiating a streamer, is refined. An expression that relates d_{crit} to the insulating material and its dielectric strength is obtained. The critical thickness divides solid insulators into thick and thin.

(3) An expression for the statistical delay of electrical breakdown in solid insulators in terms of critical thickness is derived. It has a simpler form than the conventional expression.

(4) When calculating the delay, it is suggested that an electron velocity smaller than that at the discharge formation stage be used.

REFERENCES

1. V. M. Rozhkov, Élektrotekhnika, No. 3, 37 (2000).
2. G. I. Skanavi, *Physics of Insulators (Strong Field Range)* (Fizmatgiz, Moscow, 1958).
3. A. A. Vorob'ev and G. A. Vorob'ev, *Electric Breakdown and Failure of Solid Insulators* (Vysshaya Shkola, Moscow, 1966).
4. Y. Hiroshi, F. Taniya, and S. Yasuo, J. Phys. D **26**, 1328 (1993).
5. G. A. Vorob'ev and N. S. Nesmelov, Izv. Vyssh. Uchebn. Zaved., Fiz., No. 1, 90 (1979).
6. *Handbook on Materials for Electrical Engineering*, Ed. by Yu. V. Koritskiĭ, V. V. Pasynkov, and B. M. Tareev (Énergiya, Moscow, 1974), Vol. 2.
7. Yu. N. Vershinin, Zh. Tekh. Fiz. **59** (2), 170 (1989) [Sov. Phys. Tech. Phys. **34**, 226 (1989)].
8. *Handbook of Physical Quantities*, Ed. by I. S. Grigoriev and E. Z. Meilikhov (Énergoatomizdat, Moscow, 1991; CRC, Boca Raton, 1997).
9. V. F. Korzo and V. N. Chernyaev, *Insulating Films in Microelectronics* (Énergiya, Moscow, 1977).
10. V. M. Rozhkov, Élektrotekhnika, No. 5, 54 (1997).

Translated by A. Khzmalyan

Technical Physics, Vol. 48, No. 1, 2003, pp. 52–55. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 55–59.

SOLIDS

Allowable Overheat and Limiting Current in a Superconducting Composite with Magnetic Flux Creep

V. R. Romanovskiĭ

Russian Research Centre Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia

Received December 13, 2001

Abstract—Instability arising in a superconducting composite when the current is injected with an ultimately low rate is analyzed. Under these conditions, the nonuniformity of the temperature and electric field distributions over the composite cross section is negligibly small. Equations that allow one to estimate the effect of the magnetic flux creep on the maximum current and temperature of the superconductor before the onset of instability are derived. It is shown that the allowable overheat of the composite depends on conditions of its thermal stabilization especially near the steady-state stabilization range. It is noted that conditions for the existence of stable current states may differ when the magnetic flux creep is described by power and exponential *I–V* characteristics. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

One basic current-carrying property of a superconducting composite is the (limiting) current that can be applied to the superconductor with an extremely low rate without turning into the normal state. In each specific case, this current corresponds to the maximal current load of a current-carrying element in a superconducting magnetic system. The presence of the limiting current is the direct consequence of magnetic flux creep. Namely, when the current is applied to the superconductor, a finite electrical voltage (hence, continuous heat release) appears long before the onset of instability. Under certain conditions, equilibrium between the heat being released and the heat removed by a coolant is disturbed. As a result, the electric field distribution inside the composite becomes unstable and the material turns into the normal state.

The concept of limiting current was first put forward in [1]. In [2, 3], the formation of current states preceding the onset of the limiting current state was discussed. In particular, it was concluded that, if the rate of current injection into the material is low, instability arises because of a very small allowable overheat of the composite regardless of cooling conditions (and, in general, of conditions for thermal stabilization of a superconducting composite). This overheat is called the temperature parameter of rise of the *I–V* characteristic of the superconductor. However, the balance between the heat being released and the heat being removed is disturbed when the heat transfer coefficient changes. Hence, the balance conditions must depend on the conditions for thermal stabilization of a composite superconductor. In this work, we study the thermal states of a superconducting composite and associated limiting current values in relation to the conditions of its thermal stabilization and shape of its *I–V* characteristics, which are widely used for phenomenologically describing magnetic flux creep.

PHENOMENOLOGICAL MODELS OF SUPERCONDUCTOR CURRENT–VOLTAGE CHARACTERISTICS

The dependence of the electric field on the current induced inside a superconductor by a varying current can be divided into two ranges. First, the electric field grows nonlinearly with current. Then, starting from some value of the induced current, the field strength increases linearly.

The nonlinear portion of the *I–V* characteristic is due to various factors. The major factor is the dependence of the energy of activation on the current passing through the superconductor [4, 5]. Therefore, using the equation for the potential barrier, one can easily obtain the associated *I–V* characteristic. However, such a derivation does not necessarily involve all mechanisms responsible for the current-carrying capacity of a superconductor because of their diversity and complexity. Therefore, the use of *I–V* phenomenological models constructed from many experiments is often more appropriate.

Today, power and exponential models are widely used for the phenomenological description of the *I–V* characteristics of both low- and high temperature superconductors. In the simplest cases, they have the form [6, 7]

$$E = E_C\left(\frac{J}{J_C}\right)^n, \quad (1)$$

$$E = E_C\exp\left(\frac{J - J_C}{J_\delta}\right). \quad (2)$$

1063-7842/03/4801-0052$24.00

Here, J_C is the current density determined at a given electric field strength E_C, n is the exponent of the I–V curve, and J_δ is a constant defining the curve steepness. For the I–V curve given by (1), when the flux creep is due to various spatial defects in the superconductor lattice, the current dependence of the potential barrier is logarithmic. The I–V characteristic approximated by (2) is based on the Anderson–Kim model [4, 5], which assumes a linear dependence of the potential barrier on the current. Such I–V curves are typical of superconductors with point lattice defects. There also are macroscopic reasons for the exponential rise of the I–V curves. Specifically, the rise may result from the nonuniformity of the superconductor physical parameters in the volume; namely, the normal-to-superconducting transition takes place at different values of the temperature, magnetic induction, and current density in different regions of the volume. Along with volume nonuniformity of superconductor critical parameters, it may also exhibit longitudinal nonuniformity of superconducting properties. In this case, the exponent in (2) becomes quadratic rather than linear. However, the I–V curves of such superconducting materials is also fairly well approximated with Eq. (1).

Although the above expressions do not exhaust the variety of superconductor I–V curves, they fit a wide spectrum of experimental data with a reasonable accuracy. The convenience of these formulas for the I–V curves is that they allow for the limiting transition to the model of the critical state [8] at $n \longrightarrow \infty$ and $J_\delta \longrightarrow 0$. Moreover, power and exponential equations written in forms (1) and (2) allow for the approximation of almost any dependence $E(J)$ found experimentally by properly varying the initial parameters of the I–V curve.

RESULTS AND DISCUSSION

Consider the injection of current into a cooled composite consisting of a superconductor exhibiting the I–V curve in form (1) or (2) that is uniformly distributed in a stabilizing normal-metal matrix. We will assume that the time of current injection is sufficiently large and, hence, the field and temperature nonuniformities over the cross section are negligibly small. Formally, this means that the current is injected with an infinitely low rate. Within such a model, the thermal state of the superconducting composite can be described with the simplified heat balance equation

$$EJ = \frac{hp}{S}(T - T_0),$$

where h is the heat transfer coefficient, p is the cooled perimeter of the composite, S is the cross-sectional area, T_0 is the coolant temperature, and J is the total current density in the composite. The current density J is the sum of the current densities in the superconductor, J_s, and in the matrix, J_m:

$$J = \eta J_s + (1 - \eta) J_m.$$

The current-induced electric field strength is found from the equality of the voltage drops across the superconductor and across the matrix. For a superconducting composite with a power I–V curve, this condition is written as

$$E = E_C\left(\frac{J_s}{J_C}\right)^n = J_m\rho_m;$$

for a superconducting composite with an exponential I–V curve,

$$E = E_C \exp\frac{J_s - J_C}{J_\delta} = J_m\rho_m.$$

To simplify subsequent analysis, we will assume that the current density J_C is a linear function only of temperature [6, 7]:

$$J_C = J_{C0}\frac{T_C - T}{T_C - T_0}.$$

Here, J_{C0} and T_C are given constants. Let us introduce the dimensionless variables

$$i = \frac{J}{\eta J_{C0}}, \quad e = \frac{E(1-\eta)}{\eta\rho_m J_{C0}}, \quad e_C = \frac{E_C(1-\eta)}{\eta\rho_m J_{C0}},$$

$$\Theta = \frac{T - T_0}{T_C - T_0}, \quad \delta = \frac{J_\delta}{J_{C0}}.$$

Eliminating temperature from these equations, one can easily derive the associated dimensionless dependences of the electric field on the current:

$$\frac{e}{e_C} = \begin{cases} \left(\dfrac{i-e}{1-\alpha e i}\right)^{1/n} \\ \exp\dfrac{i - e - 1 + \alpha e i}{\delta}, \end{cases}$$

where α is the thermal stability parameter [6, 7]:

$$\alpha = \frac{\eta^2 J_{C0}^2 \rho_m S}{hp(1-\eta)(T_C - T_0)}.$$

The process of current injection becomes unstable when the balance between the Joule heat release and heat transfer from the surface into a coolant breaks down. The instability arises in the limit $\partial e/\partial i \longrightarrow \infty$ [1]. From this condition, it is easy to find the limiting current and also the related voltage and overheat. For superconductors with the power I–V curve, these

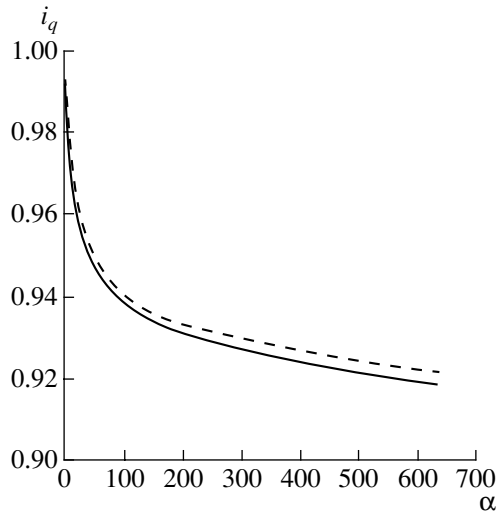

Fig. 1. Limiting current vs. thermal stability parameter. Dashed curve, *I–V* curve (1); continuous curve, *I–V* curve (2).

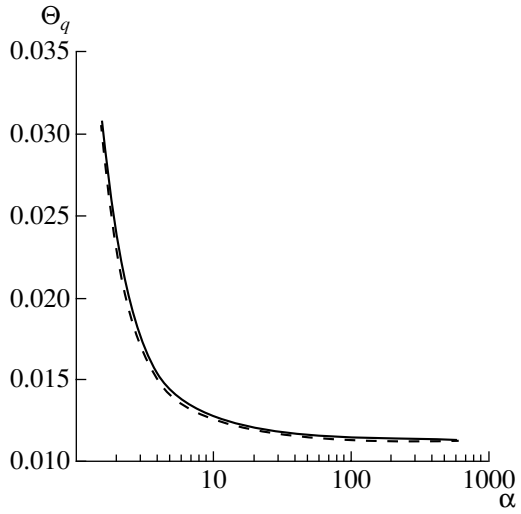

Fig. 2. Allowable overheat vs. thermal stability parameter. The curves stand for the same as in Fig. 1.

parameters are given by

$$i_q = \frac{e_q + \left(\frac{e_q}{e_C}\right)^{1/n}}{1 + \alpha e_q \left(\frac{e_q}{e_C}\right)^{1/n}},$$

$$e_q = \frac{1 - n + \alpha i_q^2(1+n) - \sqrt{[1 - n + \alpha i_q^2(1+n)]^2 - 4\alpha i_q^2}}{2\alpha i_q^2},$$

$$\Theta_q = \alpha e_q i_q;$$

for superconductors with the exponential *I–V* curve,

$$i_q = 1 - \delta - \delta \ln \frac{(\alpha i_q - 1)e_C}{\delta},$$

$$e_q = \frac{\delta}{\alpha i_q - 1}, \quad \Theta_q = \frac{\alpha i_q \delta}{\alpha i_q - 1}.$$

From these expressions, one can estimate the limiting current and the overheat preceding the breakdown of the superconductivity with regard for flux creep given by Eqs. (1) and (2).

Figures 1 and 2 show the limiting current and allowable overheat vs. thermal stability parameter, i.e., vs. the amount and properties of the stabilizing normal metal and heat removal conditions. The initial parameters were taken to be $J_{C0} = 3.5 \times 10^9$ A/m^2, $E_C = 0.6522 \times 10^{-2}$ V/m, $J_\delta = 4 \times 10^7$ A/m^2, $n = 87.5$, $T_0 = 4.2$ K, $T_C = 9$ K, $\rho_m = 2 \times 10^{-10}$ Ω m, $\eta = 0.5$, and $\alpha = 5 \times 10^{-4}$ m. The value of J_δ was selected such that the conditions $n = J_{C0}/J_\delta$ is fulfilled. In this case, the *I–V* curves calculated by (1) and (2) in the isothermal approximation ($T = 4.2$ K) touch each other at a given point $\{J_{C0}, E_C\}$ (Fig. 3). The above parameter values refer to Nb–Ti semiconductor in the copper matrix.

The results obtained demonstrate that there exists a range of the stability parameter where it greatly affects the allowable overheat and limiting current. This is true for $\alpha < 10$, i.e., near the steady-state stabilization range, where a superconducting composite has the highest current-carrying capacity. As is known, in this α range, special precautions are taken to insure the thermal stabilization of current-carrying elements and, hence, to provide trouble-free operation of large superconducting magnetic systems. In general, as α increases (for example, the heat transfer coefficient decreases), the limiting current and allowable overheat decrease monotonically, asymptotically approaching the minimal values. According to the formulas written above, with $\alpha \longrightarrow \infty$, $i_q \longrightarrow 0$ and $\Theta_q \longrightarrow 1/(\eta + 1)$ for the power *I–V* curve and $i_q \longrightarrow 0$ and $\Theta_q \longrightarrow \delta$ for the exponential *I–*

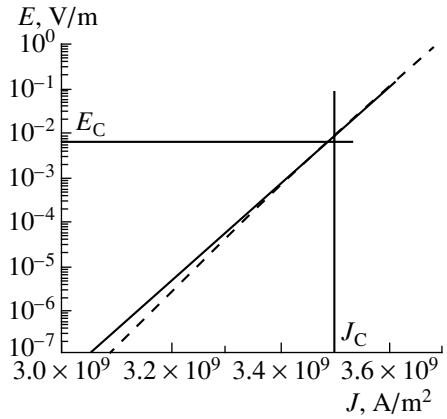

Fig. 3. Isothermal *I–V* curves for Nb–Ti superconductor calculated by different equations. The curves stand for the same as in Fig. 1.

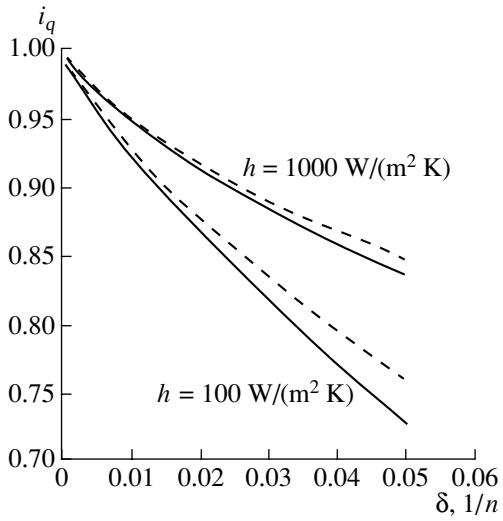

Fig. 4. Limiting current vs. parameters of rise of *I–V* curve. The curves stand for the same as in Fig. 1.

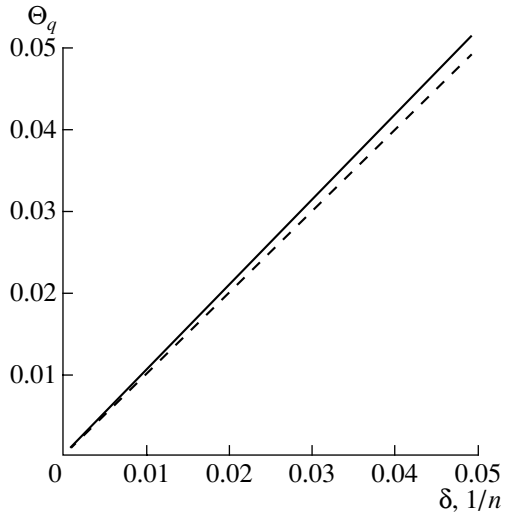

Fig. 5. Allowable overheat vs. parameters of rise of *I–V* curve. The curves stand for the same as in Fig. 1.

V curve. Thus, the conclusion that the allowable overheat does not depend on the conditions for the thermal stabilization of a composite superconductor [2, 3] cannot be applied to the entire range of the stability parameter α. This should be taken into account in order to reliably elucidate the physical reasons for stable near-critical ($i \longrightarrow 1$) limiting currents, which can be achieved even in spite of magnetic flux creep.

Figures 4 and 5 demonstrate the limiting current and allowable overheat plotted against the form and nonlinearity of the *I–V* curve. The plots suggest that the current-carrying capacity of composite superconductors decreases and the allowable overheat (preceding the onset of instability) increases nearly linearly with the nonlinearity of the *I–V* curve. Therefore, the degradation of the limiting current will inevitably grow with the "smearing" of the *I–V* curve of the superconductor.

From the aforesaid, it also follows that power and exponential *I–V* curves, strictly speaking, do not give the same values of the parameters responsible for stable current states in superconductors. The parameters coincide only at high heat transfer coefficients and within sharply ascending portions of the *I–V* curves with the nonlinearity characterized by small $1/n$ and δ. Therefore, for low-temperature superconductors, which retain superconductivity at liquid helium temperatures, creep models (1) and (2) differ insignificantly. For high-temperature superconductors, however, the calculated results for the thermal and electrodynamic states will heavily depend on whether *I–V* curve (1) or (2) is used. This should be taken into account upon processing experimental data for the *I–V* curve of high-temperature superconductors.

Thus, we analyzed stable superconducting states in the presence of flux creep up to the onset of current-induced instability. It is found that the allowable overheat of a composite superconductor depends on the thermal stabilization conditions. As a result, the composite temperature and allowable current below which the composite retains the superconductivity and current-carrying capacity are related in a nontrivial manner. It is also revealed that the states of a superconductor with flux creep described by phenomenological power and exponential *I–V* curves are generally nonequivalent. With the current injected being stable, the equivalence breaks down with increasing δ and $1/n$, which define the smearing of the superconductor *I–V* curve.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 01-02-16252).

REFERENCES

1. M. Polak, I. Hlasnik, and L. Krempasky, Cryogenics **13**, 702 (1973).
2. E. Yu. Klimenko, N. N. Martovetskiĭ, and S. I. Novikov, Dokl. Akad. Nauk SSSR **261**, 1350 (1981) [Sov. Phys. Dokl. **26**, 1180 (1981)].
3. E. Yu. Klimenko, N. N. Martovetskiĭ, and S. I. Novikov, Dokl. Akad. Nauk SSSR **282**, 1123 (1985) [Sov. Phys. Dokl. **30**, 518 (1985)].
4. P. W. Anderson, Phys. Rev. Lett. **9**, 309 (1962).
5. P. W. Anderson and Y. B. Kim, Rev. Mod. Phys. **36**, 39 (1964).
6. A. V. Gurevich, R. G. Mints, and A. L. Rakhmanov, *Physics of Composite Superconductors* (Nauka, Moscow, 1987).
7. M. N. Wilson, *Superconducting Magnets* (Oxford Univ. Press, London, 1983; Mir, Moscow, 1985).
8. C. P. Bean, Phys. Rev. Lett. **8**, 250 (1962).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 1, 2003, pp. 56–62. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 60–66.

SOLIDS

The Atomic Structure of the $1/3\langle 2\bar{1}\bar{1}0\rangle$ Superdislocation Core and Prismatic Slip in Ti_3Al

L. I. Yakovenkova, L. E. Karkina, and M. Ya. Rabovskaya

Institute of Metal Physics, Ural Division, Russian Academy of Sciences, ul. S. Kovalevskoĭ 18, Yekaterinburg, 620219 Russia

e-mail: yakovenkova@imp.uran.ru

Received February 12, 2002; in final form, July 16, 2002

Abstract—The structure of the core of a $1/6\langle 2\bar{1}\bar{1}0\rangle$ superpartial dislocation in the prism plane of the $D0_{19}$ superlattice is studied by computer simulation. Atomic interaction potentials for Ti_3Al are derived with the embedded atom method. The core of the superpartial dislocation is found to have three different configurations (of I', I, and II types) in nonequivalent prism planes $\{0\bar{1}10\}$. For screw and edge dislocations, the core is planar in type-I' prism planes and nonplanar in prism planes of type I and II. Results of simulation are compared with experimental data for the superdislocation mobility in Ti_3Al. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Plastic deformation of intermetallic compounds with different lattice structures proceeds in relation to the superdislocation core structure. It is well known that the Burgers vector of a dislocation and, hence, its elastic energy grows upon ordering. A variety of planar defects (antiphase boundaries, APBs; superlattice intrinsic stacking faults, SISFs; and complex stacking faults, CSFs) and different relationships between their energies are the reason for splitting superdislocations into various types of partial superdislocations with configurations intimately related to the configuration of the γ surface. This surface is constructed by mentally cutting a model crystallite along a given crystallographic plane and subsequently displacing the upper part relative to the lower one by a vector **f** parallel to the cut plane. The energy difference per unit surface area between a crystallite with a defect and an ideal one is the energy corresponding to the vector **f** selected (generalized stacking fault). In alloys with the $D0_{19}$ superlattice, local minima on the γ surface are lacking for a number of planes; therefore, not only the width of splitting superpartial dislocations into partial dislocations (and often the number of the latter) but also their Burgers vector will change. In this situation, computer simulation is the only way to gain detailed knowledge of the dislocation core structure and compare dislocation energies in different configurations.

In Ti_3Al, the most close-packed planes are prism planes $\{0\bar{1}10\}$, because $c/a < 1.633$. In Ti_3Al single crystals oriented for prismatic slip, the yield stress $\sigma_y\{0\bar{1}10\} \approx 150$ MPa and the plastic strain reaches ≈20% [1, 2]. The yield stress shows the usual temperature dependence, i.e., slightly declines with temperature. The same temperature dependence is also observed upon basal slip; however, the plastic strain at room temperature is much less in this case. Analysis of the slip geometry indicates that, in both cases (prismatic and basal slips), deformation is due to the motion of **a** superdislocations with the Burgers vector $1/3\langle 2\bar{1}\bar{1}0\rangle$. Electron microscopy studies [3] showed that an **a** dislocation is split into two superpartial dislocations $1/6\langle 2\bar{1}\bar{1}0\rangle$ in prism and basal planes and that these superpartial dislocations join together by the antiphase boundary. Attempts to understand the fine structure of the superpartial dislocation core have failed.

This study is a continuation of research using computer simulation of the structure and energy properties of planar defects and dislocations in the Ti_3Al intermetallic compound. In [4, 5], planar defects in (0001), $\{0\bar{1}10\}$, $\{20\bar{2}1\}$, and $\{11\bar{2}1\}$ planes were considered; in [6, 7], the core structure of the $1/6\langle 2\bar{1}\bar{1}0\rangle$ superpartial dislocation in the basal plane was studied.

As follows from calculations, in the basal plane, the γ surface, which specifies the dependence of the surface defect energy on the displacement vector, has three local minima. They are related to the antiphase boundary (the displacement vector $1/6\langle 2\bar{1}\bar{1}0\rangle$, the CSF (the displacement vector $1/6\langle 0\bar{1}10\rangle$), and the SISF (the displacement vector $1/3\langle 0\bar{1}10\rangle$). In the computer simulation of the core structure of the **a** superdislocation in the basal plane, these minima governed dislocation splitting in the initial configuration. For the superpartial dislocation with the Burgers vector $1/6\langle 2\bar{1}\bar{1}0\rangle$, which bor-

ders the antiphase boundary, the final configuration resulting after the relaxation procedure did not change. For the screw superpartial dislocation, as well as for the 60° dislocation with the Burgers vector $1/3\langle 0\bar{1}10\rangle$, which borders the SISF, the core structure was found to be nonplanar. It is shown that relative displacements near the core of these dislocations are uniformly distributed in the prism and in several basal planes simultaneously. Such a core reconfiguring during the computer simulation is explained by the fact that the split width of these superpartial dislocations in the initial configuration was on the order of the core radius.

A feature of the $D0_{19}$ superstructure is the presence of several nonequivalent sections in the set of planes $\{0\bar{1}10\}$, which differ in atom arrangement and, hence, in surface defect energies. It was found [4] that γ surfaces in $\{0\bar{1}10\}$ planes have a single low-energy minimum, which corresponds to the displacement of the antiphase boundary. This fact is responsible for the split of the **a** superdislocation with the Burgers vector $1/3\langle 2\bar{1}\bar{1}0\rangle$ in the prism plane into two superpartial dislocations bordering the antiphase boundary. The elucidation of the core structure of the $1/6\langle 2\bar{1}\bar{1}0\rangle$ superpartial dislocation in the prism plane and its energy properties is the aim of this work.

SIMULATION METHOD

The computer simulation of the $1/6\langle 2\bar{1}\bar{1}0\rangle$ superpartial dislocation in the prism planes was performed with the molecular dynamics method using *N*-atom interaction potentials for Ti_3Al found by the embedded atom method [4]. Elastic modula and planar defect energies (in the basal and prism planes) found through these potentials agree with experimental data. With the same potentials, we simulated planar defects in (0001), $\{0\bar{1}10\}$, $\{20\bar{2}1\}$, and $\{11\bar{2}1\}$ planes [4] and the $1/3\langle 2\bar{1}\bar{1}0\rangle$ superdislocation in the basal plane [5].

Upon simulating a screw dislocation, one microcrystallite edge was aligned with the $\langle 2\bar{1}\bar{1}0\rangle$ axis of the dislocation; for an edge dislocation, with the [0001] axis. Along these directions, periodic boundary conditions were imposed on atomic displacements. In other directions, the atoms were fixed in positions corresponding to the split dislocations. One of the superpartial dislocations was placed at the crystallite center. The other was outside the crystallite at a distance calculated within the elasticity theory for the antiphase boundary energy in the prism planes that corresponds to the atomic interaction potential selected, and it influenced the crystallite through its displacement fields. In the initial configuration, crystallite atoms were assigned displacements induced by the split superdislocation. The computer simulation allowed us to find the atom positions in the lattice that meet the energy minimum for the crystallite with the dislocation. The range of relative displacements of the nearest atoms that corresponds to the dislocation core was determined by the method of differential displacements. In the figures, the displacement component does not exceed $0.05a$.

SELECTION OF THE INITIAL CONFIGURATION

The computer simulation of the dislocation core in various structures shows that the results may depend on the initial configuration. This dependence is due to specific features of the γ surface. If the γ surface has a deep local minimum corresponding to a low-energy stacking fault, the Burgers vector, as a rule, coincides with the displacement vector specifying this stacking fault. If local minima are absent, one should consider several initial configurations with different axis directions, numbers of partial dislocations, and their Burgers vectors. Such a consideration is necessary for finding (after relaxation) the core structure with the minimal energy.

The superlattice $D0_{19}$ can be obtained by translating four atomic planes (II, I, I', and I) in the direction $\langle 0\bar{1}10\rangle$. These planes differ in arrangement of atoms of two sorts and in spacing between these planes (Fig. 1). In [4], the energy ζ of the antiphase boundary in prism plane II (section II in Fig. 1) was found to be 318 mJ/m^2; for prism plane I (sections I and I' in Fig. 1), it was found to be 6 mJ/m^2. For the three sections I, I', and I in Fig. 1, the antiphase energy is the same but the shape of the γ surface and, accordingly, the surface defect energies for sections I and I' differ. The γ surfaces for prism planes I and II are depicted in [4]; those for prism planes I', in Fig. 2. The point *A* in Fig. 2 corresponds to the perfect lattice; point *B*, to the displace-

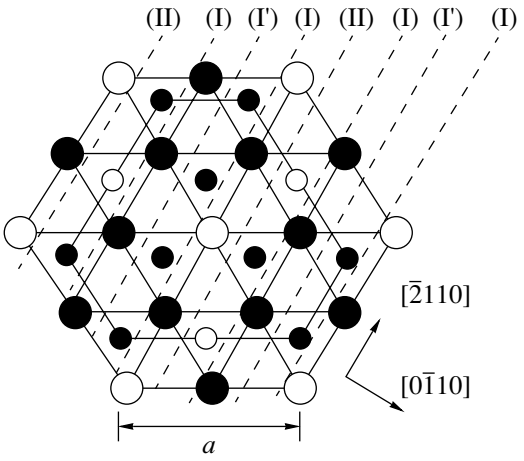

Fig. 1. Projection of the unit cell of the superlattice $D0_{19}$ onto the basal plane (●, Ti; ○, Al). Sections I, I', and II of prism planes $\{01\bar{1}0\}$ are shown.

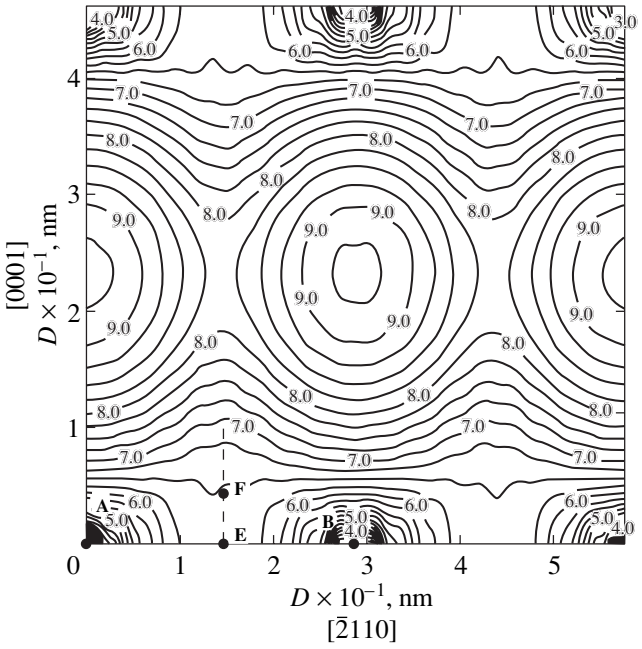

Fig. 2. γ surface of prism plane I' in Ti_3Al.

ment 1/6⟨$2\bar{1}\bar{1}0$⟩ responsible for the antiphase boundary. It was shown [4] that low energies of surface defects on the γ surface in prism plane I correlate with displacements along the direction [$2\bar{1}\bar{1}0$] from the point **A** to the point **B**. When displacements occur in other directions, resulting surface defects have much higher energies. For prism plane II [4], the height of peaks on the γ surface is much lower because of different interplanar spacings in these sections. The general view of the contour maps for prism planes I and II is the same, since the spacings between these planes are the same (Fig. 1). However, the presence of atoms of two sorts in the superlattice $D0_{19}$ makes these planes nonequivalent; therefore, for prism plane I', the energies of the surface defects are appreciably lower. The stacking fault energy is the lowest for displacement vectors in the [$2\bar{1}\bar{1}0$] direction, as well as for those from the point **E** to the point **F** (Fig. 2). In the latter case, the displacement vector is 1/12[$2\bar{1}\bar{1}x$] = 1/12[$2\bar{1}\bar{1}0$] + x/12[0001]. In this direction, the stacking fault energy varies slowly, reaching the maximum $\zeta = 650$ mJ/m^2 at $x = 1.038$ for prism plane I' and $\zeta = 910$ mJ/m^2 at $x = 1.45$ for prism plane II.

Since the Burgers vector of the superpartial dislocation 1/6⟨$2\bar{1}\bar{1}0$⟩ is large, one might expect its elastic energy to be significantly reduced when it splits into two partial dislocations. However, the mechanism of splitting of superpartial dislocations in prism planes remains uncertain, since local minima on the γ surface are absent. Yet the splitting into partial dislocations joined by an unstable stacking fault is discovered both when the dislocation core structure is treated within the elasticity theory [8] and by computer simulation [9].

Analysis of the γ surface in prism plane I shows that the stacking fault energy is the lowest for the displacement along the [$2\bar{1}\bar{1}0$] direction. Hence, the Burgers vector b_n of the partial dislocations must have a screw component that is a multiple of the Burgers vector of the superpartial dislocation: $b_n = (1/6\langle 2\bar{1}\bar{1}0\rangle)/n$. Since for other directions of the displacement vectors the stacking fault energy is much higher, the edge component of the Burgers vector of partial dislocations can be set equal to zero. The number n of partial dislocations was varied from two to ten. For any n, the total split width D of the superpartial dislocation was taken to be equal to $D \approx 5a/2$, while for $n = 10$, the spacing between two neighboring partial dislocations is $d \approx a/4$ ($a = 0.5765$ nm is the lattice parameter for Ti_3Al). Then, $d \approx r_0$, where r_0 is the dislocation core radius. Thus, the case of splitting into $n = 10$ partial dislocations is equivalent to the continuous distribution of the Burgers vector density.

Since for prism planes I' and II there exists a domain of displacement vectors (Fig. 2) where the surface defect energy is relatively low, we considered two variants of splitting the edge (or screw) 1/6⟨$2\bar{1}\bar{1}0$⟩ superpartial dislocation in these planes. The first variant, splitting into n screw (or edge) partial dislocations with the same Burgers vectors, is akin to that considered for prism plane I. The other implies splitting into two groups of partial dislocations. The total Burgers vector of the first group is equal to the displacement vector **AF**, which is associated with the local minimum along the line **EF** on the γ surface (Fig. 2); the total Burgers vector of the second group of the partial dislocations equals the displacement vector **FB**. Either group has the same number of partial dislocations with the same Burgers vectors. Thus, partial dislocations in both groups have the same screw (or edge) components of the Burgers vector, while the edge (or screw) components of the Burgers vector are equal in magnitude but opposite in sign.

SCREW AND EDGE DISLOCATIONS IN PRISM PLANE I

Figure 3 shows the energy gain when an edge superpartial dislocation splits into n partial dislocations as a result of relaxation. Curve *1* in Fig. 3 corresponds to the edge dislocation in prism plane I. It is seen that the energy is higher at $n = 2$. At $n > 3$, the energy of the dislocation configuration after relaxation only slightly depends on the number of partial dislocations in the initial configuration. It turns out that the distributions of the displacement edge components near the superpartial edge dislocation core for $n = 2$–10 are similar. In

fact, the core region $1/6\langle 2\bar{1}\bar{1}0\rangle$ of the dislocation extends in the prism plane by ≈4.5a for $n = 2$ and by ≈5.0a for $n = 10$. Along the direction that is normal to the prism plane, the core region is distributed in three neighboring prism planes. The sizes of the core region suggest that the region is planar. However, the maximal displacements for the edge dislocation are localized in type-I prism planes other than the plane where the initial superdislocation was split. Instead, the displacements were maximal in two neighboring prism planes of type I' and II. As follows from the distribution of the displacement edge component near the core of the $1/6\langle 2\bar{1}\bar{1}0\rangle$ edge superpartial dislocation in the prism plane (both for $n = 2$ and $n = 10$), the region where the relative displacements along the $[0\bar{1}10]$ plane are maximal is asymmetric. The asymmetry of the core is likely to reflect the discreteness of the distribution of the Burgers vector density, as in the case of spitting into two partial dislocations in the prism plane.

From the distribution of the displacement screw component near the core of a screw superpartial dislocation for $n = 10$, it follows that the core of the $1/6\langle 2\bar{1}\bar{1}0\rangle$ superpartial dislocation is extended along the prism plane $[0\bar{1}10]$ by a distance of ≈3.2a and along the basal plane (0001) by ≈1.7a. Although in the former case, the displacement region is almost half as much as in the latter, the dislocation core cannot be viewed as planar, since appreciable displacements are observed in eight planes that are parallel to the prism plane with the antiphase boundary. For a screw superpartial dislocation at $n = 2$, the displacement distribution is similar to the case $n = 10$. In the former case, however, the displacements are distributed along prism planes more nonuniformly, as in the case of splitting into two partial dislocations.

SCREW AND EDGE DISLOCATIONS IN PRISM PLANE I'

Curve *2* in Fig. 3 shows the energy gain when a $1/6\langle 2\bar{1}\bar{1}0\rangle$ edge superpartial dislocation splits into n partial edge dislocations in prism plane I'. The energy gain is slightly dependent on the number of partial dislocations in the initial configuration. The same dependence is observed when a $1/6\langle 2\bar{1}\bar{1}0\rangle$ dislocation splits into n partial mixed dislocations.

Figure 4 shows the core structure of a $1/6\langle 2\bar{1}\bar{1}0\rangle$ edge dislocation split into (a) two and (b) ten partial dislocations (the displacement edge component is shown). In both cases, the core can be considered as planar. However, for $n = 10$ the maximal displacements are localized only in the prism plane with the antiphase boundary, while as n decreases, displacements in parallel prism planes appear (Fig. 4a). For mixed partial dislocations, the displacement field near the core is similar.

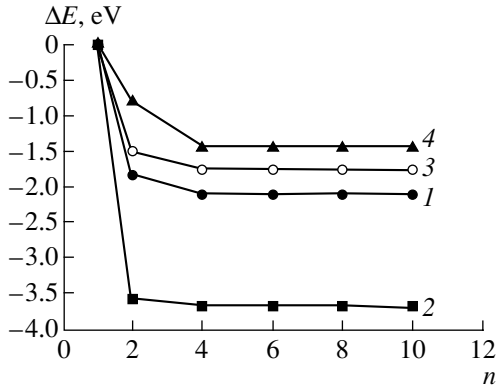

Fig. 3. Energy gain upon splitting an edge superpartial dislocation into n partial dislocations.

Figure 4c shows the core structure of a screw superdislocation for $n = 10$ (the displacement screw component is shown). The maximal sizes of the core region in the directions [0001] and $[0\bar{1}10]$ are ≈3a and ≈1.1a, respectively. Such a core can be considered as planar. For $n = 2$, the core is asymmetric, which indicates the discrete distribution of the Burgers vector density when the superdislocation splits into two partial dislocations in prism plane I'. If a $1/6\langle 2\bar{1}\bar{1}0\rangle$ screw dislocation splits into mixed partial dislocations, both the energy and the core structure change insignificantly.

SCREW AND EDGE DISLOCATIONS IN PRISM PLANE II

Curves *3* and *4* in Fig. 3 show the energy gain when an edge $1/6\langle 2\bar{1}\bar{1}0\rangle$ superpartial dislocation in prism plane II splits into n partial edge (curve *3*) and mixed (curve *4*) dislocations. The energy gain ΔE decreases with increasing number n of dislocations in the initial configuration. The core region of the edge dislocation for $n = 10$ is shown in Fig. 5a. The core sizes are nearly the same for $n = 2$–10 in both $[2\bar{1}\bar{1}0]$ and $[0\bar{1}10]$ directions. From the distribution of the displacement edge component, it follows that the core is planar although appreciable relative displacements are distributed in neighboring prism planes of type I' that are parallel to the prism plane with the antiphase boundary.

For prism plane II, we calculated the energy and structure of the core of an edge superpartial dislocation split into n mixed partial dislocations. Curve *4* in Fig. 3 is the energy gain due to dislocation splitting for this case. The displacement distribution is similar to that depicted in Fig. 5a.

From Fig. 3, it follows that for prism plane II the energy gain due to the splitting of an edge superdislocation is very low. Moreover, the splitting of a screw superdislocation was found to be energetically unfavorable. After the relaxation, the energy of the unsplit dis-

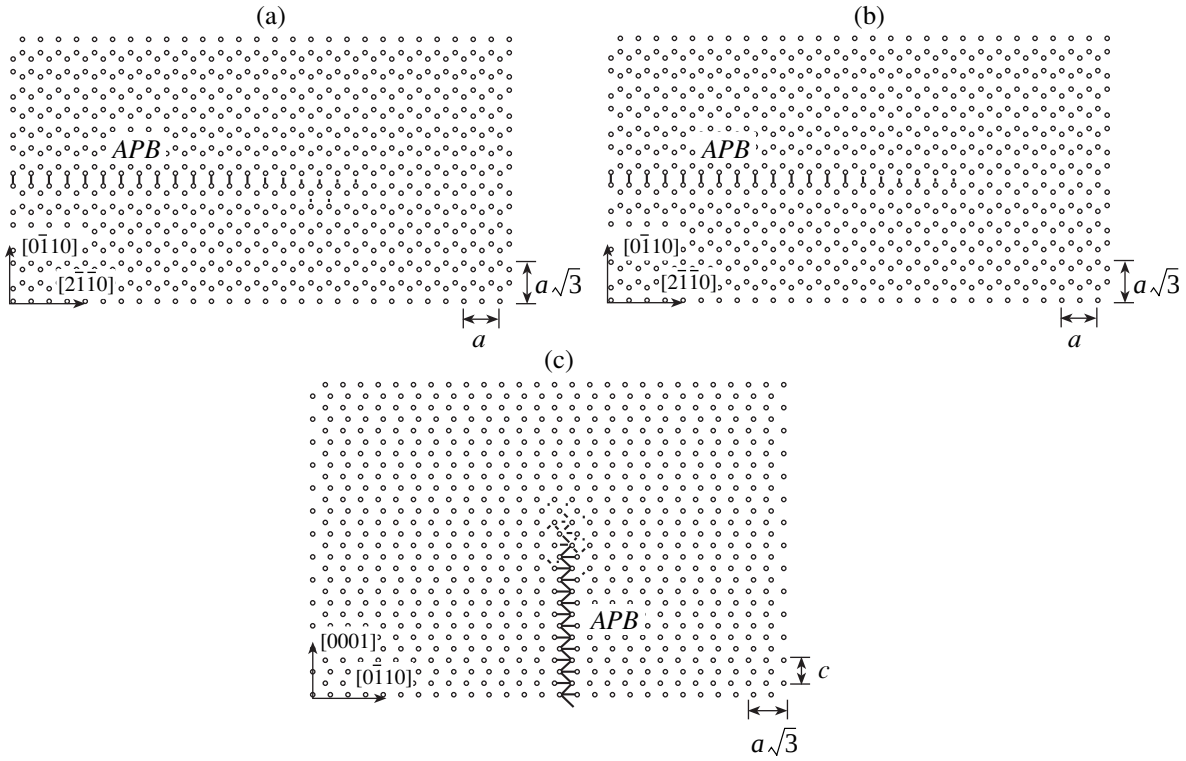

Fig. 4. Core structure of (a, b) edge and (c) screw $1/6\langle 2\bar{1}\bar{1}0\rangle$ superpartial dislocations in prism plane I'. n = (a) 2 and (b, c) 10.

location was found to be less than the energy of the n partial dislocations. Figure 5b shows the core structure after the relaxation of the screw dislocation obtained from the initial configuration of the unsplit superdislocation. For this configuration, the regions where the relative displacements in the $[0\bar{1}10]$ and [0001] directions are maximal extend to ≈2.5a and ≈2.4a, respectively. Thus, the core is not planar.

RESULTS AND DISCUSSION

The core structure of dislocations in prism planes was studied in hcp metals with the potentials for Mg, Ti, and Be [10], as well as in a model ordered alloy with a $D0_{19}$ superlattice [11]. The results of computer simulation showed that, in both hcp metals and the $D0_{19}$ superlattice, the core of an **a** superdislocation has a planar configuration. In both cases, only one configuration corresponding to prism plane I' (in terms of this work) was studied.

Our computer simulation of the core of a $1/6\langle 2\bar{1}\bar{1}0\rangle$ superpartial dislocation bordering the antiphase boundary in the prism plane showed that the energy of the split configuration of both edge and screw superdislocation in prism plane I' ($\zeta \approx 6$ mJ/m^2) is much lower than in prism planes I ($\zeta \approx 6$ mJ/m^2) and II ($\zeta \approx 318$ mJ/m^2). This relationship remains valid for any n. Obviously, the lower value of the antiphase boundary energy and of the energy of the generalized stacking fault in prism plane I' means a lower dislocation energy. The analysis of the structure of edge and screw superpartial dislocations in this plane suggests that the core is planar. In other words, the region where the displacements are maximal is localized near prism plane I', which contains the antiphase boundary. This implies that the Peierls–Nabarro stress, under which the superdislocation starts moving, is low and, hence, the superdislocation has a high mobility in prism plane I'. A high mobility of **a** dislocations explains the experimentally observed low stress yield in this plane in Ti_3Al.

An energy gain upon splitting an edge superpartial dislocation in planes I and II was found to be lower, while for a screw superpartial dislocation in prism plane II, the unsplit configuration turned out to be energetically more favorable. The core structures of these dislocations differ significantly. Both edge and screw superpartial dislocations in these prism planes have a nonplanar core (Fig. 5). The nonplanar core of screw dislocations was also discovered in fcc metals [12], bcc metals [9], ordered alloys with a $L1_2$ superlattice [12], and hcp crystals [10]. Splitting of such a type is observed when the stacking fault energy in the slip plane of a dislocation is high and, therefore, the repeat splitting of the dislocation in several planes passing through the dislocation line becomes energetically favorable.

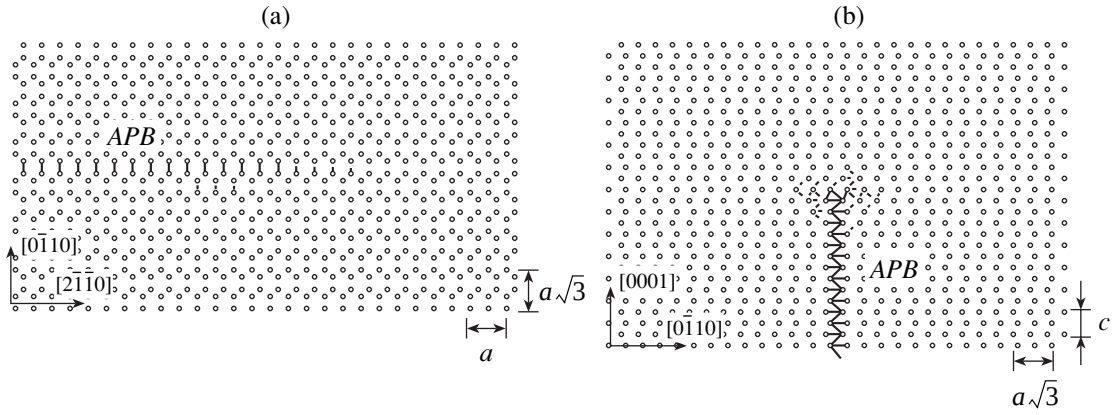

Fig. 5. Core structure of (a) edge (n = 10) and (b) unsplit 1/6$\langle 2\bar{1}\bar{1}0 \rangle$ screw superpartial dislocations in prism plane II.

Usually, nonscrew dislocations are planar. In Ti_3Al, edge dislocations were also found to be nonplanar in prism planes I and II. Such a splitting of an edge dislocation is associated with the features of the γ surface in adjacent parallel prism planes. For example, in prism plane I, the antiphase boundary energy is low ($\zeta \approx$ 6 mJ/m^2); therefore, the splitting of a 1/3$\langle 2\bar{1}\bar{1}0 \rangle$ superdislocation into two superpartial dislocations is energetically favorable. The stacking fault energy in this plane is very high, specifically, much higher than in neighboring prism planes I' and II; therefore, a 1/6$\langle 2\bar{1}\bar{1}0 \rangle$ superpartial dislocation splits not in prism plane I but in neighboring planes I' and II. For prism plane II, the reconfiguring of the **a** dislocation core is also energetically favorable. In this case, the reconfiguring is accompanied by the redistribution of displacements into the nearest parallel plane of type I' that has a lower energy of the generalized stacking fault.

The analysis of displacements near the core of an edge superpartial dislocation in prism planes I, I', and II (Figs. 4, 5a) shows that for any number n of partial dislocations, the distribution of the displacements over the prism planes is nonuniform and can be characterized as splitting into two partial dislocations. Experimentally, the splitting of a 1/6$\langle 2\bar{1}\bar{1}0 \rangle$ superpartial dislocation into two partial dislocations in prism planes was observed in [3].

The fact that the core of a superpartial dislocation in planes I and II is nonplanar implies their low mobility. Thus, the computer simulation shows that both glissile and sessile superdislocations with a Burgers vector 1/3$\langle 2\bar{1}\bar{1}0 \rangle$ can be present in prism planes $\{01\bar{1}0\}$. *In situ* experiments [2] confirm the presence of **a** dislocations with various mobilities in prism planes. According to calculations, mobile dislocations can be identified as dislocation configurations in prism plane I', while sessile dislocations can be identified as dislocations in prism planes I and/or II. For nonplanar-core dislocations to move, the displacements must be redistributed into the core region, which eventually causes the jerky motion of dislocations. Therefore, *in situ* observations of the jerky motion in experiments indicate the nonplanar structure of the core. The jerky motion was observed [2] in Ti_3Al and several other materials, e.g., bcc metals, as well as upon the low-temperature motion of a screw dislocation with a Burgers vector 1/2$\langle 111 \rangle$ [9].

Thus, our computer simulation showed that prismatic slip is associated with the motion of 1/3$\langle 2\bar{1}\bar{1}0 \rangle$ superdislocations in prism plane I'. The energy of nonplanar configurations in prism planes I and II is higher; therefore, the conditions necessary for a glissile dislocation to be reconfigured into a sessile dislocation are absent. The fact that a glissile **a** superdislocation in prism plane I' has the lowest energy and also that its reconfiguring into a sessile structure is energetically unfavorable explains the weak temperature dependence of the yield stress, which is observed in experiments on prismatic slip in Ti_3Al [1].

CONCLUSIONS

In this work, we studied the core structure of a superdislocation with a Burgers vector 1/3$\langle 2\bar{1}\bar{1}0 \rangle$ in prism planes $\{01\bar{1}0\}$.

(1) The computer simulation revealed three core configurations of a superpartial dislocation in nonequivalent prism planes (of type I, I', and II).

(2) In prism plane I', the core of both screw and edge dislocations is planar. The energy of such configurations is the lowest.

(3) In prism planes I and II, the core of both screw and edge dislocations is nonplanar.

REFERENCES

1. M. Legros, A. Couret, and D. Caillard, Philos. Mag. A **73**, 61 (1996).
2. M. Legros, A. Couret, and D. Caillard, Philos. Mag. A **73**, 81 (1996).
3. Y. Minonishi, Philos. Mag. Lett. **62**, 153 (1990).
4. L. I. Yakovenkova, V. V. Kirsanov, L. E. Karkina, *et al.*, Fiz. Met. Metalloved. **89** (3), 31 (2000).
5. L. E. Karkina and L. I. Yakovenkova, Izv. Akad. Nauk, Ser. Fiz. **65**, 807 (2001).
6. L. I. Yakovenkova, L. E. Karkina, and M. Ya. Rabovskaya, Fiz. Met. Metalloved. **93** (1), 1 (2002).
7. L. E. Karkina and L. I. Yakovenkova, Izv. Akad. Nauk, Ser. Fiz. **65**, 842 (2001).
8. F. Kroupa and V. Vitek, Czech. J. Phys. B **14**, 337 (1964).
9. V. Vitek, Cryst. Lattice Defects **5**, 1 (1974).
10. V. Vitek and M. Igarashi, Philos. Mag. A **63**, 1059 (1991).
11. I. Cserti, M. Khantha, V. Vitek, *et al.*, Mater. Sci. Eng. A **152**, 95 (1992).
12. L. E. Karkina, B. A. Grinberg, and L. I. Yakovenkova, Fiz. Met. Metalloved. **61**, 695 (1986).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 1, 2003, pp. 63–67. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 67–72.

SOLID-STATE ELECTRONICS

Memory Element Based on Si/CaF_2 Periodic Nanostructures

Yu. A. Berashevich, A. V. Korolev, A. L. Danilyuk, and V. E. Borisenko

Belarussian State University of Information Science and Radioelectronics, ul. Brovki 17, Minsk, 220013 Belarus

e-mail: julia@nano.bsuir.edu.by

Received March 13, 2002

Abstract—A memory element based on a Si/CaF_2 periodic nanostructure is proposed. In this element, information is recorded through charge capture by trap states in a CaF_2 dielectric. The high and low signal levels correspond to the current in the maximum and minimum of the negative differential resistance region, which forms as a result of the resonant tunnel distribution of charge carriers over trap levels in the dielectric. The speed of such logical elements depends on the rate of activation carrier trapping and the rate of tunnel carrier transfer from one state to another. It is shown that both Si/CaF_2-based logical elements and memory elements proposed operate at temperatures from 77 to 300 K, have a switching time of 10^{-12}–10^{-10} s, and are compatible with silicon IC technology. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Multiple-quantum-well periodic nanostructures have been of interest for several decades, because they provide a basis for designing devices with negative-differential-resistance *I–V* characteristics [1, 2]. In this respect, III–V semiconductors have been studied most extensively and good results have been achieved when resonant tunnel transfer of charge carriers was used [3]. The low operating temperatures and poor compatibility of III–V semiconductors with matured silicon IC technology have stimulated the search for new ways of designing electron devices with negative differential resistance (NDR).

The advances in silicon technology have made it possible to obtain silicon layers of nanometer thickness and, hence, to produce periodic silicon/insulator nanostructures. Along with intense luminescence, which is typical of nanodimensional silicon, these structures also exhibit hysteretic *I–V* characteristics, zero-current shift, and NDR shift at room temperature [4, 5]. Further investigations have shown that these effects are associated with charge carrier transfer via discrete trap levels in the insulator, resulting in charge accumulation in it.

Our previous works [6–8] showed that if the energy of charge carriers in a potential well coincides with the energy level of traps in the insulator, the resonant tunnel transfer of charge carries via this level can be observed. This effect was corroborated by experimental results [9–11], which demonstrated the resonant nature of charge carrier transfer in a number of silicon/insulator periodic structures in the temperature range 77–300 K.

The aim of this paper is to develop a memory element based on Si/CaF_2 periodic nanostructures and to analyze its operation.

THE EQUIVALENT CIRCUIT OF A Si/CaF_2 PERIODIC STRUCTURE

The representation of the basic device structure through its equivalent circuit suitable for standard circuit-mechanical simulation programs is an important step in simulating the characteristics of elements incorporated into an IC.

In [6–8], we considered in detail charge carrier transfer in silicon/insulator periodic nanostructures, taking into account the elastic tunneling of charge carriers through a potential barrier formed by the insulator and carrier transfer via traps in the insulator. The latter process involves the activation capture of carriers by trap states and subsequent carrier transfer from trap to trap by tunneling through the insulator. The maximum rate of carrier transfer through the structure is achieved when the carrier energy in a well coincides with the trap level, i.e., when resonant tunnel transfer via traps dominates. The identity of trap levels in the insulator implies the conservation of energy and momentum of the carriers thus transferred [8]. This results in the appearance of the NDR region in the *I–V* curves. An increase in the trap level energy or a decrease in the empty trap concentration violates the condition for resonant transfer. The simulation of resonant tunnel transfer of charge carriers in Si/CaF_2 periodic structures showed [8] that a decrease in the temperature lowers the probability of carrier trapping. As a result, the current in the NDR region decreases, the NDR maximum shifts toward higher bias voltages, and the elastic tunneling mechanism of carrier transfer becomes more pronounced. However, up to a temperature of 77 K, the ratio of the currents in the NDR maximum and minimum changes slightly for trap levels in the insulator

1063-7842/03/4801-0063$24.00

that are located near the conduction band of silicon. Below 77 K, the NDR decreases and disappears, because elastic tunneling, which slightly depends on temperature, starts dominating.

The presence of the NDR region, which is caused by resonant carrier transfer in silicon/insulator periodic nanostructures, offers possibilities for designing logic devices with two stable states in the minimum and maximum of the NDR current. In this case, critical parameters are the ratio of the currents in the NDR maximum and minimum and the rate of system switching from one state to the other as the bias changes. The simulation of resonant tunnel transfer of charge carriers in silicon/insulator periodic structures [8] showed that the above ratio grows with decreasing number of periods and is maximal in a two-period Si/CaF_2 structure. In Si/SiO_2 structures, this ratio is smaller, because the lower barrier favors carrier transfer by the competing mechanism of elastic tunneling.

Charge transfer via traps leads to charge accumulation in the insulator. The amount of the charge accumulated depends on the ratio of the times during which the traps are filled and emptied [6–8]. The charge is redistributed and polarized by the bias voltage. As a result, an internal field arises, which reduces the current through the structure. This field also shifts the zero current relative to the zero potential and makes the *I–V* characteristic hysteretic [4, 5].

The resonant nature of charge carrier transfer and the accumulation of the trapped charge, which is responsible for the capacitive properties of the structure, allow one to describe a two-period Si/GaF_2 structure with the equivalent circuit shown in Fig. 1. Here, V_{bias} is the bias voltage source; $I(V)$ is the voltage-controlled current source, which represents the active area of the structure; C is the capacitance of the two-period structure; and R is the load resistance. This device structure can be viewed as an analogue of resonant-tunneling devices based on III–V semiconductors [3].

To simulate the current passing in the structure, we make the following assumptions. First, carrier transfer by the elastic tunneling mechanism, which was taken into account in the model developed in [8], is neglected. Then, we consider transfer via two traps: one of them,

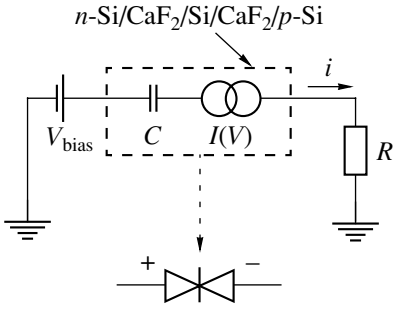

Fig. 1. Equivalent circuit of a two-period Si/CaF_2 structure.

E_{t1} = +0.08 eV (relative to the conduction band bottom), is responsible for resonant tunnel transfer, and the second, E_{t2} = +0.5 eV, specifies the current through the structure when the first trap is filled. The voltage drop across a period of the silicon/insulator structure is $V = V_{bias}/N$, where N is the number of insulating layers.

Based on these assumptions, we derived an expression for the electron current density in the structure:

$$J_n(V)/J_{n0} = (U_{(n)} - E_{t1} - Va/d_{ins})^{-0.5m} \times E_1 \frac{1}{W(V)^{-1} + E_1}\left[\frac{1}{1 + W(V)E_1} + E_2\frac{1}{E_1}\left[\frac{W(V)^{-1} + E_1}{W(V)^{-1} + E_2}\right]\left[\frac{U_{(n)} - E_{t2} - Va/d_{ins}}{U_{(n)} - E_{t1} - Va/d_{ins}}\right]^{-0.5m}\right]. \quad (1)$$

Here, d_{ins} is the insulator thickness, $a = d_{ins}/m$ is the trap spacing (m is the number of traps over the charge carrier path), and $U_{(n(h))}$ is the potential barrier height for electrons (holes),

$$E_{1(2)} = \exp(-qE_{t1(2)}/k_BT), \quad (2)$$

$$W(V) = \exp\left(\frac{qV}{k_BT}\right), \quad (3)$$

$$J_0 = J_{n0} + J_{p0} = Sv_TN_tqd_{ins}p_m(n_0 + p_N). \quad (4)$$

In (2)–(4), S is the cross section of carrier capture by a trap; v_T is the thermal velocity of charge carriers in a potential well [8]; n_0 and p_N are the concentrations of electrons and holes at the opposite injecting contacts, respectively; N_t is the trap density in the insulator; k_B is the Boltzmann constant; T is temperature; q is the electron charge; and p_m is the probability that m traps lie along the charge carrier path in the insulator [12, 13]. An equation for the hole current density is similar to Eq. (1).

The *I–V* characteristic of the two-period Si/CaF_2 structure was carried out using the parameters listed in the table. The simulation of the current source $I(V)$ in the equivalent circuit (Fig. 1) is in good agreement with the carrier transfer simulation according to the kinetic model [8] for the steady-state regime (Fig. 2).

The charge and discharge times of the capacitance C are those parameters of the equivalent circuit governing the transient process. These times are responsible for the speed of the device when a signal is applied to the input. As such signals for digital circuits, we used idealized rectangular pulses (with zero rise and fall times) with duration t_1 and spacing t_2. Then, the current through capacitance C as a function of $t_{1(2)}$ is given by

$$i_C(t) = V_{pl}C\int_0^t \exp(-t/\tau)dt. \quad (5)$$

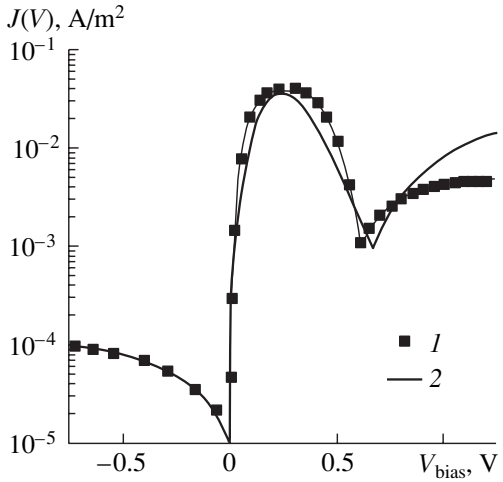

Fig. 2. *I–V* characteristic of a two-period Si/CaF_2 structure: (*1*) calculation by the model [8] and (*2*) simulation using the equivalent circuit.

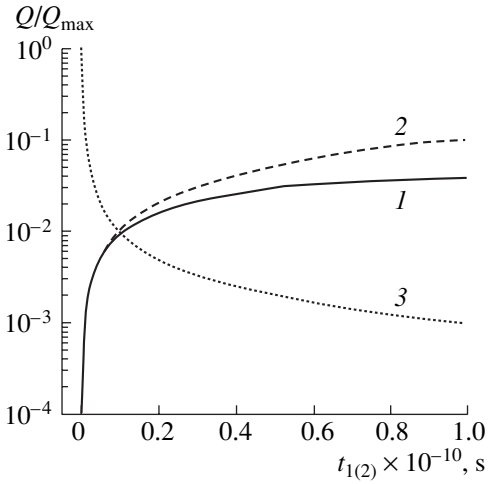

Fig. 3. Filling rate and discharge of trap states vs. duration of the pulse $V_{pl} = 0.5$ V. Q_{max} is the maximum charge density that can be trapped by the structure. (*1*, *2*) Charge of traps with the energies $E_{t1} = 0.08$ eV and $E_{t2} = 0.5$ eV, respectively, and (*3*) discharge of the traps.

Here, τ is the time constant characterizing the charge (discharge) of traps in the insulator and $t = t_1$ (t_2) when capacitance C is charged (discharged).

In the frame of the model for charge carrier transfer through a silicon/insulator periodic structure [8], the charge time of capacitance C depends on the rate of activation carrier trapping in the insulator, while the discharge time is defined by the rate of resonant tunnel transfer of charge carriers via traps. Since the time constant τ was determined from the rate of carrier trapping and transfer (via traps) (Eq. (4)), it corresponds to the total resistance of the structure. The charge and discharge of capacitance C in a two-period Si/CaF_2 structure were simulated based on the above assumptions [8]. In the charge regime, the pulse amplitude was determined from the current in the NDR peak and was taken to be equal to 0.5 V (Fig. 3). The results of simulation of transients in the structure with two trap levels ($E_{t1} = 0.08$ eV and $E_{t2} = 0.05$ eV above the conduction band bottom in silicon) are shown in Fig. 3. The time of charge accumulation in capacitance C, which depends on the trap energy in the insulator relative to the quantum well bottom, is 10^{-10} and 10^{-8} s for the first and second levels, respectively. The time of capacitance discharge and charge is about 10^{-10} s for both traps. The fact that traps with different energy levels are discharged for the same time indicates that the rate of carrier tunneling from trap to trap depends on the potential barrier width only slightly when the barrier width is comparable to the material lattice constant [8].

As the reactance of the circuit increases when capacitance C is charged, the current through the circuit decreases. As follows from the results of simulation, the current decreases nearly 20-fold for a pulsed signal duration of 10^{-10} s, which causes the ratio of the currents in the NDR maximum and minimum to diminish. As the signal duration increases, the current saturates because the first trap in the insulator, $E_{t1} = 0.08$ eV, is filled.

Thus, a two-period Si/CaF_2 nanostructure with the asymmetric *I–V* characteristic (Fig. 2), which is defined by external bias polarity, can form the basis for logic devices where two stable states of the current in the NDR maximum and minimum are used as the high and low signal levels. The equivalent circuit of the two-period Si/CaF_2 structure (Fig. 1) is suitable for the circuit simulation of this structure as a part of an IC.

Parameters used for the simulation of the equivalent circuit of a two-period Si/CaF_2 structure

Parameter	Value
Thickness of insulating and silicon layers	1.4 nm
Trap density in insulator	10^{24} m^{-3}
Electron concentration n_0 at electron contact	10^{25} m^{-3}
Hole concentration p_N at hole contact	10^{25} m^{-3}
Cross section of carrier capture by trap state	10^{-23} m^2
Barrier height for electrons and holes	3.3, 7.6 eV [14]
Number of traps lying along the carrier path in insulator	2
Position of the first and second trap levels in insulator relative to potential well bottom	0.08, 0.5 eV [5, 9]
Temperature	300 K

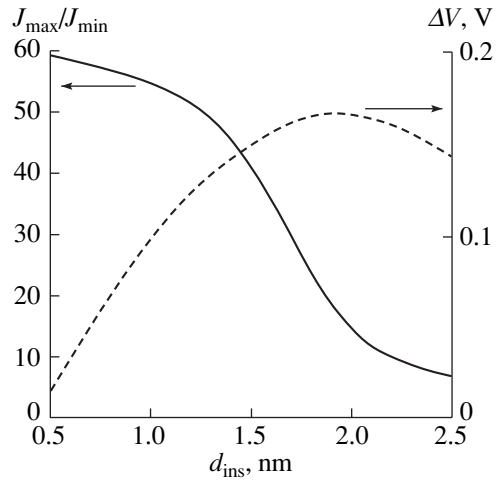

Fig. 4. Ratio of the currents in the NDR maximum and minimum and the potential ΔV as a function of the insulator thickness under the assumption that the external potential drop across insulating layers is the same.

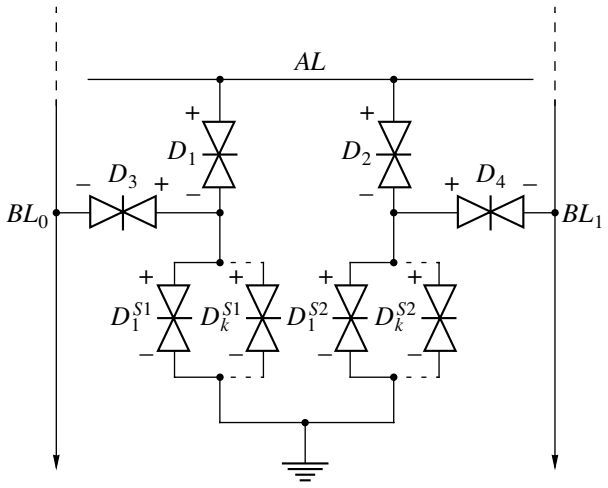

Fig. 5. Equivalent circuit of the memory cell.

A MEMORY CELL USING THE CHARGE PROPERTIES OF Si/CaF_2 PERIODIC STRUCTURES

In this paper, we describe a dynamic storage cell based on the charge properties of Si/CaF_2 structures. In such a cell, the charge of insulating layers is described by capacitance C and serves as a data bit. The presence of a charge in the layers corresponds to binary one, while the absence of a charge means binary zero. It is proposed to perform logical operations and store information in a single structure. This, however, requires certain conditions to be met for the reliable operation of the dynamic memory cell. In particular, the charge density in the cell should be increased and the charge accumulated in switching elements should be reduced.

The charge accumulated in an insulating layer depends on the density of traps, which are responsible for the maximum possible density Q_{max} of the charge trapped in the layer. Therefore, for the entire structure, this density is defined as NQ_{max} provided that the potential drop across an insulating layer remains unchanged. As the number of structure periods increases, so does the density of the stored charge. However, this generates a need to increase the bias applied to the structure. One way of removing this drawback is the parallel connection of two-period Si/CaF_2 structures. In this case, the power consumption of the memory cell does not increase.

For carrier transfer via traps in the insulator, the polarization of the trapped charge by an applied potential produces an internal field in the structure. With decreasing potential, this field governs transfer processes. Data recording into our memory cell utilizes this feature. The potential due to the internal field in the structure is expressed as

$$\Delta V = \frac{Q}{C} = \frac{2q}{\varepsilon_{ins}}[N_e w_e(d_{ins} - w_e) + N_h w_h(d_{ins} - w_h)]. \quad (6)$$

Here, w_e and w_h are the thicknesses of insulating layers where trapped electrons and holes are accumulated; ε_{ins} is the relative permittivity of the insulator; and N_e and N_h are the concentrations of electrons and holes, respectively, trapped in the insulator. These concentrations, as well as w_e and w_h, saturate when the carrier energy in a well exceeds the trap level in the insulator. The value of ΔV is maximum when trap states are partly filled, i.e., when $w_e = d_{ins}/2$ (Eq. (6)).

As follows from the experimental results, the value of ΔV increases with increasing thickness of the insulating layer [4]. On the other hand, a decrease in the insulator thickness accelerates the resonant tunnel transfer of charge carriers via traps and increases the ratio of the currents J_{max}/J_{min} in the NDR maximum and minimum [8]. The simulated dependences of J_{max}/J_{min} and potential drop ΔV (Eq. (6)) on the insulator thickness are shown in Fig. 4. The results were obtained for a two-period Si/CaF_2 structure with Eqs. (1)–(4) and (6) and the model [8].

Based on the results of simulation, we suggest the equivalent circuit of a memory cell based on a two-period resonant tunnel Si/CaF_2 structure (Fig. 5). The dependences of J_{max}/J_{min} and ΔV on the insulator thickness (Fig. 4) allow us to choose the thicknesses of insulating layers in the elements that provide data storage and perform logical operations. The thickness of the insulating layer in switching diodes D_1–D_4 is taken to be equal to 0.5–1.0 nm; in elements $D_k^{S1(2)}$ providing data storage, it is 2 nm. The value of k, which specifies the number of diodes $D_k^{S1(2)}$ necessary to record binary one, is three or four for the structure parameters selected. Note that the area of these diodes is not a crit-

ical parameter and may be as small as desired (down to several square nanometers), because the potential ΔV depends only on the insulating layer thickness.

Under the recording conditions (Fig. 5), a pulse V_{pl} with an amplitude of 1 V is applied to the address line (*AL*), while opposite pulses of about 1 and 0.2 V (corresponding to binary one and zero) are applied to the bit lines ($BL_{1,0}$). When the signal level on the bit line BL_1 corresponds to binary one, the switching diode D_4 is off and the charge is accumulated in D_k^{S2}; i.e., binary one is recorded. At the same time, the diode D_3 is on, D_k^{S1} remains uncharged, and binary zero is recorded. When information is read out from the memory cell, a signal of 0.2 V is applied to the address line and the switching diode D_4 is "on" in the presence of a potential drop $\Delta V = 0.6$–0.8 V across D_k^{S2}; hence, binary one is applied to the bit line BL_1 and binary zero, to BL_0 (Fig. 5).

The simulation shows (Fig. 3) that the transition (switching) time in two-period Si/CaF_2 structures is typically 10^{-12}–10^{-10} s. This time can be reduced by partially charging and discharging capacitance *C*. Such an approach also allows one to read out recorded information several times.

The high switching speed and nanometer sizes of the memory cell make it promising for very-high-speed IC technology.

CONCLUSION

Logical elements based on Si/CaF_2 periodic structures where the resonant tunnel transfer of charge carriers is accomplished via trap levels in the insulator are suggested. The equivalent circuit for this structure is developed, and the *I*–*V* characteristic of, and transients in, the structure are simulated. The equivalent circuit comprises a voltage-controlled current source and a capacitor. The current source is used to describe the resonant tunnel transfer of charge carriers via traps, while the capacitor takes into account charge accumulation in the insulator. The charge and discharge times of the capacitor are shown to lie in the range 10^{-12}–10^{-10} s.

In the proposed equivalent circuit of the memory cell, information is stored by means of the capacitance of resonant tunnel Si/CaF_2 structures connected in parallel. Their number is found by calculating the internal field produced by the charge trapped in the insulating layer in each of the structures.

The advantages of the memory element are the operating temperature range 77–300 K, small switching times (10^{-12}–10^{-10} s), and compatibility with silicon IC technology.

ACKNOWLEDGMENTS

This work was supported by the interuniversity program "Nanoelectronics" and the Belarussian Foundation for Basic Research (project no. T99-102).

REFERENCES

1. L. Esaki, Phys. Rev. **109**, 63 (1958).
2. R. Tsu and L. Esaki, Appl. Phys. Lett. **22**, 562 (1973).
3. C. G. Smith, Rep. Prog. Phys. **59**, 235 (1994).
4. V. Ioannou-Sougleridis, A. G. Nassiopoulou, F. Arnaud d'Avitaya, *et al.*, *European Projects: Silicon Modules for Integrated Light Engineering* (Marseille, 2000).
5. V. Ioannou-Sougleridis, T. Ouisse, A. G. Nassiopoulou, *et al.*, J. Appl. Phys. **89**, 610 (2001).
6. Yu. A. Berashevich, A. L. Danilyuk, A. N. Kholod, and V. E. Borisenko, Fiz. Tekh. Poluprovodn. (St. Petersburg) **35**, 110 (2001) [Semiconductors **35**, 112 (2001)].
7. Yu. A. Berashevich, A. L. Danilyuk, A. N. Kholod, and V. E. Borisenko, Fiz. Tekh. Poluprovodn. (St. Petersburg) **36**, 91 (2002) [Semiconductors **36**, 85 (2002)].
8. Yu. A. Berashevich, A. L. Danilyuk, and V. E. Borisenko, Fiz. Tekh. Poluprovodn. (St. Petersburg) **36**, 718 (2002) [Semiconductors **36**, 679 (2002)].
9. D. J. Paul, P. See, I. V. Zozoulenko, *et al.*, Appl. Phys. Lett. **77**, 1653 (2000).
10. M. Watanabe, T. Funayama, T. Teraji, and N. Salamaki, Jpn. J. Appl. Phys., Part 2 **39**, L716 (2000).
11. M. Watanabe, Yu. Aoki, W. Saito, and M. Tsuganazawa, Jpn. J. Appl. Phys., Part 2 **38**, L116 (1999).
12. K. J. Chen, T. Waho, K. Maezawa, and M. Yamamoto, IEEE Electron Device Lett. **17**, 309 (1996).
13. V. Ya. Kirpichenko, Zh. Éksp. Teor. Fiz. **116**, 1048 (1999) [JETP **89**, 559 (1999)].
14. M. T. Cuberes, A. Bauer, H. J. Wen, *et al.*, J. Vac. Sci. Technol. B **12**, 2646 (1994).

Translated by M. Astrov

Technical Physics, Vol. 48, No. 1, 2003, pp. 68–72. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 73–78.

SOLID-STATE ELECTRONICS

Formation of Ordered Helium Pores in Amorphous Silicon Subjected to Low-Energy Helium Ion Irradiation

V. F. Reutov and A. S. Sokhatskiĭ

Flerov Laboratory of Nuclear Reactions, Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia

e-mail: sohatsky@nrsun.jinr.ru

Received June 24, 2002

Abstract—Thin transparent (for transmission electron microscopy, TEM) self-supported Si(001) films are irradiated on the (110) end face by low-energy (E = 17 keV) He^+ ions at doses ranging from 5×10^{16} to 4.5×10^{17} cm^{-2} at room temperature. The TEM study of the irradiated Si films along the ion range shows that an a-Si layer forms in the most heavily damaged region and helium pores (bubbles) with a density of up to 3×10^{17} cm^{-3} and 2–5 nm in diameter nucleate and grow across the entire width of this layer. The growth of nanopores in the a-Si layer is accompanied by their linear ordering into chains oriented along the ion tracks. The absence of pores in the region that remains crystalline and has the maximal concentration of implanted helium is explained by the desorption of helium atoms from the thin film during the irradiation. After annealing at 600°C, the volume of immobile pores in the remaining a-Si layer increases owing to the capture of helium atoms from the amorphous matrix. Solid solution is shown to be the prevalent state of the helium implanted into the amorphous silicon. Linear features with a diameter close to 1 nm and density of about 10^7 cm^{-1} discovered in the helium-doped a-Si layer are identified as low-energy He^+ ion tracks. *© 2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

In recent years, the behavior of helium in semiconductors has been extensively studied for at least two reasons. First, helium pores are extremely efficient for gettering undesirable impurities, especially transition metals [1]. Second, the method of gas cleavage [2, 3] is used in the production of "silicon-on-insulator" structures. Moreover, there is the prospect of growing silicon crystals with an extremely high density (about 10^{18} cm^{-3}) and narrow size distribution of nanopores [4]. In this case, the resulting 3- to 5-nm-spacing between the pore walls provides luminescence from silicon nanocrystals [5]. This quantum-size phenomenon is rather attractive, since it gives a key to the design of active optoelectronic devices on silicon, which is the basic material of present-day electronics.

However, the helium ion doses required for the formation of porous structures in silicon are often above 10^{17} cm^{-2}, which causes appreciable radiation-induced damage. Under certain conditions, severe damage may lead to amorphization [6, 7].

Although the possibility of helium pores arising in amorphous silicon was discovered in [8], the behavior of helium atoms implanted into a-Si has been poorly understood. However, by analogy with crystalline silicon (c-Si), one may expect that helium atoms implanted into a-Si agglomerate into bubbles and have an anomalously high solubility as in c-Si [9, 10]. In our opinion, the essential difference in the behavior of helium in amorphous and crystalline Si must be in the migration activity of helium atoms.

In particular, it was shown by computer simulation [11, 12] that helium atoms may be stabilized in tetrahedral interstices of the Si lattice. The repulsion force acting on the interstitial helium from single vacancies is sufficient to overcome the barrier for diffusion activation (about 0.8 eV) [12]. This mechanism of diffusion activation is likely to be impossible in amorphous silicon; therefore, a much lower mobility of helium dissolved in a-Si should be expected. In this context, the formation of helium porosity in amorphous silicon is of particular interest.

In this study, we consider the formation and thermal stabilization of helium pores in amorphous layers produced in thin crystalline silicon specimens by low-energy helium ion irradiation.

EXPERIMENTAL

The TEM study of structure modifications in silicon along the path of bombarding helium ions was performed by a specially elaborated longitudinal-section technique [13, 14]. In contrast to the widely used transverse-section method (XTEM), which implies successive stages of irradiation, object preparation for XTEM, and TEM study, our technique is based on a radically different sequence of stages, where the preparation of an object by the longitudinal-section method precedes irradiation. Note that, in our method, TEM study immediately follows irradiation.

1063-7842/03/4801-0068$24.00

Test specimens prepared by the longitudinal-section method were thin self-supported Si(001) plates with the end faces parallel to the (110) crystallographic plane and a thickness t falling into the range of transparency for electrons in a transmission electron microscope; i.e., t was less than 350 nm (Fig. 1). The specimens were prepared by the electrochemical jet thinning of $4 \times 4 \times 0.35$-mm Si(001) chips with the subsequent formation of the (110) end face. To control the thickness of the specimen, we measured interference fringes of equal thickness in an optical microscope. The surface irradiated was the (110) end face of the self-supported thin Si film. The upper and lower (001) surfaces of the specimen remained intact owing to the strict collimation of the beam. The scheme of ion irradiation and TEM study by the longitudinal-section method is shown in Fig. 1.

It is reasonable to assume that the ions striking the surface of the specimen do not all stop in its bulk: many of them will scatter out of the specimen through its lateral surfaces. This circumstance should be taken into account when determining the doping profile and the profile of radiation-induced damage from the ion range in the thin specimen. To do this, we used the TRIM-98 program supplemented by a specially devised program that takes into consideration the elastic scattering of ions and recoil atoms out of a thin film of a given thickness.

Thin Si specimens were exposed to a mass-separated beam of He^+ ions with an energy E = 17 keV, intensity of 2×10^{14} cm^{-2} s^{-1}, and doses ranging from 10^{16} to 4.5×10^{17} cm^{-2}. The angle between the ion beam and the end face of the specimen bombarded was 90°, 65°, 45°, and 30°. The maximal temperature of the Si surface during the irradiation was estimated as 40°C.

The samples were annealed by rapidly heating in a dry nitrogen flow to 600°C at a rate of 10°C s^{-1} with subsequent immediate cooling (without holding). This heating rate was chosen with regard to the rate of thermal desorption of helium from silicon [9] in order that the helium implanted be completely left in the specimen.

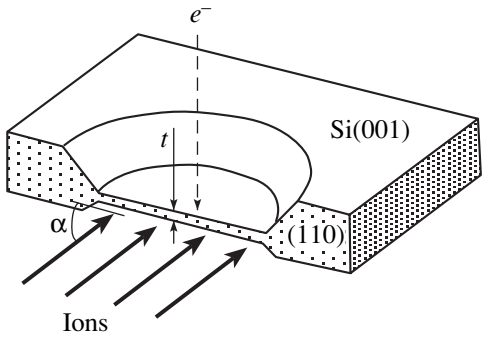

Fig. 1. Thin self-supported Si film irradiated by the longitudinal-section technique (e^- is the direction of the electron beam in a TEM).

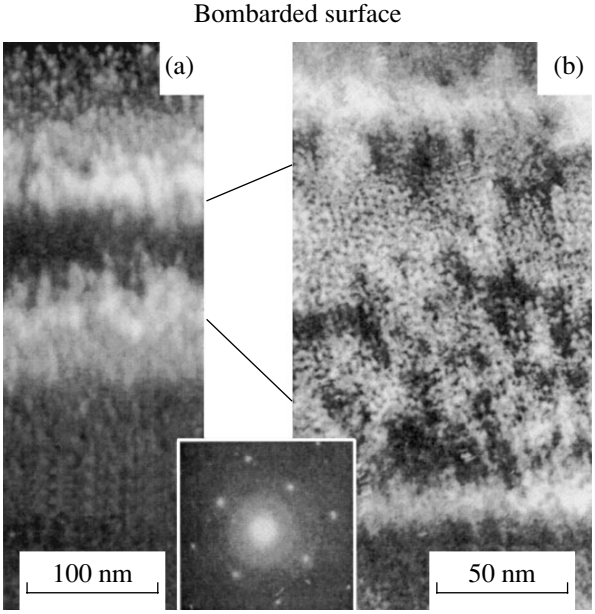

Fig. 2. Amorphous layer in the thin Si film irradiated by He^+ ions (E = 17 keV) with a dose of 10^{17} cm^{-2}. (a) Dark-field TEM image and (b) nanopores in the a-Si layer (bright field).

RESULTS AND DISCUSSION

According to the TEM results, the structure modifications along the He^+ ion tracks in the thin crystalline silicon specimen that is end-face irradiated by a dose of less than 5×10^{16} cm^{-2} are mainly clusters of radiation-induced defects, including dislocation loops. The distribution of the clusters along the ion track was found to be nonuniform, peaking at a depth of about 170 nm.

As the irradiation dose increases to 10^{17} cm^{-2}, the damage in the crystal grows, causing the amorphization of the region with the highest concentration of the clusters (Fig. 2a). An amorphous silicon layer produced at this stage contains pores of diameter 2 nm and concentration 2×10^{17} cm^{-3} (Fig. 2b). It is worth noting that the pores are linearly ordered along the direction of ion motion rather than being randomly distributed in the a-Si layer.

With a further increase in the radiation dose to 3×10^{17} cm^{-2}, the amorphous layer expands and embraces the surface bombarded. For all the doses used in the experiment, new pores nucleate in the amorphized Si layer as it widens, remaining linearly ordered along the ion beam. Both the diameter and concentration of the pores grow with dose; no pores appear outside the a-Si region.

Figure 3a shows a typical pattern of nanopores in the amorphous silicon that are induced in the thin crystal exposed to a dose of 4.5×10^{17} cm^{-2}. The resulting a-Si layer has a width of about 260 nm, which is comparable

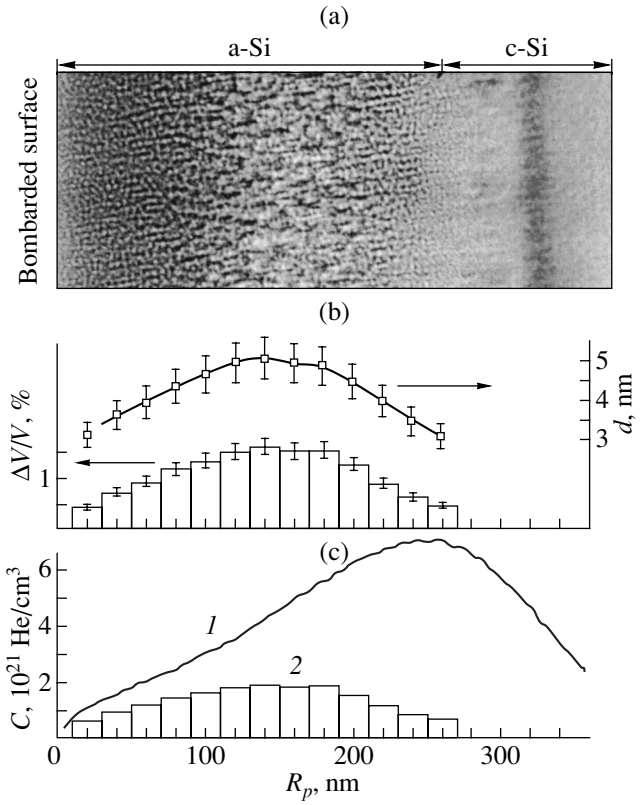

Fig. 3. (a) Ordered helium pores in the amorphous layer after the irradiation of the 150-nm-thick Si film by a dose of 4.5×10^{17} cm^{-2}, (b) the variation of the relative volume $\Delta V/V$ and diameter d of pores along the projective range R_p of He^+ ions in the a-Si layer (on the scale of the micrograph), and (c) the calculated helium concentration C (*1*) in the thin Si film (with $t = 150$ nm) that is end-face irradiated by a dose of 4.5×10^{17} cm^{-2} and (*2*) in nanopores.

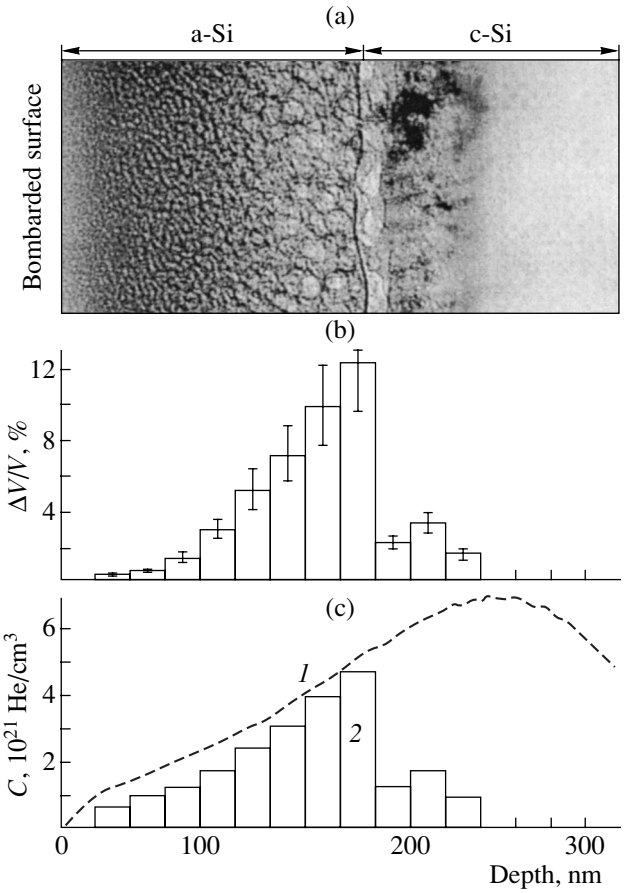

Fig. 4. (a) Pores in the amorphous silicon layer shown in Fig. 3 after annealing at 600°C (the dashed arrow shows the displacement of the front of a-Si epitaxial recrystallization), (b) depth profile of the relative pore volume (on the scale of the micrograph), and (c) comparison of (*1*) the depth profile of the implanted helium with (*2*) the calculated concentration of the helium filling the pores.

to the average projective range of He^+ ions in silicon ($\langle R_p \rangle \approx 230$ nm at $E = 17$ keV). The nanopores are observed throughout the a-Si layer, but their distribution across the layer is nonuniform.

The pores formed at the beginning and end of the ion range (in the a-Si layer) are of roughly equal average size and concentration: 3 nm and 3×10^{17} cm^{-3}, respectively. It is clearly seen that the pores are aligned to produce chains running perpendicularly to the surface bombarded. The nanopores in the chains are distributed fairly uniformly with an average spacing of about 2 nm. The density of the chains in Fig. 3 is close to 2×10^6 cm^{-1}.

In the part of the ion range corresponding to the initial formation of the amorphous layer (at a depth of about 170 nm), nanopores with a maximal size of about 5 nm and a relative volume as high as 1.6% arise (Fig. 3b). Because of the increased pore volume in this region, their alignment is partly disturbed. It should be noted that the region where the size and relative volume of the pores are maximal does not coincide with that where the peak concentration of the implanted helium is expected (Fig. 3c).

In order to elucidate the nature of the pores observed, the specimen irradiated by a dose of 4.5×10^{17} cm^{-2} was thermally annealed at 600°C with a heating rate of 15°C s^{-1}.

After the post-irradiation annealing, the amorphous Si layer still persists but its width decreases to 180 nm because of the partial epitaxial recrystallization of the layer (Fig. 4a). In the residual amorphous layer, the pores grow, with their size varying with depth. The average pore diameter ranges from 3 nm at the surface to 15 nm near the amorphous layer, whereas their concentration changes only slightly within this region. Consequently, the drastic increase in the pore volume (from 0.5 to 12%) with depth is observed (Fig. 4b). In the crystalline region, however, no pores appear within the He^+ ion range after the annealing.

It is well known that the volume of vacancy pores decreases or at least remains unchanged after anneal-

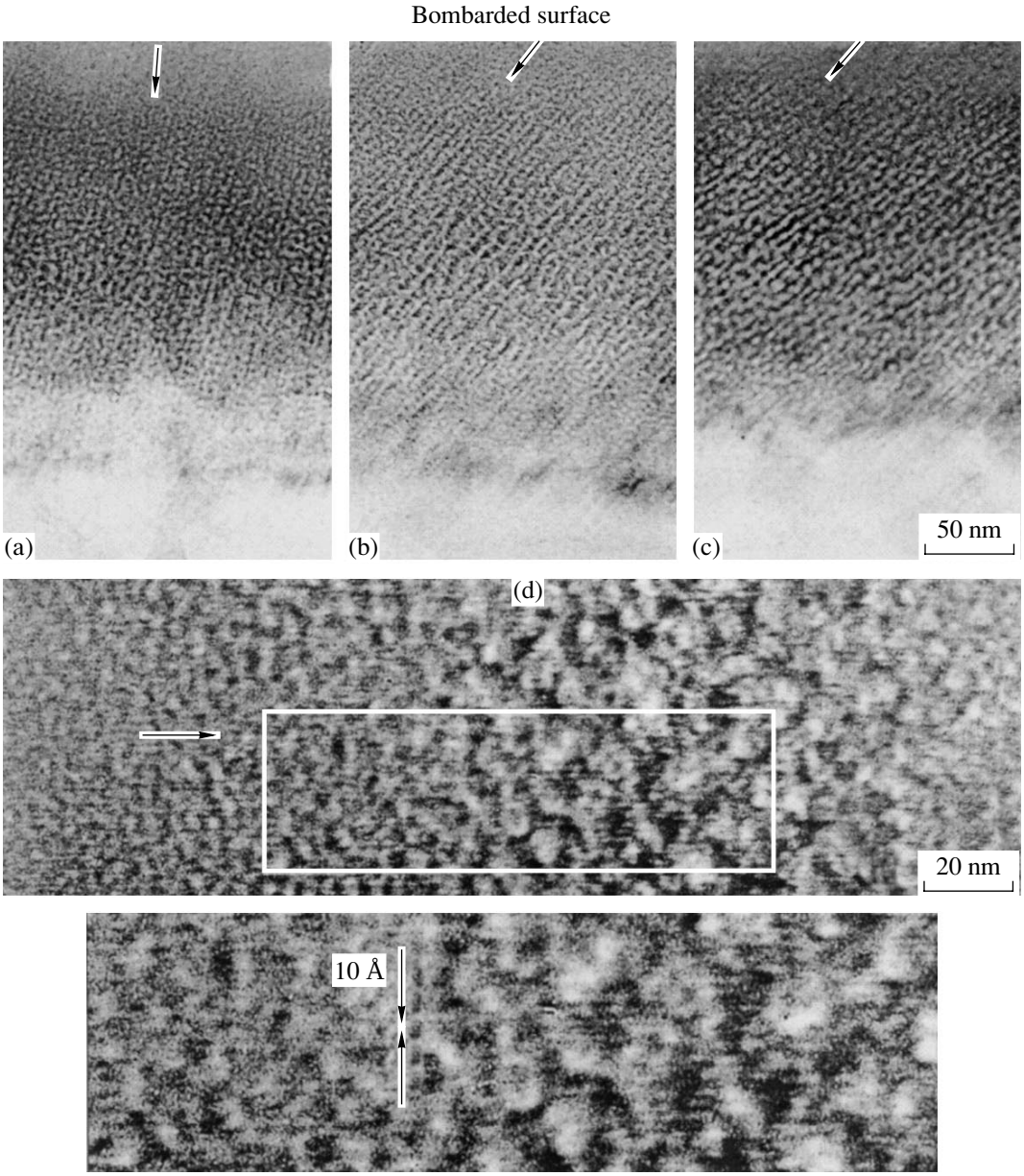

Fig. 5. Alignment of helium bubbles in the a-Si layer at different angles of incidence of the ion beam: (a) 65°, (b) 45°, and (c) 30° (the arrows indicate the direction of the incident ion beam). (d) Tracks of 17-keV He^+ ions in the helium-doped amorphous Si.

ing. In contrast, gas-filled pores (bubbles) grow and so does their volume. Thus, the observed increase in the total volume of the nanopores after the annealing indicates that they are filled with the gas. The fact that the pore concentration does not change implies that the pores do not migrate during the annealing; hence, the immobile pores grow by absorbing thermally generated vacancies, helium atoms, or both.

However, the nonuniform pore size distribution across the depth and, specifically, the constant pore size and concentration near the surface bombarded excludes the participation of thermally induced vacancies in this process. This means that the pores grow only by capturing He atoms from the amorphous matrix, clearly on the condition that the helium is nonuniformly distributed across the depth.

In order to evaluate the amount of helium atoms filling the pores after irradiation (Fig. 3c), we used the empirical isotherm for superdense helium at room temperature [16] and also the condition for the mechanical stability of a helium pore under plastic deformation of the matrix: $P = (2\gamma + Gb)/r$ [15], where P is the pressure inside the pore, $\gamma = 1400$ dyn/cm is the surface free energy, $G = 40.5$ GPa is the shear modulus, $b = 0.35$ nm is the length of the Burgers vector of edge dislocations in Si, and r is the pore radius. The concentration of helium in nanopores thus estimated is essentially lower

than the calculated concentration of helium in the specimen after the implantation.

A similar estimation of the amount of helium in pores after the annealing (Fig. 4c) using the equilibrium condition $P = 2\gamma/r$ for pores of diameter larger than 10 nm showed that, though the amount of helium in the pores increased in the heavily doped region of the amorphous layer, it did not reach the predicted concentration of the implanted helium. It seems likely that, at the initial stage of the irradiation, when the silicon is still crystalline, helium atoms may be desorbed from the thin specimen.

The above considerations suggest that, during the irradiation, the helium in the amorphous Si may be present both inside nanobubbles and as a solute.

In our opinion, the lack of pores in the thin crystalline silicon may be attributed to the desorption of the helium during the irradiation, for example, through the mechanism of radiation-induced migration [12]. Since this mechanism in amorphous silicon fails, the helium implanted into the a-Si layer is totally accumulated.

In order to check that the chains of helium bubbles are aligned with the ion beam direction, we varied the angle between the beam and the surface to be irradiated ($\alpha = 65°$, 45°, and 30°). The TEM images (Fig. 5a–5c) confirm that the ordering of the bubbles does coincide with the ion beam direction and, consequently, with the preferential direction of ion motion in the specimen. This supports the formation of helium bubbles along the ion tracks.

Along with the chains of nanopores, the thorough TEM study also revealed extended linear defects of an other type. These defects with a diameter of about 1 nm and a linear density close to 10^7 cm^{-1} were also oriented in the direction of the He^+ ion beam (Fig. 5d). The additional ion thinning of the specimen showed that they originate in the bulk. The bright (absorption) contrast from these defects suggests that the atomic density in them is lower than in the surrounding amorphous matrix. Taken together, these facts led us to conclude that these linear defects formed along the He^+ ion tracks and are cylindrical channels that are presumably filled with helium atoms.

Mechanisms underlying both the formation of the tracklike features with an "ultrasmall" energy release in the track of a slowing-down He^+ ion (less than 60 eV/nm) and the alignment of nanobubbles as chains still remain to be clarified. However, one may assume that the ionization-induced breaking of covalent bonds along the track of a charged particle may cause the decomposition of the metastable solid solution of helium in a-Si with a chain of gas bubbles arising along the track.

CONCLUSIONS

Three important and unexpected findings are worth noting.

(1) Helium nanopores in thin Si crystals form solely in the amorphous layer. This is because the mobility of helium atoms in amorphous silicon is considerably lower than in the crystalline material under irradiation.

(2) Helium nanopores are aligned with the direction of ion motion throughout the radiation-induced amorphous layer.

(3) Linear structure features with an average diameter of 1 nm and a concentration close to 10^7 cm^{-1} are observed along the He^+ ion direction in helium-saturated amorphous silicon. They are identified as the tracks of low-energy helium ions.

REFERENCES

1. V. Raineri, P. G. Fallica, G. Percolla, *et al.*, J. Appl. Phys. **78**, 3727 (1995).
2. Sh. Sh. Ibragimov and V. F. Reutov, USSR Inventor's Certificate No. 1282757 (1983).
3. A. Agarwal, T. E. Haynes, V. C. Venezia, *et al.*, Appl. Phys. Lett. **72**, 1086 (1998).
4. L. T. Canham, Appl. Phys. Lett. **57**, 1046 (1990).
5. T. Takagahara and K. Takeda, Phys. Rev. B **46**, 15578 (1992).
6. R. Siegele, G. C. Weatherly, H. K. Haugen, *et al.*, Appl. Phys. Lett. **66**, 1319 (1995).
7. V. F. Reutov and A. S. Sokhatskiĭ, Pis'ma Zh. Tekh. Fiz. **28** (14), 83 (2002) [Tech. Phys. Lett. **28**, 615 (2002)].
8. V. Raineri, S. Coffa, M. Saggio, *et al.*, Nucl. Instrum. Methods Phys. Res. B **47**, 292 (1999).
9. M. van Wieringen and N. Warmoltz, Physica (Utrecht) **22**, 849 (1956).
10. C. C. Griffioen, J. H. Evans, P. C. de Jong, and A. van Veen, Nucl. Instrum. Methods Phys. Res. B **28**, 360 (1987).
11. M. Alatalo, M. J. Puska, and R. M. Nieminen, Phys. Rev. B **46**, 12806 (1992).
12. D. Estreicher, J. Weber, A. Derecskei-Kovacs, and D. S. Marynick, Phys. Rev. B **55**, 5037 (1997).
13. V. F. Reutov and A. S. Sokhatskiĭ, Materialovedenie, No. 10, 6 (1998).
14. V. F. Reutov, A. S. Sokhatsky, V. B. Kutner, and A. N. Lebedev, Nucl. Instrum. Methods Phys. Res. B **49**, 319 (1999).
15. W. Jäger, R. Manzke, H. Trinkaus, *et al.*, Radiat. Eff. **78**, 315 (1983).
16. R. Le. Toullec, P. Loubeyre, and J.-P. Pinceaux, Phys. Rev. B **40**, 2368 (1989).

Translated by A. Sidorova

Technical Physics, Vol. 48, No. 1, 2003, pp. 7–14. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 10–18.

THEORETICAL AND MATHEMATICAL PHYSICS

Shock Wave Stability under the Spinodal Decomposition of Binary Mixtures

I. B. Krasnyuk, L. I. Stefanovich, and V. M. Yurchenko

Galkin Physicotechnical Institute, National Academy of Sciences of Ukraine, Donetsk, 83114 Ukraine

e-mail: kras@host.dipt.donetsk.ua

Received September 20, 2001; in final form, April 9, 2002

Abstract—A diffusion equation for a binary mixture and spinodal decomposition in the case of phase separation are considered. It is shown that, if the binding force between polymer chain links is weak, the diffusion equation for a binary mixture allows for the reduction to the Burgers equation with "viscosity"; that is, the coexistence of rarefaction waves and shock density waves is a possibility. The effect of strong bonds between polymer chain links on the spinodal decomposition dynamics is studied. It is demonstrated that strong bonding may cause a multiflux wave system with alternate stability to arise when the viscosity varies. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Early mathematical models in the physics of polymers were constructed for perfect multidimensional chains (see, e.g., [1]). These chains were analyzed in the self-consistent field approximation, i.e., without considering the interaction constant, which is due to monomer link bonding and/or adhesion between macromolecules (although this constant was considered phenomenologically). The aforesaid refers to the Flory–Huggins model [1]. If strong correlation in polymer systems results in small fluctuations of key parameters, one can use the medium field theory, i.e., the Flory–Huggins model.

In a number of papers (see, e.g., [2, 3]), the existence of the intermediate spatial scale $a \ll D \ll L$ (where a is the spacing between adjacent monomer links, $D = \sqrt{N}\,a$ is the diameter of a perfect polymer chain, L is the characteristic scale of density variation, and N is the degree of polymerization) has been established. It has been shown [2] that such a scale hierarchy makes it possible to perform, first, averaging on the microscale $a < R_1 < D$ and then pass on to equations averaged on the microscale $D < R_2 < L$ (this is valid for continuous description, since models on a discrete grid [4] are also available). With such a procedure, we may arrive at the diffusion equation for a binary mixture [5].

Below we will show that the diffusion equation allows for solutions of two types: rarefaction waves and shock (compression) waves. It turns out that this equation can be reduced to the Burgers equation with "viscosity," where the parameter $\varepsilon \propto \chi - \chi_c$ depending on the interaction constant in the vicinity of the spinodal decomposition point $\chi_c(T_c)$ is responsible for viscosity. Such a consideration is valid if the binding force between polymer chain links is weak (see below). For the Burgers equation, conditions are set such that viscosity can be neglected in the region of shock wave "steeping," i.e., in the region where the solution becomes many-valued and turns into a rarefaction wave (Figs. 1, 2). We will demonstrate that if the so-called entropy conditions (convexities of the spinodal decomposition curve) or Hugoniot conditions, which are well known in fluid dynamics, are met in this region, where fluctuations play a dominant role, fluctuations at the rarefaction front can be neglected and one can apply the medium field approximation. If the entropy conditions are not satisfied, the applicability of the Flory–Huggins approximation is open to question.

In what follows, we consider the classical problem for the Burgers equation. In such a statement, at $t = 0$ the mixture is separated into two phases, φ_1 and φ_2, at the critical temperature T_c, and the velocity of switching from the equilibrium state φ_1 to the equilibrium state φ_2 (the switching wave velocity) in the vicinity of the spinodal is found from formal results (see, e.g. [6–8]). It should be noted that the reduction of the starting

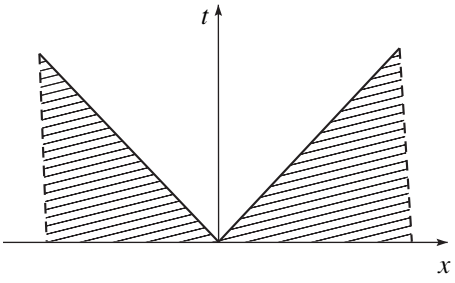

Fig. 1. Undisturbed density values (shaded region) corresponding to a rarefaction wave and a density wave propagating with a finite velocity that is generated by initial phase separation due to spinodal decomposition.

1063-7842/03/4801-0007$24.00 © 2003 MAIK "Nauka/Interperiodica"

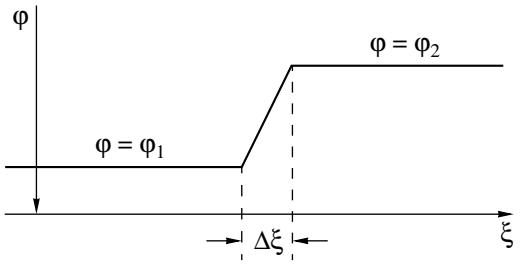

Fig. 2. Limit solution with viscosity where the interval $\Delta\xi$ is given by formula (8).

equation to the Burgers equation is justified statistically. From [6, 7] it follows that if the structure evolution of the density of a binary polymer mixture is viewed as a random process of the "Poisson clock" type [6] on a grid with a step a such that transitions $i \longrightarrow i+1$ are allowed with a probability p and transitions $i \longrightarrow i-1$ are allowed with a probability $1-p$, the mathematical expectation $\langle n_i \rangle = c_i = c(r_i, t)$ [4] satisfies the diffusion equation if $p = 1/2$ (which is proved in [6, 7]) and satisfies the Burgers equation if $p \neq 1/2$. This is the thermodynamic substantiation of the statistical problem of averaging. In this sense, the well-known Hugoniot phenomenological condition (see below) is exact.

Finally, if the binding force $K(\varphi)$ between polymer chain links is proportional to $F(\varphi)$, $K(\varphi) \propto \partial F(\varphi)/\partial\varphi$, where $F(\varphi)$ is the system free energy, one can consider, instead of the Burgers equation without viscosity, the Burgers equation where the quantity $\propto\Lambda(\varphi)\partial^2 F/\partial\varphi^2$ plays the role of viscosity (here, $\Lambda(\varphi)$ is the diffusion coefficient). This equation has solutions of shock wave or rarefaction wave type (Figs. 1, 2).

The stability of solutions to the Burgers equations was studied in [9–14]. With strong bonding between chain links taken into account, the equation for a binary mixture cannot be reduced to the Burgers equation in the general case. However, using the numerical results obtained in [8], we show that viscosity in a polymer chain competes with the binding force between its links. This may result in intriguing effects, such as the existence of multiflux shock waves and rarefaction waves. Nevertheless, there exists a situation when the binding force is such that shock waves also occur (at a certain viscosity) from the general equation for a binary mixture provided that special conditions like $g(\varepsilon, K) = \text{const}$ are met, where the parameter K defines the binding force in the vicinity of the metastable state $\varphi_-(T_c)$, which is the point of spinodal decomposition.

REDUCTION TO AN EQUATION WITH LOW VISCOSITY

Consider the set of equations for a binary mixture [5]

$$\frac{\partial\varphi}{\partial t} + \nabla I(\varphi) = 0, \quad I = -\frac{\Lambda(\varphi)}{k_B T}\nabla\frac{\delta F(\varphi)}{\delta\varphi}.$$

This set can be conveniently written as

$$k_B T\frac{\partial\varphi}{\partial t} = \frac{\partial\Lambda(\varphi)}{\partial\varphi}\nabla\frac{\delta F(\varphi)}{\delta\varphi} + \Lambda(\varphi)\nabla^2\frac{\delta F(\varphi)}{\delta\varphi}.$$

As was shown in [2], the following relationship takes place:

$$\nabla\left(\frac{\delta F}{\delta\varphi}\right) = \nabla[\mu(\varphi) - 2a^2K(\varphi)\nabla^2\varphi - a^2\dot{K}_\varphi(\varphi)(\nabla\varphi)^2], \quad (1)$$

where $\mu(\varphi)$ is the chemical potential and $K(\varphi)$ is proportional to the adhesive force between chain links [5].

If $(a/L)^2 \ll 1$, we can put $\nabla(\delta F/\delta\varphi) = \nabla\mu(\varphi)$ and rearrange the diffusion equation to the form

$$k_B T\frac{\partial\varphi}{\partial t} = \dot{\Lambda}_\varphi(\varphi)\dot{\mu}_\varphi(\varphi)\frac{\partial\varphi}{\partial x} + \Lambda(\varphi)\dot{\mu}_\varphi(\varphi)\nabla^2\varphi,$$

where $\mu(\varphi) \simeq \partial F(\varphi)/\partial\varphi$; that is, it is sufficient to consider the equation

$$k_B T\frac{\partial\varphi}{\partial t} = \dot{\Lambda}_\varphi(\varphi)\dot{\mu}_\varphi(\varphi)\frac{\partial\varphi}{\partial x} + \Lambda(\varphi)\frac{\partial^2 F(\varphi)}{\partial\varphi^2}\frac{\partial^2\varphi}{\partial x^2}. \quad (2)$$

First let us assume that the sample is homogeneous and has one unstable phase of density φ^- [1, Fig. 4.2]. We will concentrate on the spinodal curve in the plane (χ, φ). This part of the phase diagram covers those states of the system where the local concentration φ near the point $\varphi^-(T_c)$ undergoes severe fluctuations according to the formula $\partial^2 F(\varphi)/\partial\varphi^2 = TS^{-1}(q = 0, \varphi)$ [1, p. 119]. Here, $S^{-1}(q, \varphi)$ is the scattered power with a given wavevector q or, more exactly, the Fourier transform of the two-point density–density correlation function. It is known that $S^{-1}(q, t) \propto a\varphi(1-\varphi)[\chi_c(\varphi) - \chi]^{-1/2} + o(q)$, where $\chi_c(\varphi)$ corresponds to the spinodal and $o(q) \longrightarrow 0$ for $q \longrightarrow 0$. Let $\varepsilon(0, \chi, \chi_c) = TS^{-1}(0)$. Then, Eq. (2) can be written in the dimensionless variables as

$$\frac{\partial\varphi}{\partial t} + G(\varphi)\frac{\partial\varphi}{\partial x} = \varepsilon\Lambda(\varphi)\frac{\partial^2\varphi}{\partial x^2}, \quad (2')$$

where

$$G(\varphi, T) = -\frac{\dot{\Lambda}_\varphi(\varphi)}{k_B T}\dot{\mu}_\varphi(\varphi).$$

Recall that the free energy is given by

$$\frac{F(\varphi)}{k_B T} = F_0^*(\varphi) + a^2K(\varphi)(\nabla\varphi)^2, \quad \left[F_0^* = \frac{F_0}{k_B T}\right].$$

Here, k_B is the Boltzmann constant, T is temperature (hereafter, $T = T_c$, where T_c is the critical temperature at which the spinodal decomposition takes place under cooling), and $K(\varphi) = (36\varphi)^{-1} + \chi(\varphi, T_c)$. This formula, derived by de Gennes, is discussed in [15] (from the

standpoint of the validity of using the hydrodynamic approximation in [5]).

In fact, in a dilute solution, the diffusion density flux can be described by the expression $j_{D_i} = -L_{ii}(T)\nabla\bar{\mu}_i$, where the temperature is assumed to be constant. Next, it should be noted that $\bar{\mu}_i = \mu_i^0(T) + k_B T \ln\bar{\rho}_i$ and $(-\nabla\bar{j}_{D_i}) \equiv f_i$, where j_{D_i} is a random diffusion flux with regard for the number of molecules in sites i [16, p. 315]. Averaging over all lattice sites and assuming that the binding force between links in a polymer chain (the first term in the expression for $K(\varphi)$) equals $\langle f_i \rangle$, one comes to the de Gennes formula. In the general case, $j \propto \nabla^2\mu \propto \partial^3 F(\varphi)/\partial\varphi^3$; hence, the binding force includes irregular fluctuations appearing in the three-particle interaction [1, p. 77]. Geometrically, this means taking into account the curvature of the spinodal curve when the positive-curvature function $F(\varphi)$ (entropy effects dominate, making mixing favorable) changes to the negative-curvature function in passing through the critical value χ_c [1, p. 113]. This observation will be used by us later; for the moment, we assume that the inequalities

$$\left(\frac{a}{L}\right)^2 |K(\varphi^-)| \ll 1 \quad \text{and} \quad \left(\frac{a}{L}\right)^2 |\dot{K}_\varphi(\varphi^-)| \ll 1$$

are fulfilled in the vicinity of the metastable point φ^- of phase separation. Then, from relationship (1), it follows that within any (finite) time interval, the decomposition dynamics along the spinodal is defined by solutions to Eq. (2'), where $\varepsilon(\chi, \chi_c)$ is a small parameter that may change sign at $\chi = \chi_c$ and characterizes density fluctuations in the vicinity of the spinodal decomposition curve (see below).

Recall that under our conditions, the correlation function can be calculated by using the random phase approximation in terms of the Debye function $g_D(N, q)$, which describes scattering by a perfect chain of N monomers. We can write

$$S^{-1}(q) = \frac{1}{\varphi g_D(N_A, q)} + \frac{1}{(1-\varphi) g_D(N_B, q)} - 2\chi,$$

and since $q_D(N, q = 0) = N$ at $q = 0$, this relationship turns into the equality

$$0 = \frac{\partial^2}{\partial\varphi^2}\left(\frac{F}{T}\right) = \frac{1}{N_A\varphi} + \frac{1}{N_B(1-\varphi)} - 2\chi.$$

Hence, at $N_A, N_B \gg 1$, we can assume that $\chi \ll 1$. Phase separation becomes possible if $\chi > \chi_c$. Usually, it is assumed that $\chi_c = 2/N$ [1, p. 114]; hence, we can put $\chi_c = 0$ on the spinodal.

Consider the separation of the system into two phases, φ_1 and φ_2, with the statistical weights p_1 and p_2; that is, $\varphi = p_1\varphi_1 + p_2\varphi_2$, $p_1 + p_2 = 1$. Let solutions to Eq. (2') satisfy the initial conditions

$$\begin{aligned} &\varphi(-\infty, 0, \chi < \chi_c) = \varphi_1 \quad \text{and} \\ &\varphi(\infty, t, \chi > \chi_c) = \varphi_2, \quad t > 0 \end{aligned} \tag{3}$$

and boundary conditions

$$\varphi(x, 0, \chi) = \begin{cases} \varphi_1, \chi < \chi_c & \text{for } x < 0 \\ \varphi_2, \chi > \chi_c & \text{for } x > 0. \end{cases} \tag{4}$$

Conditions (4) correspond to a solution where the phases are immiscible. In this state, the solution has the density φ_1 with a probability $p_1 = 1$ and the density φ_2 with a probability $p_2 = 0$. Note that $\chi_c = \chi(T_c)$, where T_c is the critical cooling temperature of the solution.

STABILITY OF SOLUTIONS TO THE UNPERTURBED PROBLEM

Cauchy problem (2')–(4) is known as the problem of arbitrary discontinuity decay [11]. A solution to this problem is self-similar, $\varphi = \varphi(x/t)$, and leaves unperturbed values of φ in the shaded regions in Fig. 1 (φ_1 on the left and φ_2 on the right). It is conventional to assume that a perturbation from a discontinuity propagates with a finite velocity, which we will find below.

Let the value $\chi = \chi_c$ correspond to the homogeneous metastable state $\varphi_c = \varphi^-$; then, phase separation becomes possible if $\chi > \chi_c$ [1, Fig. 4.4]. A perturbation is introduced into Eq. (2') as linear viscosity; therefore, at $\varepsilon(\chi_c) = 0$, we can consider the divergent equation

$$\frac{\partial\varphi}{\partial t} + \frac{\partial f(\varphi)}{\partial x} = 0, \tag{5}$$

where conditions along the discontinuity lines have the form

$$-\nu(\varphi_1 - \varphi_2) + f(\varphi_1) - f(\varphi_2) = 0. \tag{6}$$

Here, φ_1 and φ_2 are the values of φ at the discontinuity, $\nu = dx(t)/dt$ is the discontinuity line slope, and $G(\varphi) = f'(\varphi)$.

Note that additional relationships between the desired functions on the lines of discontinuity are the laws of conservation of energy, momentum, and mass, as well as the law of entropy increase. Note that a function $\varphi(x, t)$ satisfies Eq. (6) if the equality derived in [12] is valid for any smooth contour Γ lying in the half-plane $t > 0$ and crossing the discontinuity lines of the function $\varphi(x, t)$ in a finite number of points.

If a solution $\varphi(x, t)$ to Eq. (5) is smooth, the equality from [12] is always fulfilled. Also, from the relationship

$$\iint_D \left(\frac{\partial\varphi}{\partial t} + \frac{\partial f(\varphi)}{\partial x}\right) dt\,dx = A(f),$$

where Γ is the boundary of the domain D, it follows that a function $\varphi(x, t)$ for which this relationship is satisfied is a solution to Eq. (5) in those areas of the domain D where this function is continuously differentiable [12, p. 6].

As was shown in [13], a discontinuous solution to Eq. (5) that satisfies conditions (6) is not necessarily found as the limit of continuous solutions (at $\varepsilon \longrightarrow 0$) of perturbed equation (3).

Indeed, let us restrict the analysis to solutions of form $\varphi = \varphi(\xi)$, where $\xi = x - \nu t$. Then, Eq. (2') can be written as

$$-\nu\frac{d\varphi}{d\xi} + \frac{df(\varphi)}{d\xi} = \varepsilon\Lambda(\varphi)\frac{d^2\varphi}{d\xi^2}. \qquad (5')$$

In the medium field approximation, we have $\Lambda(\varphi) \cong D_{\text{ef}}(1 - \varphi)$ provided that the diffusion coefficients of polymers and holes equal each other [5] (D_{ef} is the coefficient of self-diffusion of polymer chains). Generally speaking, the nonlinearity of $\Lambda(\varphi)$ causes the localization of the solutions; however, since the coefficient $\Lambda(\varphi)$ is linear and conditions under which the effect of fluctuations $S^{-1}(q)$ is significant only at $q \longrightarrow 0$ are looked for, we can put $\Lambda(\varphi) \equiv \Lambda(\varphi^-)$, where φ^- is the homogeneous state, which is separated into the states φ_1 and φ_2 due to perturbations $\varepsilon(\chi, T)$. Since we are interested only in the local process of spinodal decomposition within a finite time interval, such an approach is equivalent to using the method of coefficient freezing (see, e.g., [14]).

Eventually, we arrive at the equation

$$-\nu\frac{d\varphi}{d\xi} + \frac{df(\varphi)}{d\xi} = \varepsilon\frac{d^2\varphi}{d\xi^2}, \qquad (7)$$

where $\varepsilon \longrightarrow \varepsilon B(\varphi^-)$.

Assume that ν, φ_1, and φ_2 satisfy relationships (6). Let us find conditions under which Eq. (7) has a continuous solution satisfying the boundary conditions $\varphi(\xi) \longrightarrow \varphi_1$ at $\xi \longrightarrow +\infty$ and $\varphi(\xi) \longrightarrow \varphi_2$ at $\xi \longrightarrow -\infty$. These relationships are caused by the initial conditions, which can be matched to the boundary conditions "at infinity."

It was shown [11] that, if such a solution exists, it passes (at $\varepsilon \longrightarrow 0$) to a discontinuous solution to unperturbed equation (5'), which obviously has the form $\varphi(\xi) \equiv \varphi_1$ at $\xi > 0$ and $\varphi(\xi) \equiv \varphi_2$ at $\xi < 0$. Consequently, existence conditions for a continuous solution to the unperturbed equation are admissibility conditions for a discontinuous solution. For a single equation, these conditions can be stated as follows: (1) for a discontinuity φ_1–φ_2 to be admissible, it is necessary that φ_1 and φ_2 be neighboring zeros of some function $\Phi(\varphi)$ and (2) a discontinuity φ_1–φ_2 is admissible if the inequality $f'(\varphi_1) < \nu < f'(\varphi_2)$ is met. These requirements must be fulfilled simultaneously. The function $\Phi(\varphi) = -\nu\varphi + f(\varphi) + C$ is the integral of the equation $\varepsilon d\varphi/d\xi = -\nu\varphi + f(\varphi) + C$, which is derived from Eq. (7) by integration over ξ with $C(\varphi_1) = C(\varphi_2) = 0$, so that $\Phi(\varphi_1) = \Phi(\varphi_2) = 0$ owing to the conditions at the discontinuity. Usually, a continuous solution is called a rarefaction wave and a discontinuity is called a shock wave.

In [11], it was also proved that solutions to Eq. (7) satisfying condition (6) at a discontinuity are stable. To establish the stability conditions, we consider the discontinuous solution $\varphi(\xi) \equiv \varphi_1$ at $x - \nu t > 0$ and $\varphi(\xi) \equiv \varphi_2$ at $x - \nu t < 0$ and, adding a small perturbation $\delta\varphi$ to it, substitute $\varphi + \delta\varphi$ into Eq. (5). Rejecting terms on the order of $(\delta\varphi)^2$, we arrive at the equation

$$\frac{\partial(\delta\varphi)}{\partial t} + f'\frac{\partial(\delta\varphi)}{\partial x} = 0$$

with the piecewise constant coefficient $f' = f'(\varphi_1)$ at $x - \nu t > 0$ and $f' = f'(\varphi_2)$ at $x - \nu t < 0$. The solution will be stable if the perturbation δu tends to zero along the discontinuity line $x - \nu t = 0$ at $t \longrightarrow \infty$ and the inequality $f'(\varphi_1) < \nu < f'(\varphi_2)$ is fulfilled.

The "smearing" of a discontinuity arising with the introduction of viscosity was considered in [11] using the equation

$$\frac{\partial\varphi}{\partial t} + \frac{\partial f(\varphi)}{\partial x} = \varepsilon\frac{db}{dx}$$

as an example, where $b = (\partial\varphi/\partial x)^\alpha$ with $\alpha > 0$ at $\partial\varphi/\partial x > 0$ and $b = 0$ at $\partial\varphi/\partial x < 0$.

The range of φ on the axis ξ turns out to be finite and equal to

$$\Delta\xi = \varepsilon^{\frac{1}{a}}\int_{\varphi_1}^{\varphi_2}\Phi^{-\frac{1}{\alpha}}(u)du. \qquad (8)$$

At $\alpha > 1$, integral (8) converges (the solution is shown in Fig. 2). For $0 < \alpha \le 1$, the integral diverges; hence, the smearing domain is infinite: the solution $\varphi(\xi)$ tends to (φ_1, φ_2) only at $\xi \longrightarrow 0$. However, as was shown in [11], the limiting solution consists of two constants φ_1 and φ_2 in both cases. From (8), it follows that the lifetime of a shock-wave-type solution can be made as long as desired (but finite) by decreasing ε.

We pursue the stability analysis. It is known [8] that the scalar law of conservation $\varphi_t + (f(\varphi))_x = 0$ can be written in the form $G(\varphi_t) + F(\varphi)_x = 0$, where G and F are related as $F'(\varphi) = G'(\varphi)f'(\varphi)$. Let us apply this observation to the equation $\varphi_t + (f(\varphi))_x = \varepsilon\varphi_{xx}$ to tackle the question of whether there exists a traveling wave that couples the states φ_- and φ_+ such that $\varphi_+ < \varphi_-$. Let $\varphi_+ > 0$ for definiteness. Then, the solution $\varphi(x - st)$ satisfies the equation

$$-s\varphi_x + (f(\varphi))_x = \varepsilon\varphi_{xx}. \qquad (9)$$

Consider the function $G''(\varphi) > 0$. We multiply Eq. (9) by $G'(\varphi)$ and integrate the product from $-\infty$ to

$+\infty$. Integration by parts yields

$$-s[G(\varphi) + F(\varphi)] = -\varepsilon\int G''(\varphi)\varphi_x^2 dx, \qquad (10)$$

where $[A] = A(\infty) - A(-\infty)$.

From Eq. (10), we have the inequality

$$-s[G(\varphi) + F(\varphi)] \le 0,$$

which defines the limit of the propagation velocity s.

In the situation under study, it is sufficient to put $G'(\varphi) = 1$, which gives the wave velocity $s = (f(\varphi_-) - f(\varphi_+))/(\varphi_- - \varphi_+)$. According to [8], if the function $f(\varphi)$ has no more than one inflection point, to each centered shock wave $\varphi(x, t) = \varphi_-$ at $x < st$ and $\varphi(x, t) = \varphi_+$ at $x > st$, there corresponds the solution $\varphi(x, t) = \varphi_{\text{Lax}}((x - st), \varepsilon)$ to equation with viscosity (2'). This solution satisfies the condition $\varphi_{\text{Lax}}(\pm\infty) = \varphi_\pm$ and $\varphi'_{\text{Lax}}(\pm\infty) = 0$. The velocity of such a wave meets the Lax entropy conditions [8]. All waves of this kind are admissible in that they are solutions to regularized equation (5) at $\varepsilon \longrightarrow 0$.

For the Cauchy problem, the stability of solutions to the regularized equation, i.e., the equation without viscosity, against small variations of the initial function was studied in [13]. For any ξ_1 and ξ_2 from the closed interval $[a, b]$, the inequality

$$\left|\int_\xi^{\xi_1} [\varphi_1(0, x) - \varphi_2(0, x)]dx\right| \le \varepsilon$$

yields the inequality

$$\left|\int_{x_1}^{x_2} [\varphi_1(t_1, x) - \varphi_2(t_1, x)]dx\right| \le \varepsilon,$$

which is valid for any closed interval $[x_1, x_2]$ of the straight line $t = t_1$. This closed interval belongs to the domain bounded by the straight lines $t = 0$, $x - a - At = 0$, and $x - b + At = 0$, where $A = \max(|f(\varphi_1) - f(\varphi_2)|/|\varphi_1 - \varphi_2|)$.

The example discussed in [14, p. 158] demonstrates that for the Burgers equation, a solution can be taken in the form $\varphi = -c + a\tanh(a(x + ct)/2\varepsilon)$, where a and c are constants. Clearly, $\varphi(x, t) \longrightarrow -c \pm a$ at $\xi \longrightarrow \pm\infty$, where $\xi = a(x + ct)/2\varepsilon$. Taking $a = (\varphi_1 - \varphi_2)/2$ and $c = -(\varphi_1 + \varphi_2)/2$, we come to the convergence $\varphi(x, t) \longrightarrow (\varphi_1, \varphi_2)$ for $\xi \longrightarrow \pm\infty$. The initial function has the form $\varphi(x, 0, \varepsilon) \propto \tanh(ax/2\varepsilon)$; hence, at the point of discontinuity $x_0 = 0$ (but $x \ne 0$), the discontinuity is more stepwise, the smaller $\varepsilon > 0$ is. However, a substantial dependence on the initial conditions is observed: for any given $\varepsilon > 0$, the condition $x \longrightarrow 0$ makes the solution discontinuous at the point ($x = 0$, $t = 0$). Thus, in this problem, the stability due to a decrease in viscosity takes place only near the points of discontinuity (and only within a finite time interval).

THE BEHAVIOR OF SOLUTIONS TO THE GENERAL DIFFUSION EQUATION

Taking into account the possible phenomenological representation of the coefficient $K(\varphi)$, which, by definition, is proportional to the binding force between polymer chain links, we put $\dot{K}(\varphi) \propto \varphi^2$ and $\dot{\Lambda}_\varphi(\varphi)\dot{\mu}_\varphi(\varphi) = \varphi^2 - \alpha\varphi^3$. As $\Lambda(\varphi) = D_{\text{ef}}(1 - \phi)$, we find that $\dot{\mu}_\varphi(\varphi) \propto \varphi^2 - \alpha\varphi^3$. Then, assuming that $\varphi = \varphi^- = 0$ (without loss of generality), one can easily see that $\alpha \propto \chi - \chi_c$ and the plot of $F(\varphi)$ changes the sign of curvature in passing through $\chi = \chi_c$. As a result, the diffusion equation in the general form (including the term with $((\nabla\varphi)^2)$ can be written as

$$\varphi_t + (\varphi^2 - \alpha\varphi^3)_x = \varepsilon(\varphi^3\varphi_x)_x - \beta(\varphi^3\varphi_{xxx})_x, \qquad (11)$$

where α, β, and $\varepsilon > 0$. Here, $\beta \propto \alpha$ and the competition between ε and β generates shock (or rarefaction) waves (as in the Burgers equation) or additionally may cause compression waves to appear (Fig. 3). According to [8], if $(\varphi_\infty - b)$ is small, where $\varphi_\infty = \varphi_1$ and $b = \varphi_2$ with $b < \varphi_\infty$, the traveling wave is a compression wave in the sense that the characteristics of the equation $\varphi_t + (\varphi^2 - \alpha\varphi^3)_x = 0$ "cross at the end of the wave" traveling with a velocity $\lambda(b) < s < \lambda(\varphi_\infty)$ [8, p. 432]. Here, $\lambda(\varphi) = 2\varphi - 3\varphi^2$ is called the characteristic velocity.

Note at once that this phenomenon may cause density peaks to appear near the discontinuity front [8, Fig. 8]. In the case of the general equation for a binary mixture, they arise near the point $\partial^2 f(\varphi)/\partial\varphi^2 = 0$ [8, Fig. 7]. This means that even if density fluctuations near the point $\varphi = \varphi^-$ are small ($\varepsilon \ll 1$), the binding force between polymer chain links enhances density fluctuations. In spite of this, it appears that there exist conditions where fluctuations due to density diffusion are compensated for by fluctuations due to bonds between chain links. This, however, takes place only for special initial distributions of the binary mixture density.

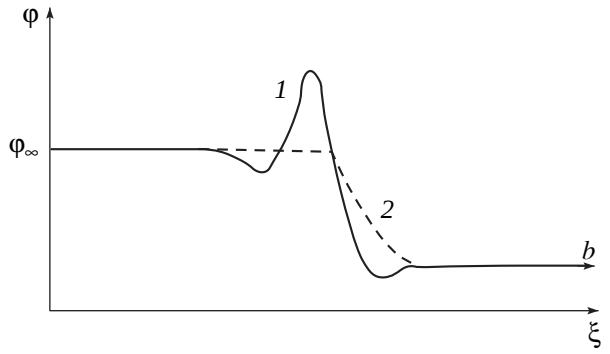

Fig. 3. Solution to the equation for a binary mixture (*1*) at $\beta = 0$ (no bonding between links) and (*2*) at $|\varepsilon| \ll 1$ and $\beta > 0$ (incompressible wave).

If φ_∞ increases, there appear several traveling waves that approximate Lax shock waves. In this range of φ_∞, the equation $\varphi_t + (\varphi^2 - \varphi^3)_x = -(\varphi^3\varphi_{xxx})_x$ has asymptotically stable solutions consisting of two waves traveling with different velocities. The slower wave is a Lax wave connecting (on the phase portrait [8]) φ_∞ and the point $\varphi_{uc} > \varphi_\infty$ (with φ_{uc} being independent of φ_∞) in such a way that the shock (compression) intensity decreases with increasing φ_∞. The faster wave is an incompressible shock wave connecting the points φ_{uc} and *b*. The other waves are incompressible in the sense that the Lax entropy conditions are violated.

With diffusion taken into account in (2), one might expect that the result will be the same at relatively small $\varepsilon > 0$, since lower order perturbations modify the vector field in a "soft" way (the shock wave corresponds to the separatrix running from saddle to saddle).

When two waves appear (Fig. 3), the weaker compressed one should have a lesser effect on fluctuations (for $\varepsilon \gg 1$) in comparison with the one compressed more strongly. As was shown [8, Sect. 3.1], at $\varepsilon \longrightarrow \infty$, incompressible waves (that is, those changing their profile insignificantly) are absent [8, Fig. 3].

It is of interest to study the transition for a finite ε. Preliminary calculations [8] show that for $\varepsilon \ll 1$ and a high density (the inclusion of the cubic term in the functional F_0), a low viscosity has a minor effect. Conversely, if the density is low (that is, the Burgers equation applies) and the viscosity is low, the quantity $(\varphi_\infty - b)$ is small and, hence, a solution to the general equation behaves in the same way as solutions to the equation with diffusion, since the equation without diffusion (but with the fourth derivative taken into account) yields rarefaction waves and shock waves under these conditions. Thus, with low densities, the equation for a binary mixture allows for strict reduction to the Burgers equation. For $\varepsilon \longrightarrow \infty$ and high densities, one can restrict the function analysis to considering the fourth-order equation when studying a binary mixture that is separated into phases near the spinodal curve.

A new result here is the possibility of switching from a rarefaction wave to a shock (incompressible) wave and vice versa in passing through the spinodal. This necessitates studying the Burgers equation with negative viscosity. For $\varepsilon < 0$ and $\varepsilon \longrightarrow 0$, the results for low densities are strict; for high densities, the effect of fluctuations near the spinodal ($\varepsilon < 0$) has not been investigated even numerically. At the same time, for $\varepsilon > 0$ (the region over the spinodal) in the case of the general equation for a binary mixture, the effect of density fluctuations is quite clear even for finite $\alpha > 0$ and $\varepsilon > 0$, as follows from numerical results [8] and the behavior of solutions to the Burgers equation at a low viscosity [10–14].

STABILITY OF SOLUTIONS TO THE INITIAL PROBLEM WITH VISCOSITY

Solutions to the Burgers equation with $\varepsilon = 0$ are distinctive in that the wave front steepens and then reverses at the time instant $t_0 = (\max[d\nu(x, t_0)/dx])^{-1}$, $(d\nu(x, t_0)/dx < 0)$, where $\nu(x)$ is the wave velocity. Actually, the steeping may generate multiflux motions with subsequent wave front reversal [15, p. 188].

Solutions to the Burgers equation with $\varepsilon > 0$ show how viscosity prevents the reversal. In fact, for $t \longrightarrow t_0$, the front steepness grows and so does the derivative $d\nu(x, t)/dx < 0$. Eventually, the terms $\varepsilon\varphi_{xx}$ and $\varphi\varphi_x$ become of the same order (even at low viscosities); in other words, we are dealing with the competition of two processes: steeping due to nonlinearity and decay because of viscosity. Note that stationary motion may also occur without competition if the Hugoniot condition (when characteristics along which the shock wave front propagates do not intersect) is fulfilled in the range of solution uncertainty.

Let us show that the solution branches φ_1 and φ_2 outside the range $\Delta\xi$ (Fig. 2) are Lyapunov stable against perturbations that are as small as desired. The solutions are sought in the form $\tilde{\varphi}(x, t, \varepsilon) = \varphi_1^\varepsilon + h\varphi(x, t)$, where $0 \le h \ll 1$ and ε is a given parameter (not necessarily small). Then, the Burgers equation can be written in the form

$$\frac{\partial\varphi_1^\varepsilon}{\partial t} + h\frac{\partial\varphi}{\partial t} + (\varphi_1^\varepsilon + h\varphi)\left(\frac{\partial\varphi_1^\varepsilon}{\partial x} + h\frac{\partial\varphi}{\partial x}\right) = \varepsilon\frac{\partial^2\varphi_1^\varepsilon}{\partial x^2} + \varepsilon h\frac{\partial^2\varphi}{\partial x^2};$$

hence, the perturbations satisfy the equation

$$\frac{\partial\varphi}{\partial t} + \varphi_k\frac{\partial\varphi}{\partial x} = \varepsilon\frac{\partial^2\varphi}{\partial x^2} \quad (k = 1, 2)$$

up to terms $o(h^2)$, which may also depend on viscosity. We assume, however, that they are viscosity independent. This equation can be conveniently represented as

$$\frac{1}{2}\frac{\partial\varphi^2}{\partial t} + \frac{1}{2}\varphi_k\frac{\partial\varphi^2}{\partial x} = \varepsilon\varphi\frac{\partial^2\varphi}{\partial x^2} \quad (k = 1, 2).$$

Integration at $k = 1$ yields

$$\frac{1}{2}\frac{\partial}{\partial t}\int_{-\infty}^{x_1}\varphi^2 dx + \frac{1}{2}\varphi_1\varphi^2\Big|_{-\infty}^{x_1} = \varepsilon\varphi\frac{\partial\varphi}{\partial x}\Big|_{-\infty}^{x_1} - \varepsilon\int_{-\infty}^{x_1}\left(\frac{\partial\varphi}{\partial x}\right)^2 dx,$$

where x_1 is the left extreme of the interval $\Delta\xi = x_2 - x_1$ of the rarefaction wave (Fig. 2) (for the shock wave, one can put $\Delta\xi = 0$).

Let the boundary conditions

$$\frac{\partial\varphi}{\partial x}\Big|_{x = \pm\infty} = 0 \quad \text{and} \quad \varphi = \begin{cases}\varphi_1 & \text{for} \quad x \longrightarrow -\infty \\ \varphi_2 & \text{for} \quad x \longrightarrow +\infty\end{cases}$$

and the additional requirement $\varphi_1 - \varphi_2 > 0$ be fulfilled

for the shock wave. Then, if the conditions

$$\varphi^2(x_1) > \varphi^2(-\infty) \quad \text{and} \quad \varphi^2(+\infty) > \varphi^2(x_2)$$

are satisfied outside the range $\Delta\xi$, the inequality

$$\frac{1}{2}\frac{\partial}{\partial t}\int_{-\infty}^{x}\varphi^2(s, t)ds < 0$$

and the same inequality within the interval $(x_2, +\infty)$ are obviously valid. Since the functional

$$F(\varphi) = \int_{-\infty}^{x_1}\varphi^2 ds$$

takes only positive values and decreases (as a function of t) for any function $\varphi(x, t)$, the relationship

$$\liminf_{t\to\infty} F(\varphi(x, t)) = 0 \quad \text{for} \quad -\infty < x < x_1$$

must be satisfied; hence,

$$\liminf_{t\to\infty} \varphi^2(x, t) = 0 \quad \text{for} \quad -\infty < x < x_1$$

according to the Du Bois-Reymond lemma (see, e.g., [16–18]). Here, inf is the exact lower limit of a function. Thus, perturbations decay with time.

If we require the Hugoniot condition

$$\nu[\varphi_h^+(\varepsilon) - \varphi_h^-(\varepsilon)] = f(\varphi_h^+(\varepsilon)) - f(\varphi_h^-(\varepsilon))$$

to be satisfied for any perturbations φ_h^+ and φ_h^- at relatively small $h > 0$ and each given ε (here, $h = h(\varepsilon)$) in the vicinity of the stationary points φ_1^ε and φ_2^ε, respectively, the characteristics bounding the domain of definition of the rarefaction wave will shift parallel to the initial ones. For the shock wave, the only discontinuity front will shift parallel to the initial front. In this sense, we can argue that the rarefaction wave is stable, while outside the rarefaction region, it is asymptotically stable. If conditions of type $\varphi_1 - \varphi_2 > 0$ fail, perturbations depend on viscosity. In the published data [10–13], such a situation has not been considered (these conditions are usually fulfilled, since they follow from the Hugoniot entropy conditions, which have a clear physical meaning).

DETERMINATION OF THE SHOCK (RAREFACTION) WAVE VELOCITY

Let us recast Eq. (5) in the form

$$\frac{\partial\varphi}{\partial t} + p\frac{\partial}{\partial x}[S(\varphi)] = 0, \qquad \varphi(x, 0) = \varphi_0(x). \tag{5"}$$

Here, in the medium field approximation, $p = D_{\text{ef}}/k_B T$ and $S(\varphi) = \mu(\varphi)$, where $dF(\varphi)/d\varphi = \mu(\varphi)$ is the generalized chemical potential and F is the free energy.

The entropy condition is stated as follows: $p > 0 \longrightarrow p\mu(\varphi)$ convex function $\longrightarrow$ inequality $\varphi^- < \varphi^+$ is fulfilled. A similar condition can be written for $p < 0$. Then, the solution to problem (5") takes the form [8]

$$\varphi(x, t) = \begin{cases} \varphi_2 & \text{for} \quad -\infty < x < p\mu'(\varphi_1)t \\ (\mu')^{-1}\left[\dfrac{x}{pt}\right] & \text{for} \quad p\mu'(\varphi_1)t \le x \le p\mu'(\varphi_2)t \\ \varphi_1 & \text{for} \quad p\mu'(\varphi_2)t < x < +\infty. \end{cases} \tag{12}$$

Here, $(\mu')^{-1}$ is the function reciprocal to the function μ'. Solution (12) is depicted in Fig. 4. It was found in [6–8] with the entropy condition

$$\frac{F(\varphi_1) - F(\varphi_2)}{\varphi_1 - \varphi_2} = \max_{c\in[\varphi_1\cap\varphi_2,\,\varphi_2\cup\varphi_1]}\frac{F(\varphi_1) - F(c)}{\varphi_1 - c} \tag{13}$$

fulfilled, where

$$\varphi_1(x, t) = \lim_{y\downarrow x}\varphi(y, t) \quad \text{and} \quad \varphi_2(x, t) = \lim_{y\uparrow x}\varphi(y, t).$$

Here, $y \downarrow x$ and $y \uparrow x$ means that y approaches the point x from the right and left, respectively. The designation $c \in \varphi_2 \cup \varphi_1$ means that we approach the discontinuity line "along" point c either from the right (of φ_2) or from the left (of φ_1), while the designation $c \in \varphi_2 \cap \varphi_1$ implies that point c may belong to the discontinuity line. Here, $F'(\varphi) = G(\varphi)$ as before.

Comparing entropy condition (13) with admissibility condition (6) (the latter, as is known, is derived from

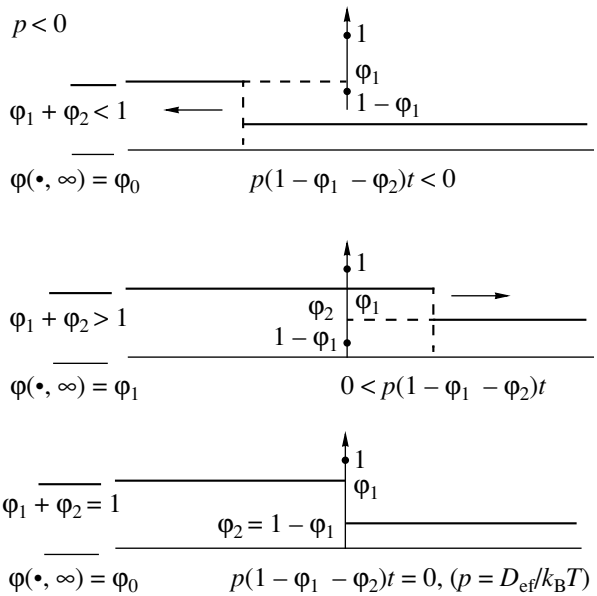

Fig. 4. Possible limit density distributions vs. the relationship between the stable phase (φ_1 and φ_2) amplitudes and the lability parameter $p = D_{\text{ef}}/k_B T$.

the Hugoniot integral condition in the vicinity of the discontinuity front), one can see that they coincide if

$$\nu = \frac{dx}{dt} = \max_{c \in [\varphi_1 \cap \varphi_2, \varphi_2 \cup \varphi_1]} \frac{\int_{\rho}^{\varphi_1} G(\varphi)d\varphi - \int_{\rho}^{c} G(\varphi)d\varphi}{\varphi_1 - c}, \quad (14)$$

where $\rho \in \Delta\xi$ (i.e., ρ belongs to the set of points where the function $G(\varphi)$ is many valued).

A similar inequality must be fulfilled with φ_1 changed to φ_2 in relationship (14); hence, we come to the area rule

$$\int_{\varphi_1}^{c} G(\varphi)d\varphi = \int_{c_1}^{\varphi_2} G(\varphi)d\varphi \quad (15)$$

with $\rho \longrightarrow c$. Then, from (6), (14), and (15), it follows that, generally speaking, Hugoniot relationship (15) does not take into account fluctuations in the range $\Delta\xi$, i.e., near the discontinuity front. This observation, in particular, has led Kruzhkov to the elaboration of the entropy theory for quasi-linear hyperbolic equations (see, e.g., [9]). Relationship (15) represents, in essence, the law of conservation of mass for the density distribution: a discontinuity of a steeping wave should be placed at point c, where its artificial appearance does not change the area between the plot of the wave $\varphi(x, t)$ and the x axis [19, p. 42].

Note that if the solutions in Figs. 1 and 4 are viewed as the limits of the perturbed equation, the case $p < 0$ is physically meaningful. In fact, let us consider Eq. (2') for the case when the curve $F(\varphi)$ has a negative-curvature region $\delta^2F/\delta\varphi^2 < 0$ in the interval (φ_1, φ_2). Within this interval, the system separates into two phases with concentrations φ_1 and φ_2 [1, p. 113]. In this case, an expression obtained by substituting $-t$ for t in (1) and proceeding to the limit in a small parameter will be the limit relationship to Eq. (2'): we will arrive at Eq. (5) with $p < 0$. This meets the situation considered in [1, p. 112] and takes place when the interaction constant χ exceeds some critical value χ_c that corresponds to the negative-curvature region in the curve $F(\varphi)$.

The aforesaid means that in the vicinity of the spinodal decomposition curve $\delta^2F/\delta\varphi^2 < 0$, there exists a narrow region where fluctuations can be ignored. Accordingly, spinodal decomposition can be simulated by divergent equation (5), which has the form of the law of conservation with $p > 0$ over the spinodal and $p < 0$ below the spinodal. However, the decomposition changes qualitatively: strictly in going through the spinodal curve, the propagation of the shock wave (Fig. 2) changes to the generation of the rarefaction wave (Fig. 3).

ACKNOWLEDGMENTS

The authors thank A.F. Tedeev and A.E. Shishkov for fruitful discussions.

This work was financially supported in part by the Ministry of Education and Science of Ukraine (project no. 2M/71-2000).

REFERENCES

1. P. G. de Gennes, *Scaling Concepts in the Physics of Polymers* (Cornell Univ. Press, Ithaca, 1979; Mir, Moscow, 1982).
2. V. S. Mitlin and L. I. Manevich, Vysokomol. Soedin., Ser. A **31**, 1020 (1989).
3. V. S. Mitlin and L. I. Manevich, Vysokomol. Soedin., Ser. A **30**, 9 (1988).
4. V. G. Vaks, C. V. Beiden, and V. Yu. Dobretsov, Pis'ma Zh. Éksp. Teor. Fiz. **61**, 65 (1995) [JETP Lett. **61**, 68 (1995)].
5. V. S. Mitlin, L. I. Manevich, and I. Ya. Erukhimovich, Zh. Éksp. Teor. Fiz. **88**, 495 (1985) [Sov. Phys. JETP **61**, 290 (1985)].
6. A. Benassi and J. P. Fouque, Ann. I. H. P. **29**, 189 (1988).
7. A. Benassi and J. P. Fouque, Ann. Prob. **15**, 546 (1987).
8. A. L. Bertozzi, A. Münch, and M. Shearer, Physica D (Amsterdam) **134**, 431 (1999).
9. S. N. Kruzhkov, Mat. Sb. **81**, 228 (1978).
10. O. A. Oleĭnik, Usp. Mat. Nauk **12**, 3 (1957).
11. I. M. Gel'fand, Usp. Mat. Nauk **14**, 87 (1959).
12. O. A. Oleniĭ, Usp. Mat. Nauk **14**, 159 (1959).
13. O. A. Oleniĭ, Usp. Mat. Nauk **14**, 166 (1959).
14. D. Henry, *Geometric Theory of Semilinear Parabolic Equations* (Springer-Verlag, Berlin, 1981; Mir, Moscow, 1987).
15. R. Z. Sagdeev, D. A. Usikov, and G. M. Zaslavsky, *Nonlinear Physics: From the Pendulum to Turbulence and Chaos* (Nauka, Moscow, 1988; Harwood, Chur, 1988).
16. P. G. De Gennes, J. Chem. Phys. **72**, 4756 (1980).
17. J. Keiser, *Statistical Thermodynamics of Nonequilibrium Processes* (Springer-Verlag, Heidelberg, 1987; Mir, Moscow, 1990).
18. L. Young, *Lectures on the Calculus of Variations and Optimal Control Theory* (Saunders, Philadelphia, 1969; Mir, Moscow, 1974).
19. M. J. Lighthill, *Waves in Fluids* (Cambridge Univ. Press, Cambridge, 1978; Mir, Moscow, 1981).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 1, 2003, pp. 73–79. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 79–85.

OPTICS, QUANTUM ELECTRONICS

Optical Limitation of Mid-IR Radiation in Vanadium Dioxide Films

O. B. Danilov*, V. A. Klimov, O. P. Mikheeva*, A. I. Sidorov*, S. A. Tul'skiĭ*, E. B. Shadrin**, and I. L. Yachnev***

** Research Institute of Laser Physics, St. Petersburg, 199034 Russia*

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

Received May 24, 2002

Abstract—Theoretical and experimental results on the limitation of pulsed radiation in the spectral range 3.8–10.6 μm in vanadium dioxide films are presented. The effect of the film structure on the shape of the temperature hysteresis loop is studied. The film thickness and the structure of an interferometer with a vanadium dioxide film are optimized to improve the radiation limitation efficiency. The spatial dynamics of vanadium film switching under the action of a radiation pulse is investigated. It is shown experimentally that the radiation attenuation coefficient under the limitation conditions can be as high as 10^4 or more. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Nonlinear optical effects that attenuate the transmitted radiation intensity as the incident intensity increases are applied for controlling the shape and amplitude of laser pulses, as well as for protecting eyes and photodetectors against blinding and radiation-induced failure, respectively. The latter field of application is of great practical importance, since today's radars and pollution prevention systems use high-sensitivity radiation detectors, the reliable operation of which specifies the reliability of the system as a whole. The field of application imposes specific requirements for radiation limiters. These are minimal radiation losses in the linear transmission range, low energy threshold of limitation, attenuation coefficient exceeding 1000 under the limitation conditions, high speed of response, radiation resistance, and wide spectral range of limiter operation.

In the visible and near-IR ranges, there exists a variety of materials and nonlinear optical effects that can be used for radiation limitation. Among them are nonlinear optical effects in semiconductors arising upon two- and one-photon absorption [1, 2] resonance effects in composites [3], effects associated with two-photon absorption in liquid crystals [4], effects arising at singlet–triplet transitions in fullerenes [5], and many others. In the mid-IR range, the number of effects and materials appropriate for radiation limitation decreases sharply. Basically, this is due to the lack of materials transparent in the mid-IR range. Another reason is that the photon energy in this range is low and cannot excite most nonlinear optical effects. However, for several widely used lasers, such as chemical (HF, DF) lasers and CO (CO_2) lasers, lasing is observed in precisely this mid-IR range. Therefore, the development of radiation limiters for this spectral range seems to be a topical problem.

Vanadium dioxide (VO_2), which exhibits the reversible semiconductor–metal phase transition (PT) [6, 7], is a promising material for mid-IR radiation limiters [8, 9]. This is largely because the variation of the VO_2 optical constants is the greatest in the mid-IR range when the PT takes place [10]. The aim of this work is to elucidate factors governing the limitation of radiation with $\lambda = 3$–11 μm in polycrystalline VO_2 films and to find ways of improving the limitation efficiency.

MECHANISMS BEHIND THE RADIATION-INDUCED PHASE TRANSITION IN VO_2 FILMS

In vanadium dioxide, the semiconductor–metal phase transition of the first order takes place at $t_c = 67$°C. Above t_c, VO_2 is a metal with the tetragonal lattice; below t_c, it is a semiconductor with the monoclinic lattice and an energy gap $E_g \approx 0.7$ eV [6].

According to the photon energy $h\nu$, the PT in VO_2 may be initiated by either band-to-band electron transitions at $h\nu > E_g$ or thermal processes at $h\nu < E_g$. The first case corresponds to the visible and near-IR ranges; the latter, to the mid-IR range.

In the first mechanism of the semiconductor–metal transition, an induced increase in the free carrier concentration causes the Coulomb screening of electron–electron correlations, which reduces stepwise the correlation energy in VO_2 [11]. In this case, the total energy of free carriers falls into the conduction band and VO_2 acquires metallic conductivity. Simultaneously, the

1063-7842/03/4801-0073$24.00

high-frequency permittivity, i.e., the optical constants of the material, changes. Note that in VO_2 the electronic order parameter changes markedly starting from 6°C, that is, 60°C below the PT point [12], while in other materials, such a screening occurs near the PT temperature.

Hence, absorbing a radiation pulse that generates free carriers, the electronic subsystem passes into the excited state. As the number of absorbed pulses increases, the free carrier concentration rises and the Mott electron PT starts. Then, the crystal lattice, responding to the changes in the electronic subsystem, also rearranges and the thermodynamic potential of the crystal grows. These processes change the optical constants. If, however, the absorbed energy is insufficient to overcome the potential barrier between the semiconducting and metallic phase, the crystal quickly returns to the initial state, which corresponds to the minimal thermodynamic potential of the semiconducting phase. In this case, the complete PT, when both the electronic and lattice subsystems are rearranged and the material fixes the new (metallic) state, is not observed.

In other words, the early stage of the semiconductor–metal PT at $h\nu > E_g$ involves the fast rearrangement of the valence electron subsystem (an additional transfer of the electron density to the ligand in terms of [11, 13]) because of a sharp increase in the free carrier concentration in the conduction band when an optical pulse is absorbed. This may take place at temperatures much lower (23°C) than the PT temperature (67°C) owing to the temperature dependence of the electronic order parameter of VO_2 [12]. Because of the electronic nature of the PT [15], it may be completed within a very short time: from several picoseconds [16] to one hundred femtoseconds [17].

The process described above is akin to ultra-high-speed vibron PTs in Si, GaAs, or InSb semiconductors arising under the action of a powerful optical pulse [14, 18]. However, for the ultra-high-speed PT to occur, the softening of the phonon spectrum is necessary. In these semiconductors, the softening of the spectrum is due to the sharply increased density of the free carrier plasma. In our case, the phonon spectrum softens irrespective of the plasma density as the temperature approaches the semiconductor–metal PT temperature [12, 19]. The softening starts with a temperature of 6°C, i.e., 60°C below the temperature of the transition.

Thus, a VO_2 film may switch to the metallic state within several fractions of a picosecond under the action of an optical pulse from the visible or near-IR range. However, since the semiconducting phase of VO_2 is highly absorbing and its optical constants vary insignificantly in this spectral range, VO_2 films can hardly compete with other limiters that are more effective at $\lambda < 2$ μm.

When a VO_2 film is exposed to a photon of energy less than the energy gap of the VO_2 semiconducting phase and two-photon absorption is absent (which is valid for the mid-IR range), the PT is initiated by VO_2 heating due to radiation absorption.

For this type of radiation, the transition-initiating mechanism can be described at length in terms of the electron (Mott) semiconductor–metal PT in VO_2 [15].

When exposed to mid-IR radiation, a VO_2 film is heated first as a result of radiation absorption by a free carrier plasma [20], which heats up the crystal lattice during the relaxation of the stored energy. Second, the film is heated because of the direct absorption of the radiation by the lattice, since the absorption of mid-IR radiation by the VO_2 semiconducting phase effectively generates phonons [21].

The heating of the material by an optical pulse increases the concentration of free carriers, which can leap the energy gap of the VO_2 semiconducting phase. Because of the increase in the concentration, the Coulomb screening of electron correlations at the same lattice site appears, which is typical of the Mott transition [13]. The screening partially reduces the electron correlation energy [11], narrowing the Mott energy gap of semiconducting VO_2 [7]. In turn, the narrowing of the energy gap favors the thermal transfer of carriers over the energy gap, which narrows it still further. Thus, a specific positive feedback mechanism that causes the energy gap of the semiconducting phase to collapse and the material to pass to the metallic state is triggered. Simultaneously, the high-frequency permittivity (optical constants) of the material changes. Note once again that the electronic order parameter can vary appreciably starting from 6°C. This means that an optical pulse heating the VO_2 film may initiate the superconductor–metal PT and carry the film to the high-absorptivity state even at room temperature.

THE EFFECT OF POLYCRYSTALLINE VO_2 FILM STRUCTURE ON THE PHASE TRANSITION PARAMETERS

The radiation limitation parameters of VO_2 films correlate appreciably with the temperature dependences of the reflection, *R*, and transmission, *T*, coefficients. If the steepness of the associated curves is large, the speed of the limiter rises; if the steepness is small, the dynamic range of limitation expands. The time it takes for the initial transmission to recover depends on the steepnesses of the $R(t)$ and $T(t)$ curves, as well as on the width of the temperature hysteresis loop.

Experiments show that the cooling and heating curves for the VO_2 physical parameters diverge, indicating the temperature hysteresis of the material properties (Fig. 1). It is noteworthy here that the hysteresis loop for the reflection coefficient of a VO_2 film interferometer is several times wider than that for the same parameter of VO_2 single crystals (15°C vs. 2°C), has extended (20–30°C along the temperature axis)

branches, and is shifted toward low temperature by 10–20°C (Fig. 1b).

The reason for such a difference is that the hysteresis loop of a polycrystalline VO_2 film consists of elementary loops from individual grains [15]. The temperatures t_{ci} and the widths Δt_i of elementary loops differ (i is the number of a group of grains with similar properties). Therefore, hysteresis effects in the film are described by the grain temperature (t_{ci}) distribution and by the grain distribution over the deviations Δt_i from t_{ci} (this deviation is necessary for the PT to take place in a given grain). Then, it follows that (1) the width of the total loop for a VO_2 film depends on the widths of the elementary loops for the grains that are responsible for the peak of the distribution of the elementary loops over their widths Δt_i, (2) the temperature extent of the branches of the total loop depends on the width of the temperature (t_{ci}) distribution of the elementary loops, and (3) the position of the total loop branches depends on the position of the peak of this temperature distribution. It is assumed here that each grain has a single elementary loop and elementary loops are "vertical" relative to the temperature axis; that is, once the temperature of the transition (e.g., to the metallic state) has been reached, a nucleus of the new metallic phase grows through the semiconducting grain with the speed of sound [15].

The PT temperature and the shape of the temperature hysteresis loop are affected by the presence of impurities in VO_2. They also depend on the film nonstoichiometry (oxygen excess or deficiency) [6, 7]. For example, the PT temperature in a tungsten-doped VO_2 film can be reduced by 20–30°C in comparison with an undoped film. The oxygen nonstoichiometry (both oxygen excess and deficiency) extends the PT temperature interval. Also, the absorption of the semiconducting phase grows in this case.

From the aforesaid, it follows that a highly extended and still narrow temperature hysteresis loop can be provided if the elementary loop distribution over widths Δt_i is as narrow as possible and its peak is near small Δt_i values. Simultaneously, the elementary loop distribution over their temperature positions t_{ci} must be as wide as possible. If it is necessary to obtain a narrow total hysteresis loop with short branches, the elementary loop distribution over t_{ci} must also be narrow. Both approaches are applied to designing optical limiters.

OPTICAL PROPERTIES OF LAYER SYSTEMS WITH A VO_2 FILM

The semiconductor–metal PT in layer structures with a VO_2 film (VO_2 mirrors) greatly changes the reflection and transmission coefficients [10, 22]. The optical properties of film coatings forming a VO_2 mirror define the parameters of radiation limiters such as the linear transmission coefficient, spectral operating

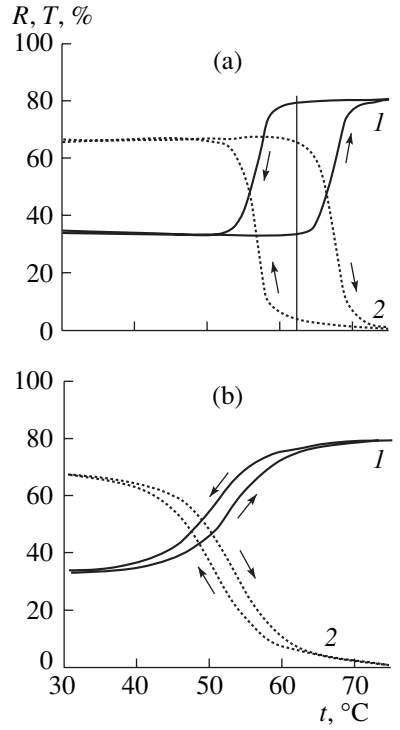

Fig. 1. Temperature dependences of the (*1*) reflection, R, and (*2*) transmission, T, coefficients for 0.25-μm-thick VO_2 films with (a) narrow and (b) wide grain size distribution. $\lambda = 10.6$ μm.

range, attenuation coefficient under radiation limitation conditions, and radiation resistance. The PT changes the VO_2 complex refractive index n^* ($n^* = n - i - k$) from $3.1 - i0.002$ to $4 - i6$ for $\lambda = 3.4$ μm [22] and from $2.55 - i0.08$ to $8 - i9$ for $\lambda = 10.6$ μm [10]. Since after the PT the absorption coefficient k of the film increases 100–3000 times, even a single 0.2- to 0.3-μm-thick VO_2 film applied on a transparent substrate can limit mid-IR radiation. Yet the VO_2 film thickness d should be optimized to attain the maximal transmission of this system at an operating wavelength of the limiter for $t < t_0$ (t_0 is the temperature of the onset of the PT). Figure 2a shows the spectral dependence of the transmission coefficient for a VO_2 film on a germanium substrate at $t < t_0$ and $t > t_c$. At $d = 1$ μm and $t < t_0$ (curve *1*), the dependence $T(\lambda)$ has two peaks: at $\lambda = 4$ μm ($T = 82\%$) and $\lambda = 10.5$ μm ($T = 85\%$). The transmission maxima are associated with the antireflection property of the VO_2 film, since its optical width (nd) equals $(3/4)\lambda$ and $\lambda/4$ for these wavelengths. When the film passes to the metallic state, its transmission diminishes, respectively, to 10^{-5} and 3×10^{-4}%. Thus, prior to the PT, a 1-μm-thick VO_2 film provides a high transmission (more than 80%) in the spectral ranges 4.0–4.5 and 8.0–11.5 μm; after the PT, a high attenuation (more than 10^5) is observed in the range 3–12 μm. For comparison, Fig. 2a shows the transmission spectrum for a VO_2 film

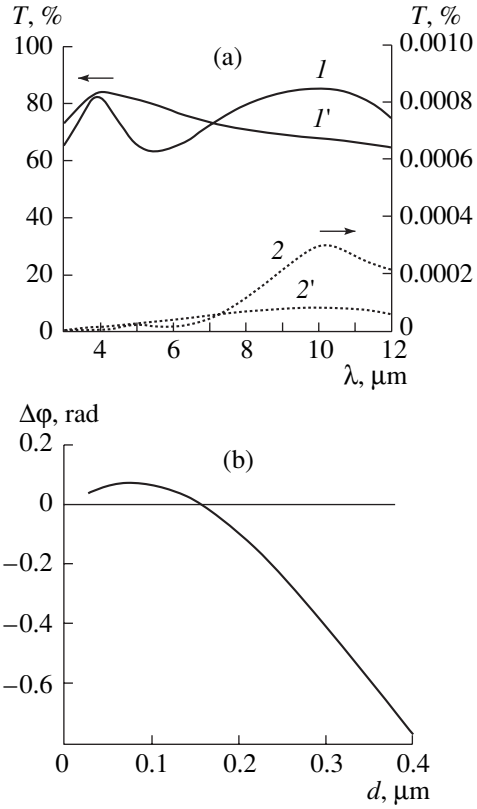

Fig. 2. Transmission spectra for the VO_2 film on the germanium substrate. d = (*1*, *2*) 1 and (*1*', *2*') 0.3 μm; t = (*1*, *1*') 20 and (*2*, *2*') 70°C.

of thickness 0.3 μm. Since the optical thickness of this film before the PT is λ/4 for a wavelength of ≈4 μm, it offers a high transmission in the range 3.5–5.5 μm. After the PT, the attenuation coefficient of this film exceeds 10^3 for λ = 3–12 μm.

When limiters are used for protecting the photodetectors of laser radars, the phase distortions of the optical signal should be minimized in some cases. This is necessary, for example, when the detector analyzes the radiation wave front. Figure 2b demonstrates how the thickness of the VO_2 film on the germanium substrate affects the phase shift $\Delta\varphi = \delta\varphi_1 - \delta\varphi_2$ of the transmitted radiation before and after the PT. Here, $\delta\varphi_1$ and $\delta\varphi_2$ are the phase shifts of the transmitted radiation before and after the PT, respectively. The dependence $\Delta\varphi(d)$ was calculated for λ = 10.6 μm with the VO_2 optical constants given above. It follows from Fig. 2b that for $d <$ 0.35 μm, the phase shift after the PT changes by no more than π/5, while for $d \approx 0.17$ μm, the phase shift remains unchanged. Thus, a relatively thin film may attenuate the radiation with the least phase distortion. However, the other optical parameters are not optimal in this case. For example, at d = 0.2 μm, the transmission coefficient of the film before and after the PT is 64 and 3%, respectively.

The reflection coefficient of a VO_2 film thicker than 0.25 μm after the PT is about 80% for λ = 5–11 μm. This means that such films absorb roughly 20% of the incident radiation after the PT. If the film breaks down by the thermal mechanism, the breakdown threshold decreases with increasing fraction of the energy absorbed by the metallic phase [23]. For a 0.3-μm-thick VO_2 film, the radiation-induced breakdown energy density at λ = 10.6 μm was found to be 6–7 J/cm^2 for a pulse duration of 20 μs. The breakdown threshold of the limiter can be raised by increasing the reflection coefficient of the VO_2 mirror for $t > t_c$. To this end, a multilayer thin-film interferometer with a VO_2 film as a control element can be used. It was shown that in a Ge(d = 0.66 μm)–ZnSe(1.1 μm)–VO_2(0.25 μm)–ZnS(1.0 μm)–Ge(substrate) four-layer interferometer, the reflection coefficient for λ = 10.6 μm at $t > t_c$ is 95%. Hence, for such a VO_2 mirror, the fraction of the absorbed energy at $t > t_c$ is four times lower than for a single VO_2 film. It was shown experimentally that VO_2 mirrors with a reflection coefficient more than 95% ($t > t_c$) have a breakdown threshold of 20–25 J/cm^2 [23]. For an interferometer of such a design, the transmission coefficient equals 0.22% at $t > t_c$ and 60% at $t < t_0$. It should also be noted that the above-mentioned optical parameters of the interferometer are observed in the narrow spectral range $\Delta\lambda \approx 0.1$ μm.

Multilayer structures can also be used to increase the transmission of VO_2 mirrors at $t < t_0$. For example, in a ZnS(1.1 μm)–VO_2(0.4 μm)–Ge(substrate) two-layer structure, the transmission coefficient at $t < t_0$ equals 95%. However, the transmission and reflection coefficients at $t > t_c$ equal, respectively, 1.4 and 38%, which impairs the limitation efficiency and decreases the breakdown threshold.

Thus, the use of a VO_2 film in various interference systems makes it possible to improve (in comparison with a single VO_2 film) one of the optical parameters of radiation limiters, for example, the radiation resistance or transmission in the linear mode. At the same time, the other optical parameters of the limiters degrade. Therefore, the design of a VO_2 mirror should be chosen according to specific requirements for a limiter.

SWITCHING OF A VO_2 FILM UNDER THE ACTION OF A RADIATION PULSE

A feature of a VO_2 film as a nonlinear optical element is the relatively narrow temperature interval of nonlinearity: from several degrees to several tens of degrees. Both before and after the PT, the transmission of the film does not depend on the radiation intensity and the attenuation coefficient after the film has been switched to the metallic state is $1/T_c$, where T_c is the transmission coefficient of the film after the PT. Such a situation, however, takes place only if the VO_2 mirror surface is uniformly heated in the static approximation.

Under real conditions, the switching of a VO_2 film under the action of a laser pulse gives rise to several effects owing to which both the dynamic range of nonlinearity and the attenuation factor after the switching differ substantially from those under static conditions.

First, recall that the limiter is considered reversible if the initial transmission coefficient is recovered after the action of a laser pulse. This condition is met when the VO_2 mirror substrate is kept at the temperature $t = t_0 - \Delta t$, where Δt is the width of the temperature hysteresis loop for a VO_2 film (Fig. 1a). The presence of the early linear portion in the characteristic of the limiter causes a switching delay and decreases its integral attenuation coefficient. The delay depends on the value of Δt, radiation intensity, and thermophysical properties of the VO_2 mirror. For an intensity of ≈ 1 MW/cm^2, the delay is several tens of nanoseconds.

Second, with a pure thermal mechanism of VO_2 switching under the action of a laser pulse, the power of a heat source in the film is defined by the instantaneous incident intensity $I_0(r, \tau)$ and instantaneous reflection and transmission coefficients:

$$P(r, \tau) = I_0(r, \tau)[1 - R(r, \tau) - T(r, \tau)].$$

In turn, $R(r, \tau)$ and $T(r, \tau)$ depend on the profile of the instantaneous film temperature $t(r, \tau)$. Because of this, the rate of film switching $dT/d\tau$ varies during the PT and attains a maximum in the middle of the heating branch of the hysteresis loop [24]. If the VO_2 film is heated by a spatially nonuniform laser pulse, the switching dynamics depends on both the time waveform and spatial profile of the pulse. Figure 3 shows the calculated spatial distribution of the radiation transmitted through the VO_2 film when the incident radiation ($I_0 = 0.5$ MW/cm^2) has a Gaussian profile. The laser pulse has a rectangular front. The self-consistent thermal problem was solved in the approximation of a thin plate (VO_2 film) that is in ideal thermal contact with a semi-infinite body (germanium substrate). As follows from Fig. 3, at $\tau = 30$ ns, the limitation is absent because of the delay of VO_2 film switching. With increasing pulse duration, the switching occurs in the central part of the pulse, where the intensity is the highest. The region of limitation expands with time from the center to the periphery. At the edges of the distribution, the limitation is absent because here the intensity is insufficient for the VO_2 film to pass into the metallic state.

Thus, the spatially nonuniform distribution of the incident radiation causes additional effects associated with the spatial dynamics of VO_2–radiation interaction. While in the highest-intensity region the attenuation coefficient is the highest (corresponding to static conditions), the integral (over time and cross section) attenuation coefficient may be several tens or even several hundreds of times lower. Moreover, the spatial dynamic processes occurring when the beam is spatially nonuniform expand the energy range of VO_2 film nonlinearity and, thus, the dynamic range of limitation.

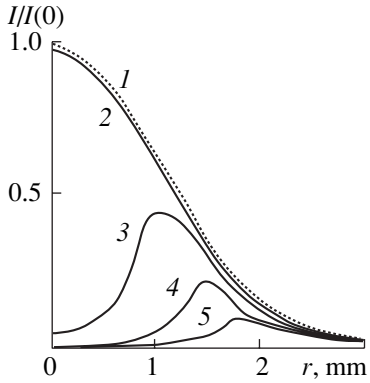

Fig. 3. Spatial distribution of the radiation passed through the VO_2 film under the limitation conditions. $I_0 = 0.5$ MW/cm^2. (*1*) Incident beam profile. τ = (*2*) 30, (*3*) 300, (*4*) 600, and (*5*) 1000 ns.

The time it takes for the transmission to recover after the end of the pulse depends on the incident intensity, which specifies the maximal temperature of the VO_2 film, the hysteresis loop width, and the thermal conductivity of the substrate. For a VO_2 film on a germanium substrate, this time lies within 10–100 μs according to the radiation intensity.

Experiments with a simple limiter consisting of a 0.3-μm-thick VO_2 film on a germanium substrate were carried out at a wavelength of 10.6 μm (TEA CO_2 laser, pulse duration $\tau_0 = 2$ μs) and in the spectral range 3.8–4.2 μm (DF laser, $\tau_0 = 0.25$ μs). The substrate temperature was kept at 55°C. The energy threshold of radiation limitation (the onset of nonlinear transmission) was found to equal roughly 5 mJ/cm^2 in both cases. Figure 4 shows the waveforms of the incident pulse and transmitted radiation for $\lambda = 10.6$ μm at different incident energy densities Q_0. At $Q_0 = 400$ mJ/cm^2, the switching delay is seen to be about 50 ns; that is, the limitation occurs even at the front of the incidence radiation. The fall time of the transmitted radiation is defined by the spatial dynamic processes described above and equals several hundreds of nanoseconds.

LIMITATION OF 10.6-μm RADIATION IN A THREE-PASS LIMITER

The energy threshold of limitation can be reduced and the dynamic range of limitation can be extended when a focused laser beam passes successively through several VO_2 films. However, the optimal design is that where the focused radiation crosses the same VO_2 film many times [24]. The optical scheme of such a limiter is shown in the inset to Fig. 5. The limiter consists of a reflecting telescope with entrance (*1*) and exit (*2*) mirrors between which nonlinear optical element *3* (a VO_2

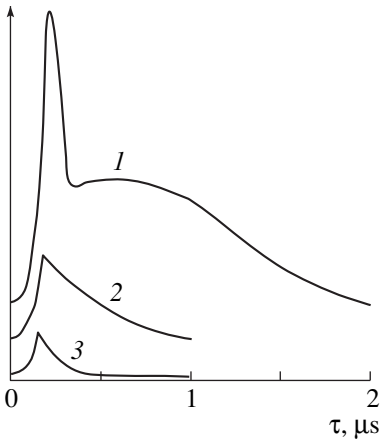

Fig. 4. Waveform of the radiation pulse transmitted through the VO_2 film under the limitation conditions. $\lambda = 10.6$ μm. (*1*) Incident pulse; Q_0 = (*2*) 100 and (*3*) 400 mJ/cm^2.

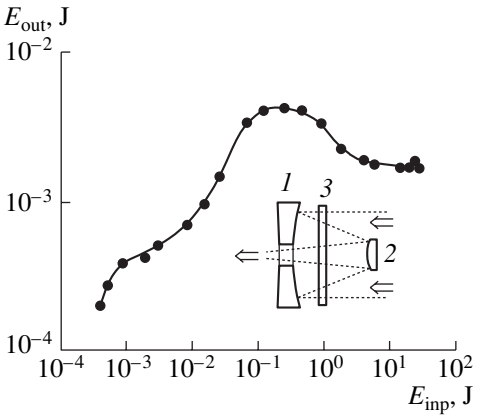

Fig. 5. Output energy vs. input energy dependence. The inset shows the optical scheme of the limiter. $\lambda = 10.6$ μm, $\tau_0 = 20$ μs.

film on a germanium substrate with an antireflection coating) is placed. The focal spots of the mirrors are shifted relative to each other along the optical axis to form a convergent beam at the exit of the limiter. The thickness of the VO_2 film is 0.75 μm. The transmission of the nonlinear optical element per one pass is 80% for $\lambda = 10.6$ μm at $t = 20$°C. The entrance and exit apertures of the limiter are, respectively, 35 and 10 mm.

The incident radiation passes through the VO_2 film three times while simultaneously being focused. Under the limitation conditions, the central part of the film switches over first (third pass of the radiation). Then, the regions corresponding to the second and first passes switch over. After the second-pass region has been switched, the radiation load on the central part of the film is reduced, protecting it against radiation-induced breakdown. Similarly, the switching-over of the first-pass region reduces the load on the second- and third-pass regions of the VO_2 film. The integral attenuation coefficient of the limiter equals the product of the partial integral attenuation coefficients at each of the passes.

The limitation experiments were carried out at a wavelength of 10.6 μm for a laser pulse duration of 20 μs. The radiation source was an electroionization CO_2 laser. The temperature of the germanium substrate supporting the VO_2 film was kept at 55°C. Figure 5 shows the experimental dependence of the output energy on the input energy. At $E_{inp} \geq 0.8$ mJ, the VO_2 film switches over (the smooth portion of the limitation curve). After the central part of the film has been completely switched to the metallic state, the dependence $E_{out}(E_{inp})$ becomes nearly linear. Then, at $E_{inp} \geq 100$ mJ, the second and third switching stages arise, when the edge regions of the film switch over. The energy threshold of limitation is 0.8 mJ, or 0.08 mJ/cm^2 in terms of incident energy density. For an incident energy of 26.5 J, the integral attenuation coefficient is 1.5×10^4. Analysis indicates that the attenuation coefficient may be raised further by optimizing the limiter design.

CONCLUSION

Our results demonstrate that polycrystalline vanadium dioxide films are a promising material for mid-IR radiation limiters. Good limitation characteristics can be obtained by optimizing the film structure and its interference properties. Under static conditions, the switching contrast may exceed 10^6. However, under dynamic conditions (pulse-induced switching-over), the integral attenuation coefficient decreases appreciably due to dynamic effects. Yet, attenuation coefficient as high as 10^4 can be achieved with a three-pass limiter. The basic disadvantage of a limiter with a VO_2 film is the need for thermostatic control.

Radiation limiters based on the reversible semiconductor–metal PT in a VO_2 film can be used for protecting mid-IR photodetectors against blinding and radiation-induced breakdown.

ACKNOWLEDGMENTS

This work was financially supported by the Interindustry Research Center (grant no. 1454).

REFERENCES

1. T. E. Boggess, A. L. Smirl, S. C. Moss, *et al.*, IEEE J. Quantum Electron. **21**, 488 (1985).
2. I. V. Bagrov, A. P. Zhevlakov, A. I. Sidorov, *et al.*, Opt. Zh. **69** (2), 15 (2002) [J. Opt. Technol. **69**, 77 (2002)].
3. R. D. Averitt, S. L. Westcott, and N. J. Halas, J. Opt. Soc. Am. B **16**, 1814 (1999).
4. V. V. Danilov, E. N. Sosnov, O. V. Chistyakova, *et al.*, Opt. Spektrosk. **90**, 429 (2001) [Opt. Spectrosc. **90**, 371 (2001)].

5. V. P. Belousov, I. M. Belousova, E. A. Gavronskaya, *et al.*, Opt. Zh. **68** (12), 13 (2001) [J. Opt. Technol. **68**, 876 (2001)].
6. A. A. Bugaev, B. P. Zakharchenya, and F. A. Chudnovskiĭ, *Phase Metal–Semiconductor Transition and Its Application* (Nauka, Leningrad, 1979).
7. W. Bruckner, H. Opperman, W. Reichelt, *et al.*, *Vanadiumoxide* (Akademic, Berlin, 1983).
8. O. B. Danilov, V. P. Belousov, I. M. Belousova, *et al.*, Proc. SPIE **3263**, 124 (1998).
9. J. P. Segaud, B. Dillemann, and O. Giraudo, Nonlinear Opt. **21**, 175 (1999).
10. O. P. Konovalova, A. I. Sidorov, and I. I. Shaganov, Opt. Zh. **66** (5), 13 (1999) [J. Opt. Technol. **66**, 391 (1999)].
11. T. Altanhan, J. Phys. C **20**, L949 (1987).
12. V. S. Vikhnin, I. N. Goncharuk, V. Yu. Davydov, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **37**, 3580 (1995) [Phys. Solid State **37**, 1971 (1995)].
13. N. F. Mott, *Metal–Insulator Transitions* (Taylor & Francis, London, 1974; Nauka, Moscow, 1979).
14. V. I. Emel'yanov and D. V. Babak, Fiz. Tverd. Tela (St. Petersburg) **41**, 1462 (1999) [Phys. Solid State **41**, 1338 (1999)].
15. E. B. Shadrin and A. V. Il'inskiĭ, Fiz. Tverd. Tela (St. Petersburg) **42**, 1092 (2000) [Phys. Solid State **42**, 1126 (2000)].
16. I. A. Balashenkov, P. A. Solunin, I. A. Khakhaev, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **36**, 2032 (1994) [Phys. Solid State **36**, 1347 (1994)].
17. A. Cavalleri, C. Toth, C. W. Siders, *et al.*, Phys. Rev. Lett. **87**, 237401 (2001).
18. A. L. Roĭtburd, Usp. Fiz. Nauk **113**, 69 (1974) [Sov. Phys. Usp. **17**, 326 (1974)].
19. I. A. Khakhaev, F. A. Chudnovskiĭ, and E. B. Shadrin, Fiz. Tverd. Tela (St. Petersburg) **36**, 1643 (1994) [Phys. Solid State **36**, 898 (1994)].
20. V. S. Vavilov, *Effects of Radiation on Semiconductors* (Fizmatgiz, Moscow, 1963; Consultants Bureau, New York, 1965).
21. A. S. Barker, H. W. Verleur, and H. J. Guggenheim, Phys. Rev. Lett. **17**, 1286 (1966).
22. F. C. Case, Appl. Opt. **30**, 4119 (1991).
23. O. B. Danilov, A. P. Zhevlakov, A. I. Sidorov, *et al.*, Opt. Zh. **67** (6), 31 (2000) [J. Opt. Technol. **67**, 526 (2000)].
24. O. P. Mikheeva and A. I. Sidorov, Opt. Zh. **68** (4), 48 (2001) [J. Opt. Technol. **68**, 278 (2001)].

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 1, 2003, pp. 80–84. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 86–90.

ELECTRON AND ION BEAMS, ACCELERATORS

Characteristic Features of the Generation of Large-Diameter Low-Energy Electron Beams from a Penning Plasma Source

V. N. Borisko and A. A. Petrushenya

Kharkov National University, 61108 Kharkov, Ukraine

e-mail: Borisko@pht.univer.kharkov.ua

Received January 10, 2002

Abstract—A wide-aperture plasma source of low-energy electron beams on the basis of an electrically asymmetric reflex discharge is studied experimentally. The characteristic features of the generation of electron beams are investigated, and the boundaries of the generation regime are determined. The influence of the dimensions and shape of the discharge electrodes on the source characteristics is examined. A method for efficiently controlling the electron-beam current is proposed. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

At present, large-diameter electron beams are widely used for annealing metals and semiconductors [1] and for pumping electroionization lasers. They also find wide application in radiation and electron-beam technologies, plasma chemistry, and other fields [2]. In most cases, when generating large-diameter electron beams, it is expedient to use plasma electron sources (PESs), because they are capable of producing electron beams with a wide range of parameters. The main requirements for these sources are the following: the production of a nonuniform plasma with an increased density in the emission region, a low flux of working-gas neutrals into the acceleration gap, efficient control over the electron-beam current, minimum power expenditure for plasma production, and a fairly high operational resource. PESs based on reflex discharges with cold cathodes meet these requirements in many respects. Sources of this kind can efficiently operate with chemically active working gases under conditions of strong ion bombardment and periodic contacts with the atmosphere, guarantee stable source operation at small discharge currents, operate at low gas flow rates, have a simple design, and are reliable in operation.

In most of the existing PESs based on reflex discharges with cold cathodes, electrons are extracted from the discharge plasma by applying a high voltage between the plasma and the extracting electrode. In these sources, electrons are emitted from the discharge into the accelerating gap through a small aperture, which hampers the use of these sources for the production of large-diameter beams. However, at certain parameters of a discharge of this kind, the ejection of "anomalous" electrons in the axial direction can occur when the negative space charge is dominant in the discharge gap [3, 4]. In this case, electrons can be emitted without using an extracting electrode only due to energy they acquire when interacting with RF oscillations excited in the anode layer. In the absence of an extracting electrode, it is possible to extract low-energy electron beams and increase the emission aperture to the anode diameter. However, such regimes of a reflective discharge still remain poorly investigated.

In this paper, we present the results of experimental studies of the influence of the external discharge parameters on the generation of low-energy electron beams from a wide-aperture plasma electron source based on an electrically asymmetric reflex discharge.

EXPERIMENTAL TECHNIQUE

The experiments were carried out with two configurations of Penning electrodes (Fig. 1), which were imbedded in a uniform magnetic field directed along the system axis. The first electrode configuration (Fig. 1a) consisted of a cylindrical anode and two reflex electrodes, one of which was plane (*1*) and the second was cylindrical (*3*). In the second configuration (Fig. 1b), all the electrodes were cylindrical. The cylindrical electrodes were made from stainless steel, whereas the plane electrode was made from aluminum.

In experiments, we used discharge electrodes with diameters d = 18, 30, and 80 mm. The following condition was always satisfied: the distance between the cathode and anode and the length of the cylindrical cathode were equal to the electrode radius, and the anode length was equal to the electrode diameter. The experiments were carried out with steady-state discharges. The discharge electrodes were supplied through either an electrically symmetric or electrically asymmetric circuit. In the former case, cathodes *1* and *3* were grounded, whereas anode *2* was at the positive potential. In the latter case, cathode *3* was grounded, and the voltage ΔU between cathodes *1* and *3* was varied in the range from –500 to +500 V.

1063-7842/03/4801-0080$24.00

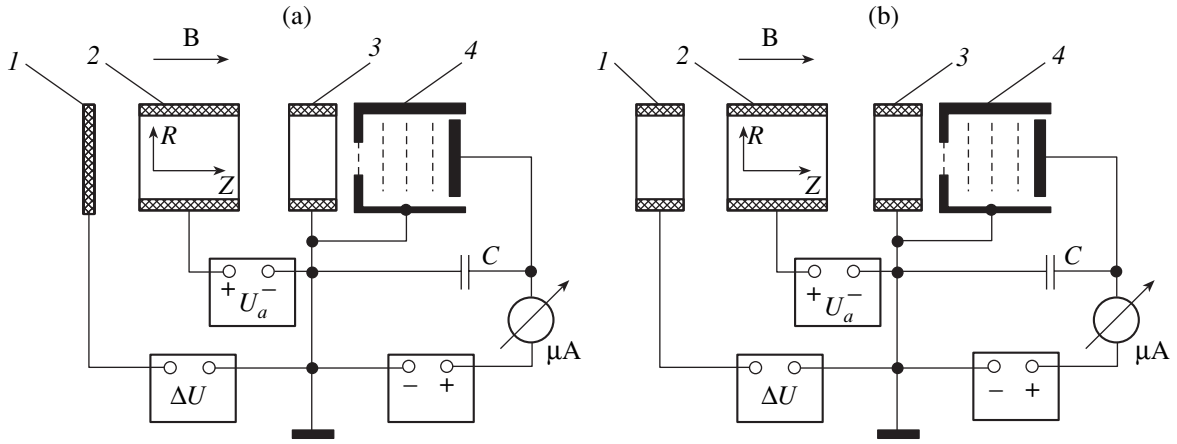

Fig. 1. Schematic of an electron plasma source: (a) electrode configuration with a plane cathode and (b) electrode configuration with cylindrical cathodes.

The working gas was hydrogen or air at a pressure of $P = 0.01–1 \times 10^{-3}$ torr, the magnetic field was $B = 0.01–0.1$ T, the anode voltage was $U_a = 0.5–3.5$ kV, and the discharge current was $I_p = 0.1–100$ mA. The plasma density $n_e \sim (0.1–2) \times 10^{10}$ cm^{-3} and the electron temperature $T_e = 5–20$ eV were measured by probes. The residual pressure in the vacuum chamber was no higher than 4×10^{-6} torr.

The electron beam was extracted from the side of cathode *3* in the axial direction. The radial current-density profile of the ejected electron beam and the longitudinal energy of the beam electrons were measured with the help of a movable multigrid electrostatic energy analyzer *4*.

Wave processes in the plasma were investigated with the use of single shielded cylindrical probes at the floating potential; the load was a capacitive divider with a high impedance in this frequency range.

EXPERIMENTAL RESULTS

The existence of different discharge regimes significantly affects not only the reflex discharge parameters, but also the properties of the plasma that emits electrons [2]. The positions of the boundaries separating the discharge regimes depend on the anode voltage, the magnitude of the external magnetic field, the sort and pressure of the working gas, and the geometric parameters of discharge [5]. For this reason, we experimentally studied the interrelation between different discharge regimes, the discharge geometry, and the regimes of electron-beam generation. The domains of the external discharge parameters (the anode voltage U_a and the working gas pressure P) corresponding to these regimes are shown in Fig. 2. In all cases, the generation of electron beams was observed only in the high-voltage discharge regime. When we used a lighter working gas (hydrogen instead of air) or decreased the diameter of the discharge electrodes, the boundary of the regime of electron-beam generation extended toward higher working-gas pressures and anode voltages. Similar dependences were also observed when the reflex electrodes were supplied through an electrically asymmetric circuit at negative values of ΔU. For electrode configurations with a plane electrode and an electrode diameter of 18 mm, the generation of electron beams was observed for both sorts of the working gas in the entire domain of the high-voltage discharge regime (Figs. 2e, 2f).

The studies of radial potential and plasma-density profiles in the regime with electron-beam generation showed that this regime is characterized by a sharp decrease in the plasma potential (Fig. 3a, curves *1–3*) and a simultaneous increase in the plasma density (Fig. 3b, curves *1–3*) in the central region of the discharge as compared to the regime without electron-beam generation (Fig. 3, curves *4*, *5*). When the reflex electrodes were supplied through an asymmetric circuit ($\Delta U = –500$ V), the plasma potential was lower (Fig. 3a, curve *2*) and the plasma density was higher (Fig. 3b, curve *2*) than in the case with a symmetric circuit. When air was used as a working gas, the plasma potential was higher (Fig. 3a, curve *3*) and the plasma density was lower (Fig. 3b, curve *3*) than in the case with hydrogen. In the regime with electron-beam generation, intense high-frequency drift oscillations in the frequency range $f \sim 10–55$ MHz were observed in the anode sheath. The oscillation frequency depended on both the external discharge parameters and the sort of the working gas. The smaller the diameter of the discharge electrodes or the molecular mass of the working gas, the lower the oscillation frequency.

For all of the electrode systems, the ejected electrode beam was radially nonuniform. The beam current density was maximum at a certain distance from the system axis (Fig. 4). The position and intensity of this maximum depended on the magnitude of the external

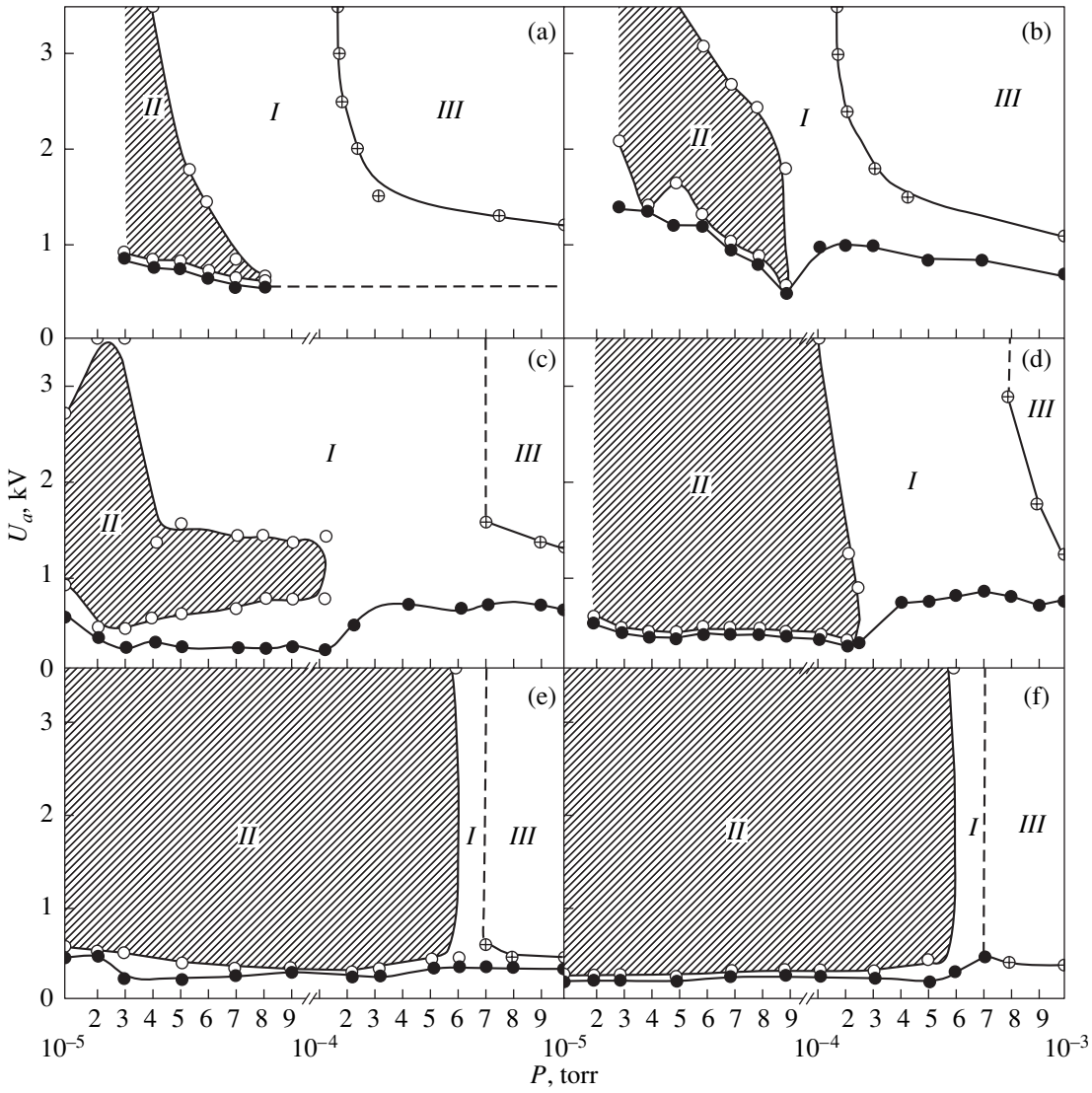

Fig. 2. Regimes of discharge operation and electron-beam generation: (*I*) high-voltage regime, (*II*) electron-beam generation regime, and (*III*) high-current regime. Shown are the boundary of the domain of the existence of a discharge (●), the boundary of the electron-beam generation regime (○), and the boundary of the high-current discharge regime (⊗) for (a, b) cylindrical electrodes 80 mm in diameter, (c, d) cylindrical electrodes 18 mm in diameter, and (e, f) an electrode configuration with a plane cathode 18 mm in diameter. (a, c, e) The working gas is air, (b, d, f) the working gas is hydrogen, (a, b, d, e, f) $B = 0.06$ T and $\Delta U = 0$, and (c) $B = 0.06$ T and $\Delta U = -500$ V.

magnetic field H, the anode voltage U_a, the voltage between the cathodes ΔU, the pressure P, the sort of the working gas, and the shape and diameters of the discharge electrodes. The electron beam was closer to uniform when we used plane cathodes or electrodes with smaller diameters (Fig. 4b). The electron-beam current density increased and the maximum shifted toward the system axis when the lighter working gas (hydrogen) (Fig. 4b, curves *1*, *2*) was used instead of air (Fig. 4b, curve *3*). The use of an electrically asymmetric circuit ($\Delta U = -500$ V) allowed as to obtain even higher electron-beam current densities. In this case, at optimum values of the external discharge parameters, an electron-beam current density of $J = 30$–40 mA/cm^2 was attained. The energy of the beam electrons was $E = 5$–10 eV, the energy efficiency of the source was $h = 0.3$–0.6 mA/W, and the electron extraction efficiency was $\alpha = 70$–90%.

We note that the plasma potential in the central region of the discharge decreases with increasing electron-beam current density. This circumstance evidences that the ejection (or more likely efflux) of electrons in the axial direction is also of electrostatic nature.

In all cases, the electron-beam current was modulated by high-frequency drift oscillations.

The most probable electron energy in the beam was 5–80 eV, depending on the geometry of the discharge electrodes and the sort of the working gas. This energy

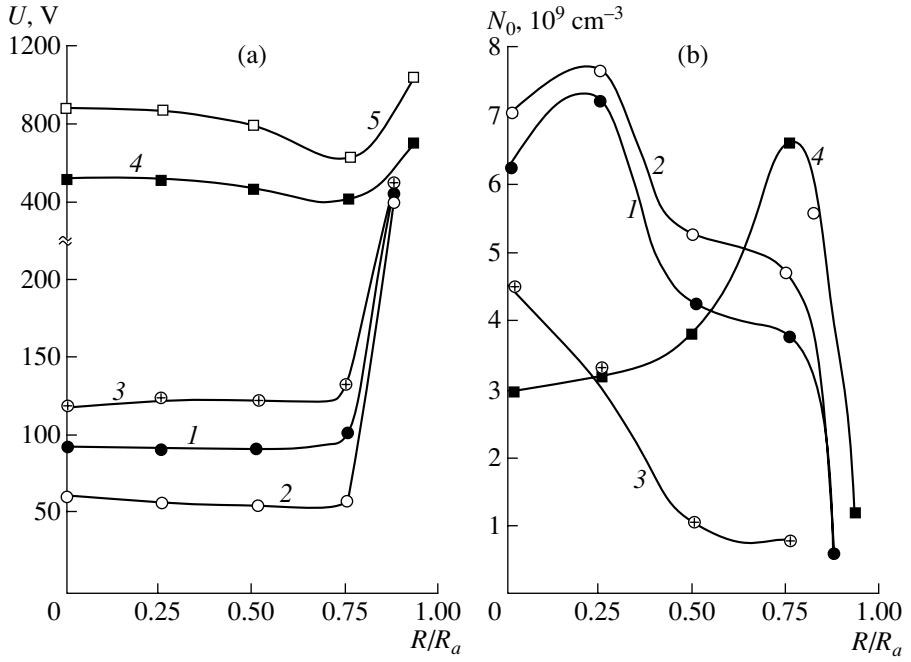

Fig. 3. Radial profiles of (a) the potential and (b) plasma density on the side of cathode *3* for cylindrical electrodes 80 mm in diameter: (*1*) the working gas is hydrogen, $P = 9 \times 10^{-5}$ torr, $B = 0.06$ T, $U_a = 1.5$ kV, and $\Delta U = 0$; (*2*) the working gas is hydrogen, $P = 9 \times 10^{-5}$ torr, $B = 0.06$ T, $U_a = 1.5$ kV, and $\Delta U = -500$ V; (*3*) the working gas is air, $P = 4 \times 10^{-5}$ torr, $B = 0.06$ T, $U_a = 1.5$ kV, and $\Delta U = 0$; (*4*) the working gas is hydrogen, $P = 10^{-4}$ torr, $B = 0.06$ T, $U_a = 1.5$ kV, and $\Delta U = 0$; and (*5*) the working gas is hydrogen, $P = 9 \times 10^{-5}$ torr, $B = 0.023$ T, $U_a = 1.5$ kV, $\Delta U = 0$.

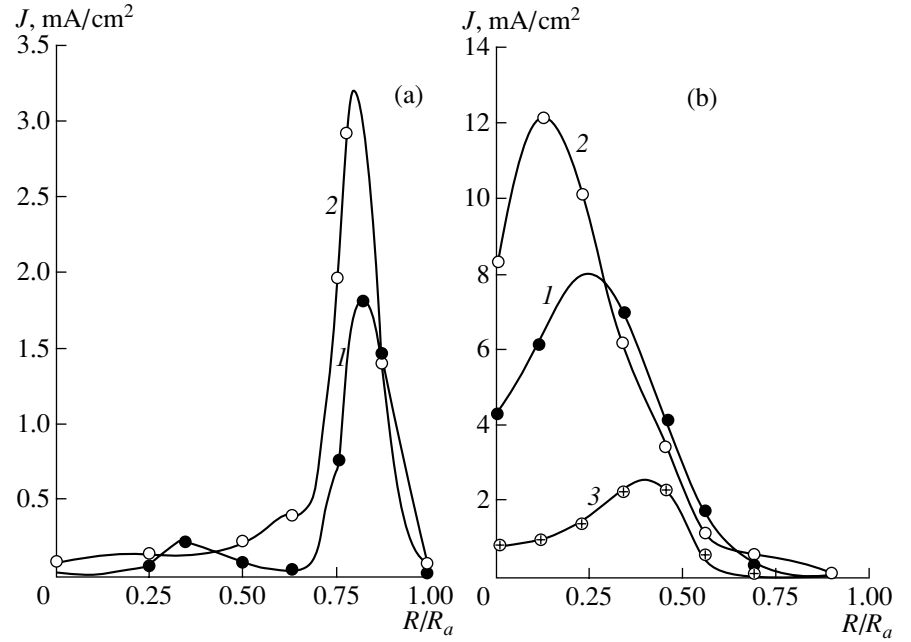

Fig. 4. Radial profiles of the electron-beam current density behind cathode *3* (see Fig. 1) for (a) cylindrical electrodes 80 mm in diameter, $P = 9 \times 10^{-5}$ torr, $B = 0.06$ T, and $U_a = 1.5$ kV and (b) an electrode configuration with a plane cathode 18 mm in diameter, $P = 9 \times 10^{-5}$ torr, $B = 0.08$ T, and $U_a = 1.5$ kV at $\Delta U =$ (*1*, *3*) 0 and (*2*) –500. (*1*, *2*) The working gas is hydrogen, and (*3*) the working gas is air.

was lower for hydrogen (5–10 eV) than for air (40–80 eV). When plane cathodes were used, electrons with energies corresponding to the potential difference between the cathodes ΔU were observed in the energy spectra. Their fraction in the energy spectrum decreased with decreasing electrode diameter.

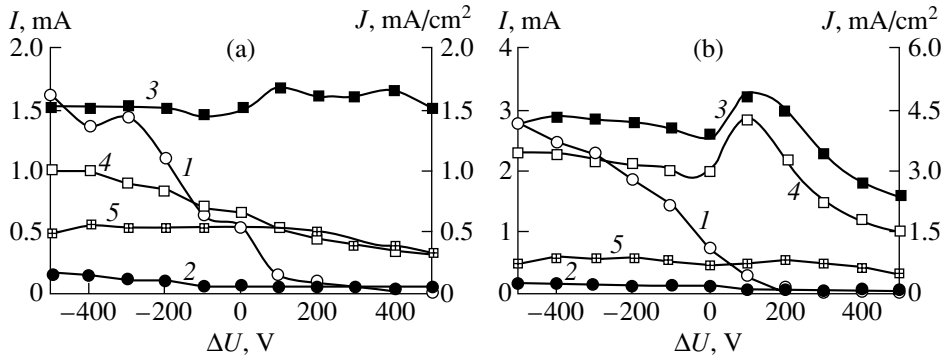

Fig. 5. Current densities of (*1*) the electron and (*2*) ion beams, (*3*) the discharge current, (*4*) the current at cathode *1*, and (*5*) the current at cathode *3* as functions of ΔU at $P = 9 \times 10^{-5}$ torr, $U_a = 1.5$ kV, and $B = 0.08$ T for (a) cylindrical electrodes 18 mm in diameter and (b) a plane and a cylindrical electrode 18 mm in diameter. The working gas is air.

We also investigated how the asymmetry of the electric circuit affected the current densities of the electrons and ions ejected simultaneously from the discharge. As can be seen from Fig. 5, the electron-current density in the range of negative ΔU values substantially exceeds the ion-current density. Varying the value of ΔU, we could efficiently control the electron-beam current. In this case, the discharge current for systems with cylindrical electrodes depended only slightly on the voltage ΔU between the cathodes (Fig. 5a), whereas for a configuration with plane electrodes, this dependence was observed in the range $\Delta U = -100$ to 500 V (Fig. 5b). Such behavior was typical of all the electrode systems and all the working gases.

As can be seen from Fig. 5, at certain positive values of ΔU, we observed the current compensation of the ion beam ejected simultaneously with the electron beam. The position of the current compensation region in the longitudinal direction depends on the value of ΔU. The larger the value of ΔU, the larger the distance between the compensation region and the emission aperture.

CONCLUSION

Regimes of the generation of electron beams from Penning plasma sources have been investigated for different geometries of the discharge electrodes. The influence of geometric factors (such as the shape and diameters of the discharge electrodes), the sort of the working gas, and the asymmetry of the electric circuit on the boundaries of the regime of electron-beam generation has been examined. It has been shown that electron-beam generation is observed only in the high-voltage discharge regime. The electron-beam generation regime is characterized by lower plasma potential and higher plasma density in the central region of the discharge as compared to the regimes without electron-beam generation. As the diameters of the discharge electrodes decrease or when a lighter working gas is used, the boundaries of the electron-beam generation regime extend to higher working gas pressures and higher anode voltages. It has been proposed that the electron-beam current be controlled by varying the voltage between the cathodes. Electron beams with current densities of $J = 30$–40 mA/cm^2 and electron energies of 5–10 eV have been obtained. An energy efficiency of the source of $h = 0.3$–0.6 mA/W and an electron extraction efficiency of $\alpha = 70$–90% have been achieved.

REFERENCES

1. Yu. E. Krendel', N. I. Lebedeva, V. Ya. Martens, *et al.*, Pis'ma Zh. Tekh. Fiz. **8**, 1465 (1982) [Sov. Tech. Phys. Lett. **8**, 630 (1982)].
2. Yu. E. Krendel', *Plasma Sources of Electrons* (Atomizdat, Moscow, 1977).
3. É. M. Barzudarov, N. A. Kervalishvili, and V. P. Kortkhondzhiya, Zh. Tekh. Fiz. **42**, 1904 (1972) [Sov. Phys. Tech. Phys. **17**, 1526 (1972)].
4. N. A. Kervalishvili, Zh. Tekh. Fiz. **60** (2), 78 (1990) [Sov. Phys. Tech. Phys. **35**, 182 (1990)].
5. É. M. Reĭkhrudel', G. V. Smirnitskaya, and G. A. Egizaryan, Zh. Tekh. Fiz. **43**, 130 (1973) [Sov. Phys. Tech. Phys. **18**, 83 (1973)].

Translated by N. Larionova

Technical Physics, Vol. 48, No. 1, 2003, pp. 85–89. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 91–95.

ELECTRON AND ION BEAMS, ACCELERATORS

Space Charge Oscillation in Electron Cooling Devices

S. G. Konstantinov, V. V. Parkhomchuk, and V. B. Reva

Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 11, Novosibirsk, 630090 Russia

e-mail: reva@csd.inp.nsk.su

Received April 4, 2002

Abstract—Electron cooling is used for improving the parameters of ion beams. The cooling efficiency depends drastically on the space charge fluctuation intensity in the beam. The fluctuations present in the cooling region cause the stochastic heating of the ions, which adversely affects the cooling efficiency and may even annihilate the ion beam. The space charge fluctuation intensity as a function of various operating parameters of a cooler is studied experimentally. A mechanism of fluctuation generation is suggested, and the effect of fluctuations on the ion beam parameters is estimated. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Electron cooling is used to improve the parameters of ion beams in a storage ring [1]. In this method, a cool electron beam and ion beam copropagate in a rectilinear section of the trajectory, the electron velocity being equal to the mean velocity of the ions. Because of electron–ion collisions, the energy of the chaotic motion of the ions is lost. A relatively short cooling time is provided if the electron beam is sufficiently intense (with currents of several amperes and voltages from several keV [2] to several MeV [3]). As a rule, such beams are formed in energy-recuperation systems, where the beam is first accelerated to the maximal energy and then the electron energy is given up to the source in the decelerating structure. The slowing-down is accomplished to the least possible value (1–5 keV), which depends on the space charge of the beam. The efficiency of electron beam capture by the collector is usually high; however, a small fraction of the electron beam may reflect from the collector to the cooling region. In this case, the space charge field may break into oscillation with the subsequent heating of the ion beam. If the oscillation (noise) is sufficiently intense, the electron beam may cause heating (instead of cooling) and the loss of a number of ions. In this work, we study space charge oscillation generated in a setup with a rectilinear magnetic field. The experiment was carried out on the bench for testing the gun and collector that are used in electron cooling equipment in the Institute of Modern Physics (Lanzhou, China).

EXPERIMENTAL SETUP AND MEASURING TECHNIQUE

The setup is schematically shown in Fig. 1. An electron beam is formed by an electron gun (*1–4*) [4] placed in a magnetic field of 700–1000 G and is transported along the field to the electron collector (*4–6*). The diameter of the beam was varied between 2 and 3 cm.

The noise characteristics were measured with two schemes. One of them is passive. Noise due to the beam space charge is measured by a spectrum analyzer or rf oscilloscope through a special loop that is placed near the beam and has a capacitance of 1.5 pF relative to the beam. The other scheme is active. Oscillation in the beam is initiated by applying an ac voltage to the control electrode (the signal and beam copropagate) or to the near-collector electrode (the signal and beam counterpropagate) with the subsequent recording of the signal from the collector and frame, which was used as a pickup.

Figure 2 shows the electrical potential profile with regard for the space charge with U_{coll} = 3 kV, U_{sup} = 0.5 kV, U_{cath} = 7 kV, U_{an} = 3.5 kV, U_{cont} = +0.6 V, and J_e = 900 mA. The cathode voltage U_{cath} is measured relative to the ground potential; the collector (U_{coll}) and anode (U_{an}) potentials, relative to the cathode potential. The suppressor potential U_{sup} is set by a voltage divider

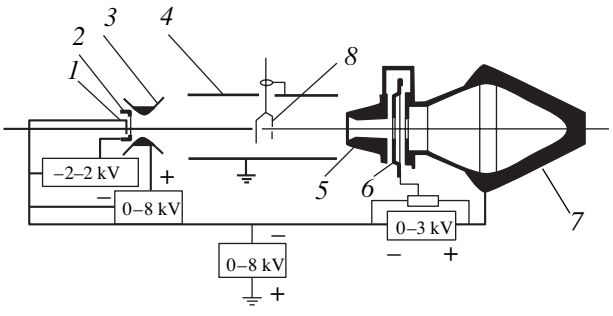

Fig. 1. Schematic representation of the setup for studying space charge oscillation. *1*, cathode; *2*, control electrode; *3*, anode; *4*, vacuum chamber; *5*, near-collector electrode; *6*, suppressor; *7*, collector; and *8*, probe.

1063-7842/03/4801-0085$24.00 © 2003 MAIK "Nauka/Interperiodica"

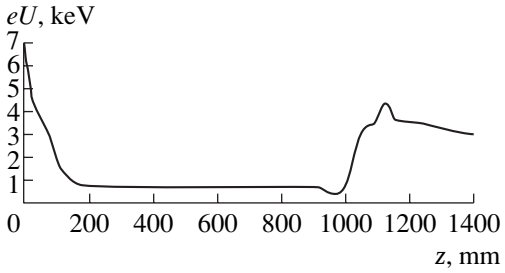

Fig. 2. Longitudinal profile of the potential along the setup axis with due account for the influence of the space charge.

inserted between the collector and cathode. The position 0 (1) of the divider corresponds to the case when the suppressor voltage equals the cathode (collector) voltage. The simulation was performed with the SAM software suite.

The potential profile in Fig. 2 makes possible the accumulation of secondary electrons in the collector–anode spacing. They may fall into this spacing by reflecting from the collector or as a result of residual gas ionization. Elastically or inelastically scattered electrons fall into the central region, where they accumulate. As the space charge exceeds a certain (critical) value, instability develops and the secondary electrons diffuse.

INSTABILITY

The experimental results on instability development are as follows.

(1) The "explosive" character of instability development. Figure 3 shows the instability waveform recorded in the open-collector mode, where $U_{coll} = U_{sup}$. The signal was applied to the high-resistivity input of an oscilloscope series-connected to a 1000-pF capacitor. A burst of instability 20 μs wide (between 70 and 95 μs, curve *2*) is seen. Curve *1* depicts the signal arising when a current pulse of amplitude 1 A and duration 280 μs passes through the loop. At the end of the burst (t = 100 μs, curve *2*), the signal decreases by 30% relative to the initial level (t = 50 μs, curve *2*) of the basic beam (curve *1*). This indicates that, during the burst of instability, some of secondary electrons bearing a charge of up to 30% of the basic beam charge were removed from the central region. The lower part of Fig. 3 demonstrates the development of instability when the signal was picked up from a load resistance of 50 Ω. In this case, the signal varies in proportion to dn/dt; therefore, charge accumulation is not observed but oscillation due to instability is clearly seen. The instability rise time is ≈5 μs, and the quasi-stationary range, during which the excessive space charge is removed from the central region, lasts ≈10 μs. The fundamental frequency is 70–100 mHz.

The burst mean repetition rate is 100–1000 Hz; hence, the mean current of the accumulated charge in this mode varies within 30–300 nA.

(2) Beam current threshold of instability. Figure 4 shows the space charge oscillation spectra vs. beam current at a suppressor voltage of –600 V relative to the collector. The spectra were taken with an SK4-59 spectrum analyzer.

It is seen that natural oscillations in the beam are absent up to 300 mA; at higher currents, instability develops very rapidly. When the current exceeds the threshold insignificantly, individual peaks of harmonics

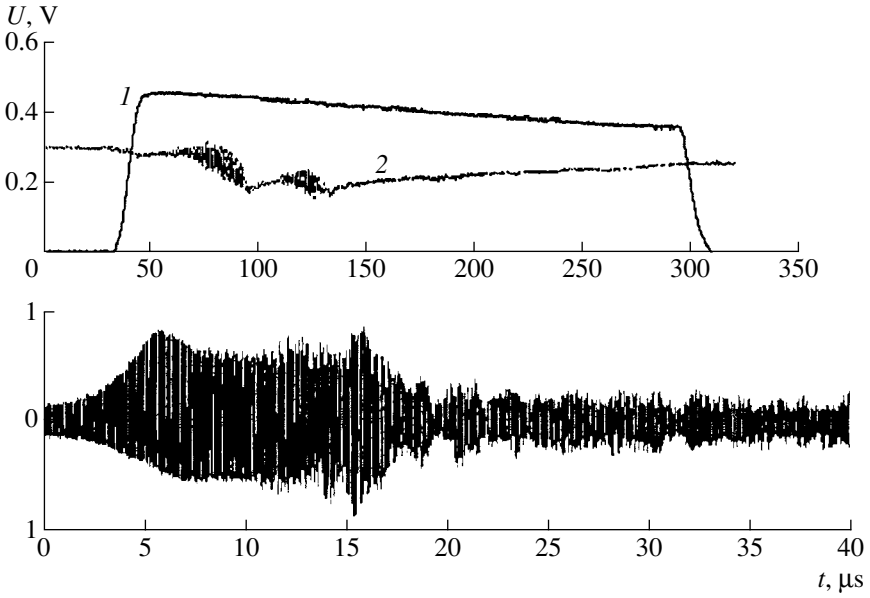

Fig. 3. Variation of the space charge n and dn/dt near the loop when the instability develops.

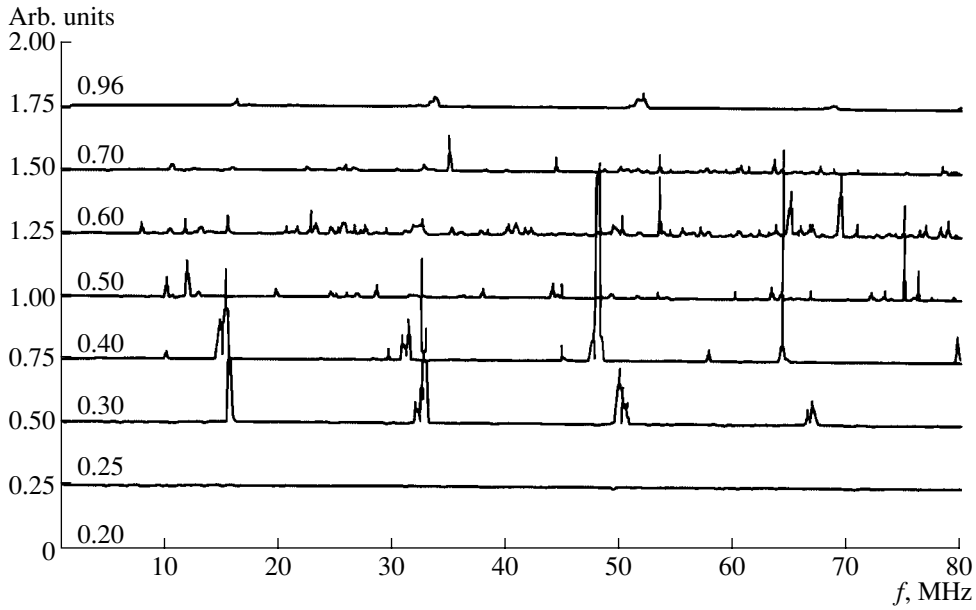

Fig. 4. Oscillation intensity vs. beam current. On the vertical axis is plotted the oscillation amplitude. Figures by the curves are current values in amperes.

are distinct. Subsequently, the number of peaks increases substantially and the spectrum becomes noisy. As the current grows further, both the amplitude and number of the peaks decrease and eventually the instability virtually disappears. The change in the magnetic field from 700 to 1050 G changes the instability threshold from 270 to 420 mA, without affecting the threshold behavior.

(3) The effect of the collector parameters on instability. The instability is most sensitive to the suppressor voltage. When this voltage equals the collector voltage (open-collector regime), the oscillation intensity is the highest and exceeds the intensity observed in other regimes by more than ten times. However, a low (about 30 V) electron-blocking voltage applied to the suppressor radically improves the situation. A further rise in the blocking voltage (up to $U_{cath} = U_{sup}$) has practically no influence on the instability. In the open-collector regime, the oscillation cannot be suppressed by decreasing the beam current either. Space charge oscillation in the beam is observed even at a current of 50 mA. Moreover, the situation does not change when the magnetic field, cathode voltage, and collector voltage vary. The open-collector regime was found to be the most noisy, all other things being equal.

The secondary electron flow can be affected by appropriately selecting the collector operating mode (Fig. 5). By decreasing the collector voltage to the level close to the virtual cathode formation, one can block most of the reflected fast electrons because of a high space charge.

(4) Induced oscillations. Intriguing data were obtained upon observing induced oscillations in the near-threshold range of instability. The exciting field was applied to near-collector electrode *5* (Fig. 1). The induced signal was picked up from the probe. We recorded the propagation of the external signal in the direction opposite to that of the basic beam. In the frequency interval 75–100 MHz (Fig. 6), the response due to secondary electrons from the collector is seen. In the low-frequency range near the threshold, the induced excitation of the natural oscillations of a secondary electron column is observed. In the vicinity of the threshold, the amplitude of the peaks grows.

CHARGE ACCUMULATION MECHANISM

Thus, it can be supposed that true secondary electrons with an energy inside the collector of ≈10 eV play a dominant role in space charge oscillation in the open-collector mode. Having been born in the collector, they pass over the collector surface, fall into the central region, and there accumulate. When the density of secondary electrons exceeds a certain threshold, instability arises and the secondary electrons are ejected from the central region. A low negative (relative to the collector) voltage applied to the suppressor blocks most of the secondary electrons. Owing to the electrode geometry, it is precisely near the processor the potential changes most significantly. Its effect on the beam center is much weaker. A further rise in the blocking voltage applied to the suppressor does not influence the flow of true secondary electrons, and elastically and inelastically reflected fast (>50 eV) electrons make a major contribution to the particle flow from the collector.

The contribution of electrons due to ionization seems to be weak (it is estimated at ≈50 nA in our experiment). A twofold variation of the pressure in the vacuum chamber changes the oscillation intensity and

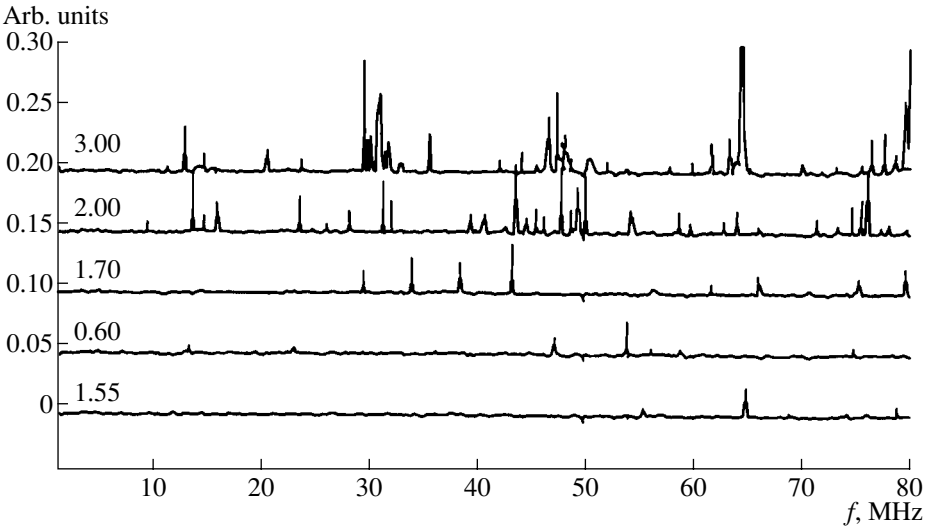

Fig. 5. Oscillation intensity vs. collector operating mode. On the vertical axis is plotted the oscillation amplitude. Figures by the curves are voltage values in kV.

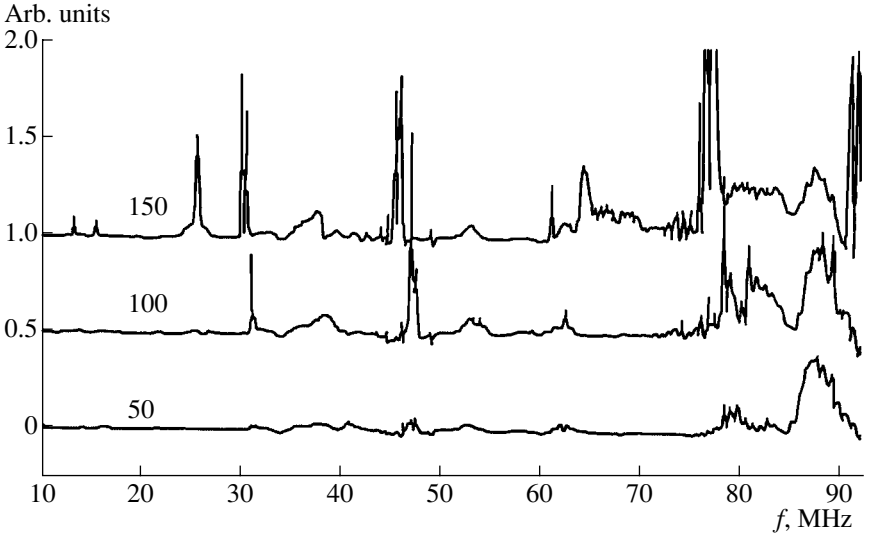

Fig. 6. Variation of the induced oscillation spectrum vs. beam current. On the vertical axis is plotted the oscillation amplitude. Figures by the curves are current values in mA.

spectrum to a minor extent. The possibility of decreasing the oscillation intensity by appropriately choosing the collector voltage also indicates that the collector is the basic source of electrons falling into the central region. An increase in the space charge density in the collector influences the reflected electron flow but does not affect the flow of ionization electrons.

As an instability model, the instability of trapped particles in the beam can be considered. When the frequency of the space charge wave in the beam, $\omega = kv_0$, approaches the natural frequency of trapped electrons (i.e., $\omega = n\omega_0$), growing oscillations may appear. Here, $\omega_0 = 4\pi L/v_0$, L is the size of the electrostatic well, v_0 is the oscillation frequency of trapped electrons, and n is an arbitrary integer (a set of harmonics appears when the equation of steady motion for an electron moving between electrostatic mirrors is expanded in harmonic functions). Solving the dispersion relation for this model, one finds unstable solutions for given values of the wave vector. The resulting instability increment grows linearly with mode number n. Its typical value under our experimental conditions is $\mathrm{Im}\omega = 4 \times 10^{-5}\ \mathrm{s}^{-1}$ ($n = 1$).

COOLED APERTURE

The cooling time for an ion oscillating in the storage ring is given by [5]

$$\tau^{-1}(a) = \frac{2\pi(J_e/q)r_e r_i \mathrm{Ln}_c \eta}{\beta^4\gamma^5(\sqrt{a^2 + a_{\mathrm{eff}}^2}/\beta_\perp)^3}\begin{cases} 1/(\pi a_e^2) & a < a_e \\ 1/(\pi a^2) & a > a_e. \end{cases} \quad (1)$$

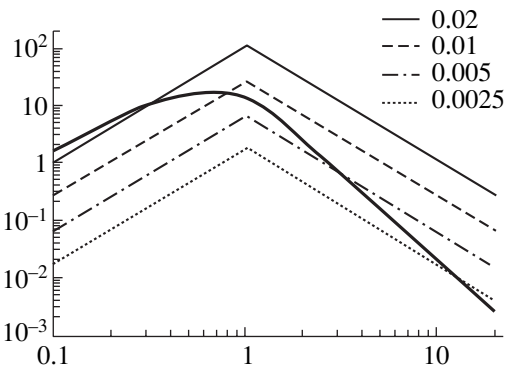

Fig. 7. Cooling and heating powers (vertical axis, cm^2/s) for electron current fluctuations $\Delta Je/Je$ = 2, 1, 0.5, and 0.25%. On the abscissa axis is plotted the ion oscillation amplitude (cantimeters). The thick continuous curve stands for cooling.

Here, J_e is the electron beam current, q is the charge of an electron, r_e and r_i are the respective classical electron and ion radii, Ln_c is the Coulomb logarithm, η is the ratio of the cooling region length to the ion path length, β and γ are beam parameters, a_{eff} is the effective amplitude (ions with a lesser amplitude cool down for the minimal time), a_e is the electron beam radius, $\beta_\perp$ is the value of the β function within the cooling region, and f_0 is the rotation frequency in the storage ring.

In the simplest model, where the electron beam noise is associated with current fluctuations (note that the fluctuations are random during each ion rotation), the rate of amplitude growth can be written in the form

$$\frac{da^2}{dt} = f_0\left[\frac{2\langle J_e\rangle r_i l_{\mathrm{cool}}\beta_\perp}{q\beta^3\gamma^3 c}\right]^2\begin{cases} a^2/a_e^4 & a < a_e \\ 1/a^2 & a > a_e, \end{cases} \qquad (2)$$

where $\langle J_e\rangle$ is the rms fluctuation of the electron beam current or the related change in the beam space charge.

From (1), it follows that the cooling time increases rapidly with the oscillation amplitude (outside the beam, it increases as a^5); hence, the cooling power ($a^2/\tau(a)$) drops more quickly than the power of beam-noise-induced heating. Starting from a certain critical amplitude, the cooling no longer compensates for the heating and the amplitude of the ions will grow infinitely until the ions are lost on the walls of the vacuum chamber. By way of example, Fig. 7 compares (on the logarithmic scale) the cooling and heating powers for the parameters of the cooler similar to that described in [4].

Figure 7 demonstrates that, even if electron beam fluctuations are as low as 2%, only ions with amplitudes less than 3 mm will be cooled; other ions will be heated and lost. Normal operation of the cooler is provided if the electron beam noise is no higher than 0.1%; in this case, the critical amplitude exceeds 10 cm.

The effect of space charge fluctuations on the cooling efficiency was estimated from the measured rms fluctuations of the space charge density. To use estimator (2), the rms fluctuation were converted to the equivalent fluctuation of the beam total current ($\delta Je = eV_0\delta nS$). For a suppressor value of 0.8, the relative current fluctuation was $\langle\delta Je\rangle/Je \approx 3\times10^{-4}$; for a suppressor value of 1.0 (open collector), $\langle\delta Je\rangle/Je \approx 2\times10^{-2}$. This corresponds to an ion beam lifetime of 10^4 s with the blocking voltage on the suppressor and 10 s in the case of an open collector. The latter value is clearly inappropriate for the operation of electron cooling equipment. Thus, under unfavorable conditions, space charge oscillation may cause high losses of the ion beam.

REFERENCES

1. G. I. Budker, At. Énerg. **22**, 246 (1967).
2. M. Steck, L. Groening, K. Blasche, *et al.*, Nucl. Instrum. Methods Phys. Res. A **441**, 175 (2000).
3. S. Nagaitsev, A. Burov, A. C. Crawford, *et al.*, Nucl. Instrum. Methods Phys. Res. A **441**, 241 (2000).
4. E. I. Antokhin, V. N. Bocharov, A. V. Bubley, *et al.*, Nucl. Instrum. Methods Phys. Res. A **441**, 87 (2000).
5. V. V. Parkhomchuk and A. N. Skrinskiĭ, Usp. Fiz. Nauk **170**, 473 (2000) [Phys. Usp. **43**, 433 (2000)].

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 1, 2003, pp. 90–95. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 96–101.

ELECTRON AND ION BEAMS, ACCELERATORS

Characteristics of a Hot Ion Collector Placed in a Weak Ion Cyclotron Resonance Magnetic Field

A. I. Karchevskiĭ and E. P. Potanin

Institute of Molecular Physics, Russian Research Centre Kurchatov Institute, pl. Akademika Kurchatova 1, Moscow, 123182 Russia

e-mail: potanin@imp.kiae.ru

Received April 15, 2002

Abstract—The separation characteristics of a planar collector system placed in the weak magnetic field region of a facility that separates isotopes of stable metals by the ion-cyclotron-resonance method are calculated. An increase in the gyroradius of the accelerated ions in the weak field allows the designer to increase the collector plate spacing and decrease undesired screen losses. It is theoretically shown that placing the collector in the week field increases the degree of separation. Also, the transverse-to-longitudinal ion energy conversion makes it possible to extend the product deposition area on the collector plates and, thus, to reduce the thermal load on the collector and facilitate the long-term production of the target material in large amounts. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

As is known, stable isotopes are presently produced by the electromagnetic method [1] or with the help of gas centrifuge cascades [2]. The latter are only capable of separating ions of elements whose gaseous compounds have a sufficiently high vapor pressure at room temperature. There are about 20 such elements in the periodic table. The electromagnetic method, though well developed, is only used to produce relatively small amounts of the target product when the element has no gaseous compounds under normal conditions. A high ion current in an electromagnetic separator is difficult to generate, because the ion beam density is limited by the repulsion of like-charged particles. The neutralization of the space charge by the residual gas has limitations of its own. In this respect, the plasma method, which is based on the selective ion cyclotron resonance (ICR) heating of ions of the target component [3–6], has advantages over the electromagnetic method in separating metal isotopes, because here limitations on the ion fluxes are virtually absent. The ICR method, which is today in the development stage, is expected to reduce energy consumption and improve significantly the yield of a single module.

SEPARATION CHARACTERISTICS OF THE ION COLLECTOR

At present, the ICR method of isotope separation in a plasma is the most promising for industrial use, specifically, for producing large amounts (tens of kilograms or more) of the elements that can now be separated only in electromagnetic separators. This method relies on the selective heating of target component ions in a weakly collisional plasma. An ICR device is schematically shown in Fig. 1. A steady plasma stream is directed to the region of a uniform magnetic field produced by solenoid *1*. This region contains RF antenna *2*, which generates a heating field. The heating region is followed by a product collection system, consisting of equispaced plates *3* (which collect hot ions *7* and are protected by screens *4*) and plate *5*, placed perpendicularly to the stream and intended for depositing cold ion component *6* of the plasma. A positive potential *U*, which repulses the cold ions, can be applied to the plates to improve the separation of the isotope mixture. Figure 2 illustrates a fragment of the hot fraction collector. The fragment shows a screen (its full width in the direction perpendicular to the plate is denoted 2*a)*, which protects the collector plate from the cold plasma stream. The condition for the selective heating of the target component is given by

$$\Delta\omega_{1/2} \ll \frac{\Delta m}{m}\omega_c, \qquad (1)$$

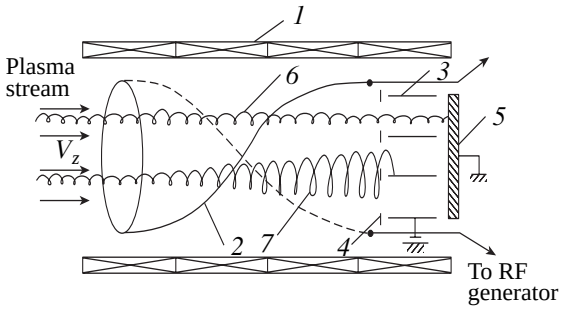

Fig. 1. ICR device.

1063-7842/03/4801-0090$24.00

where $\Delta\omega_{1/2}$ is the half-width of the resonant ion energy curve and ω_c is the cyclotron frequency, which varies in proportion to the intensity B_z of the longitudinal magnetic field.

It is clear from this formula that the mass resolution increases in high magnetic fields, which may decrease the gyroradius of the target ions below the critical value (even if their energy prevents material scattering at the collector) specified by the design and manufacturing requirements. It is known that the screen and plates must be cooled and cannot be too small or too thin. At the same time, the optimal plate spacing d, which equals about two gyroradii R_g of hot ions (Fig. 2), decreases with increasing magnetic field intensity. That is why the screen losses of the material grow at strong magnetic fields. Also, in high magnetic fields, the pitch L of the helical ion trajectory becomes small (as follows from Fig. 2, $L \approx V_z T_c$, where V_z is the average longitudinal ion velocity and T_c is the cyclotron gyration period, which is inversely proportional to the magnetic field intensity) and the associated decrease in the longitudinal size of the deposition area may cause local thermal overloads and rapid accumulation of the material on the front edge of the collector plate. This will adversely affect the collection process.

To reduce the above effects and thus improve the device efficiency, it has been proposed that the collector be placed near the end face of the ICR device, where the magnetic field is weaker [7, 8]. In this case, electromagnetic forces appearing in the intermediate region of the decaying field convert the transverse energy of the ions into longitudinal energy and the gyroradius of the particles increases, tending to conserve the adiabatic invariant. Because of this effect, the plate spacing and the screen size can be increased, thereby making it possible to produce cooled collector plates and front screens of optimal size.

Target ions, which are accelerated in the transverse direction, acquire a much higher energy in this process than the weakly heated component. The pitch of the helical trajectory of the hot ions becomes significantly larger than for the cold ion trajectory, causing the nontarget component to deposit largely on the front part of the collector plate and the hot component, over a much wider area. This additionally improves the longitudinal separation on the collector plate. Moreover, in this case, the material deposits more uniformly along the collector; as a result, the deposition cycle can last for a longer time without interrupting the separation process.

Consider a magnetic field varying in only two directions (refer to Fig. 2, where z is the direction along the plasma stream and the y coordinate is perpendicular to the plates). Let the collector flat plates be placed in the region of the weak uniform field immediately behind the region where the field declines. Unlike [7, 8], we will study the effect of the front screen and repulsive collector potential, as well as the axial variation of the plasma density. Since the magnetic field is assumed to be constant over the collector, we can use the early technique for simulating the separation of lithium isotopes [9–11] when the collector is placed as usual. In this technique, the isotope fluxes are calculated by integrating the flux density over the admissible range of transverse and longitudinal ion velocities ($V_\perp$, V_z) and over admissible coordinates of their guiding centers y_0 before the collector plate with regard to the known velocity distribution function $f(V_\perp, V_z)$ (for details, see [10, 11]):

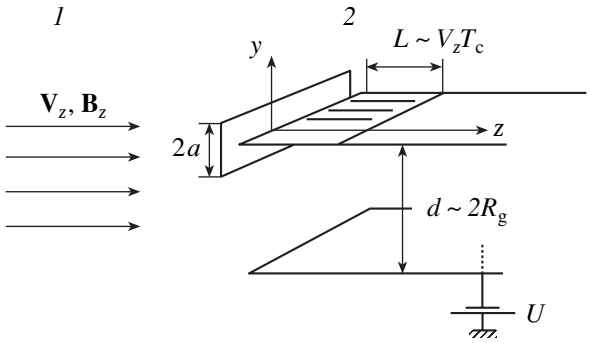

Fig. 2. Scaled-up fragment of the collector: (*1*) front screen and (*2*) deposition area (a is the screen height and R_g is the gyroradius of hot ions).

$$j = \int f \omega_c V_\perp dV_\perp dV_z dy_0. \quad (2)$$

We will apply such an approach to the case of separating a ^{6}Li–^{7}Li isotope mixture. Let the collector be placed in the region of the weak uniform field with a magnetic induction B_1. To create such a magnetic field configuration at the end of the magnetic system, special magnetic coils must be used. We introduce the field attenuation coefficient $R = B_0/B_1$, where B_0 is the magnetic induction in the heating zone. If this parameter is close to unity, the implementation of the magnetic system is rather easy. As the initial parameters of the ^{6}Li–^{7}Li isotope mixture, we take the transverse and longitudinal energies of ions of the sixth and seventh isotopes after passage through the region of selective acceleration of the target isotope ^{6}Li: $^6W_{\perp 0} = 40$ eV, $^6W_{\parallel 0} = 10$ eV, $^7W_{\perp 0} = 5$ eV, and $^7W_{\parallel 0} = 10$ eV [11].

Due to the conservation of the adiabatic invariant [12], the final ion energies can be calculated as

$$^6W_\perp = {}^6W_{\perp 0}/R, \quad {}^6W_\parallel = ({}^6W_{\perp 0} + {}^6W_{\parallel 0}) - {}^6W_\perp,$$

$$^7W_\perp = {}^7W_{\perp 0}/R, \quad {}^7W_\parallel = ({}^7W_{\perp 0} + {}^7W_{\parallel 0}) - {}^7W_\perp.$$

Let $R = 1.2$. Then, the ion energies in the region of the weak magnetic field are $^6W_\perp \cong 33$ eV, $^6W_\parallel \cong 17$ eV, $^7W_\perp \cong 4$ eV, and $^7W_\parallel \cong 11$ eV.

In what follows, these quantities will be referred to as the effective temperatures of the components.

First, let us evaluate the ion fluxes onto a collector plate and the concentration of the target isotope (^{6}Li) as a function of the longitudinal coordinate for the zero repulsive potential. Following [9], we ignore electrostatic field distortions caused in the plasma by ion collection. Let the ion transverse velocity distributions be Maxwellian with different effective temperatures $T_{\perp 6}$ and $T_{\perp 7}$ for the sixth and seventh lithium isotopes, respectively. Assume first that the longitudinal velocities are described by the shifted semi-Maxwellian distribution. In this case, there are no particles moving in the backward direction toward the plasma source and we obtain the following normalized complete distribution function for particles of any sort K:

$$F_K = 2n_K(m_K/2\pi kT_{\perp K})(m_K/2\pi kT_{\parallel K})^{1/2} \times \exp\left(-\frac{m_K}{2k}\left(\frac{V_\perp^2}{T_{\perp K}} + \frac{(V_z - V_{0K})^2}{T_{\parallel K}}\right)\right),$$

where $V_z \geq V_{0K}$, n_K is the particle density, m_K is the particle mass, and V_{0K} is the addition to the longitudinal velocity due to the transverse-to-longitudinal energy conversion.

The longitudinal velocity distribution function can then be written as

$$f_K(V_z) = 2\left(\frac{m_K}{2kT_{\parallel K}}\right)^{1/2} \exp\left(-\frac{m_K(V_z - V_{0K})^2}{2\pi kT_{\parallel K}}\right) \quad \text{for} \quad V_z \geq V_{0K},$$
$$f_K(V_z) = 0 \quad \text{for} \quad V_z < V_{0K}. \tag{3}$$

Let us define the concentration of the component of the first sort as the ratio of its density n_1 to the total density $(n_1 + n_2)$ of the mixture. Then, the concentration C of the material deposited on the collector can be determined from the relationship

$$\frac{C}{1-C} = \frac{C_f J_6}{(1-C_f)J_7}(T_{\perp 6}m_7/T_{\perp 7}m_6)^{1/2}, \tag{4}$$

where C_f is the concentration in the stream in front of the collector, $J_K = j_K/j_0$ are the dimensionless transverse ion fluxes toward the collector, and $j_0 = 2n_K(2kT_{\perp K}/m_K)^{1/2}\pi^{-3/2}$.

In the formula for j_0, n_K are the component densities in the region before the collector. They may differ from the corresponding values n_{K0} before the region of magnetic field attenuation. Let us take into account that the total longitudinal particle flux is constant along the axis of the system. The ion density behind the acceleration region must then decrease both because the cross section of the plasma filament increases (because of the decrease in the magnetic induction) and because the average longitudinal stream velocity grows. The variation in the ion density is then expressed as

$$n_6 = n_{60}\frac{1}{a_6 R}, \quad n_7 = n_{70}\frac{1}{a_7 R}, \tag{5}$$

where

$$a_6 = a_{60}(1 + (a_{60} - 1)\sqrt{\pi}),$$
$$a_7 = a_{70}(1 + (a_{70} - 1)\sqrt{\pi}),$$
$$a_{60} = \sqrt{\frac{T_{\parallel i6}}{T_{\parallel 60}}}; \quad a_{70} = \sqrt{\frac{T_{\parallel i7}}{T_{\parallel 70}}}.$$

The concentration C_f of the target isotope in the stream before the collector is

$$C_f = \frac{C_0}{C_0 + (1 - C_0)\dfrac{a_6}{a_7}},$$

where C_0 is the concentration in the starting product.

According to [10], the expression for J_K takes the form

$$J_K = \int_{2X_K^*}^{\infty} dX\left[\int_{A_1}^{A_2} 2a_K^* \cot(b_K z/X) D_K YT(X, Y)dY + \int_{A_2}^{\infty}(Y - a_K^* + a_K^* \cot(b_K z/X)D_K)YT(X, Y)dY\right] + \int_{X_K^*}^{2X_K^*} dX \int_{A_2}^{\infty}[Y - a_K^* + a_K^* \cot(b_K z/X)D_K]YT(X, Y)dY, \tag{6}$$

where

$$a_K^* = \frac{a\omega_{cK}}{2}(m_K/2kT_{\parallel K})^{1/2},$$
$$D_K = (Y^2 \sin^2(b_K z/X)/a_K^{*2} - 1)^{1/2},$$
$$A_1 = a_K^*/\sin(b_K z/X), \quad A_2 = a_K^*/\sin^2(b_K z/X),$$
$$X = V_z/(2kT_{\parallel K}/m_K)^{1/2},$$
$$X_K^* = (\omega_{cK} z/2000\pi)(m_K/2kT_{\parallel K})^{1/2},$$
$$Y = V_\perp/(2kT_{\perp K}/m_K)^{1/2},$$
$$b_K = X_K^*\pi/z, \quad T(X, Y) = \exp(-(X - X_{0K})^2 + Y^2);$$

a is the height of the front screen, ω_{cK} is the cyclotron frequency of ions of the Kth sort, z is the longitudinal coordinate (measured in millimeters), and $X_{0K} = a_{0K} - 1$.

Using formulas (3)–(6) and numerically integrating over all allowable velocities and coordinates of the particle's guiding center near the front screen, we arrive at the concentration of the target isotope as a function of the z coordinate (in millimeters). The coordinate is reckoned from the front edge of the collector plate. Figure 3 (curves *1–4*) plots the concentration C of the ^{6}Li isotope versus z coordinate for the front screen heights a = 0, 1, 2, and 3 mm, respectively. The dashed line shows the z dependence of the concentration at $R = B_0/B_1 = 1$ and $a = 0$. As follows from these results, with the collector placed in the region of the weak magnetic field, the degree of separation increases significantly. It is also seen that the ion energy conversion greatly extends the product deposition area on the collector plates. This must reduce the thermal load on the collector and decrease the growth rate of the thickness of the product deposited on the front part of the collector plate.

Let us evaluate the separation power per unit area of the collector plate [7]:

$$\delta U_0 = \tau_0 \frac{(\alpha - 1)}{\alpha + 1} \ln \alpha, \qquad (7)$$

where τ_0 is the flux per unit area of the collector plate and

$$\alpha = \frac{C/(1-C)}{C_0/(1-C_0)}$$

is the coefficient of mixture separation with respect to the initial composition.

Solid lines *1–4* in Fig. 4 plot the dimensionless separation power

$$\delta U_0/((n_{60} + n_{70})(2kT_{\perp 6}/m_6)^{1/2}/4\pi) \qquad (8)$$

versus z coordinate for the front screen height of a = (*1*) 0, (*2*) 1, (*3*) 2, and (*4*) 3 mm. The dashed lines show the separation power as a function of z at $R = B_0/B_1 = 1$ and $a = 0$. The significant increase in the separation power that is associated with the use of the weak field region is observed. Note that an increase in the height of the screen lowers the separation power for all z.

Now we evaluate the effect of the additional acceleration of ions when a positive repulsive potential U is applied to the collector plate. In this case, the expression for the flux J_K in the absence of the front screen takes the form

$$J_K = \int_{2X_k^*}^{\infty} dX \left[\int_{A_1}^{A_2} 2Y(Y^2 - U_K^*)^{1/2} T(X, Y) dY \right.$$

$$\left. + \int_{A_2}^{\infty} [(Y^2 - U_K^*)^{1/2} + Y\cos(\alpha_K/2)] YT(X, Y) dY \right] \qquad (9)$$

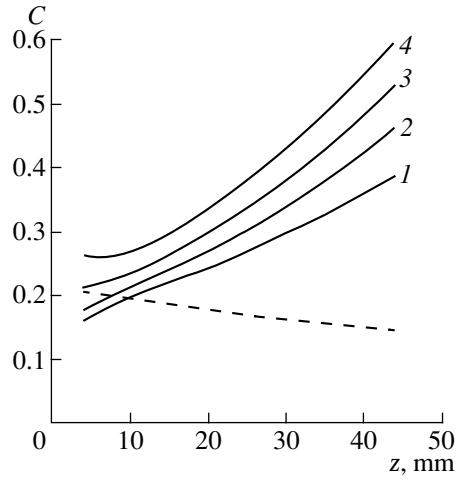

Fig. 3. Mole concentration $C(z)$ of ^{6}Li along the z coordinate for the shifted semi-Maxwellian distribution of the longitudinal velocity.

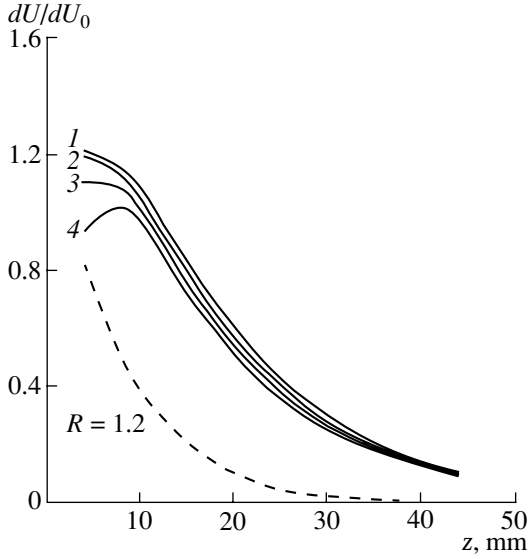

Fig. 4. Nondimensional separation power along the coordinate z for front screen half-heights a = (*1*) 0, (*2*) 1, (*3*) 2, and (*4*) 3 mm.

$$+ \int_{X_K^*}^{2X_K^*} dX \int_{A_2}^{\infty} [(Y^2 - U_K^*)^{1/2} + Y\cos(\alpha_K/2)] YT(X, Y) dY,$$

where $\alpha_K = b_K z/X$, $U_K^* = eU/kT_{\perp K}$, and A_1 and A_2, unlike formula (6), are defined as

$$A_1 = (U_K^*)^{1/2}, \quad A_2 = (U_K^*)^{1/2}/\sin(\alpha_K/2).$$

Figure 5 shows the concentration of the product deposited on the plate versus longitudinal coordinate for different repulsive potentials U. The degree of separation increases fairly rapidly with the repulsive

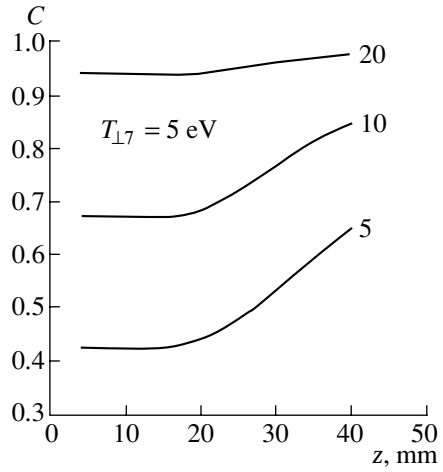

Fig. 5. Deposit concentration on the plate along the z coordinate at repulsive potentials U = 5, 10, and 20 V.

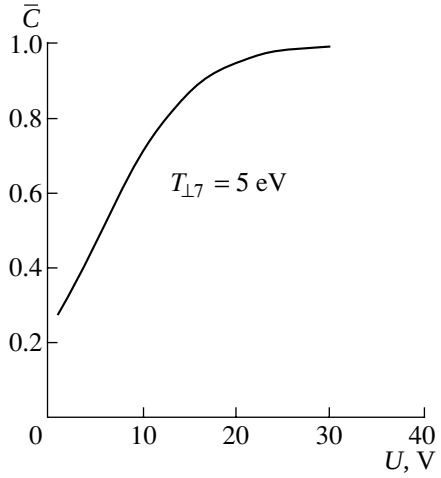

Fig. 6. Concentration averaged over a distance of 30 mm, $\bar{C}$, versus effective potential U.

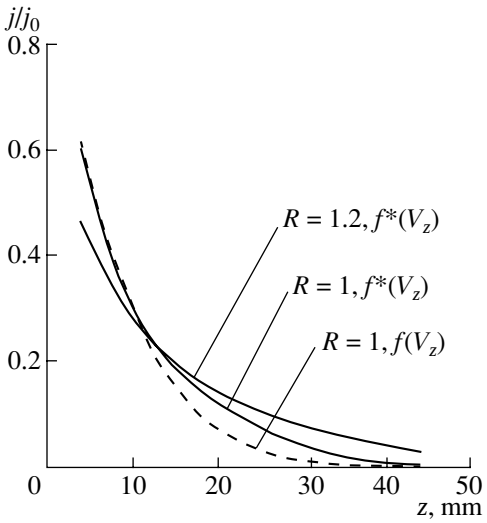

Fig. 7. Dimensionless flux of ^{6}Li ions, j/j_0, toward the collector plate versus longitudinal coordinate z for different transverse ion velocity distributions at R = 1 and 1.2 with a = 0.

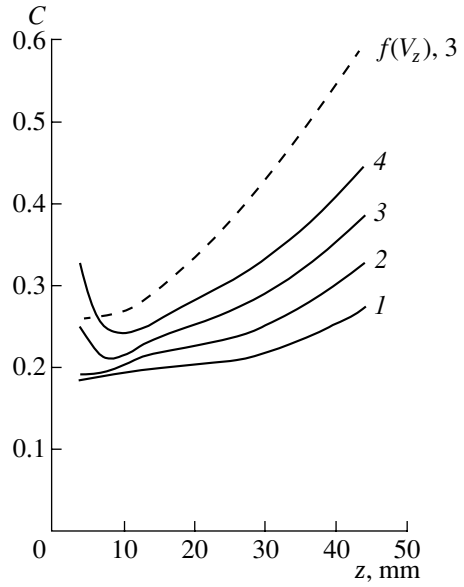

Fig. 8. Concentration of ^{6}Li versus coordinate z at a = (*1*) 0, (*2*) 1, (*3*) 2, and (*4*) 3 mm for the model longitudinal velocity distribution $f^*(V_z)$. The dashed line refers to the distribution function $f(V_z)$ given by (3).

potential and the longitudinal coordinate z. Figure 6 shows the concentration averaged over a distance of 30 mm, $\bar{C}$, versus repulsive potential U = 30 V.

Finally, let us evaluate the effect of the shape of the distribution function on the separation characteristics of the hot particle collector for the zero repulsive potential. Above, we assumed that the transverse velocity distribution is a shifted Maxwell function (Eq. (3)). Below, we assume that the transverse velocity distribution has the model form

$$f_K^*(V_z) = \frac{m}{kT_{\|K}} V_z \exp\left(-\left(\frac{mV_z^2}{2kT_{\|K}}\right)\right). \qquad (10)$$

A characteristic feature of this function is that, unlike function (3), it allows for the presence of particles with low longitudinal velocities. In this case, the expression for J_K becomes

$$J_K = \int_{2X_K^*}^{\infty} \sqrt{\pi} X dX \left[\int_{A_1}^{A_2} 2a_K^* \cot(b_K z/X) D_K YT(X, Y) dY \right.$$

$$\left. + \int_{A_2}^{\infty} (Y - a_K^* + a_K^* \cot(b_K z/X) D_K) YT(X, Y) dY \right]$$

$$+\int_{X_K^*}^{2X_K^*}\sqrt{\pi}XdX\int_{A_2}^{\infty}[Y-a_K^*+a_K^*\cot(b_Kz/X)D_K]YT(X,Y)dY.$$

Figure 7 plots the dimensionless flux of ^{6}Li ions onto the collector plates versus z coordinate for the field attenuation coefficient $R = 1$ and 1.2 without the screen (solid curves). The dashed line shows the same dependence for distribution function (3). First, as could be expected, the presence of the acceleration region extends the deposition area in the longitudinal direction. Second, distribution function (10), which assumes the presence of slow particles in the plasma stream, provides a more favorable distribution of the deposit along the collector plate length.

In Fig. 8, the solid lines depict the concentration C of the ^{6}Li isotope as a function of the longitudinal coordinate for different heights a of the front screen at $R = 1.2$. The dashed line shows the same dependence for distribution function (3) at $a = 3$ mm. Unlike distribution (3), at short distances z from the plate's edge, the concentration increases with the screen height. This is explained by the decrease in the transverse flux of the target component onto the collector plate near the screen in the case of model distribution (10).

CONCLUSION

The separation characteristics of the hot ion collector placed in the region of the weak magnetic field in an ICR device are calculated. It is shown that the longitudinal acceleration of transversely heated particles in the region of the nonuniform magnetic field must increase both the degree of separation and separation power and also expand the product deposition area on the collector plates. The latter effect reduces the thermal load on the collector and decreases the growth rate of the deposit thickness on the front part of the collector plate, thus extending the operating time without replacing the collector.

The concentration of the deposit on the plate is studied as a function of the longitudinal coordinate at different repulsive potentials. The use of a repulsive potential is efficient when it is necessary to obtain the high concentration of the target isotope. The effect of different longitudinal particle velocity distributions on the separation properties of the collector is estimated.

ACKNOWLEDGMENTS

The authors thank Yu.A. Muromkin for valuable remarks.

REFERENCES

1. R. N. Kuzmin, V. G. Bondarenco, and V. M. Tchesnocov, in *Synthesis and Applications of Isotopically Labelled Compounds* (Elesevier, Amsterdam, 2001), Vol. 7, p. 10.
2. E. I. Abbakumov, V. A. Bazhenov, Yu. V. Verbin, *et al.*, At. Énerg. **67**, 255 (1989).
3. G. A. Askar'yan, V. A. Namiot, and A. A. Rukhadze, Pis'ma Zh. Tekh. Fiz. **1**, 820 (1975) [Sov. Tech. Phys. Lett. **1**, 356 (1975)].
4. J. M. Dawson, H. C. Kim, D. Arnush, *et al.*, Phys. Rev. Lett. **37**, 1547 (1976).
5. A. C. La Fontaine and P. Louvet, *Compte rendu des Journees sur les Isotopes Stables* (Saclay, 1993), pp. 332–336.
6. A. I. Karchevskiĭ, V. S. Laz'ko, Yu. A. Muromkin, *et al.*, Preprint No. 5239/7, IAÉ (Institute of Atomic Energy, Moscow, 1990); A. I. Karchevskiĭ, V. S. Laz'ko, Yu. A. Muromkin, *et al.*, Fiz. Plazmy **19**, 411 (1993) [Plasma Phys. Rep. **19**, 214 (1993)].
7. A. I. Karchevskiĭ and E. P. Potanin, in *Proceedings of the 3rd All-Russia Scientific Conference "Physicochemical Processes in Selection of Atoms and Molecules," Zvenigorod, 1997*, p. 84.
8. A. I. Karchevskiĭ and E. P. Potanin, Fiz. Plazmy **25**, 301 (1999) [Plasma Phys. Rep. **25**, 273 (1999)].
9. A. L. Ustinov, Preprint No. 5354/6, IAÉ (Institute of Atomic Energy, Moscow, 1991).
10. A. I. Karchevskiĭ and E. P. Potanin, Fiz. Plazmy **21**, 416 (1995) [Plasma Phys. Rep. **21**, 394 (1995)].
11. A. I. Karchevskiĭ and E. P. Potanin, Fiz. Plazmy **22**, 1146 (1996) [Plasma Phys. Rep. **22**, 1044 (1996)].
12. K. D. Sinel'nikov and B. N. Rutkevich, *Lectures on Plasma Physics* (Kharkov Univ., Kharkov, 1964), p. 62.

Translated by A. Khzmalyan

Technical Physics, Vol. 48, No. 1, 2003, pp. 96–99. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 1, 2003, pp. 102–105.

ELECTRON AND ION BEAMS, ACCELERATORS

On the Design of Temporal Cerenkov Detectors That Have Optimal Algorithms for the Unique Reconstruction of Muon Tracks

V. S. Kinchakov

Computer Center, Far East Division, Russian Academy of Sciences, Khabarovsk, 680063 Russia
e-mail: kinchakov@as.fe.ru
Received April 12, 2002

Abstract—A general condition providing the unambiguous reconstruction of an azimuthal muon track, which is necessary in the design of both temporal and amplitude detectors, is found analytically. Two designs of temporal Cerenkov detectors that allow one (unlike conventional strings of detectors) to recover all the parameters of the muon track and admit the partition of the obtained set of equations into two independent sets of lower dimensions are proposed. Optimal algorithms for calculating the muon track parameters are found with this approach. One of the designs proposed allows for an analytical solution to the set of equations. *© 2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

As has been shown in [1], the use of time differences between Cerenkov radiation signals coming from different modules of a string of detectors improves the accuracy of determination of the muon track parameters by an order of magnitude compared with that provided by the amplitude analysis method [2, 3]. However, such a vertical string of detector modules does not allow one to reconstruct the muon track [1, 2]. The two designs of Cerenkov detectors that are proposed in this paper are free of this disadvantage. The more symmetric design (design A) consists of twelve modules placed in the vertices of four identical equilateral triangles that have parallel planes and sides and are spaced vertically at the same distance H. This design allows for an analytical solution to the set of equations for the track parameters obtained in this work. The less symmetric design B requires the use of numerical methods to solve this set of equations.

A general condition providing the unambiguous reconstruction of an azimuthal muon track that is vital for designing both temporal and amplitude detectors was found. It turned out that the projections of at least three detector modules onto the horizontal plane must not belong to the same straight line.

As in the case of a string [1], the problem of recovering the coordinates of a relativistic muon from its Cerenkov radiation with the detectors under consideration is solved under the following assumptions: in the region where the Cerenkov radiation is detected, the muon track is a straight line; the time resolution of the detectors is independent of the signal amplitudes; and the detector modules are considered as mathematical points.

A SET OF EQUATIONS FOR THE DETERMINATION OF THE MUON TRACK PARAMETERS

Let us place the coordinate origin at the center of the top triangle of the detector modules and direct the z axis vertically. A muon track can be described by four independent parameters: θ and ϕ are, respectively, the zenith and azimuth angles of the directing vector **a** of the track and x_1 and y_1, the two coordinates of the point of intersection between the muon track and the plane $z = 0$. Then, the shortest distance from the track to an ith detector is given by

$$l_i = [(z_i^0)^2(a_x^2 + a_y^2) + (x_1 - x_i^0)^2(a_y^2 + a_z^2) + (y_1 - y_i^0)(a_x^2 + a_z^2) + 2(y_1 - y_i^0)a_y \times (a_z z_i^0 - a_x(x_1 - x_i^0)) + 2a_x a_z(x_1 - x_i^0)z_i^0]^{1/2}. \quad (1)$$

Here, x_i^0, y_i^0, and z_i^0 are the Cartesian coordinates of the radius vector $\mathbf{r}_i^0$ specifying the position of an ith photoelectric multiplier (PEM). It is worth noting that expression (1) is nonlinear with respect to the desired parameters θ, ϕ, x_1, and y_1. Note also that the substitution $\phi \longrightarrow k\pi + \psi_i^0 \pm (\phi - \psi_i^0)$ does not change l_i (here, $k = 0$, 1 and ψ_i^0 is the polar angle of the vector with the Cartesian coordinates ($x_1 - x_i^0$ and $y_1 - y_i^0$)). This means that such tracks are fundamentally indistinguishable in both temporal and amplitude detectors, since the parameter recovery algorithm in an amplitude detector is also based on expression (1). From this symmetry, it follows that, in order to unambiguously recover the

parameter ϕ of a muon track, any detector must have at least three modules whose projections onto the plane $z_i^0 = 0$ do not satisfy two equations of a straight line (or one equation with an arbitrary angular coefficient)

$$\begin{aligned} y_i^0 &= y_1 - \tan\phi(x_1 - x_i^0), \\ y_i^0 &= y_1 + \cot\phi(x_1 - x_i^0). \end{aligned} \tag{2}$$

The symmetry noted in [3] is a particular case of the general symmetry found here: $k = 0$ and $\psi_i^0 = \pi/4$.

The symmetry $\phi \longrightarrow \pi - \phi$ of the solutions to the equations for a muon track found in [1] for a string is also a particular case at $k = 0$.

As is easily seen, the time of arrival of the Cerenkov radiation from the track at an ith PEM is

$$t_i = (l_1\cot\alpha + l_i\tan\alpha + (\mathbf{r}_{1i}\cdot\mathbf{a}))/c, \tag{3}$$

where

$$\mathbf{r}_{ji} = \mathbf{r}_j^0 - \mathbf{r}_i^0, \tag{4}$$

$$\cos\alpha = 1/n, \tag{5}$$

c is the velocity of light in a vacuum, and n is the refractive index of the medium.

Correspondingly, the difference between the time instants the Cerenkov radiation arrives at ith and jth detectors is

$$t_{ij} = t_i - t_j = ((l_i - l_j)\tan\alpha + (\mathbf{r}_{ij}\cdot\mathbf{a}))/c, \tag{6}$$

where the vectors in the parentheses mean the scalar product.

The variables t_{ij} are measurable and depend on the muon track parameters θ, ϕ, x_1, and y_1. Since the track equation has four independent parameters, they can be determined with at least four measured time shifts, i.e., the detector has to incorporate at least five PEMs. It is convenient to rearrange (6) into the form

$$\gamma^2[ct_{i1} - (\mathbf{r}_{1i}\cdot\mathbf{a})]^2 + 2\gamma l_1[ct_{i1} - (\mathbf{r}_{1i}\cdot\mathbf{a})] = S_{1i}, \tag{7}$$

where

$$S_{1i} = S_{1i}^{x_1}x_1 + S_{1i}^{y_1}y_1 + S_{1i}^0, \tag{8}$$

$$S_{1i}^{x_1} = -2[(a_y^2 + a_z^2)(x_i^0 - x_1^0) + a_xa_y(y_1^0 - y_i^0) - a_xa_zz_i^0], \tag{9}$$

$$\begin{aligned} S_{1i}^{y_1} &= -2[(a_x^2 + a_z^2)(y_i^0 - y_1^0) + a_xa_y(x_1^0 - x_i^0) - a_ya_zz_i^0], \end{aligned} \tag{10}$$

$$\begin{aligned} S_{1i}^0 &= (a_y^2 + a_z^2)[(x_i^0)^2 - (x_1^0)^2] \\ &+ (a_x^2 + a_z^2)[(y_i^0)^2 - (y_1^0)^2](a_x^2 + a_y^2)(z_i^0)^2 \\ &- 2[a_xa_y(x_i^0y_i^0 - x_1^0y_1^0) + a_za_yy_i^0z_i^0 + a_za_xx_i^0z_i^0], \end{aligned} \tag{11}$$

$$\gamma = \cot\alpha. \tag{12}$$

ANALYTICAL DETERMINATION OF THE MUON TRACK PARAMETERS FOR DESIGN A

In this case, the Cartesian coordinates of modules have the form

$$\begin{aligned} x_i^0 &= D\cos\phi_i^0/2, \quad \phi_i^0 = 2\pi(i-1)/3 + 2\pi/6, \\ y_i^0 &= D\sin\phi_i^0/2, \quad z_i^0 = -HINT((i-1)/3), \end{aligned} \tag{13}$$

where D is the detector diameter and $INT(C)$, the integer part of the argument C.

Linearly combining Eqs. (7), one can eliminate the variables x_1 and y_1 and form a set of two equations in the parameters θ and ϕ alone:

$$\begin{aligned} &T_2[P_1 - (\mathbf{a}\cdot\mathbf{R}_{T1}) + a_xa_zA_{xz}^1 + a_ya_zA_{yz}^1] \\ &= T_1[P_2 - (\mathbf{a}\cdot\mathbf{R}_{T2}) + a_xa_zA_{xz}^2 + a_ya_zA_{yz}^2], \\ &T_4[P_3 - (\mathbf{a}\cdot\mathbf{R}_{T3}) + a_xa_zA_{xz}^3 + a_ya_zA_{yz}^3] \\ &= T_3[P_4 - (\mathbf{a}\cdot\mathbf{R}_{T4}) + a_xa_zA_{xz}^4 + a_ya_zA_{yz}^4], \end{aligned} \tag{14}$$

where

$$T_j = \sum_i^j t_{i1}, \tag{15}$$

$$P_j = \gamma^2c^2\sum_i^j t_{i1}^2, \tag{16}$$

$$\mathbf{R}_{Tj} = 2\gamma^2c^2\sum_i^j t_{i1}\mathbf{r}_{1i}, \tag{17}$$

$$A_{xz}^j = 2\sum_i^j[x_i^0z_i^0 + \gamma^2x_{1i}z_{1i}], \tag{18}$$

$$A_{yz}^j = 2\sum_i^j[y_i^0z_i^0 + \gamma^2y_{1i}z_{1i}], \tag{19}$$

and $\sum^j$ involves the following terms:

$$\begin{aligned} &\sum_i^1 t_{i1} = t_{41} + t_{21} - t_{51}, \quad \sum_i^2 t_{i1} = t_{41} + t_{31} - t_{61}, \\ &\sum_i^3 t_{i1} = t_{71} + t_{21} - t_{81}, \quad \sum_i^4 t_{i1} = t_{71} + t_{31} - t_{91}. \end{aligned} \tag{20}$$

In order to analytically determine the angle θ, it is convenient to use one of the three strings in our construction. For example, for the second string at $t_{11,2} \neq 3t_{52}$, we have from (7)

$$\cos\theta = [-B \pm \sqrt{B^2 - 4AC}]/(2A), \tag{21}$$

where

$$A = 2H^2(1+\gamma^2)(3t_{82} - 3t_{52} - t_{11,2})/F, \quad (22)$$

$$B = 2H\gamma^2 c[t_{11,2}(4t_{52} - t_{82}) - 3t_{52}t_{82}]/F, \quad (23)$$

$$C = \gamma^2 c^2[t_{11,2}^2 - t_{82}^2 - t_{52}^2 - G(t_{11,2}^2 - 3t_{52}^2)/F] + 2H^2(3G/F - 2), \quad (24)$$

$$F = t_{11,2} - 3t_{52}, \quad G = t_{11,2} - t_{82} - t_{52}. \quad (25)$$

If $t_{11,2} = 3t_{52}$,

$$\cos\theta = \cos^2\alpha \times (ct_{11,2} \pm \sqrt{9H^2\sec\alpha - c^2 t_{11,2}^2}/\gamma)/(3H). \quad (26)$$

The multiple-valuedness of $\cos\theta$ found from Eqs. (21) and (26) is easily eliminated by finding two other values of $\cos\theta$ with another string and choosing two coinciding values. When θ is determined, set (14) will turn into a set of two linear equations in two unknown variables, $\cos\phi$ and $\sin\phi$. This makes it possible to calculate these variables by the Cramer formula. Thus, the angle ϕ can be found uniquely.

Then, in order to determine the parameters x_1 and y_1 of the muon track, we choose any four equations from (7). For instance, we take $i(j) = 2(3)$ and $4(10)$ and pairwise eliminate l_1 from these equations. Eventually, we arrive at the set of two linear equations in x_1 and y_1

$$x_1(T_i' S_{1j}^{x_1} - T_j' S_{1i}^{x_1}) + y_1(T_i' S_{1j}^{y_1} - T_j' S_{1i}^{y_1}) = B_{ij}, \quad (27)$$

where

$$T_i' = 2\gamma[ct_{i1} - (\mathbf{r}_{1i} \cdot \mathbf{a})], \quad (28)$$

$$B_{ij} = T_i[T_j'^2/4 - S_{1j}^0] - T_j[T_i'^2/4 - S_{1i}^0]. \quad (29)$$

DETERMINATION OF THE MUON TRACK PARAMETERS FOR DESIGN B

Design B is less symmetric; nevertheless, the parameters x_1 and y_1 can be eliminated from Eq. (7) in this case too. The coordinates of the detector modules are

$$x_i^0 = D_0\cos\phi_i^0/2, \quad y_i^0 = D_0\sin\phi_i^0/2,$$
$$\phi_i^0 = 2\pi(i-1)/3 + 2\pi INT((i-1)/3)/6,$$
$$z_i^0 = -3HINT((i-1)/3); \quad i = 1, 2, 3, 4, 5, 6,$$
$$x_i^0 = D_k\cos\phi_i^0/2, \quad y_i^0 = D_k\sin\phi_i^0/2, \quad (30)$$
$$\phi_i^0 = 2\pi(i-1)/3 + 2\pi/6, \quad k = INT((i-4)/3),$$
$$z_i^0 = -HINT((i-4)/3); \quad i = 7, 8, 9, 10, 11, 12.$$

The desired set of equations in parameters θ and ϕ somewhat differs from set (14):

$$T_2[P_1 - (\mathbf{a} \cdot \mathbf{R}_{T1}) + a_x a_y A_{xy}^1 + a_y a_z A_{yz}^1] = T_1[P_2 - (\mathbf{a} \cdot \mathbf{R}_{T2}) + a_x a_y A_{xy}^2 + a_y a_z A_{yz}^2],$$
$$T_4[P_3 - (\mathbf{a} \cdot \mathbf{R}_{T3}) + a_x a_y A_{xy}^3 + a_y a_z A_{yz}^3] = T_3[P_4 - (\mathbf{a} \cdot \mathbf{R}_{T4}) + a_x a_z A_{xz}^4 + a_y^2 A_{xx}^4 + a_z^2 A_{zz}^4 - a_y^2(B_x^4 + B_z^4)]. \quad (31)$$

Here, T^j, P^j, $\mathbf{R}_{tj}$, A_{xz}^j, and A_{yz}^j have the same definitions (14)–(18) and

$$A_{xx}^4 = \gamma^2 \sum_i^4 x_{1i}x_{1i} - B_x^4, \quad (32)$$

$$A_{zz}^4 = \gamma^2 \sum_i^4 z_{1i}z_{1i} - B_z^4, \quad (33)$$

$$B_x^4 = \sum_i^4 x_i^0 x_i^0, \quad B_z^4 = \sum_i^4 z_i^0 z_i^0, \quad (34)$$

$$A_{xy}^j = 2\sum_i^j [x_i^0 y_i^0 + \gamma^2 x_{1i}y_{1i}]; \quad (35)$$

however, in (15)–(19) and (32)–(35), $\sum^j$ means

$$\sum_i^1 t_{i1} = t_{21} + t_{16} - t_{31} - t_{41},$$
$$\sum_i^2 t_{i1} = t_{21} - t_{31} + D_0(t_{91} - t_{71})/D_1,$$
$$\sum_i^3 t_{i1} = t_{21} - t_{31} + D_0(t_{12,1} - t_{10,1})/D_2, \quad (36)$$
$$\sum_i^4 t_{i1} = t_{51} - D_0 t_{81}/D_1.$$

Note that this set of equations is valid only if

$$D_0/D_1 = z_5^0/z_8^0 \quad (37)$$

is satisfied.

Since design B does not involve a string, the set of equations in the parameters θ and ϕ should be solved numerically [4]. This requires a considerably higher computational resource than in the case of design A. The multiple-valuedness of the sign of $\cos\phi$ is easily eliminated by calculating all the differences between the times of arrival of the Cerenkov radiation with (6).

With θ and ϕ determined, the parameters x_1 and y_1 of the muon track can be found in the same way as for design A.

CONCLUSION

It is found analytically that, in order to uniquely determine the azimuthal angle of the muon track, the coordinates x_i^0 and y_i^0 of at least three detector modules must not satisfy Eqs. (2) of straight lines. This statement is valid for both temporal and amplitude detectors.

As is shown, detectors of both types are designed in such a way that the set of equations in parameters θ, ϕ, x_1, and y_1 can be divided into two independent lower dimension systems. One is used to find θ and ϕ; the other, to find x_1 and y_1. In this sense, these designs are optimal for determining the parameters of the muon track. Note that, in contrast to a conventional string, the detector designs proposed make it possible to uniquely recover all the parameters of the muon track. The more symmetric design A seems to be preferable from the computational standpoint, since it allows for an analytical solution to the obtained set of equations.

As for the stability of the solution to the equations found here against temporal fluctuations of the Cerenkov radiation (the relevant analysis was conducted in [1] by the standard technique), from general considerations it follows that the solution for a three-dimensional structure of detectors is more stable than that for a one-dimensional string.

REFERENCES

1. V. P. Pustovetov, Preprint No. 146, FIAN (Institute of Physics, Academy of Sciences of USSR, Moscow, 1986).
2. V. I. Aĭnutdinov, I. A. Danil'chenko, and P. Ya. Yashin, in *Proceedings of 1st All-Union Conference "Study of Muons and Neutrinos in Large Water Volumes," Alma-Ata, 1983*, p. 195.
3. V. S. Kinchakov, Zh. Tekh. Fiz. **71** (11), 101 (2001) [Tech. Phys. **46**, 1448 (2001)].
4. G. E. Forsythe, M. A. Malcolm, and C. B. Moler, *Computer Methods for Mathematical Computations* (Prentice-Hall, Englewood Cliffs, 1977; Mir, Moscow, 1980).

Translated by M. Fofanov