Technical Physics, Vol. 49, No. 10, 2004, pp. 1249–1279. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 1–34.
Original Russian Text Copyright © 2004 by Gomoyunova, Pronin.

Photoelectron Spectroscopy of Atomic Core Levels on the Silicon Surface: A Review

M. V. Gomoyunova and I. I. Pronin

Ioffe Physicotechnical Institute, Russian Academy of Sciences,
ul. Politekhnicheskaya 26, St. Petersburg, 194021 Russia
e-mail: Marina.Gomoyunova@mail.ioffe.ru

Received March 11, 2004

Abstract—Recent studies of the atomic structure of the single-crystal silicon surface (both clean and covered by adsorbates) that are performed by high-resolution core-level photoelectron spectroscopy using synchrotron radiation are reviewed. The physical principles of the method, experimental techniques, the spectrum processing procedure, and the procedure of determining the energy shifts of the core levels in the subsurface layer are outlined. Emphasis is placed on the surface modes of silicon 2*p* spectra, which are observed for the main types of silicon surface reconstruction (Si(111)-7 × 7 and Si(100)-2 × 1), and on a correlation between these modes and the atomic structure of the (111) and (100) surfaces. Also, particular attention is given to the studies of the Ge/Si system, which is viewed as a promising material of nanoelectronics, as well as to those concerned with metal and gas adsorption on basic (low-index) silicon faces. These studies clearly demonstrate that core-level photoelectron spectroscopy provides extremely detailed information on the structure of adsorbed layers and on the adsorption-stimulated reconstruction of the substrate surface. *© 2004 MAIK “Nauka/Interperiodica”.*

INTRODUCTION

The binding energy (E_i) of core (inner-shell) electrons is unique for each chemical element and is a basic parameter characterizing the elemental composition of solid surfaces [1–5]. Basically, the binding energy depends on the physicochemical state of an atom; therefore, measurement of core level energy shifts is used for identifying chemical compounds [1]. The amount of this effect may reach several (sometimes 10 or even more) eV. The detection of such shifts does not require a very high energy resolution and is accomplished by means of standard X-ray spectrometers.

A finer effect is observed when the atoms are localized in the subsurface region of a solid that is several monolayers thick [6]. Because of translation symmetry break at the crystal boundary, the state of the surface atoms differs from that of the atoms in the volume and is characterized by energy shifts lying in the interval from several hundredths of an electron volt to ≈0.5 eV. Analysis of such shifts requires a still higher energy resolution and becomes possible owing to powerful sources of synchrotron radiation. The systematic investigation of core electron spectra on solid surfaces with these sources has got under way in the last decade. Much progress in this field had been achieved to date, and core-level spectroscopy is now viewed as a promising tool for studying the atomic structure of crystal surfaces.

The elementary semiconductors, silicon and germanium, show the greatest distinction between the surface and volume structures. Their surfaces are reconstructed, and the atoms occupying different sites in the layer reconstructed are characterized by a set of lines with different energy shifts. In this review, we analyze a great body of publications devoted to photoelectron spectroscopy of silicon atom 2*p* levels on both clean and adsorbate-coated surfaces. In the latter case, we also report the data for the energy shifts of electron levels in the adsorbate atoms. Before considering these results, we briefly present the physical principles of core-level spectroscopy and describe the structure of the Si(100)-2 × 1 and Si(111)-7 × 7 reconstructed surfaces. Among the adsorption systems, particular attention is placed on the Ge/Si one, which is of great independent interest. Interaction of metals with the silicon surface is also considered in detail. The adsorption of gases, specifically oxygen, on the silicon surface is considered in the last section but one. Basic inferences are drawn in the last section.

1. PHYSICAL BASIS OF THE METHOD

In the method of photoelectron spectroscopy, a test material is irradiated with a beam of monochromatic photons and then the energy spectrum of the photoelectrons is analyzed. Core-level spectroscopy, a variant of this method, is based on the optical excitation of core electrons. The spectrum of core electrons is discrete. Each of its lines is related to an electron excited from a particular atomic level. In a solid, unlike a free atom, the binding energy E_i of a core electron is measured from the Fermi level, rather than from the level of vacuum. An event of photoexcitation for this case is sche-

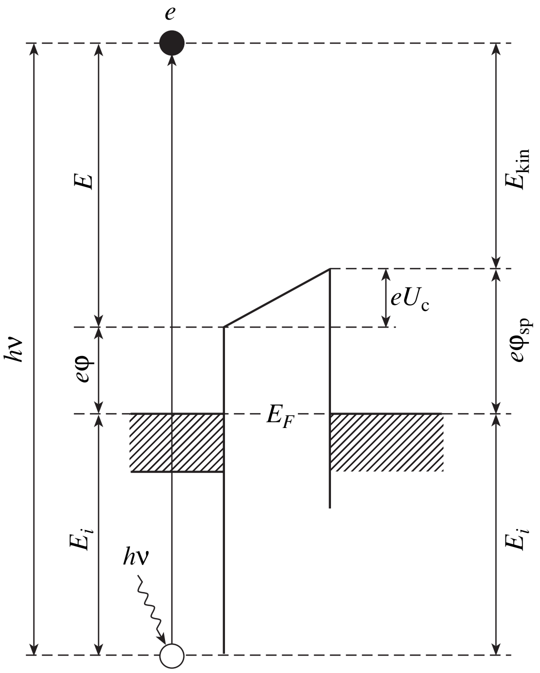


Fig. 1. Energy diagram illustrating the photoexcitation of a core electron.

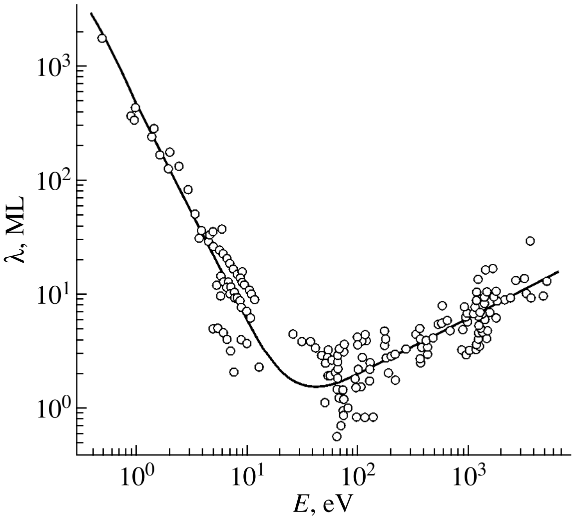


Fig. 2. Energy dependence of mean path length λ of an electron [7].

matically shown in the left-hand side of the energy diagram shown in Fig. 1. As follows from the diagram, the kinetic energy of the photoelectron escaping from the *i*th level that is ionized by a photon of energy $h\nu$ is

$$E = h\nu - E_i - e\varphi, \qquad (1)$$

where $e\varphi$ is the work function of the solid and E is measured from the level of vacuum.

Formula (1) implies that the binding energy E_i of a core electron is uniquely related to its kinetic energy, with $h\nu$ and $e\varphi$ being known. However, when energy E is measured in the gap between the emitter and spectrometer, i.e., under the experimental conditions, the field of contact potential difference U_c arises, which changes the energy of the photoelectrons by eU_c. It is known that $eU_c = e\varphi - e\varphi_{\text{sp}}$, where $e\varphi_{\text{sp}}$ is the work function of the material which the spectrometer is made of. With this in mind, formula (1) can be recast into the form from which the binding energy of a core electron is found:

$$E_i = h\nu - E_{\text{kin}} - e\varphi_{\text{sp}}, \qquad (2)$$

where E_{kin} is the photoelectron's kinetic energy being measured (Fig. 1).

As follows from formula (2), E_i can be determined without knowing the work function of the test specimen. As for φ_{sp}, it is usually found by calibration of the spectrometer with a reference specimen for which the binding energies of core electrons are known.

The model depicted in Fig. 1 assumes that the electrons emitted from the solid do not experience energy-decreasing inelastic collisions with the material. Such a situation takes place if the electrons are emitted from a sufficiently thin layer whose thickness d does not exceed the inelastic scattering mean free path of the electrons (λ). The value of λ depends on the electron energy, this dependence being nearly the same for all materials (Fig. 2): it features a dip at energies of several tens of electron volts [7].

If the energy of the electrons emitted is such that their lossless escape depth far exceeds thickness d_s of the subsurface layer that microscopically has a structure differing from the bulk structure, a major contribution to a specific spectral line is made by the electrons of interior atoms. In this case, we are dealing with the volume-sensitive spectrum. If thickness d of the subsurface layer is comparable to d_s, surface atoms also contribute significantly to the spectrum (the surface-sensitive spectrum). Clearly, the surface sensitivity of the spectrum will be the highest when the energy of the photoelectrons corresponds to the minimum of the dependence $\lambda(E)$. This is achieved, as a rule, by properly selecting the photon energy. The volume-sensitive spectrum is usually recorded using the descending branch of the curve $\lambda(E)$, where λ rapidly drops with increasing energy. To switch to the volume-sensitive spectrum, it is usually sufficient to decrease the photon energy by only 20 or 30 eV. Also, the surface components of the spectrum grow as the polar angle of departure θ_e of the electrons increases. The larger θ_e (measured from the normal to the surface), the smaller d and, accordingly, the greater the contribution from the surface atoms.

The implementation of the method requires sophisticated equipment. First of all, a source of intense monochromatic radiation that makes it possible to controllably vary the photon energy is necessary. Today, only synchrotron radiation sources meet these requirements. Then, a high-energy-resolution electron spectrometer must be at hand. The overall resolution of the instrument, which depends on the monochromatism of synchrotron radiation and on the resolving power of the energy analyzer, should be ≈100 meV or higher. Otherwise, the spectral lines will be unacceptably broad. Their starting breadth depends, first, on the lifetime of the excited state (the lifetime of the hole on the core level) and, second, on the phonon broadening. For example, the natural breadth of silicon 2*p* lines, which depends on the former factor, is usually assumed to be equal to 70–80 meV. The phonon broadening is much larger and varies between 140 and 350 meV, according to the published data. The influence of this broadening may be somewhat smoothed by taking the spectra at lower temperatures. Yet, the starting breadth of the lines is, as a rule, comparable to the energy shifts of surface atom levels. Therefore, surface spectral modes can be revealed by resolving the spectra into components through computer-aided simulation.

It should also be noted that core-level spectra are usually presented as the dependence of the electron emission intensity on the binding, rather than kinetic, energy. The binding energy of the bulk electrons is taken as the zero energy, and only a narrow spectral range near a line selected is depicted. In this case, the surface atoms for which the absolute values of the binding energies of core levels are higher (lower) than the corresponding binding energies of interior (volume) atoms show positive (negative) energy shifts.

2. SINGLE-CRYSTAL SILICON SURFACE

(i) 2*p* spectra for Si(111)-7 × 7 and Si(100)-2 × 1. The works concerned with the study of the single-crystal silicon surface, emphasis is on the 2*p* spectral line ($E_i \approx 100$ eV), which is the most intense. This level is a spin–orbit doublet consisting of the $2p_{3/2}$ and $2p_{1/2}$ sublevels. This doublet was first resolved in 1973, when a small dip between the components was detected [8]. The further refinement of the experimental technique allowed the researchers to reliably resolve the sublevels (Fig. 3a) and determine the amount of the spin–orbit split ($\Delta E \approx 0.6$ eV) [9]. The intensity of one component is twice as high as that of the other, according to the populations of the sublevels.

The majority of recent works used synchrotron radiation at photon energies from 105 to 150 eV. The photons with energies $h\nu = 105$–115 eV, to which mean electron escape depths of 25–30 Å correspond, were used to take the volume-sensitive spectra. The surface-sensitive spectra were taken in the energy range $h\nu = 130$–150 eV ($d = 2.5$–4.0 Å).

The test objects were primarily two reconstructed silicon surfaces: Si(111)-7 × 7 [10–23] and Si(100)-2 × 1 [10, 21–34]. The typical surface-sensitive spectra for these faces are shown in Figs. 3b and 3c. The curves are seen to diverge considerably and differ from the volume-sensitive spectrum (Fig. 3a), which clearly demonstrates the features of the atomic and electronic structures of the given surfaces. Before proceeding to an analysis of the results, we will take a look at the structure of these single-crystal silicon faces.

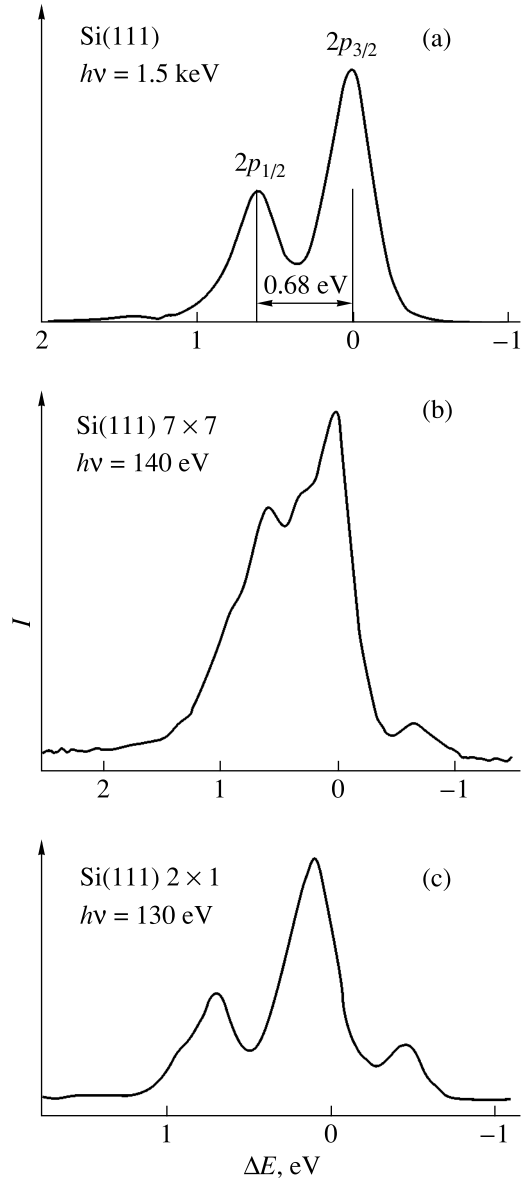


Fig. 3. Si 2*p* electron spectra for (a) Si(111) [9], (b) Si(111)-7 × 7 [20], and (c) Si(100)-2 × 1 [28].

(ii) The structure of reconstructed silicon surfaces. The driving force for semiconductor surface reconstruction is a reduction of the surface free energy when dangling bonds of the surface atoms close on themselves. This changes the angles between the bonds and bond lengths, causing strains. Atom relaxation reduces the strains, so that the reconstructed surface structure depends on the energy balance between these two factors [35].

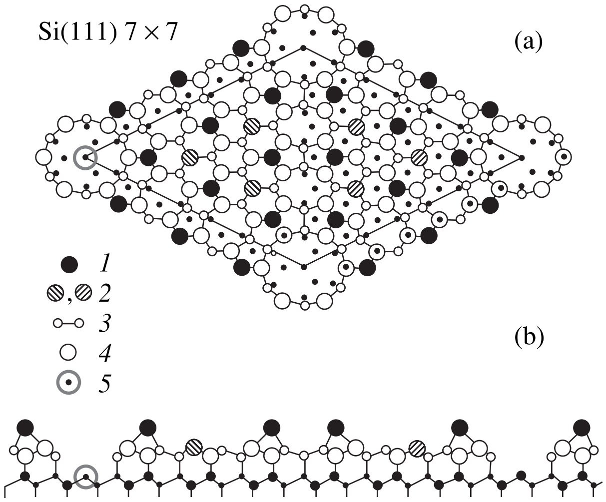


Fig. 4. Schematic representation of the atomic configuration on the Si(111)-7 × 7 reconstructed surface: (a) top view and (b) side view.

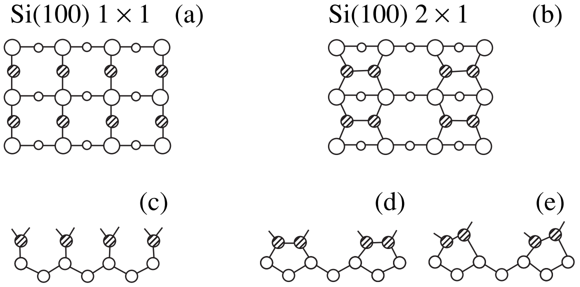


Fig. 5. Structures of the (a, c) Si(100)-1 × 1 virgin surface and (b, d, e) reconstructed Si(100)-2 × 1 surface. (a, b) Top view, (c–e) side view, (d) model of symmetric dimers, and (e) model of asymmetric dimers.

The bonds close through the formation of specific structural units or building blocks on the surface. The basic building blocks are dimers, adatoms, and quasi-one-dimensional atomic chains with π bonds. The dimers arise when two neighboring atoms on the surface approach each other and close their dangling bonds. The adatoms on the reconstructed surface usually appear between three surface atoms each having one dangling bond and close these atoms on themselves, producing one (instead of three) dangling bonds localized on the adatom. The chains with π bonds arise when the sp_3-hybridized electronic configuration of the surface atoms (or some of the surface atoms) turns into the sp_2 planar configuration and p_z orbital. The closure of p_z orbitals of neighboring silicon atoms produces π bonds and, thereby, saturates the dangling bonds.

The reconstruction of the Si(111) surface, Si(111)-7 × 7, has attracted the most interest, and a variety of models explaining the atomic reconfiguration on this face have been suggested. The model commonly accepted today assumes the formation of adatoms and dimers, as well as stacking faults (the so-called dimer–adatom–stacking fault (DAS) model), in the layer reconstructed [36]. Subsequently, this model was strengthened experimentally by using a scanning tunnel microscope [37]. The essence of the model is schematically shown in Fig. 4. A unit cell (7 × 7) of the reconstructed surface includes 49 unit cells (1 × 1) of the virgin silicon surface. The top-most atoms of a (7 × 7) cell are 12 adatoms *1*. Each of them is bonded to three underlying (pedestal) atoms *4*. Six rest atoms *2*, which occupy the same position as on the virgin surface, are situated slightly below the adatoms. One more rest atom *5* is in the corner hole of the cell. Finally, the atoms producing nine dimers *3* are located at the edges of the cell along its minor diagonal. Note also that the right- and left-hand sides of the cell are not equivalent (the latter has a stacking fault compared with the volume, Fig. 4).

In this model, only the adatoms and rest atoms have dangling bonds. The total of these atoms within the cell is 19 instead of 49 on the virgin Si(111) surface, where each of the atoms has a dangling bond directed normally to the surface. Such a considerable (more than twofold) decrease in the number of dangling bonds as a result of the reconstruction and such a profound atomic reconfiguration on the surface are a unique case.

On the Si(100) surface, the reconstruction is less significant (Fig. 5). The structure of the virgin (100) surface, Si(100)-1 × 1, is shown in Figs. 5a and 5c. This face has a quadratic unit cell of side 3/2*a* (where *a* is the lattice constant of the unit cell of silicon), which runs along the [110] direction. Each of the surface atoms has two dangling bonds making an angle with the surface. As the dangling bonds of neighboring atoms close on themselves during the reconstruction, dimers aligned with the ⟨110⟩ directions and separated by recesses form on the surface (Fig. 5b).

The dimeric model of Si(100) surface reconstruction, Si(100)-2 × 1, was suggested as early as in 1959 [38] and later was corroborated by the method of scanning tunnel microscopy (STM) [39]. The major issue discussed in the context of this model was whether the dimers are symmetric or asymmetric [40–45]. Their structure is specified by the type of bonds between silicon atoms. Symmetric dimers have covalent bonds (Fig. 5d), while asymmetric dimers also have the ionic component, as follows from the model of asymmetric dimers [40] (Fig. 5e). The concepts of symmetric and asymmetric dimers were also discussed in the works that used the method of core-level spectroscopy [21–33]. To date, the model of asymmetric dimers has gained wider recognition. On the Si(100) surface reconstructed, the density of dangling bonds decreases twofold and the unit-cell symmetry on the surface changes (there appears a 2 × 1 structure, whose period in one of the ⟨110⟩ directions is doubled).

The study of the Si(100)-2 × 1 surface by low-energy electron diffraction (LEED) showed that the region where the crystal structure is modified covers several atomic layers and is in the strained state [46, 47]. Owing to the fourfold symmetry of the (100) face, the (2 × 1) and (1 × 2) structures turn out to be physically equivalent, and both types of domains usually coexist on the surface in order to relieve the strain in the layers reconstructed. However, there are special processing techniques with which one can obtain a Si(100) surface where the domains of one orientation dominate [48, 49]. Note that, in Fig. 5, all the dimers have the same orientation. It was shown [50] that the energy of such a surface structure is not minimal. The energy decreases when neighboring dimers become antiparallel. The diffraction pattern $c(4 \times 2)$, which is observed on this face at low temperatures, is associated with a specific alignment of the dimers. As the crystal temperature rises to ≈200 K, phase transition due to dimer disordering takes place and the pattern (2 × 1) is observed [51]. The core-level spectroscopy study done in [28] showed that the local atomic structure of the dimers persists at this phase transition and that the phase transition is due to the higher order reconstruction mentioned above, which, in turn, is related to the mutual orientation of the dimers.

(iii) Decomposition of the spectra. Basically, each group of the atoms occupying identical sites in the subsurface layer reconstructed can be assigned a spectral mode. However, these modes are unobservable when the linewidth and energy shift are comparable to each other. Therefore, detection and identification of surface components in the spectra are a challenge. As a preliminary, one should perform comparative analysis of the volume- and surface-sensitive spectra (note that it is desirable to take the spectra at low temperatures and different angles of departure of electrons). By way of example, Fig. 6 shows such spectra for Si(111)-7 × 7 at T = 110 K [19]. The surface-sensitive spectrum ($h\nu$ = 131 eV) exhibits three features that are absent in the volume-sensitive spectrum ($h\nu$ = 106 eV). Feature (1) is to the right of the line $2p_{3/2}$ and is characterized by the energy shift $\Delta E \approx -0.7$ eV. Mode (2) is at $\Delta E \approx 0.28$ eV, and surface component (3) appears as a short shoulder on the left-hand branch of the line $2p_{1/2}$.

For the Si(100)-2 × 1 surface, the situation is similar (Fig. 7) [33]. Here, the surface-sensitive spectrum has a low-energy component ($\Delta E \approx -0.5$ eV), which is, however, much more pronounced than in the previous case. In the spectra recorded at low temperatures and grazing angles to the surface, its intensity may be even comparable to the intensity of the volume doublet $2p_{1/2}$. On the contrary, two other surface components for this face are much weaker and show up only in a decrease in the dip between the lines $2p_{1/2}$ and $2p_{3/2}$ of the volume-sensitive spectrum and also in a considerable broadening of one of these peaks.

Mathematical simulation with mode parameter variation provides more comprehensive and accurate data for the surface modes of the spectrum. In this case, comparison with the experiment for the purpose of finding an optimal model is based on the least-squares method [52–54]. The spectrum is routinely represented as a sum of a volume doublet and several spin–orbit doublets with the same split between the $2p_{3/2}$ and $2p_{1/2}$ sublevels and with an initial ratio of the corresponding intensities of two. All these components are described in terms of the Voigt functions, which are the convolu-

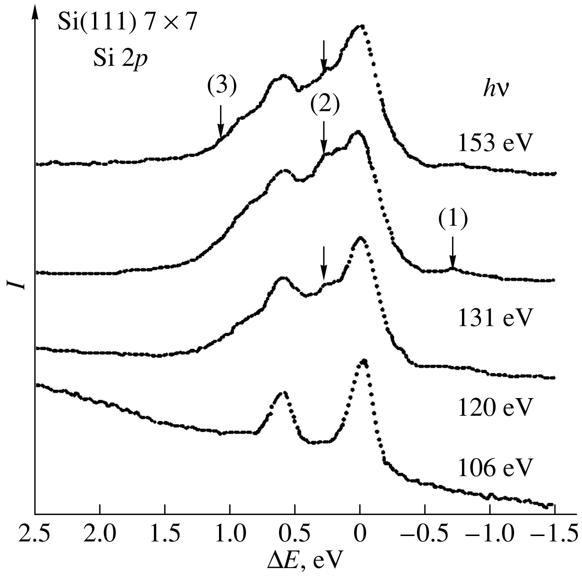


Fig. 6. Si 2p photoelectron spectra taken from the Si(111)-7 × 7 surface at different electron energies [19].

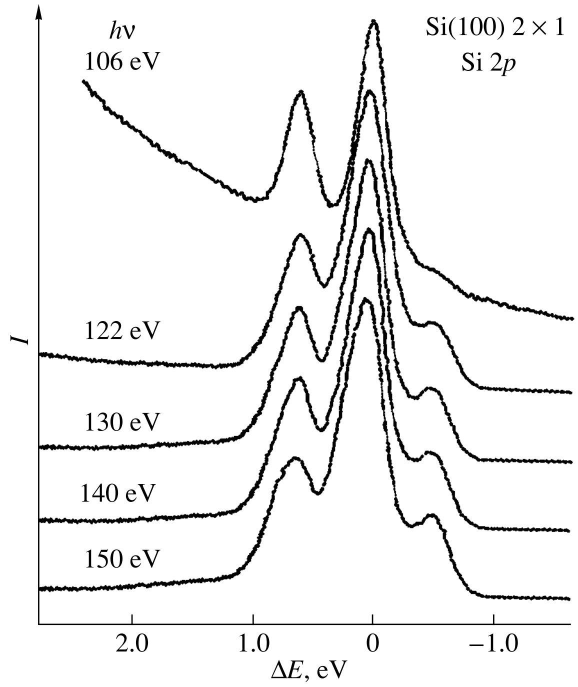


Fig. 7. Si 2p photoelectron spectra taken from the Si(100)-2 × 1 surface at different electron energies [33].

tions of the Lorentz and Gauss functions. The former approximate the intrinsic linewidth, which depends on the lifetime of a hole on a core level; the latter, the phonon broadening and the energy resolution of an instrument. For metals, the former function may be asymmetric, because the energy is lost on the excitation of conduction electrons. Therefore, some authors use the Doiach–Sunjic functions with appropriate asymmetry parameters [55].

As the input for mathematical simulation, the researchers usually specify the number of surface modes, their energy shifts, and the half-widths of the functions describing various spectral components. Preliminary values of these parameters are usually selected based on visual analysis of the spectra. From the volume-sensitive spectrum, the parameters of the Voigt function, which approximates the bulk component, are found. As for the functions describing surface modes, the half-widths of the Gaussian distributions involved in these functions usually far exceed the corresponding values for the bulk component. Next, the functions selected are added up and the result is compared with the experiment. To provide the best fit to the experimental curves, the energy positions of the lines selected, as well their intensities and widths, are varied. If the result is unsatisfactory, an additional spectra component is included and the entire procedure is repeated. If the spectrum is resolved adequately, the energy shifts of the components found must be independent of neither the photon energy nor the angles at which the photoelectrons are detected.

It should be noted that, when the analytical curves are compared with the experiment, they are contrasted with the spectrum minus the background. The background is associated, first, with the faster photoelectrons that lose a part of their energy when escaping the crystal (and also with secondary electrons excited by these photoelectrons) and, second, with the very core-level photoelectrons being analyzed that experience minor energy losses. The first component is usually extrapolated by polynomials; the other, by the Shirley method [56].

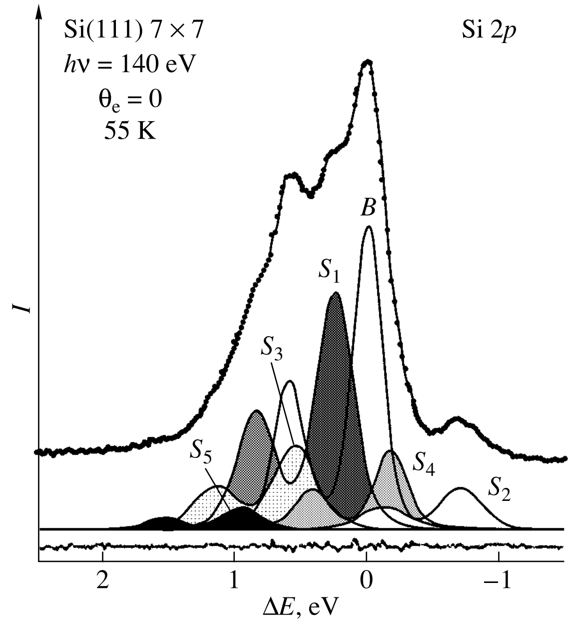


Fig. 8. Decomposition of the spectrum of photoelectrons emitted from the Si(111)-7 × 7 surface into the bulk and surface components [20].

The 2p spectrum for the Si(111)-7 × 7 surface that is resolved by the above technique is shown in Fig. 8. The spectrum is seen to consist of bulk component B and five extra components S_1–S_5 [20]. Components S_1–S_4 are surface modes (their interpretation will be considered in the following subsection). Component S_5 is of another nature: it is related to the excitation of valence electrons either immediately during the photoexcitation of 2p electrons or when 2p electrons leave the crystal.

The decomposition of the spectra for the Si(100)2 × 1 surface was performed in [28] (Fig. 9a) and [33] (Fig. 9b). The authors of the former work gave a satisfactory description of the spectra through the superposition of the volume and four surface components. One more surface mode was revealed more recently [33, 34].

(iv) Identification of the surface modes. To identify the surface spectral components, it is necessary that a number of conditions be satisfied. First, the intensities of different modes must correlate with the number of their associated atoms in the unit cell and also adequately reflect the distribution of the atoms across the depth (the intensity must exponentially drop with depth). Second, the energy positions of different components and the signs of their energy shifts must correlate with the nature of the components. It is known [6] that two effects are mainly responsible for the shifts. The major one is the chemical shift, which arises when the initial level of a core electron of the atom shifts in response to a change in the state of its valence electrons. This shift may be negative or positive, according to whether the effective charge of the atom increases or decreases, respectively. The second (relaxation) effect arises when the final states of the electron subsystem of an ionized atom change and also when an elastic force field is generated around this atom because of the nonequilibrium state of the excited system.

The chemical shifts of surface spectral components are due to electron bond break at edge (surface) atoms, which generates surface bands and effectively charges the atoms. For the reconstructed Si(111)-7 × 7 surface, three such bands were reported [18]: S_1, S_2, and S_3 (Fig. 10). The first two, S_1 and S_2, are due to dangling bonds at the adatoms and rest atoms; the third one is associated with the valence bonds of the adatoms that are directed into the crystal. Since band S_1 is above band S_2, the rest atoms gain a part of the charge from the adatoms. As a result, the former and latter become positively and negatively charged, respectively.

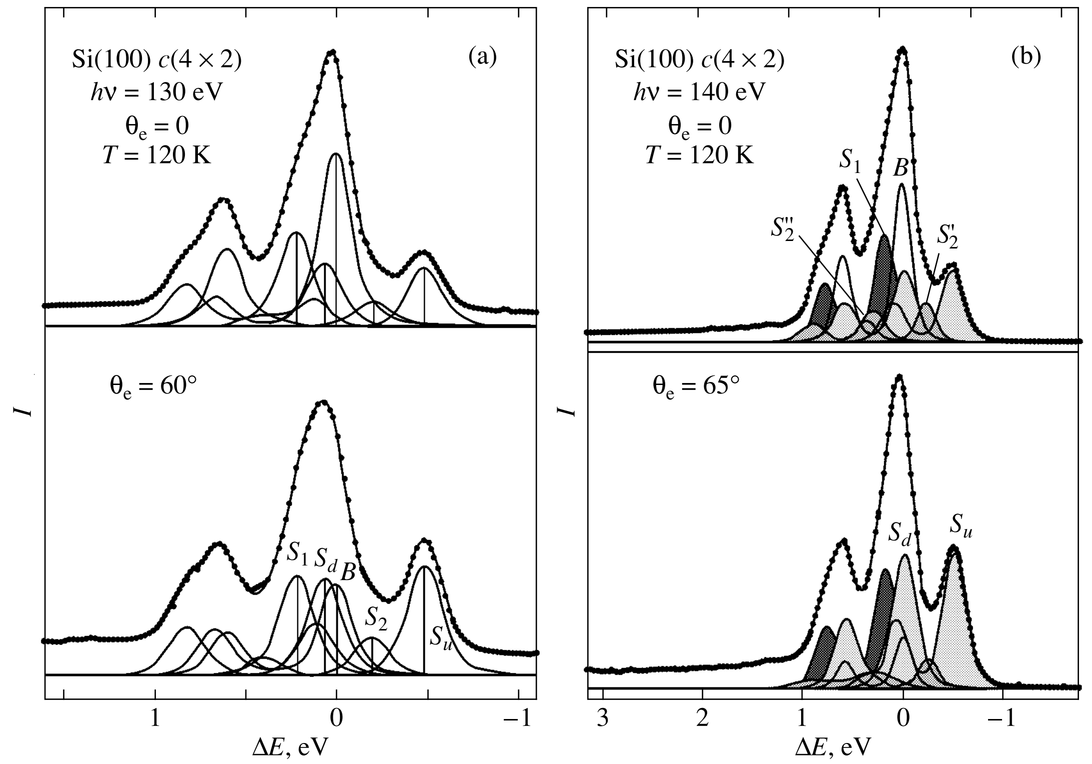


Fig. 9. Decomposition of the spectrum of 2*p* photoelectrons emitted from the Si(100)-2 × 1 surface into the bulk and surface components: (a) data from [28] and (b) data from [33].

Accordingly, the 2*p* electrons of the adatoms (rest atoms) acquire a negative (positive) chemical shift. The relaxation shifts for the Si(111)-7 × 7 surface are usually ignored.

According to the above considerations, low-energy surface mode S_2, which is observed at $\Delta E = -700$ meV for Si(111)-7 × 7, is attributed to the rest atoms, while mode S_3, which is featured by the greatest positive shift ($\Delta E = 550$ meV), to the adatoms.[1] The fact that the intensity ratio for these modes equals two strengthens such an interpretation of the spectrum. Indeed, the number of adatoms in the unit cell (12) is roughly twice as large as the number of rest atoms.

Now let us pass to other spectral modes shown in Fig. 8. Note that the intensity of mode S_1, which was observed by different authors in the shift range 244–280 meV [17–20], exceeds that of the mode S_3 of the adatoms by nearly three times. Based on this observation, mode S_1 is assigned to the pedestal atoms, the number of which (36) is three times that of the adatoms. There is some disagreement on the fourth spectral component (S_4), specifically, on its position and identification. In [18, 20], it is assigned to the dimers, while in [19], to the first-layer atoms and also to those lying immediately under the adatoms.

In the context of the problem considered, of great importance is work [20], where the spectra were taken at 55 K. Although the temperature was taken as extremely low in order to reduce the phonon broadening of the lines, the attempt to considerably sharpen the 2*p* spectrum in comparison with that recorded at 120 K [18] failed. The failure was explained by the complex atomic structure of the Si(111)-7 × 7 surface. Actually, however, none of the atomic groups to which spectral

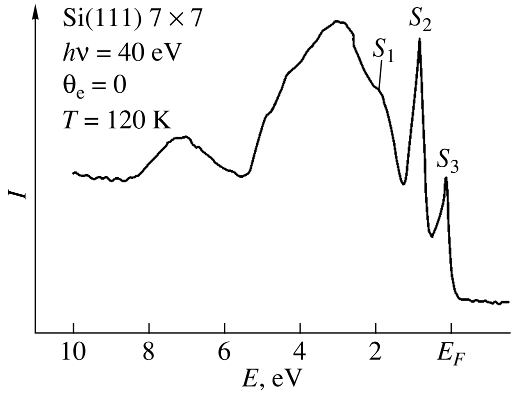


Fig. 10. Photoelectron spectrum of valence electrons emitted from the Si(111)-7 × 7 surface [18].

---

[1] According to [17], component S_2 is split by the crystal filed into tow modes; however, no confirmation of this effect was found in other publications.

components are assigned are fully homogeneous. For example, central and corner adatoms with dissimilar structural environments are distinguished. The environments of the atoms occupying the right- and left-hand halves of the unit cell on the surface are also somewhat different, etc. The superposition of the closely spaced lines corresponding to these atoms makes the spectrum diffuse, and a minor decrease in the phonon broadening goes unnoticed. It was therefore speculated [20] that the ultimate experimental resolution of the 2*p* spectrum for the Si(111) surface has already been reached.

Let us now turn to the spectra for Si(100)-2 × 1. In the early studies of this surface, the spectra were taken at room temperature and only two hardly identifiable surface modes (ΔE = –400…–500 and 220–340 meV) were revealed as a rule. For example, in [13], the first mode was attributed to the dimers (which were considered symmetric), while the other one was assigned to the atoms of the second monolayer. At the same time, the authors of [27], who adhered to the model of asymmetric dimers, assigned the low-energy mode to the upper atoms of the dimers and the high-energy mode to their lower atoms. As was noted above, the model of asymmetric dimers is today dominant. Therefore, at present, the mode of the silicon 2*p* spectrum that has the highest negative shift (its intensity corresponds to half a monolayer) is also attributed to the upper atoms of the dimers [28, 33, 34] (mode S_u in Fig. 9). The lower atoms of the dimers are assigned mode S_d, whose intensity is the same as that of mode S_u and varies identically with photon energy and angle of departure of the electrons. The sign and value of the shift of mode S_d still remain unclear. According to [28], the energy shift of mode S_d is positive, 60 meV (Fig. 9a). The authors of [34] also argue that this shift is positive but were unable to give its accurate value and indicated only its range (30–130 meV). Finally, in [33], the shift of this mode was found to be negative (–16 meV, Fig. 9b), which is explained by the relaxation processes discussed in [57, 58].

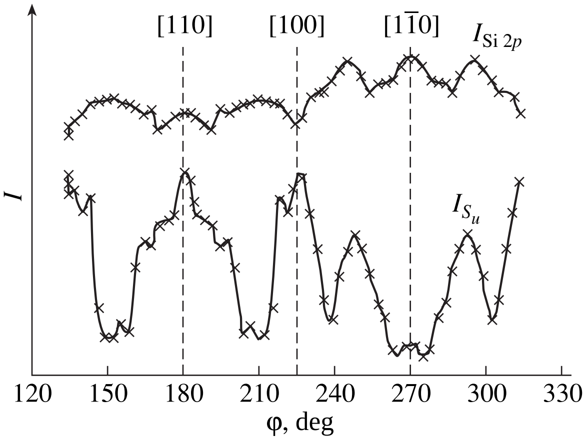


Fig. 11. 2*p* electron emission intensity vs. the azimuth angle of departure for the Si(100)-2 × 1 surface at a polar angle of departure of 45° [31]. $I_{\text{Si}_{2p}}$ is the total intensity of the line; I_{S_u} is the intensity of surface component S_u.

The third component of the spectrum, S_1, is invariably assigned to the atoms of the second monolayer based on the analysis of its intensity. The least intense modes (S_2 in [28], as well as S_2' and S_2'' in [33, 34]) call for further investigation, and we will not consider them in detail. Note only that the same surface components are present in the 2*p* electron spectrum both for the crystal cooled to 120 K and for the crystal kept at room temperature [28]. This suggests that the local structure of the asymmetric dimers on the silicon surface is identical at these two temperatures. The change $(2 \times 1) \longrightarrow c\,(4 \times 2)$ of the diffraction patterns, which is observed as the temperature decreases, is explained by the change in the mutual orientation of the dimers (see above).

In [31], quantitative data for the local configuration of the asymmetric dimers were obtained using core-level spectroscopy. To examine their configuration, the intensities of component S_u and the entire 2*p* spectral line were measured against the azimuth emission angle of electron with the polar angle equal to 45°. The first dependence was found to have a well-defined structure, while the other was irregular (Fig. 11). The structure of the first dependence is due to the diffraction of the photoelectrons by the atoms surrounding the emitter. The simulation of this dependence with allowance for multiple electron scattering showed that agreement with the experiment is observed only for the electrons leaving the upper atoms of the dimers. Varying the structural parameters of the model and comparing the analytical and experimental results with the help of the reliability factor, the authors of [31] found the inclination of the dimers to the surface (19°) and the length of the bond between dimer atoms (2.25 Å). These values are close to those obtained in [59] (18.3° and 2.26 Å).

3. THE GERMANIUM/SILICON SYSTEM

The Ge/Si system is a typical representative of semiconductor/semiconductor systems and is of great interest in the context of heterojunction formation. The primary goal of the investigations performed to date was finding optimal conditions for the epitaxial growth of Ge-on-Si films with an abrupt film–substrate interface. Another goal was to clarify the diffusion of germanium atoms into silicon. Both issues have been studied intensively by using core-level spectroscopy since early in the 1990s. Emphasis has been on the Ge/Si(100)-2 × 1 system, which is of special importance for the further development of semiconductor technology. The 2*p* electron spectra of silicon have been examined along with the 3*d* electron spectra of germanium (the binding energy of 3*d* electrons in germanium is near 30 eV). The Ge 3*d* level is a spin–orbit doublet and consists of the $3d_{3/2}$ and $3d_{5/2}$ sublevels 585 meV apart.

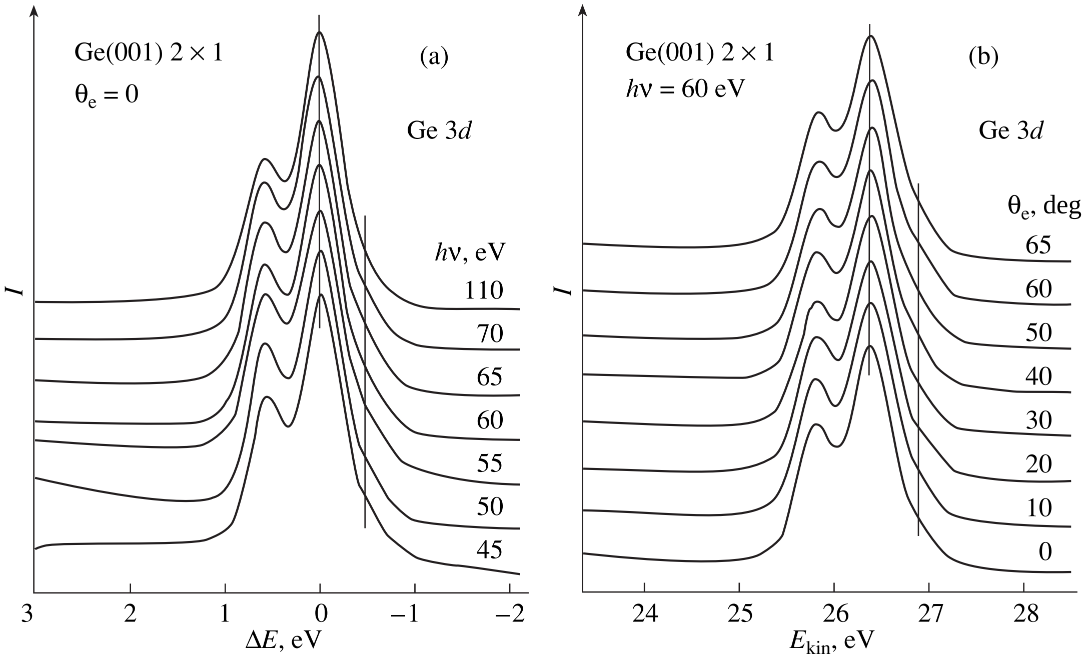


Fig. 12. Germanium photoelectron spectra taken from the Ge(100)-2 × 1 surface at different (a) electron energies and (b) angles of departure [65].

(i) 3*d* spectrum of the Ge(100)-2 × 1 surface. In order to treat the Ge 3*d* spectrum in the Ge/Si system, one must know its features for a clean germanium surface. The (100) surface of single-crystal germanium has the same structure as Si(100), i.e., 2 × 1. The 3*d* photoelectron spectra for this surface were studied in [22, 52, 60–65]. The typical results are shown in Fig. 12. It is seen that the surface modes of germanium are less pronounced than those of silicon. Specifically, in the germanium spectra, only one surface component (marked by the vertical line) is observed. Also, the spectra are less sensitive to the angle of departure and photon energy. These factors make the interpretation of the data difficult.

The decomposition of the spectra into components is shown in Fig. 13, where the results obtained in two most comprehensive studies are summarized. In one of them, three surface modes with energy shifts of –530, –240, and 190 meV were detected [64]. In the other, the values of the shifts were found to be –442, –183, and 93 meV [65]. The first modes, which have the lowest binding energy, are assigned, as in the case of Si(100)-2 × 1, to the upper atoms of the asymmetric dimers. The other modes are identified variously. In [64], the mode with the highest binding energy is assigned to the lower atoms of the dimers and the third component, to the atoms of the second monolayer. However, the authors of [65] hold the opposite viewpoint. Thus, a commonly accepted interpretation of the germanium 3*d* spectrum is lacking.

(ii) Adsorption of germanium on the Si(100)-2 × 1 surface. The Ge/Si(100) system was studied in [17, 66–70]. The most interesting results were obtained for submonolayer coverages. In this thickness range, the formation of germanium dimers, which substitute for silicon dimers, takes place, as observed in a scanning tunnel microscope [71]. According to [70], the Ge 3*d* spectra (Fig. 14) distinctly show only one surface mode *U* early in the evaporation of germanium (up to one tenth of a monolayer), which dominates up to half a monolayer (ML). If germanium dimers were formed in this case, two modes would be seen in the spectra with regard to the dimer symmetry. The presence of only one mode in the Ge 3*d* spectrum, as well as the characteristic variation of the Si 2*p* spectra during germanium evaporation (a decrease in the intensity of the mode associated with the upper atoms of the silicon dimers), suggests that the upper atoms of the silicon dimers are substituted for by germanium atoms. New components (*S* and *D*) in the germanium spectrum, which have binding energies higher than the binding energy of the first mode, are distinctly seen starting from 0.3 ML, and their intensity grows with adsorbate dose (Fig. 14) Mode *D* is due to the germanium atoms substituting for the lower silicon atoms in the dimers that were initially present on the substrate. Thus, the method of core-level spectroscopy revealed mixed (Ge–Si) dimers very early in their formation and made it possible to trace their subsequent transformation into dimers consisting of only germanium atoms.

In [70], the *S* component of the spectrum is assigned to the Ge atoms that penetrated into the second subsurface monolayer. If so, this means that Ge atoms start diffusing into the silicon before the formation of the Ge

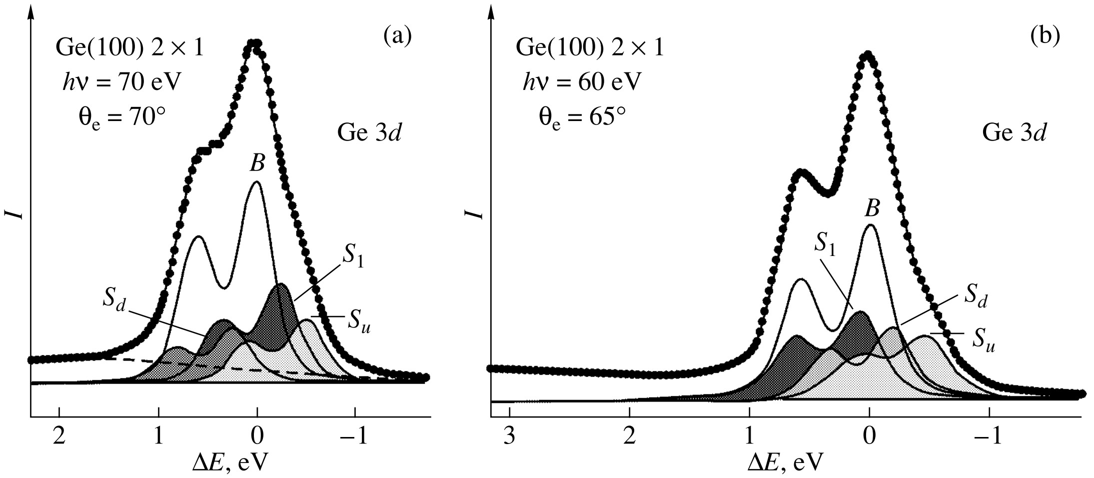


Fig. 13. Germanium photoelectron spectra taken of the Ge(100)-2 × 1 surface, and their decomposition into the bulk and surface components: (a) data from [64] and (b) data from [65].

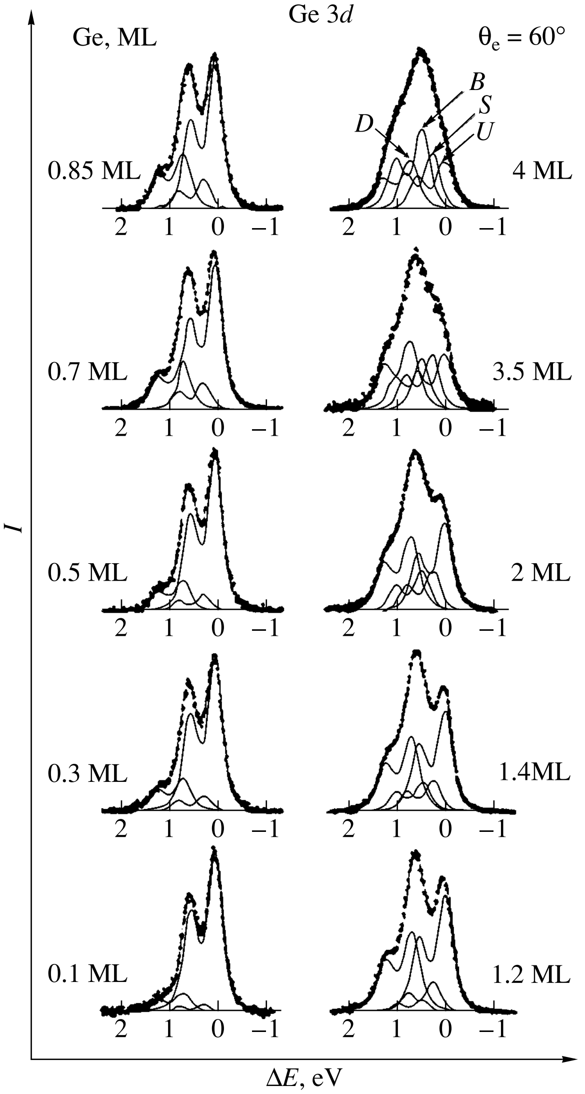


Fig. 14. Germanium core-electron spectra taken from the Si(100)-2 × 1 surface at increasing Ge doses and their decomposition into components [70].

dimers has been completed. The changes observed in the silicon 2*p* spectra count in favor of such a supposition. It should be noted that other methods have not detected Ge diffusion into the substrate in this system (see, e.g., [72, 73]). According to [72, 73], the silicon surface is first covered by two or three pseudomorphous germanium layers, which give rise to germanium islands once an adsorbate is deposited (the Stranski–Krastanov growth mechanism).

Finally, the last component of the spectrum (*B*), which is observed between modes *S* and *D*, appears when the coverage exceeds a monolayer. Its energy position corresponds to the bulk mode of single-crystal silicon (Fig. 13); accordingly, it is also considered as a bulk mode. Moreover, the Ge atoms that penetrated into the third and deeper layers of the substrate are also qualified as volume atoms, since the electronegativities of silicon and germanium are close to each other. The results obtained for thick (up to 20 MLs) coverages agree with the Stranski–Krastanov growth mechanism mentioned above, as indicated by the presence of surface components in the silicon 2*p* electron spectra throughout the coverage thickness range.

(iii) Epitaxial growth of germanium on silicon. Layer-by-layer growth of germanium films on silicon is often facilitated with surfactants, such as Group-V elements As [74, 75] and Sb [62, 76–80]. When deposited on the silicon surface, As and Sb close the dangling bonds of surface atoms of the substrate, forming one-ML-thick ordered structures, which prevent surface reconstruction and reduce the free surface energy of the system. After subsequent application of germanium, the surfactant and adsorbate atoms change places: the surfactant emerges on the surface and closes the dangling bonds of new germanium atoms. The latter become deposited on the unreconstructed silicon surface and occupy those sites providing the epitaxial growth of germanium. Then, the process is repeated. It was reported that not only thick Ge/Si films, but also

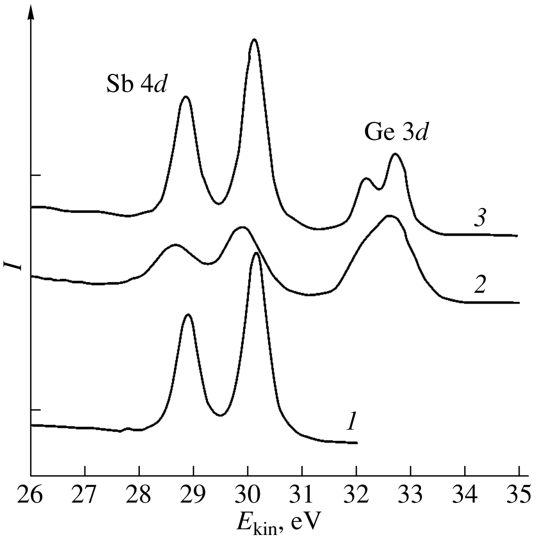


Fig. 15. Spectra of the Sb and Ge photoelectrons for the Sb/Ge/Si(100) system [78]. *1*, Si(100)-1 × 1 surface covered by an ordered Sb monolayer; *2*, Sb/Si(100)-1 × 1 system covered by a Ge monolayer; and *3*, Sb/Ge/Si(100)-1 × 1 system annealed at 500°C.

Si/Ge films and even Ge/Si superlattices, were grown [62, 78].

The above process as deposited to the Ge/Sb/Si(100) system was studied in [76, 78]. One-monolayer-thick ordered Sb/Si films were prepared either by applying Sb atoms on the substrate heated to ≈500°C or by annealing the film deposited at room temperature. Information about the structure of the films was gained from the 4*d* spectra of Sb and 3*d* spectra of Ge (see Fig. 15). The presence of only one component in the Sb spectrum (curve *1*) indicates that only one type of adsorption site for the surfactant atoms is available. The subsequent room-temperature application of a Ge monolayer noticeably decreases the intensity of the Sb line (curve *2*) and causes its shift. However, the annealing of the specimen completely restores both the intensity and shape of the line (curve *3*). This means that a new surfactant monolayer appears on the surface. The variation of the Ge 3*d* spectrum is consistent with the above scenario.

The variation of the Si 2*p* spectra in this system was studied most comprehensively in [76]. As follows from Fig. 16, a Sb monolayer deposited on the surface eliminates the surface components of the asymmetric silicon dimers, which means that the substrate surface loses its initially reconstructed state. Remaining surface components *S* and *C* are assigned to the silicon atoms occupying three upper monolayers of the substrate. As more and more Ge atoms are deposited, these modes are gradually and almost completely quenched. Eventually, bulk mode *B* alone remains in the spectrum, which confirms the layer-by-layer (epitaxial) growth of the Ge film.[2]

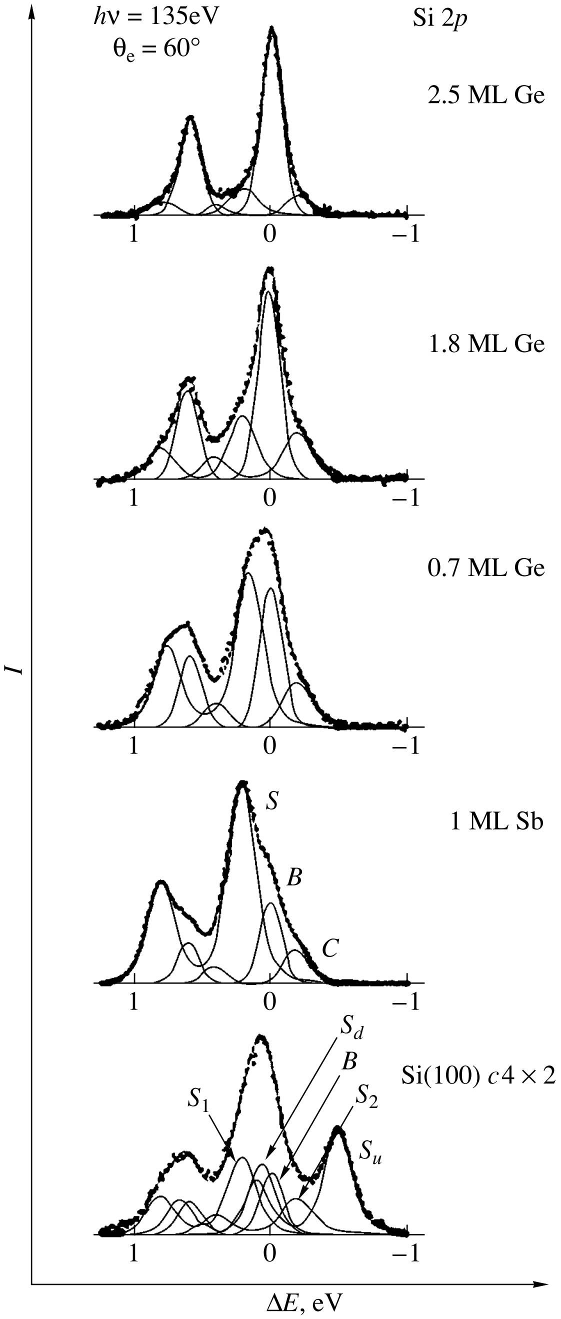


Fig. 16. Si photoelectron spectra for the Ge/Si(100)-2 × 1 system and the result of decomposition [76].

4. METAL/SILICON SYSTEMS

Core-level spectroscopy as a tool for studying metal adsorption on the silicon surface was first deposited in the 1980s and has enjoyed wide application since the mid-1990s. To date, core-level spectroscopy has been used to study the adsorption of alkali metals [33, 81–100], alkaline-earth metals [101–105], noble metals [87,106–111], Group-III metals [112–114], transition metals [115–123], rare-earth metals [124–129], and others. Attention has been drawn to many aspects of the

---

[2] Note that the intrinsic linewidth of this mode is unprecedentedly narrow for photoelectron spectroscopy of silicon: 20 meV.

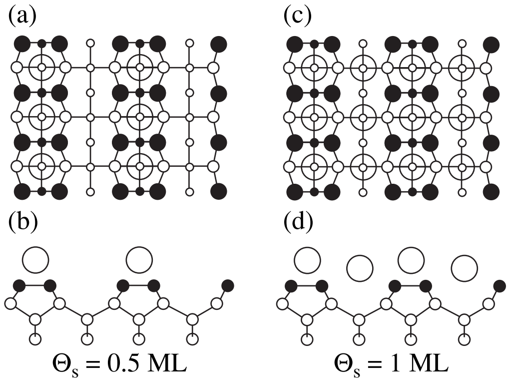


Fig. 17. Models of alkali metal adsorption on the Si(100)-2 × 1 surface: (a, b) the Levin model of one-dimensional metallic chains [29] and (c, d) the double layer model [30].

process. In some works, the general scenario of the adsorption was the focus of attention. The issues under study were which sites on the substrate surface are appropriate for the adsorbate atoms; how many adsorption sites are typical of a specific adsorption system; in what sequence these sites are occupied; whether the atomic reconstruction of the substrate surface takes place; if so, how this reconstruction proceeds; etc. In other works, one or several structures that include particular adsorbates and are formed under certain conditions were studied in detail. A great body of data concerning many systems cannot be covered in this review. Therefore, we will concentrate on the most typical and, in our opinion, most interesting results.

4.1. Interaction of Metals with the Si(100)-2 × 1 Surface

(i) Adsorption of alkali metals. The electronic constitution of alkali metals is simple; therefore, an alkali metal/silicon system can be considered as a model in investigating the formation of the metal–semiconductor interface. The related systems have been examined in a variety of works using various surface-sensitive techniques. It has been shown that the Si(100)-2 × 1 surface usually retains its configuration when covered by alkali metal submonolayer coverages. Two models to explain this fact have been suggested (Fig. 17). According to the model of one-dimensional atomic chains [130], the metal atoms are adsorbed on dimers (Figs. 17a, 17b) and lose their valence electron to the substrate, producing ionic bonds with the substrate. In this case, the saturated coverage corresponds to half a monolayer (by a monolayer, we mean the atomic concentration that equals the atomic concentration on the virgin Si(100) face: 6.78×10^{14} atoms/cm^2). Under these conditions, the surface band of electron energies turns out to be half-filled and the substrate surface is metallized. The second model put forward in [131] assumes the formation of the double adsorbate layer (Figs. 17c, 17d). As in the previous model, half of this layer consists of atoms adsorbed on the dimers, while the other atoms are adsorbed between the chains of the dimers. In this case, the saturated coverage equals a monolayer and the specimen surface retains the semiconductor properties. Each of the models has its own pros and cons, and such issues as adsorption sites, the type of atom–substrate bonds, and the concentration of adatoms in the saturated coverage have been the subject of much investigation. The method of core-level spectroscopy has certainly provided fresh insight into these problems.

The K/Si(100) system has been most extensively studied by this method [23, 90–96], and we will consider it first. Potassium was adsorbed both at room temperature and at low temperatures. At room temperature, the saturated potassium coverage does not exceed one monolayer. Lower temperatures make it possible to study a much wider range of the coverages. The photoelectron spectra of the adsorbate were examined together with the core-electron spectra of the substrate. Typical 3*d* electron spectra of potassium that were taken during potassium evaporation on the Si(100)-2 × 1 surface at room temperature are shown in Fig. 18. At doses below 0.5 ML, component *D* alone is present in the spectrum. This means that incident potassium atoms occupy the same adsorption sites that are localized between the chains of the substrate dimers. As the coverage grows, the second component (*C*) appears in the spectrum. This component is related to the potassium atoms resting on the other adsorption sites, i.e., on dimers. Both modes grow until the coverage becomes saturated ($\Theta_s = 0.9$ ML, where Θ_s is the saturation coverage of the surface), which agrees with the double layer model (Figs. 17c, 17d).

The Si 2*p* spectra show that potassium adsorption quenches the S_u component of the upper atoms of the silicon dimers and causes component *S* to arise. New component *S* is between S_u and bulk component of silicon. The intensity of component *S* reaches a maximum when the coverage becomes saturated (about one monolayer of silicon atoms). This line is associated with symmetric silicon dimers, into which asymmetric dimers on the initial Si(100)-2 × 1 surface transform during potassium adsorption. Such an idea is also consistent with the second model mentioned above.

When potassium is deposited at low temperatures [95], the process goes much in the same way as at room temperature as long as the coverage remains in the submonolayer range; that is, the 3*p* electron spectrum of the potassium shows two components (*C*, *D*) as in the previous case. However, when the material dose corresponds to the room-temperature saturation coverage Θ_s, the third component appears in the spectrum. As the potassium dose grows further, the intensity of this line increases and the line bifurcates into components *S* and *B* (Fig. 19). The data obtained at different angles of departure of the photoelectrons suggest that mode *S* is a surface mode, while mode *B* is a bulk mode. Modes *C* and *D* are present in the potassium spectrum throughout

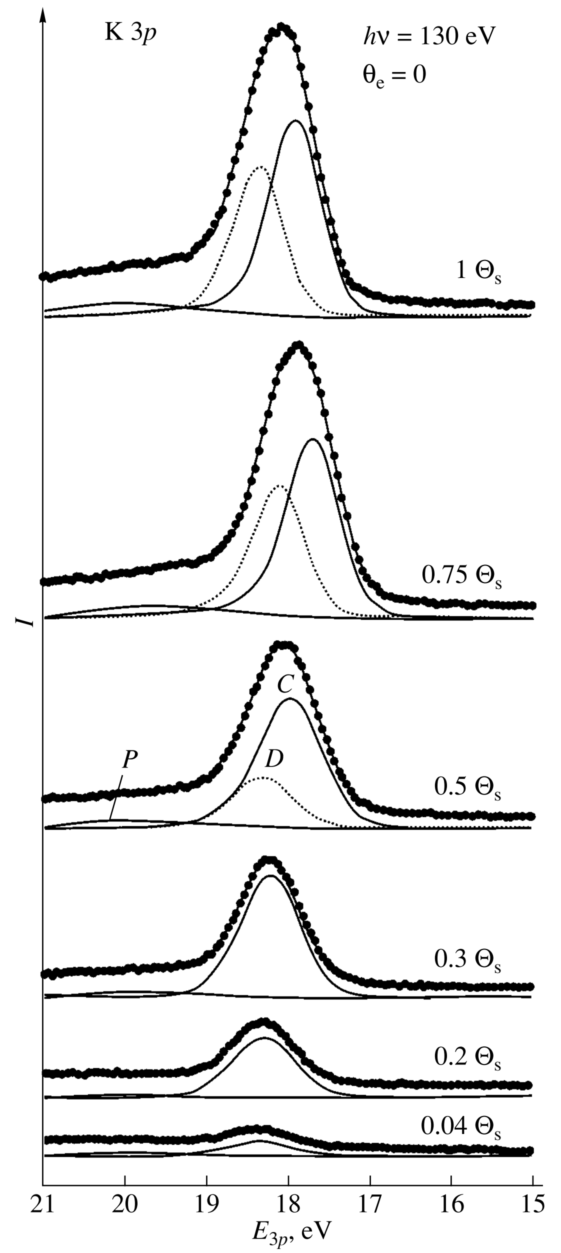


Fig. 18. Potassium photoelectron spectra taken when a potassium submonolayer was deposited on the Si(100)-2 × 1 system at room temperature and the decomposition into components *D* and *C* [94].

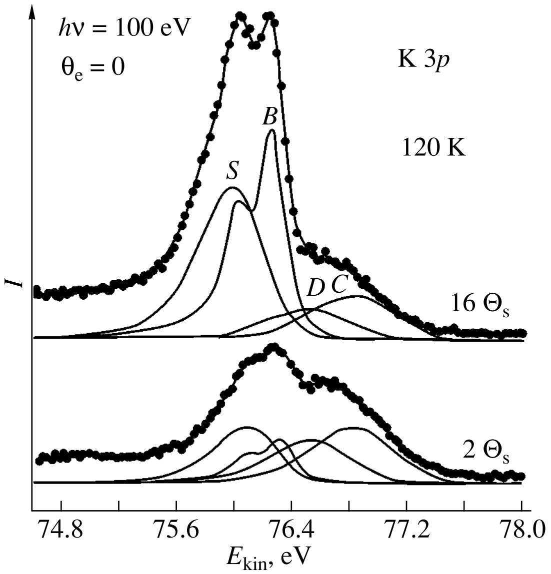


Fig. 19. Potassium photoelectron spectra taken when a thick potassium film was deposited on the Si(100)-2 × 1 system at low temperatures and the decomposition into components *D*, *C*, *S*, and *B* [95].

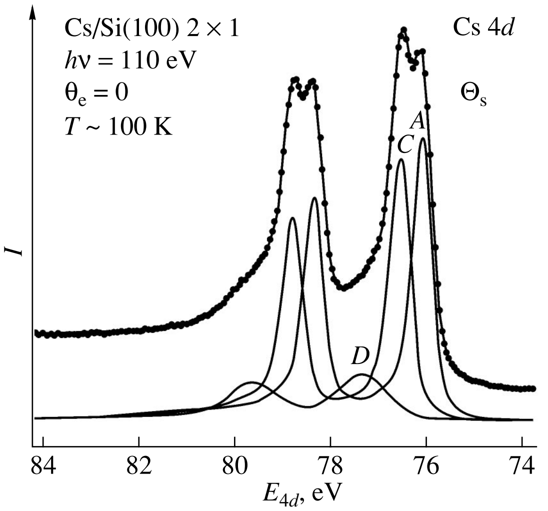


Fig. 20. Cesium 4*d* photoelectron spectrum taken when the saturated cesium coverage was deposited on the Si(100)-2 × 1 surface at room temperature and the decomposition into components [99].

the coverage thickness range (up to 20 ML). Taking into account that the surface modes of silicon persist, the authors of [95] relate modes *S* and *B* to the formation of metal islands on the potassium monolayer (according to the Stranski–Krastanov mechanism).

In the Cs/Si(100) system [99] (as in the K/Si system), the Cs 4*d* electron spectrum first (at the very beginning of adsorption) has a single component (*A*); then, the second one appears (*C*). When the coverage reaches saturation, the intensities of both modes become almost the same. The results are summarized in Fig. 20, where the low-temperature spectrum for the saturated coverage is shown.[3] It is seen that cesium atoms may be adsorbed on sites of two types, so that this system can also be described by the double layer

---

[3] It should be noted that the third component (*D*) of the Cs 4*d* spectrum, which exhibits a considerable positive energy shift (≈1 eV, Fig. 20), is due to photoelectron energy losses through plasmon excitation in the cesium layer adsorbed.

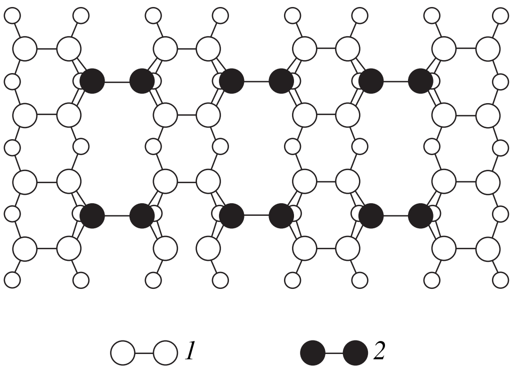


Fig. 21. Model of indium adsorption on the Si(100)-2 × 1 surface for a coverage of 0.5 ML. *1*, substrate dimers; *2*, adsorbate dimers.

model. However, mode S_u, which is associated with the asymmetric dimers, does not disappear from the 2*p* silicon electron spectrum in this case. At the early stage of Cs deposition, this mode broadens considerably, because the surface becomes irregular. However, it returns to its initial position (not counting an insignificant negative energy shift) as the coverage becomes one monolayer thick. Thus, cesium adsorption, unlike potassium adsorption, does not transform the asymmetric dimers into symmetric. This distinction is related to a decrease in the alkali atom–substrate bond strength as the atomic size of alkali metals increases (a well-known fact discovered by the method of thermal desorption spectroscopy) [132, 133].

Lithium atoms are much smaller than potassium and, especially, cesium atoms. Therefore, their interaction with the Si(100)-2 × 1 surface is much different. Specifically, a multilayer lithium coverage can be deposited on the silicon surface even at room temperature [81]. Moreover, when the coverage is less than one monolayer thick, the Li 1*s* electron spectrum has only one component; that is, only one type of adsorption site for adatoms is available. At the same time, the Si 2*p* spectrum exhibits a mode other than mode S_u, which is assigned to the upper atoms of the asymmetric dimers. When one lithium monolayer is formed on the surface, the intensity of the new mode roughly corresponds to one silicon monolayer. The authors of [81] assume that each of the dimers adsorbs two lithium atoms, which close its dangling bonds. Concurrently, the asymmetric silicon dimers become symmetric.

When the lithium coverage is thicker than one monolayer, the dimers collapse to form a lithium silicide. The appearance of new components in both the Li 1*s* and Si 2*p* spectra is an indication of the silicide formation reaction [81]. Indeed, as the amount of the lithium deposit grows, lithium-rich phases form. At low temperatures, the interaction of lithium atoms with the Si(100)-2 × 1 surface proceeds in a similar way. Silicide formation in this system seems to be associated with the small size of lithium atoms, which may penetrate into and react with the substrate. Note that lithium silicides also form when lithium is deposited on the Si(111)-7 × 7 surface [82].

Thus, the method of core-level spectroscopy allowed the researchers to determine the number of adsorption sites for each of the adsorbates in the alkali metal–silicon system, establish the sequence of occupying these sites as the coverage thickness grows, detect substrate surface reconstruction during the adsorption, reveal the trends in the adsorption process with increasing atomic number of the element deposited, and also detect the silicide formation reaction.

(ii) Adsorption of alkaline-earth metals. The adsorption of alkaline-earth metals on the Si(100)-2 × 1 surface will be demonstrated with barium [105]. In this case, the 2*p* electron spectrum of silicon retains the component S_u of the upper atoms of the dimers. This means that these atoms remain stable during barium adsorption. If, however, the coverage thickness approaches one monolayer, this component disappears and a new mode related to the symmetric dimers arises. When the amount of the barium deposit is large, the spectra of both the substrate and adsorbate exhibit extra modes due to the silicide phases of barium as in the case of lithium. Thus, the application of alkaline-earth metals gives rise to a new intriguing effect: the asymmetric substrate dimers transform into symmetric ones in a very narrow range of coverage thickness near one monolayer.

(iii) Adsorption of Group-III metals. The interaction of Group-III metals (Al, Ga, and In) is also to some extent specific. It was shown (by methods other than core-level spectroscopy) that atoms of these metals are adsorbed on the Si(100)-2 × 1 surface as dimers, which are aligned between the chains of substrate dimers, thereby producing a double dimeric layer. For coverages of one-third, two-fifths, and half of a monolayer, such adsorbed layers show (2 × 3), (2 × 5), and (2 × 2) diffraction patterns, respectively. The arrangement of the substrate and adsorbate dimers in this situation (whether they are parallel or orthogonal to each other) has been the subject of wide speculation. Theoretical and experimental results [134, 135] have confirmed the validity of the second model, which is schematically shown in Fig. 21 for the (2 × 2) structure.

The core-level spectroscopy data for the systems considered will be illustrated for the (2 × 3) and (2 × 2) structures appearing on the Si(100)-2 × 1 surface upon indium adsorption [112]. Typical In 4*d* electron spectra are depicted in Fig. 22. Note that the energy $h\nu$ = 108 eV, which was used to take these spectra, is appropriate for recording the surface-sensitive spectrum of indium. The presence of only one doublet in the spectra is a direct indication of the fact that all adsorbate atoms occupy adsorption sites of one type in both cases. The structures are distinguished only by a small energy shift of 4*d* levels. Comparing the 2*p* spectra (Fig. 23) for clean silicon and for the (2 × 3) structure, one can see

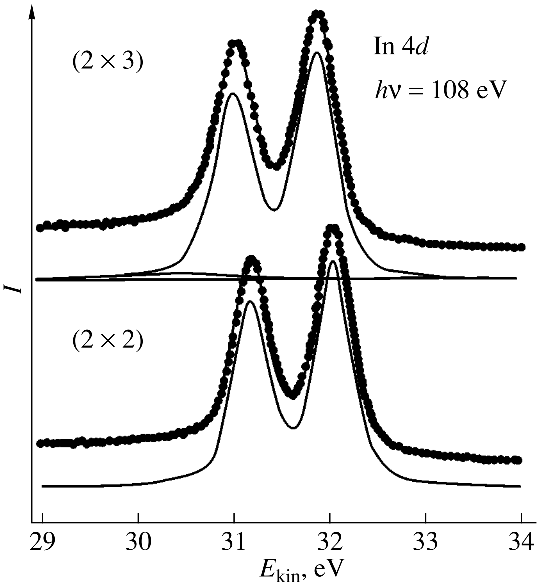


Fig. 22. Indium 4*d* photoelectron spectra for the In/Si(100)-2 × 3 and In/Si(100)-2 × 2 structures (upper curves) and the result of simulation after subtracting the background (lower curves) [112].

that, when the latter forms, the component S_u of the upper atoms of the asymmetric dimers decreases drastically and new component *S* appears between S_u and the bulk mode *B* of silicon. In the (2 × 2) structure, mode S_u is absent, while component *S* becomes much more intense. By analogy with the systems considered above, these findings are treated in terms of the adsorption-induced transformation of the asymmetric silicon dimers into symmetric ones. The absence of mode S_u for the (2 × 3) structure is explained by the fact that, when the coverage equals 1/3 ML, some of the dimers do not adsorb indium atoms and the asymmetric configuration of these dimers holds. When the coverage equals 0.5 ML, all Si dimers adsorb indium atoms (Fig. 21), becoming symmetric. In this case, all the dangling bonds of substrate atoms on the surface close on the indium atoms. Thus, indium adsorption also initiates atomic reconstruction on the substrate and indium atoms occupy adsorption sites of only one type in both systems studied.

(iv) Adsorption of transition metals. In this case, the situation on the substrate surface becomes much more complicated. A characteristic feature of the transition metal–silicon interface is its high reactivity. Under certain conditions, this property shows up in adsorbate–substrate interdiffusion and in the formation of silicides. Let us consider the interaction of cobalt, a 3*d* metal, with the Si(100)-2 × 1 surface [115–123].

The final product of the cobalt–silicon reaction is cobalt disilicide. The similarity of the Si and $CoSi_2$ crystal structures, as well as the fact that their lattice constants are close to each other, makes it possible to grow $CoSi_2$ epitaxial layers on silicon with an abrupt and perfect interface. Since $CoSi_2$ offers metallic properties, such systems, on the one hand, serve as a model in studying metal–semiconductor contacts and, on the other hand, are finding wide application in solid-state electronics as ohmic and barrier contacts; hence, particular emphasis on them is seen today. Although these systems have been the subject of extensive research, the formation of the $CoSi_2$/Si interface is not clearly understood. This is especially true for the initial stage of $CoSi_2$ growth in the Co/Si(100) system, which is of special interest for process designers. Specifically, it remains unclear whether $CoSi_2$ nucleates at room temperature within a monolayer. The method of core-level spectroscopy sheds light on this question.

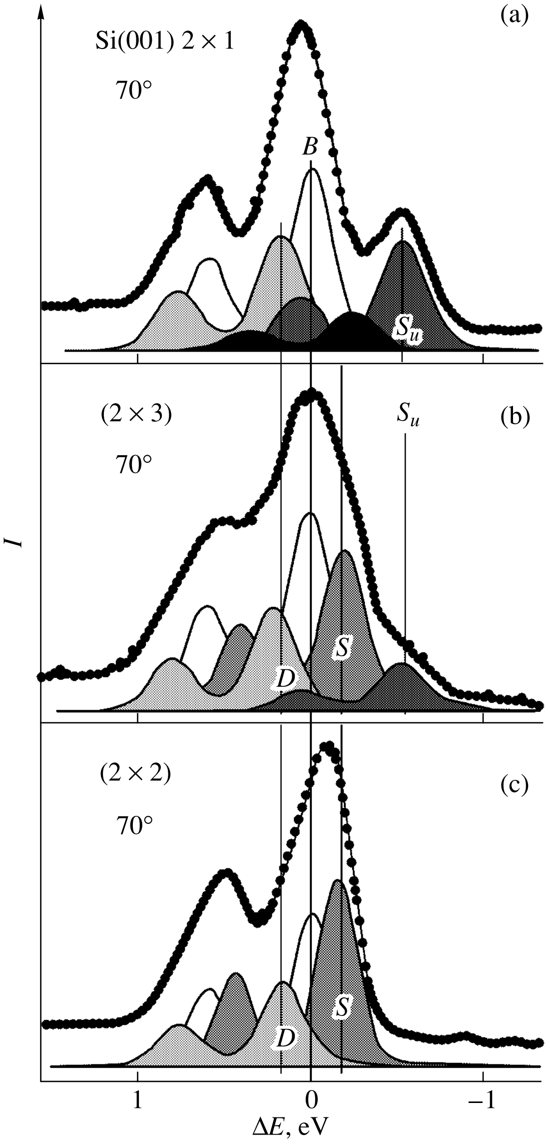


Fig. 23. Silicon 2*p* photoelectron spectra for (a) the clean Si(100)-2 × 1 surface, (b) the In/Si(100)-2 × 3 structure, and (c) the In/Si(100)-2 × 2 structure and the decomposition of the spectra into components [112].

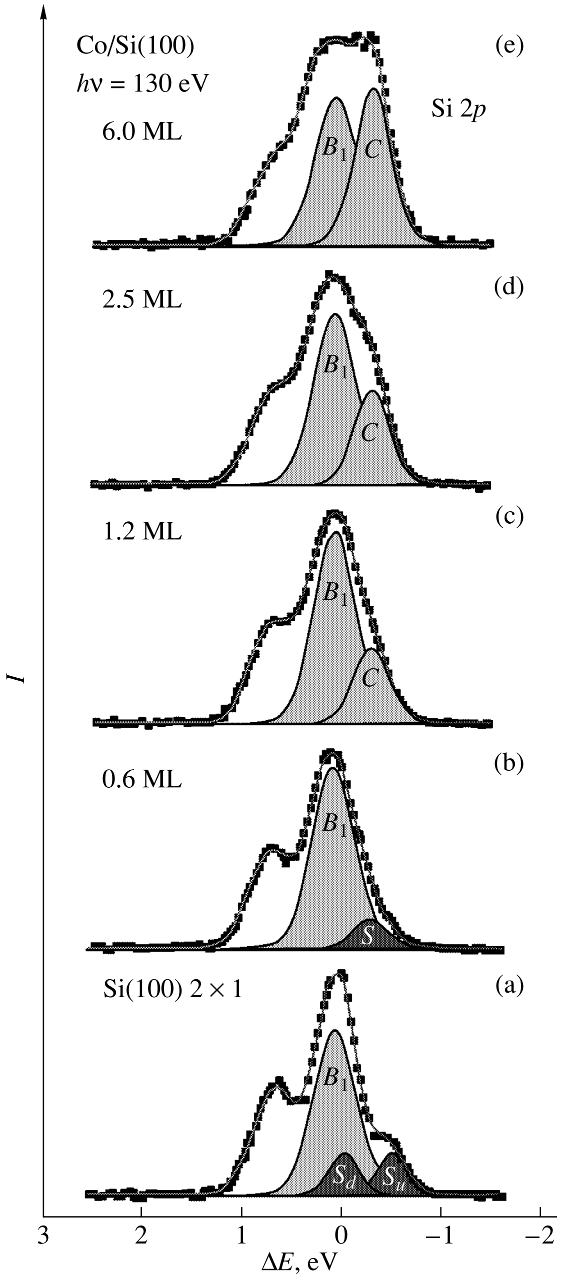


Fig. 24. Variation of the silicon 2p electron spectrum for the Si(100)-2 × 1 surface with increasing Co dose at room temperature and the decomposition into components [121]. (a) Clean surface, (b) 0.6 ML Co, (c) 1.2 ML Co, (d) 2.5 ML Co, and (e) 6 ML Co.

Various $CoSi_2$ phases are characterized by different shifts of the Si 2p lines relative to the lines of pure silicon. For example, the $CoSi_2$ stable phase (with the CaF_2 structure) exhibits a positive energy shift of ≈250 meV [117–119, 123]. A still larger shift is observed for the $CoSi_2$ metastable phase (with the CsCl structure): 350–400 meV relative to the stable phase [136]. Negative shifts were found for Si atoms in the Co–Si solid solution [115–119, 121, 123], in the so-called adamantine silicide [137], and in amorphous cobalt silicide [118, 123]. It was also shown that different faces of the $CoSi_2$ stable phase have different surface components [138].

To demonstrate the formation of the $CoSi_2$/Si(100)-2 × 1 interface during solid-state epitaxy (cobalt atoms are deposited on the substrate surface at room temperature, and the system is then annealed), consider the results reported in [120, 121]. The variation of the Si 2p spectra at both stages of the process is shown in Figs. 24 and 25. These figures also show the decomposition of the spectra into components.[4] It is seen that the adsorption of 0.5–0.6 ML of cobalt completely quenches surface modes S_u and S_d and gives rise to mode S. Mode S is of low intensity and so cannot be assigned to symmetric dimers, into which the asymmetric dimers of the substrate might transform under the action of the adsorbate (Fig. 24b). Unlike the cases discussed above, this effect is explained by the fact that the silicon surface is no longer reconstructed: chemisorbed cobalt atoms are built in the upper layer of silicon atoms and saturate their dangling bonds. The (2 × 1) diffraction pattern in the submonolayer range persists, suggesting that the cobalt atoms are aligned in the same manner as the substrate dimers. These data convincingly demonstrate that silicides do not form at room temperature.

As the adsorbate thickness increases, mode B_1 decreases and component S is substituted for by new component C, which features a negative shift of −300 meV. Subsequently, the intensity of the new mode grows. With regard to the shifts of core levels in the various phases of cobalt disilicide (see above), the appearance of this mode may indicate the formation of the Co–Si solid solution. The solid-state reaction between Si and Co to produce cobalt disilicide starts at ≈250°C. This follows from the appearance of mode D after annealing (Fig. 25), which is characterized by a positive shift typical of cobalt disilicide. The disilicide formation reaction is complete at about 350°C, when the surface component S_D of cobalt disilicide becomes detectable. Finally, heating of the system to 600°C again gives rise to the initial spectral components, S_u and S_d. This means that the substrate surface has become partly uncovered and that $CoSi_2$ islands (rather than a continuous film) are growing.

Consider also the interaction of cobalt with the Si(111)-7 × 7 surface. Here, the component with the positive energy shift, which is typical of $CoSi_2$, is observed even at the very beginning of cobalt deposition at room temperature [115]. This reaction slows down only at high coverages, when a film of the Co–Si solid solution starts growing over the silicide layer. As for the Si(100)-2 × 1 surface, this film transforms into

---

[4] For clarity, only components S_u and S_d, which are the most sensitive to adsorption, are shown in Fig. 24a, although five surface modes were found in the spectrum of pure silicon in accordance to [33, 34]. The other modes are combined into mode B_1.

an epitaxial layer of cobalt disilicide upon annealing as a result of the solid-phase reaction.

A somewhat different scenario of the process for the Co/Si(111) system was observed in [122, 123]. The room-temperature silicide phase forming at coverages of ≈1 ML here is identified with amorphous silicide. The phase that forms upon annealing is considered as metastable cobalt silicide with a CsCl structure. Upon subsequent heating to ≈500°C, the metastable silicide passes to the stable phase of $CoSi_2$.

The distinctions between the results obtained by different authors for the Co/Si system are likely to be explained by the complexity of the silicide formation process, which is sensitive to many factors, specifically, to the initial condition of the single-crystal silicon surface [139].

4.2. Interaction of Metals with the Si(111)-7 × 7 Surface

Interest in the adsorption properties of the Si(111)-7 × 7 surface was accelerated in the 1990s, when the researchers stated investigating the atomic configuration of (n × 1)-like structures, which were detected in the submonolayer range of metal deposits. It was found that, unlike the Si(100)-2 × 1 surface, metal adsorption on the Si(111)-7 × 7 surface may radically change its atomic configuration, specifically, cause quasi-one-dimensional chains of silicon atoms to form. The self-organization of these chains is of special interest, since it offers possibilities for studying the physical properties of 1D structures, which may be promising for applications. Some of these studies will be the focus of our consideration.

(i) Adsorption of alkali and alkaline-earth metals. A great deal of publications are devoted to the Si(111)-3 × 1 structure, which appears when metals are adsorbed on the Si(111)-7 × 7 surface at elevated temperatures or at room temperature with subsequent heating. This structure was first observed upon alkali metal adsorption [140]. Later, it was found that this structure passivates the silicon surface, preventing its oxidation [141]. On the other hand, it is well known that the room-temperature adsorption of an alkali metal on Si(111) increases the oxidizability of silicon [142]. The coverage corresponding to this surface structure equals 1/3 ML [143, 144].

(3 × 1) structures were also observed when alkaline-earth (Mg [145], Ca [146], and Ba [147]), noble (Ag [148]), and rare-earth [124, 125, 128] metals were deposited on the Si(111)-7 × 7 surface. It was shown using LEED that the I–V curves for the different adsorbates are similar to each other [145, 149]. Also, it was found by angular-resolution photoelectron spectroscopy that the electronic configuration of surface states is almost independent of the metal deposited and that the (3 × 1) surface has semiconductor properties [83, 87, 101, 150]. Nearly identical images of this surface covered by various adsorbates were obtained by STM [151–153]. Finally, the $2p$ spectra of core electrons for silicon covered by various metals also turned out to be very much alike [83–85, 87]. The observations listed suggest that, in essence, we are dealing with the same structure whose properties depend on the configuration of the substrate surface rather than on the adsorbed layer structure. The adsorbate in this case stabilizes this structure.

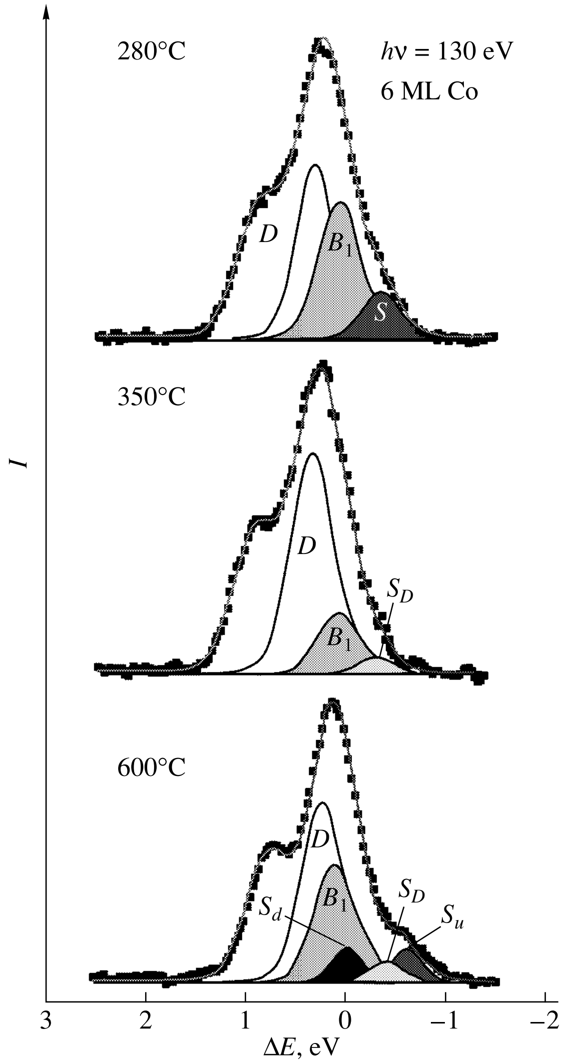


Fig. 25. Silicon $2p$ electron spectra for the Si(100)-2 × 1 surface coated by a 6-ML-thick cobalt film and subjected to step annealing and the decomposition of the spectra into components [121].

A number of models have been put forward to describe the structure being discussed. However, the only model that is in complete agreement with the experimental findings appears to be that where quasi-one-dimensional chains of atoms are considered as building blocks of the surface. This model relies on the fact that some of the properties of the (3 × 1) structure and reconstructed Si(111)-2 × 1 surface arising upon cleavage of Si(111) single crystal in ultrahigh vacuum

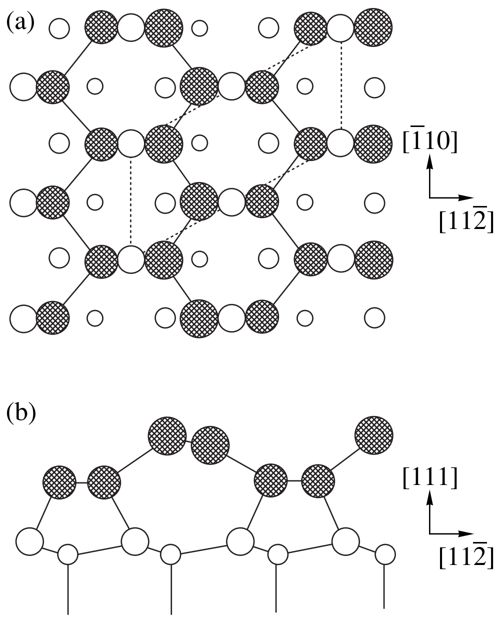


Fig. 26. Chain model of the atomic configuration of the metastable Si(111)-2 × 1 surface [154]. (a) Top view and (b) side view. The atoms of the surface layer reconstructed are hatched. The unit cell on the surface is outlined by the dashed line.

are alike. The configuration of this metastable silicon surface was explained earlier through the formation of zigzag chains (Fig. 26) in the ⟨110⟩ directions from two upper silicon monolayers, the atoms of these chains linked by π bonds [154].

Several models for the (3 × 1) structure were suggested. The honeycomb chain–channel model (HCC model) [155–157] is today the commonly accepted one (Fig. 27). Surface silicon atoms *1–4* produce honeycomb chains lying in the plane parallel to the surface in the ⟨110⟩ directions. Outer atoms *1* and *4* have the sp_3 tetrahedral configuration, while the configuration of inner atoms *2* and *3* is dehybridized into the sp_2 configuration and p_z orbital. The sp_2 bonds localized in the plane of the surface are responsible for the formation of two interrelated zigzag chains of Si atoms. This quasi-one-dimensional structure is stabilized with π bonds appearing between atoms *2* and *3* owing to the interaction of p_z orbitals. Concurrently, the bonds of these atoms with atom *5* in the underlying layer loosen. The metal atoms, which are ionized and localized in the channels between adjacent honeycomb chains, also produce chains.

Consider now the results obtained for the (3 × 1) structure by core-level spectroscopy. While in the early works only two surface modes of the Si 2*p* spectra were detected [83, 86, 101], the authors of recent work [96], concerned with potassium–silicon interaction at 100 K, five surface components were distinguished. The related volume- and surface-sensitive spectra are depicted in Fig. 28. The resolution of one of the spectra is shown in Fig. 29. Three negative-shift components (S_1, S_2, S_4) and two positive-shift ones (S_3, S_5) are present in the spectrum. Modes S_1 and S_2 markedly grow with increasing angle of departure of the photoelectrons. Conversely, modes S_3–S_5 weaken. According to the HCC model (Fig. 27), surface-sensitive modes S_1 and S_2 are related to atoms *1* and *4* of the unit cell on the surface. The energy shifts of these modes are negative, because the charge is transferred from the potassium

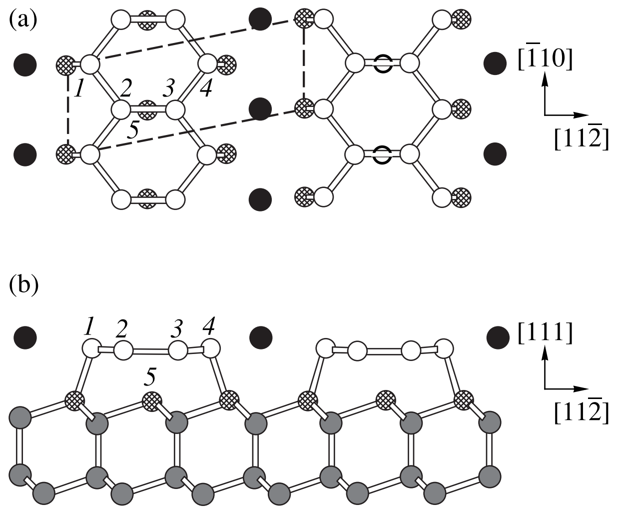


Fig. 27. HCC model for the metal/Si(111)-3 × 1 surface [155]. (a) Top view and (b) side view. Adsorbate atoms are shown by filled circles. The unit cell on the surface is outlined by the dashed line.

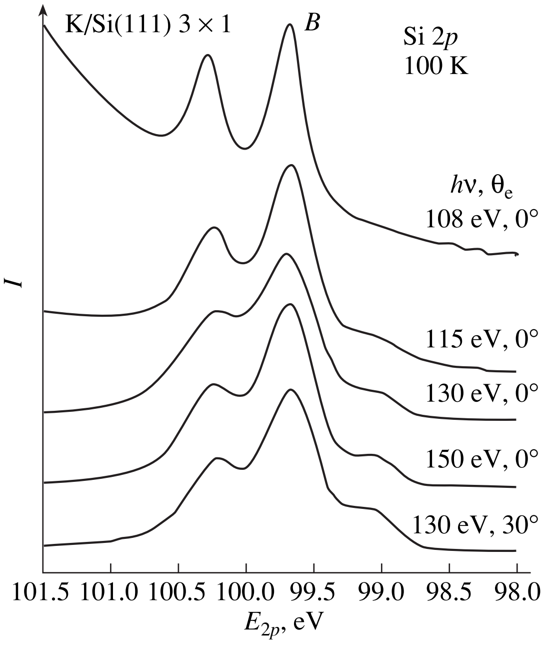


Fig. 28. Silicon photoelectron spectra for the K/Si(111)-3 × 1 structure at different photon energies [96].

atoms to the dangling bonds of atoms *1* and *4*. However, atom *1* interacts with one potassium atom, while atom *4*, with two potassium atoms. The configurations of these atoms differ from each other; hence, the difference in the energy shifts of the modes considered [157]. The intensities of modes S_1 and S_2 are nearly the same, which is consistent with the above approach. Atoms *2* and *3*, which are in the sp_2 hybridized state and linked by a π bond, have an excess charge and are expected to have a positive energy shift of the 2*p* level: ≈370 meV as estimated by the authors of [157]. The energy shift of mode S_3 is close to this value. However, the dependence of its intensity on the angle of departure is contradictory to such an interpretation. It was therefore proposed that the atoms of the second and third monolayers of the crystal also contribute to mode S_3. The large half-width of component S_3, which exceeds the half-widths of the other spectral components, counts in favor of its complex nature. The remaining components, S_4 and S_5, are assigned to those atoms of the second and third monolayers not contributing to component S_3.

The situation is still more complicated when alkaline-earth metals, specifically, calcium, are adsorbed on the silicon surface. Although the (3 × 1) structure is also typical of calcium coverages [146], it is stable only at elevated temperatures, as follows from more recent papers. As the temperature drops to the room value or below, the (3 × 1) structure configures into the (3 × 2) or *c*(6 × 2) structure [158]. The 2*p* spectra of these structures were found to be nearly the same as those for the (3 × 1) structure in the case of the systems with the alkali metals; hence, they also can be treated in terms of the HCC model [103]. Indeed, the STM data showed that the quasi-one-dimensional chains of silicon atoms in these structures are arranged in the same way as in the systems with the alkali metals. The period of the (3 × 2) structure is twice as large as that of the (3 × 1) structure in the ⟨110⟩ directions, since the concentration of the Ca atoms is half the concentration of alkali metals (1/6 ML versus 1/3 ML). The value 1/6 ML was obtained in [103]. The occurrence of the *c*(6 × 2) structure seems to be associated with the displacement of the Ca atoms in neighboring recesses. As the temperature rises, the regular arrangement of the Ca atoms breaks and they do not produce a diffraction pattern. Therefore, the (3 × 1) structure observed under these conditions is due to substrate surface reconstruction alone. Similar conclusions were drawn in [159, 160], where the (3 × 2) structure was studied on the silicon surface coated by barium and magnesium, respectively.

In [130], the (3 × 2) structure was studied together with other structures observed for submonolayer calcium coverages: (5 × 2), (7 × 2), (9 × 2), and (2 × 1). These structures sequentially give way to each other as the coverage thickness grows. The results are illustrated in Fig. 30. Each of the structures has its own 2*p* spectrum reflecting its specific atomic configuration.

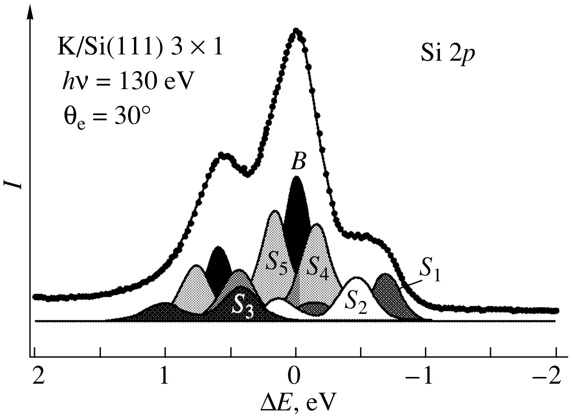


Fig. 29. Silicon photoelectron spectra for the K/Si(111)-3 × 1 structure and the decomposition into components [96].

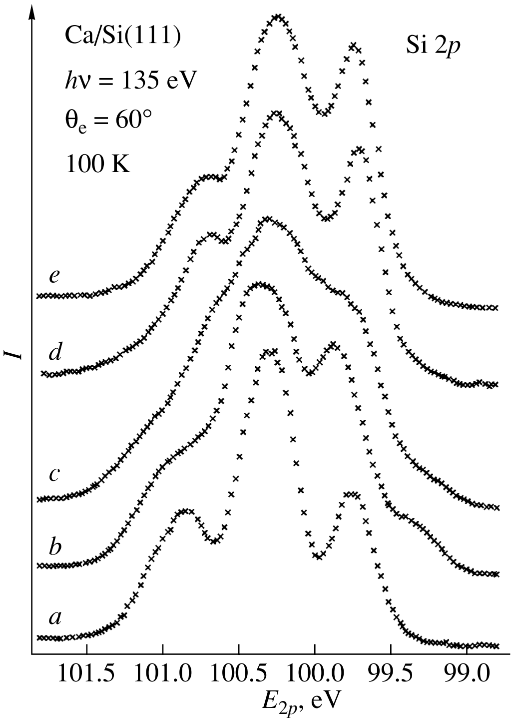


Fig. 30. Silicon photoelectron spectra for the (a) Ca/Si(111)-3 × 2, (b) Ca/Si(111)-5 × 2, (c) Ca/Si(111)-7 × 2, (d) Ca/Si(111)-9 × 2, and (e) Ca/Si(111)-2 × 1 structures [103].

Let us consider the (2 × 1) structure more closely. This structure, arising when half a calcium monolayer is deposited on the silicon surface, was first studied in [104]. The associated results are demonstrated in Fig. 31.The Si 2*p* electron spectra are seen to exhibit three surface components S_1, S_2, and S_3. They are treated in terms of the Seiwatz model [161], which is a simpler version of the model of quasi-one-dimensional chains.[5] The atomic arrangement on the silicon surface is sche-

---

[5] Earlier, this model was suggested to be deposited for the identification of the (3 × 1) structure [85, 104, 144].

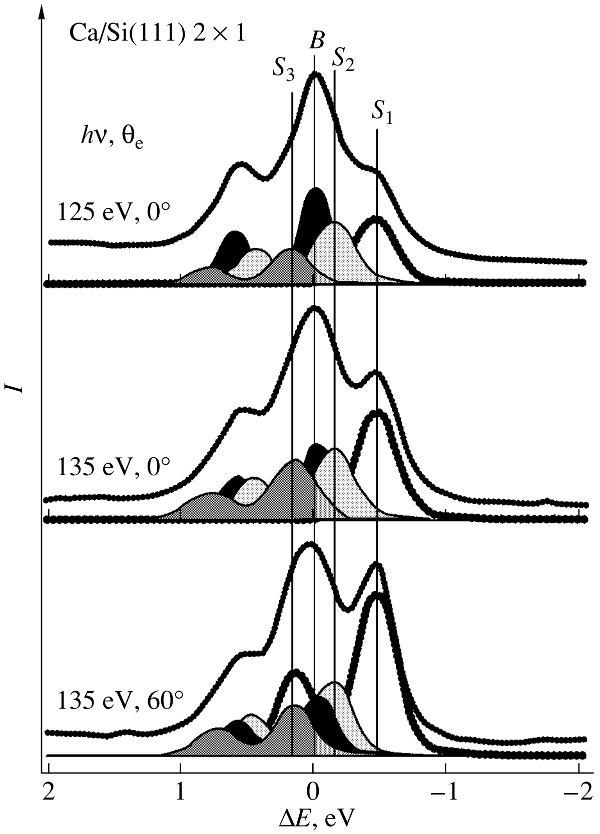


Fig. 31. Surface (S_1, S_2, S_3) and bulk (B) modes of the silicon photoelectron spectra for the (2 × 1) structure that was observed upon calcium adsorption on the Si(111)-7 × 1 surface [104].

matically shown in Fig. 32. By analogy with the HCC model, modes S_1 and S_2 with negative shifts are assigned to atoms *1* and *2* of the silicon chains. These atoms are negatively charged because of charge transfer to their dangling bonds from the adsorbate atoms. The configurations of the bonds between atoms *1* and *2* and the calcium atoms are also different in the framework of this model; hence, different energy shifts of modes S_1 and S_2. The third surface component (S_3) with a positive energy shift is due to atoms *3* of the second monolayer, which are located immediately below the adatoms (Fig. 32).

(ii) Adsorption of Group-III metals. The most dramatic example of self-organization of the quasi-one-dimensional chains on the Si(111)-7 × 7 surface is the (4 × 1) structure, which is observed when an indium monolayer is adsorbed on this surface. The detailed study of this structure by X-ray diffraction [162] showed that it is formed by four rows of the adsorbate atoms placed between the zigzag chains of silicon atoms (Fig. 33). Calculations based on the first principles [163] confirmed the validity of this model.

The investigation of the (4 × 1) structure by core-level spectroscopy was performed in [112, 114]. The 4*d* electron spectra for indium and the 2*p* electron spectra for silicon that were obtained in [114] at room temperature and low temperatures are shown in Fig. 34. At room temperature, the indium spectrum has two surface modes of nearly equal intensities (α, β), whose energy shifts differ by 500 meV. The presence of these two modes indicates that the indium adatoms occupy two types of adsorption sites and are in different electronic states. The states with a lower binding energy (mode β) are assigned to the indium atoms of the inner chains

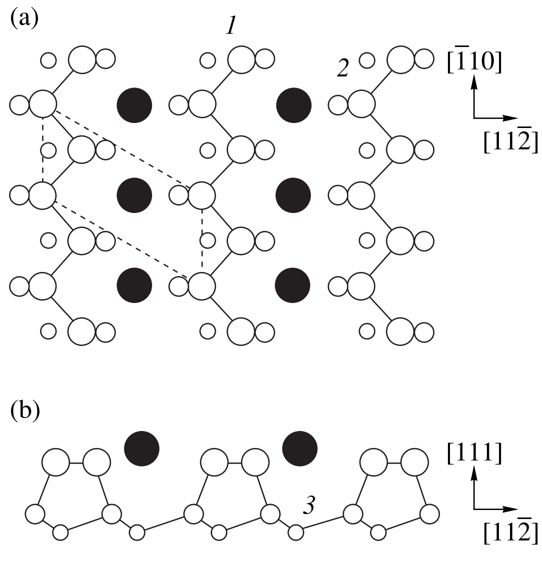


Fig. 32. Atomic arrangement on the Ca/Si(111)-2 × 1 surface [104]: (a) top view and (b) side view. Empty circles, Si atoms; filled circles, Ca atoms. The unit cell on the surface is shown by the dashed line.

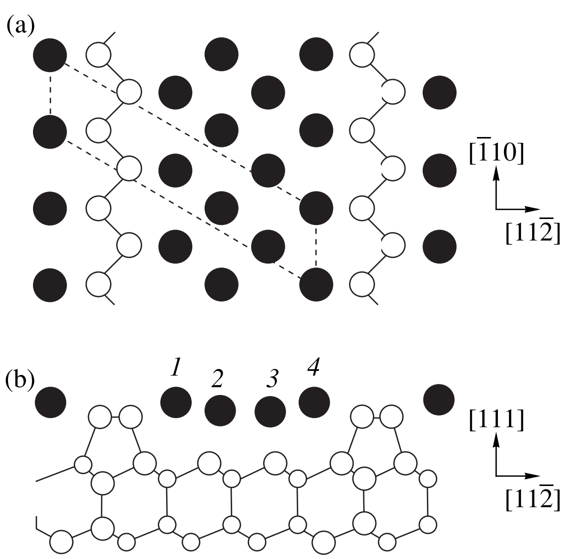


Fig. 33. Structural model of the In/Si(111)-4 × 1 surface [163]: (a) top view (shown are only the atoms of the upper monolayer) and (b) side view. Empty circles, Si atoms; filled circles, In atoms. The unit cell on the surface is shown by the dashed line.

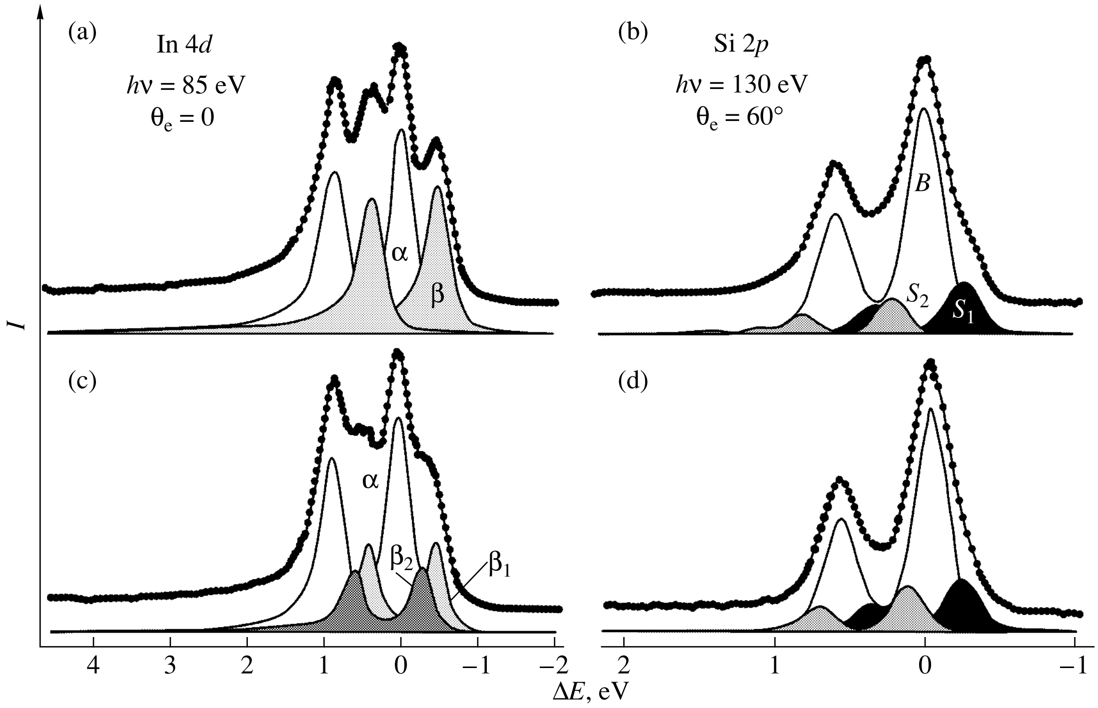


Fig. 34. (a, c) Indium and (b, d) silicon photoelectron spectra taken from the In/Si(111)-4 × 1 structure at (a, b) room temperature and (c, d) 70 K and the decomposition into components [114].

(atoms *2, 3*), while the states associated with mode α are due to the atoms of the outer chains (atoms *1, 4*). The noticeable asymmetry of these spectral lines suggests that these low-dimensional structures are of metallic conduction. The higher symmetry of mode β means that the inner chains have more pronounced metallic properties than the outer ones, which are in contact with silicon atoms. The lower binding energies observed in this case are explained by intense relaxation processes during the photoexcitation of the core levels of inner-chain atoms.

The Si2*p* electron spectra also exhibit two surface components (S_1, S_2), which are characterized by a negative and positive energy shift, respectively (Fig. 34c). The authors of [114] relate the former mode to the silicon atoms that produce the zigzag surface chains; the origin of the latter was not discussed.

Cooling of the crystal splits the component β of the In 4*d* spectrum into modes β_1 and β_2 separated by an interval of 200 meV. These modes are fairly asymmetric (Fig. 34b). Mode α changes insignificantly. The Si 2*p* spectrum also changes slightly (Fig. 34d). Unfortunately, the effect of cooling remained unexplained. At the same time, it was noted that a decrease in the temperature of this system stimulates the structural phase transition showing up in doubling the period along the atomic chains, generating charge-density waves, and appearing an energy gap at the Fermi level [164], which is typical of 1D metallic systems.

(iii) Adsorption of noble metals. The adsorption of noble metals, specifically, silver and gold, on the Si(111)-7 × 7 surface causes a variety of surface structures. The corresponding Si 2*p* electron spectra are, as a rule, rather complex and contain several components (see Fig. 35 for gold). The spectra of the surface structures differ greatly both in the number of modes and in their energy position and intensity, reflecting the difference in the atomic and electronic configurations.

In the case of silver, the (3 × 1) structure considered above coexists with another structure, $(\sqrt{3} \times \sqrt{3})$–*R*30°, which was also explored thoroughly by various techniques (see, e.g. [108]), core-level spectroscopy among them. In general, the data for noble metals, along with the above data for the other metals, are further evidence for the idea that the binding energy of the core electrons of surface silicon atoms and the binding energy of adatoms are highly sensitive to the microscopic properties of low-dimensional structures that form on the silicon surface.

5. ADSORPTION OF GASES ON SILICON

Core-level spectroscopy also provides ample information about gas adsorption on the single-crystal silicon surface. It gave a clearer insight into the interaction of silicon with hydrogen [13, 165–168], oxygen [169–188], chlorine [67, 189, 190], acetylene [191], water vapor [186, 192], and other molecules. In some of these systems, considerable energy shifts of the Si 2*p* spectral

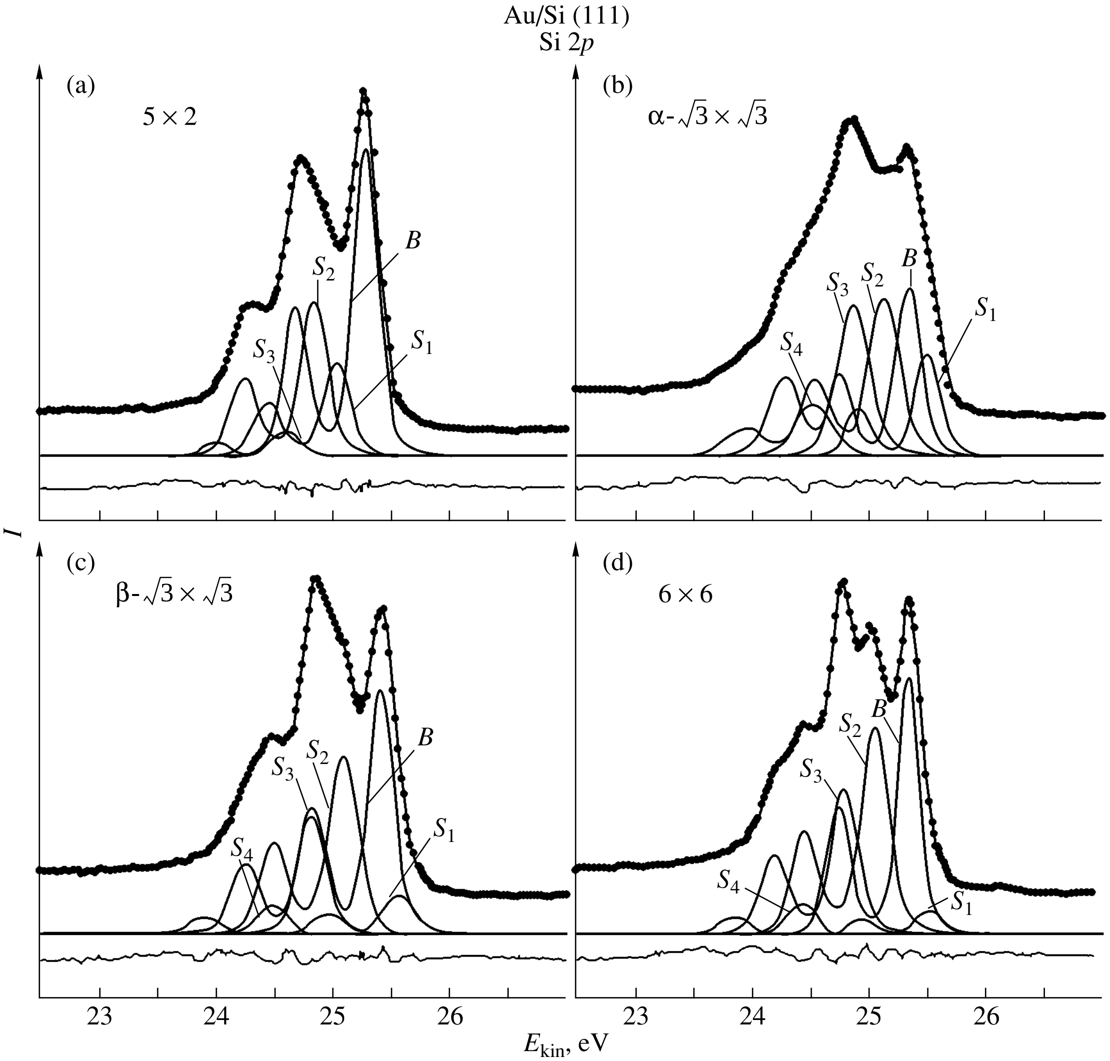


Fig. 35. Silicon photoelectron spectra for the (a) Si(111)-5 × 2, (b) Si(111)-α-√3 × √3, (c) Si(111)-β-√3 × √3, and (d) Si(111)-6 × 6 structures formed on the Si(111)-7 × 7 surface upon the deposition of gold and the decomposition into components [111].

lines may be observed because of a great difference in the electronegativities of the adsorbate and silicon atoms, which simplifies the identification of surface modes. This effect is the most dramatic in the case of oxygen, and oxygen adsorption on silicon will be considered first of all.

(i) Oxygen adsorption on the Si(100)-2 × 1 surface. As is known, oxygen–silicon interaction produces a film of silicon dioxide (SiO_2). The SiO_2/Si structure is a basic building block of advanced solid-state electronics and has been the subject of much investigation (see, e.g., [169, 170]) with emphasis on the SiO_2/Si(100) structure. To date, it has been established that an intermediate oxide layer of complex composition arises between Si and SiO_2. In spite of extensive studies in this field, the structure and stoichiometric composition of this layer are as yet imperfectly understood, and the formation of the SiO_2/Si interface continues to be a central preoccupation of the researchers. The method of core-level spectroscopy, which is capable of detecting various oxidation states of silicon atoms, has proved to be an efficient tool for studying the early stage of silicon oxidation. It should be noted, however, that the data obtained by different authors are sometimes contradictory.

The early stage of the process was investigated at low temperatures [177, 179, 180], room temperature [172, 181, 184, 185], and elevated temperatures [173, 176, 178, 181–183, 186–188]. In the first two cases, which are primarily of purely scientific interest, silicon single crystals were usually subjected to subsequent annealing. High-temperature oxidation of the silicon surface is routinely used in the semiconductor technology to prepare SiO_2/Si(100) structures; so, investigation of this process is of great practical value.

The behavior of the Si(100)-2 × 1 surface under the conditions of high-temperature oxidation will be illustrated with the data obtained in one recent work [183] (Fig. 36). The oxidation was performed at 600°C and an oxygen pressure of 1×10^{-10}–5×10^{-9} Torr. The total exposure of the specimen to oxygen reached 100 langmuirs (L). Such an exposure provided the formation of the saturated (under the given conditions) coverage of thickness 13 Å (as estimated in [183]). As follows from Fig. 36, the component S_u of the upper atoms of the substrate dimers fades out early in the process. At the same time, three new oxygen-induced modes (*2*–*4*) arise. They show positive energy shifts up to 4 eV. The greater the energy shift of the mode, the higher its intensity, and such a correlation persists with increasing exposure.

The spectral modes considered, which were first detected and identified in [171], are associated with different oxidation states of the silicon atoms. The positive sign of their energy shifts is due to Si–O ionic bonds, which form when the charge is transferred from the silicon atoms to the oxygen atoms adsorbed. The degree of oxidation of the silicon atoms is responsible for the chemical shift value. Four oxidation states of silicon atoms are known, where the charge varies from +1 to +4. Accordingly, the atoms are bonded to one, two, three, or four oxygen atoms. The first three oxide phases have an intermediate valence and are called suboxides, and the phases where silicon atoms are tetravalent is called silicon dioxide. Possible different-valence oxidation states of silicon that may arise early in the oxidation of the Si(100)-2 × 1 surface are shown in Fig. 37. Although only three oxidation modes (states) are seen in Fig. 36, this spectrum also contains the fourth component, which corresponds to the Si^{1+} state. This mode becomes distinct after computer decomposition of the spectrum (Fig. 38).

The fact that the four components due to different oxidation states are revealed even at the early stage of oxidation directly testifies that the oxide layer growing on the surface is of complex composition. On the other hand, the virtual constancy of the intensity ratios for the different oxide phases and fading-out of the mode S_u of the substrate during the oxidation process are a demonstration that the arising oxide complexes rapidly reach their ultimate thickness and then spread over the surface. A characteristic feature of the forming oxide layer is its inhomogeneity across the depth [183]: the nearer silicon atoms are to the surface, the higher their oxidation state.

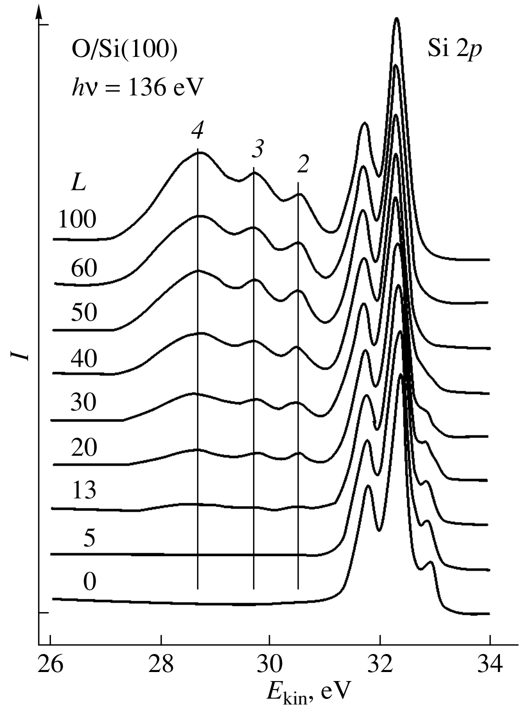


Fig. 36. Silicon photoelectron spectra taken during thermal oxidation of the Si(100)-2 × 1 surface [183]. The numbers by the curves indicate oxide components: (*2*) Si^{2+}, (*3*) Si^{3+}, and (*4*) Si^{4+}.

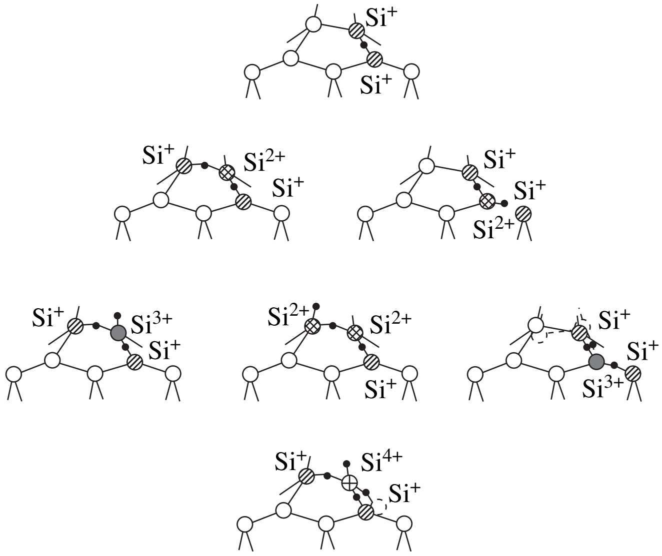


Fig. 37. Silicon oxide states that may form at the early stage of oxidation of the Si(100)-2 × 1 surface [184].

Similar results were obtained in other works concerned with thermal oxidation of silicon. The energy shifts reported for various oxidation states of silicon atoms are in fairly good agreement. In [182], the Si(100)-2 × 1 surface was oxidized in oxygen at 600°C and an exposure of 150 L, the intensities of the spectral components associated with different oxidation states against the polar angle of departure of the photoelectrons were studied. This dependence was found to become more pronounced as the degree of oxidation increases (Fig. 39). The close examination of these results made it possible to determine the depth profiles of the oxide phases. It was shown that the intermediate oxide layer has a graded structure and consists of three sublayers. The one nearest to the silicon incorporates

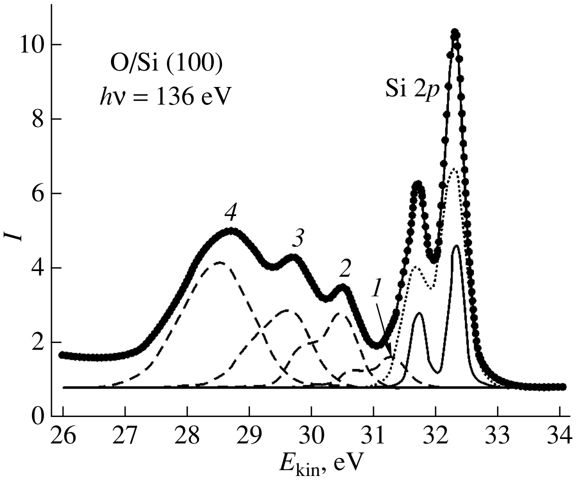


Fig. 38. Resolution of the Si 2p electron spectrum taken from the oxidized silicon surface (T = 600°C, exposure 100 L) [183]. The numbers by the curves indicate oxide components: (*1*) Si^+, (*2*) Si^{2+}, (*3*) Si^{3+}, and (*4*) Si^{4+}.

the suboxides of Si^{1+} and Si^{2+}, while the constituents of the second and third sublayers are the suboxides of Si^{3+} and Si^{4+}. Each of the sublayers has a specific concentration of the related oxide components. The structural model of the SiO_2/Si(100) interface that was suggested in [182] based on these findings is schematically shown in Fig. 40. Note that several structural models of this interface were also proposed in theoretical works [192–196]. However, they are in poor agreement with the core-level spectroscopy data, according to [182].

The low-temperature adsorption of oxygen on the silicon surface is much alike the thermal oxidation of silicon. Figure 41 refers to the initial stage of oxygen adsorption (0.1–0.5 ML) on the Si(100)-$c(4 \times 2)$ surface. As in the case considered above, the adsorption of oxygen suppresses the mode S_u of the upper atoms of the substrate dimers and generates new modes with positive energy shifts. However, the ratios of their intensities differ from the previous case. For example, the component due to Si^{4+} is hardly noticeable, while that due to Si^{1+} markedly grows. The intensities of the Si^{2+} and Si^{3+} modes are close to each other (which is typical of low-temperature adsorption), and their energy shifts decrease by 0.1 and 0.2 eV, respectively. The spectra taken from the Si(100)-2 × 1 surface oxidized at room temperature vary in a similar manner. In this case, the Si^{1+} mode is also of highest intensity and the intensities of the other oxide modes usually drop with increasing Si atom valence. Another feature of the oxide structures grown at room temperature and low temperatures is discontinuity, which was observed in a number of works.

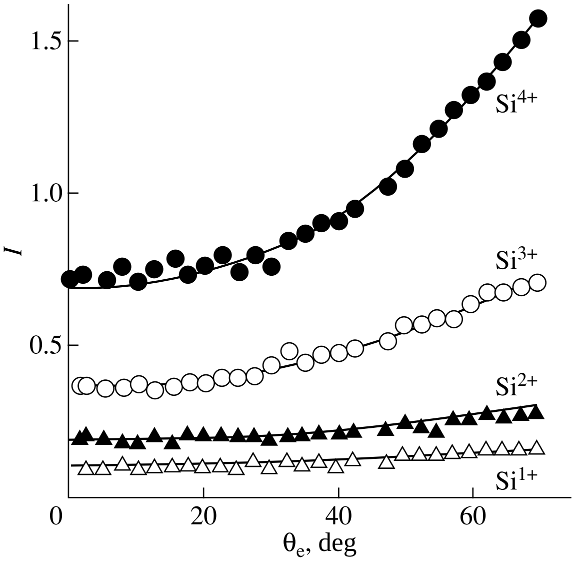


Fig. 39. Intensities of the different oxide components in the Si 2p spectrum taken from the oxidized Si(100)-2 × 1 surface vs. the polar angle of departure. The intensities are normalized by the total intensity of all the oxide components [182].

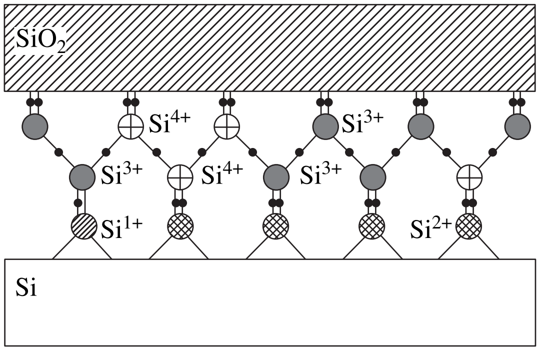


Fig. 40. Model of the SiO_2/Si(100) interface that forms under thermal oxidation of silicon (side view) [182].

The annealing of the single crystals exposed to oxygen at low temperatures or room temperature considerably modifies the subsurface oxide layer. This is especially true for annealing temperatures above 500 K, when the shape of the spectrum approaches that observed at thermal oxidation (Fig. 42). Also, the shifts of the Si^{2+} and Si^{3+} modes increase to the values typical of the thermal oxidation. It seems that the surface phases responsible for these modes are metastable at low temperatures because of elastic stresses arising in the oxide layer. The stresses relax during the annealing, and the phases switch to the stable states. The effect of elastic stresses on the formation of the SiO_2/Si(100) interface was analyzed in [197].

Thus, in all the cases considered, the interaction of oxygen with the Si(100)-2 × 1 surface gives rise to silicon atoms that are chemically bonded to one, two, three, or even four oxygen atoms. These bonded atoms serve as nuclei for an oxide phase of complex composition. Eventually, an oxide layer consisting of all types

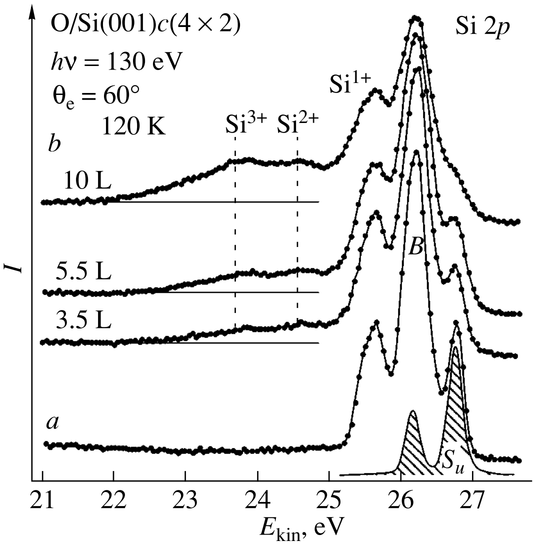


Fig. 41. Silicon photoelectron spectra observed under low-temperature adsorption of oxygen [179]. (*a*) Specimen with the clean Si(100)-$c(4 \times 2)$ surface (component *S* due to the upper atoms of the dimers is separated) and (*b*) specimens exposed to oxygen under different exposure conditions.

of silicon oxide phases grows. Under thermal oxidation, this layer is intermediate between Si(100) and an amorphous film of pure silicon dioxide SiO_2 [169, 170]. It is noteworthy that the theoretical works devoted to oxygen adsorption on silicon usually consider only the monovalent states of silicon atoms [198–200].

The formation of different oxide phase is also characteristic of oxygen adsorption on the Si(111)-7 × 7 surface, although the composition of the oxide layer is other than in the previous case [169, 171, 175, 176, 182, 185]. From the dependences of different oxide modes in the silicon spectra on the polar angle of departure of the electrons, it was concluded [170, 171] that the SiO_2/Si(111) interface is abrupt and consists virtually of one monatomic layer, yet including all the oxidation states of silicon. The formation of such an interface was explained in terms of the model of statistical closure. According to this model, dangling bonds on the virgin (unreconstructed) Si(111) surface and on the surface of the amorphous SiO_2 film close on each other. Note that the photoelectron diffraction data [187] also suggest that the SiO_2/Si(111) interface is abrupt and the SiO_2/Si(100) interface is graded.

Thus, core-level spectroscopy studies provided much insight into the complex chemical constitution of the oxide layer at the SiO_2/Si interface and revealed distinctions between the structures of this interface for different silicon faces.

(ii) Adsorption of chlorine. Chlorine, like oxygen, has a much higher electronegativity than silicon. Interest in chlorine adsorption on silicon has considerably quickened in recent years. It has been established to date that, at room temperature, chlorine molecules, when adsorbed on the Si(100)-2 × 1 surface, dissociate and do not break the two-dimensional periodicity. However, such issues as which adsorption sites are occupied by the adatoms and whether any surface reconstruction occurs upon the adsorption still remain vague. Two models of adsorption are considered in this case. In one model, it is assumed that chlorine atoms are adsorbed on both atoms of the substrate dimers, making them symmetric [201]. The other point of view is that, after chlorine molecules have dissociated, the atoms are adsorbed only on the upper atoms of the asymmetric dimers without distorting their structure [202].

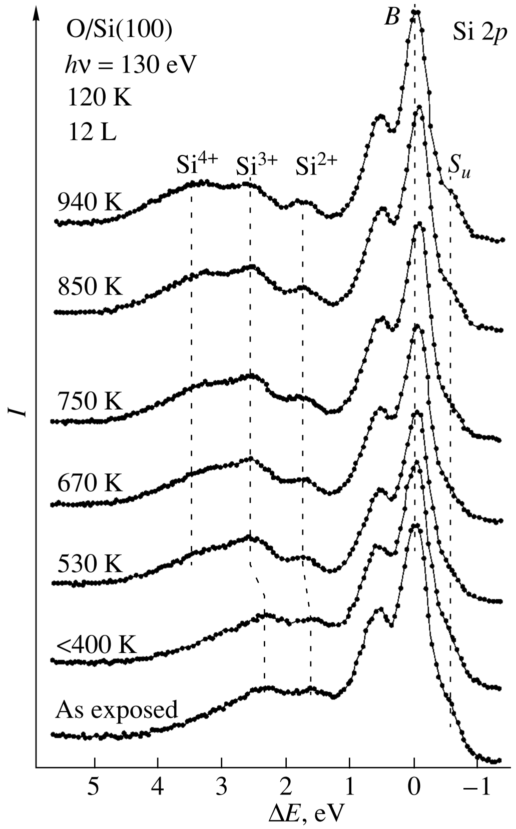


Fig. 42. Silicon photoelectron spectra taken from the annealed specimen exposed to oxygen at *T* = 120 K (exposure 12 L) [180].

The method of core-level spectroscopy was deposited to examine the Cl/Si(100)-2 × 1 system in [188, 189]. The variation of the Si 2*p* electron spectra during chlorine adsorption at room temperature is illustrated in Fig. 43 [189]. The resolution of these spectra is shown in Fig. 44a. Similarly to hydrogen adsorption, component S_u, due to the upper atoms of the asymmetric substrate dimers, fades out as the chlorine dose increases. Simultaneously, new component *S* with a positive shift of 0.9 eV builds up. For the saturated coverage, its intensity corresponds to one silicon monolayer. On the one hand, this indicates that only one type of adsorption

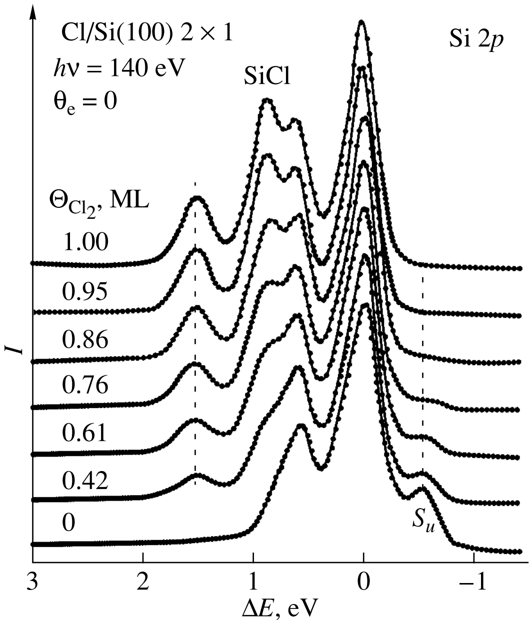


Fig. 43. Silicon photoelectron spectra taken from the clean Si(100)-2 × 1 surface and after the specimen has been exposed to the chlorine atmosphere [189].

site for the adsorbate is available; on the other hand, one can infer that the asymmetric substrate dimers become symmetric. It is these symmetric dimers on which chlorine atoms are adsorbed and close all dangling bonds of surface silicon atoms when the coverage is saturated (Fig. 44b). As for oxygen, the large energy shift of mode S is due to charge transfer from the atoms of the dimers to those of the adsorbate, which produces Cl–Si polar bonds. Thus, the core-level spectroscopy data support the validity of the first model of chlorine adsorption on the Si(100)-2 × 1 surface.

The surface structure of the Cl/ Si(100)-2 × 1 system that is shown in Fig. 44b persists up to 540°C. At higher temperatures, the Si 2p electron spectra exhibit an extra component with an energy shift of 1.5 eV. It is supposed that the new mode is a result of bond rupture between the atoms of the asymmetric dimers followed by additional chlorine adsorption. However, $SiCl_2$ adsorption complexes occupy only a small fraction (about 20%) of surface silicon atoms. Thus, thermal processes initiate a new stage of chlorine–silicon interaction.

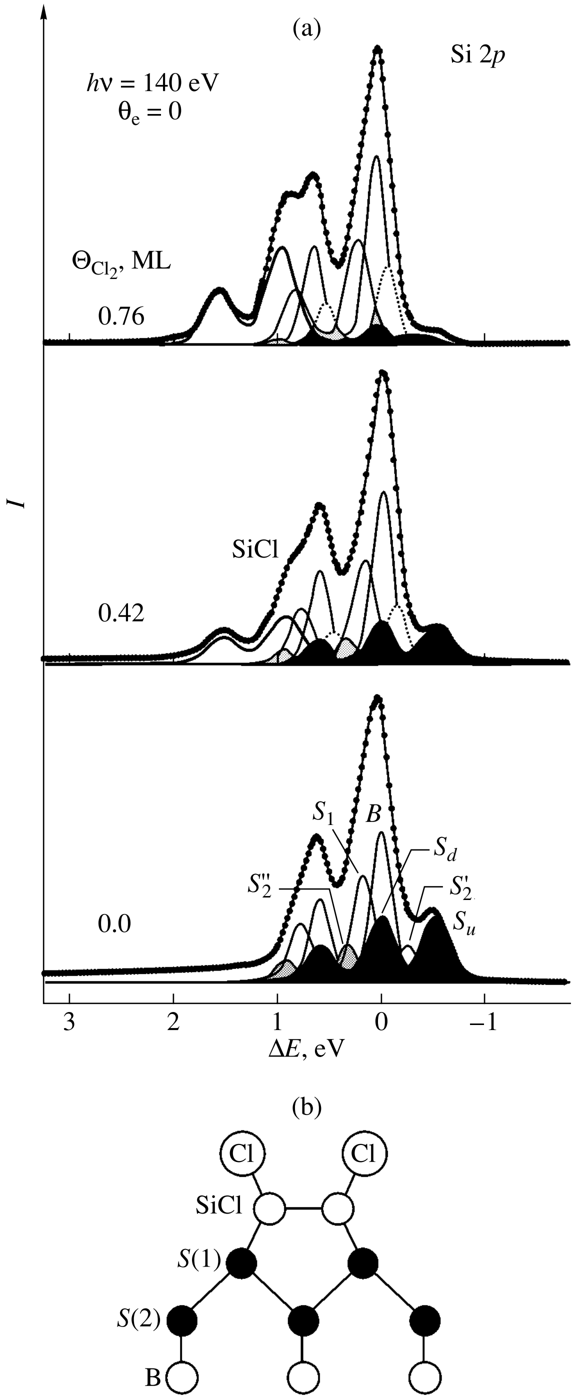


Fig. 44. (a) Decomposition of the Si 2p electron spectra taken from the clean Si(100)-2 × 1 surface and from the Si(100)-2 × 1 surface covered by submonolayer chlorine coverages and (b) the model of chlorine adsorption on silicon (side view) [189].

(iii) Adsorption of hydrogen. A great deal of core-level spectroscopy data for gas adsorption on silicon were devoted to the H_2/Si, more precisely, to the H_2/Si(111), system. An intriguing feature of this system is that the crystal surface remains unreconstructed under certain conditions, as indicated by (1 × 1) diffraction patterns. In this case, each dangling bond of surface silicon atoms is saturated by a hydrogen atom. Such a monohydride structure passivates the silicon surface. The chemical way of forming this structure turned out to be the most efficient [167]. The Si 2p electron spectrum taken from the passivated surface (Fig. 45) differs substantially from the spectrum for the clean reconstructed Si(111)-7 × 7 surface (Fig. 4). The former demonstrates a clear-cut dip between the 2$p_{3/2}$ and 2$p_{1/2}$ sublevels of the silicon doublet, because the surface modes of adatoms and rest atoms (these modes

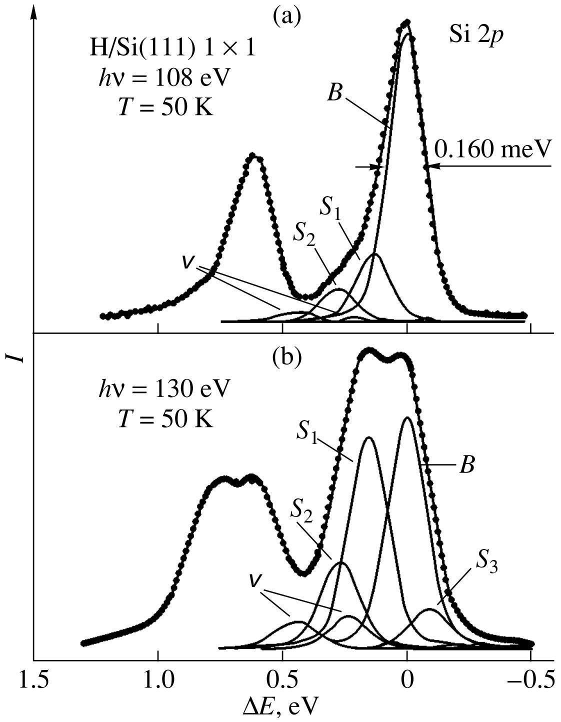


Fig. 45. (a) Volume-sensitive and (b) surface-sensitive Si 2*p* electron spectra taken from the surface of the H/Si(100)-1 × 1 structure at $T = 50$ K and the decomposition into components. The results are shown only for the $2p_{3/2}$ sublevel of the Si 2*p* doublet [167].

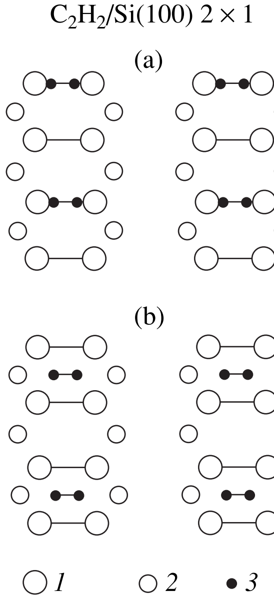


Fig. 46. Models of acetylene adsorption on the Si(100)-2 × 1 surface (top view) [190]: (a) di-σ model and (b) tetra-σ model. (*1*) Atoms of substrate dimers, (*2*) atoms of the second layer of the substrate, and (*3*) carbon atoms of C_2H_2 molecules (hydrogen atoms are omitted).

characterize the reconstructed surface) are absent [168]. Thus, hydrogen adsorption may significantly modify the properties of the Si(111)-7 × 7 surface.

(iv) Adsorption of acetylene. Today, core-level spectroscopy is coming into use for the study of polyatomic gas (such as acetylene) adsorption on silicon. The C_2H_2/Si system seems promising for growing silicon carbide films by thermally decomposing carboniferous molecules adsorbed. It has been shown that, at room temperature, C_2H_2 molecules are adsorbed on the Si(100)-2 × 1 surface without dissociation, so that the initial reconstruction of the substrate surface is retained. The molecules dissociate only at elevated temperatures. As regards adsorption sites, two models, the di-σ model and tetra-σ model, have been discussed [190, 203]. According to the former, C_2H_2 molecules are adsorbed on bridges between the dangling bonds of separate substrate dimers. In the latter model, acetylene molecules are adsorbed on pedestals between two adjacent dimers. For the saturated coverage (0.5 ML), both models are schematically shown in Fig. 46. The method of core-level spectroscopy has provided compelling evidence in favor of the second model [190].

The Si 2*p* electron spectra taken during C_2H_2 molecules on the Si(100)-2 × 1 surface, as well as the result of their resolution into components, are demonstrated in Fig. 47. As in the other systems, the component S_u

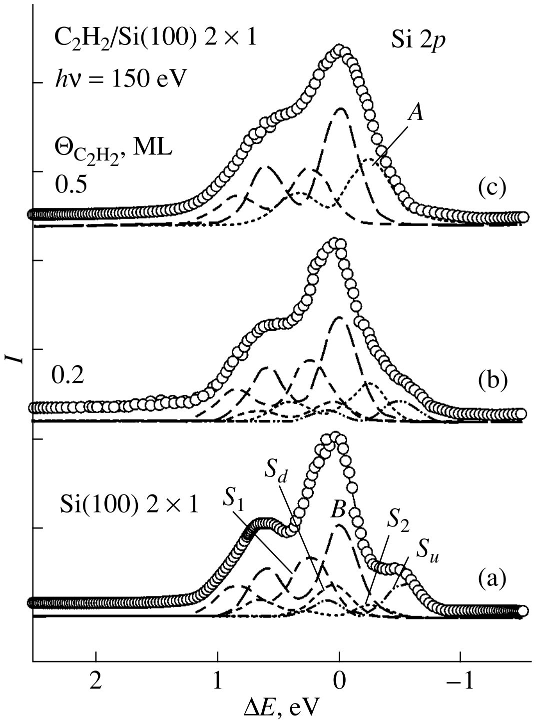


Fig. 47. Si 2*p* electron spectra taken from the (a) clean Si(100)-2 × 1 surface and from the surface covered by (b) 0.2 and (c) 0.5 ML of acetylene molecules adsorbed [190]. The dashed curves show different modes of the spectra.

due to the upper atoms of the asymmetric substrate dimers fades out (for a saturated coverage of 0.5 ML) and new component *A* with a negative energy shift builds up. For the saturated coverage, the intensity of the latter is twice as high as the intensity of the mode S_u of the uncovered silicon surface. The other spectral components change insignificantly. In combination, these findings indicate that C_2H_2 adsorption affects only the upper monolayer of the substrate dimers, which lose symmetry. The fact that only one type of new upper-atom state is induced by the adsorption argues into accepting the second model. If the process went in accordance with the model depicted in Fig. 46a, the spectra would have components of two types, i.e., those due to uncovered dimers and dimers covered by acetylene molecules.

The data obtained for the specimen coated by the saturated C_2H_2 coverage and heated to 700°C indicate that, in this system, the decomposition of acetylene into C_2H_x ($x = 1, 0$) and H is followed by the formation of SiC clusters with the participation of silicon atoms.

CONCLUSIONS

The material presented in this review clearly demonstrates the wide potentialities of core-level photoelectron spectroscopy for investigating atomic-level processes on the silicon surface. Based on the energy shifts of Si 2*p* electrons, as well as of the core levels of atoms adsorbed, one can (i) gain insight into the atomic configuration of the reconstructed surface of single-crystal silicon and determine the charge states of surface atoms; (ii) detect various adsorption phases and associated adsorption sites; (iii) determine the adsorbate concentration in these phases and the variation of this concentration with increasing coverage; (iv) study the adsorption-induced reconstruction of the substrate surface; (v) investigate thin-film growth mechanisms and also trace the diffusion of adsorbate atoms into the substrate and the diffusion of silicon atoms into the growing film; (vi) clarify the atomic constitution of the interface; and (vii) detect various chemical compounds on the crystal surface and elucidate their nature and phase composition.

With this method, a large body of data that are of fundamental interest for the physics and chemistry of surface have been collected. Specifically, it has been clearly demonstrated that (i) the dimers on the Si(100)-2 × 1 surface are asymmetric, (ii) the Si(111)-7 × 7 surface undergoes radical reconstruction when covered by a number of metals (the formation of quasi-one-dimensional structures made up of silicon atoms), (iii) four silicon oxide phases arise early in the oxidation, etc. These observations support the efficiency of the method and lend hope for its further application.

ACKNOWLEDGMENTS

We thank A.D. Buravlev, who prepared this review for publication.

This work was supported by the Russian Foundation for Basic Research (grant no. 01-02-17288) and the Ministry of Industry, Science, and Technology of the Russian Federation (contract no. 40.012.1.1.1152).

REFERENCES

1. K. Siegbahn, K. Nordling, A. Fahlman, *et al.*, in *Proceedings of International Conference on Electron Spectroscopy, Asilomar, California, 1971*, Ed. by D. A. Shirley (Mir, Moscow, 1971; North-Holland, Amsterdam, 1972).
2. *Practical Surface Analysis by Auger and X-ray Photoelectron Spectroscopy,* Ed. by D. Briggs and M. P. Seach (Wiley, New York, 1983; Mir, Moscow, 1987).
3. D. P. Woodruff and T. A. Delchar, *Modern Techniques of Surface Science* (Cambridge Univ. Press, Cambridge, 1986; Mir, Moscow, 1989).
4. M. V. Gomoyunova, Usp. Fiz. Nauk **136**, 105 (1982) [Sov. Phys. Usp. **25**, 58 (1982)].
5. S. Hufner, *Photoelectron Spectroscopy: Principles and Applications* (Springer, Berlin, 1995).
6. J. W. F. Egelhoff, Surf. Sci. Rep. **6**, 253 (1987).
7. M. P. Seach, Surf. Interface Anal. **9**, 85 (1986).
8. K. Siegbahn, *Atomic Physics* (Plenum, New York, 1973), Vol. 3, p. 514.
9. U. Gelius, L. Asplund, E. Basilier, *et al.*, Nucl. Instrum. Methods Phys. Res. B **229**, 85 (1984).
10. E. J. Himpsel, P. Heimann, T.-C. Chaing, *et al.*, Phys. Rev. Lett. **45**, 1112 (1980).
11. T. Miller, T. C. Hsich, and T.-C. Chiang, Phys. Rev. B **33**, 6983 (1986).
12. K. Hricovini, G. Le Lay, M. Abrahan, *et al.*, Phys. Rev. B **41**, 1258 (1990).
13. C. J. Karlsson, E. Landemark, L. S. O. Johansson, *et al.*, Phys. Rev. B **41**, 1521 (1990).
14. K. O. Magnusson, S. Wiklund, R. Dudde, *et al.*, Phys. Rev. B **44**, 5657 (1991).
15. J. A. Carlisle, M. T. Sieger, T. Miller, *et al.*, Phys. Rev. Lett. **71**, 2955 (1993).
16. G. Le Lay and M. Fontaine, Phys. Rev. Lett. **72**, 3740 (1994).
17. G. Le Lay, M. Gothelid, T. M. Grekh, *et al.*, Phys. Rev. B **50**, 14277 (1994).
18. C. J. Karlsson, E. Landemark, Y.-C. Chao, *et al.*, Phys. Rev. B **50**, 5767 (1994).
19. J. J. Paggel, W. Thies, K. Horn, *et al.*, Phys. Rev. B **50**, 18686 (1994).
20. R. I. G. Uhrberg, T. Kaurila, and Y.-C. Chao, Phys. Rev. B **58**, R1730 (1998).
21. D. H. Rich, T. Miller, and T.-C. Chiang, Phys. Rev. B **37**, 3124 (1988).
22. D.-S. Lin, T. Miller, and T.-C. Chiang, Phys. Rev. Lett. **67**, 2187 (1991).

23. D.-S. Lin, J. A. Carlisle, T. Miller, *et al.*, Phys. Rev. Lett. **69**, 552 (1992).
24. X. Yang, R. Cao, J. Terry, *et al.*, J. Vac. Sci. Technol. B **10**, 2013 (1992).
25. R. Cao, X. Yang, J. Terry, *et al.*, Phys. Rev. B **45**, 13749 (1992).
26. R. J. Himpsel, F. R. McFeely, A. Taleb-Ibrahimi, *et al.*, Phys. Rev. B **38**, 6084 (1988).
27. G. K. Wertheim, D. M. Riffe, J. E. Rowe, *et al.*, Phys. Rev. Lett. **67**, 120 (1991).
28. E. Landemark, C. J. Karlsson, Y.-C. Chao, *et al.*, Phys. Rev. Lett. **69**, 1588 (1992).
29. J. E. Rowe and G. K. Wertheim, Phys. Rev. Lett. **69**, 550 (1992).
30. F. J. Himpsel, Phys. Rev. Lett. **69**, 551 (1992).
31. E. L. Bullock, R. Gunnella, L. Patthey, *et al.*, Phys. Rev. Lett. **74**, 2756 (1995).
32. C.-P. Cheng, I.-H. Hong, and T.-W. Pi, Phys. Rev. B **58**, 4066 (1998).
33. T.-W. Pi, I.-H. Hong, C.-P. Cheng, *et al.*, J. Electron Spectrosc. Relat. Phenom. **107**, 163 (2000).
34. H. Koh, J. W. Kim, W. H. Choi, *et al.*, Phys. Rev. B **67**, 073306 (2003).
35. A. A. Backi, S. C. Erwin, and L. J. Whitman, Surf. Sci. **392**, 69 (1997).
36. K. Takayanagi, Y. Tanishiro, S. Takahashi, *et al.*, Surf. Sci. **164**, 367 (1985).
37. R. J. Hamers, R. M. Tromp, and J. E. Demuth, Phys. Rev. Lett. **56**, 1972 (1986).
38. R. F. Schlier and H. E. Farnsworth, J. Chem. Phys. **30**, 917 (1959).
39. R. J. Hamers, R. M. Tromp, and J. E. Demuth, Phys. Rev. B **34**, 5343 (1986).
40. D. J. Chadi, Phys. Rev. Lett. **43**, 43 (1979).
41. R. Kaplan, Surf. Sci. **93**, 145 (1980).
42. W. S. Yang, F. Jona, and P. N. Marcus, Phys. Rev. B **28**, 2049 (1983).
43. B. W. Holland, S. D. Duke, and A. Paton, Surf. Sci. **140**, L269 (1984).
44. N. Jedrecy, M. Sauvage-Simkin, R. Pinchaux, *et al.*, Surf. Sci. **230**, 197 (1990).
45. P. Koke, A. Goldmann, W. Monch, *et al.*, Surf. Sci. **152–153**, 1001 (1985).
46. R. M. Tromp, R. G. Smeenk, F. W. Saris, *et al.*, Surf. Sci. **133**, 137 (1983).
47. Z. Zhu, N. Shima, and M. Tsukada, Phys. Rev. B **40**, 11868 (1989).
48. R. D. Bringans, R. I. G. Uhrberg, and M. A. Olmstead, Phys. Rev. B **34**, 7447 (1986).
49. T. Abukawa, T. Okane, and S. Kono, Surf. Sci. **256**, 370 (1991).
50. J. Ihin, D. H. Loe, J. D. Joannopoulos, *et al.*, Phys. Rev. Lett. **51**, 1872 (1983).
51. T. Tabata, T. Aruga, and Y. Murata, Surf. Sci. **179**, L63 (1987).
52. T. Miller, E. Rosenwinkel, and T.-C. Chiang, Solid State Commun. **47**, 935 (1983).
53. P. H. Mahowald, D. J. Friedman, G. P. Carey, *et al.*, J. Vac. Sci. Technol. A **5**, 2982 (1987).
54. L. I. Johansson, K. L. Hakansson, P. L. Wincott, *et al.*, Phys. Rev. B **43**, 12355 (1991).
55. S. Doniach and M. Sunjoc, J. Phys. C **3**, 285 (1970).
56. D. A. Shirley, Phys. Rev. B **5**, 4709 (1972).
57. E. Pelke and M. Scheffler, Phys. Rev. Lett. **71**, 2338 (1993).
58. M. Rohlfing, P. Krüger, and J. Pollmann, Phys. Rev. B **56**, 2191 (1997).
59. A. Ramstad, G. Brocks, and P. J. Kelly, Phys. Rev. B **51**, 14504 (1995).
60. R. D. Schnel, F. J. Himpsel, A. Bogen, *et al.*, Phys. Rev. B **32**, 8052 (1985).
61. G. Le Lay, J. Kanski, P. O. Nilsson, *et al.*, Phys. Rev. B **45**, 6692 (1992).
62. R. Cao, X. Yang, J. Terry, *et al.*, Phys. Rev. B **45**, 13749 (1992).
63. E. Landemark, C. J. Karlsson, L. S. O. Johansson, *et al.*, Phys. Rev. B **49**, 16523 (1994).
64. A. Goldoni, S. Modesti, and D. R. Dhanok, Phys. Rev. B **54**, 11340 (1996).
65. T.-W. Pi, J.-F. Wen, C.-P. Ouyang, *et al.*, Phys. Rev. B **63**, 153310 (2001).
66. T.-W. Pi, R.-T. Wu, C.-P. Ouyang, *et al.*, Surf. Sci. **461**, L565 (2000).
67. T.-W. Pi, C.-P. Ouyang, J.-F. Wen, *et al.*, Surf. Sci. **514**, 327 (2002).
68. L. Patthey, E. L. Bullock, T. Abukawa, *et al.*, Phys. Rev. Lett. **75**, 2538 (1995).
69. K.-H. Huang, T.-S. Ku, and D. S. Lin, Phys. Rev. B **56**, 4878 (1997).
70. R. Larciprete, P. D. Padova, C. Quaresima, *et al.*, Phys. Rev. B **61**, 16006 (2000).
71. F. Iwawaki, M. Tomitori, and O. Nishikama, Surf. Sci. **266**, 285 (1992).
72. Y.-W. Mo and D. E. Savage, Phys. Rev. Lett. **65**, 1020 (1990).
73. D. J. Euglesham and M. Cerullo, Phys. Rev. Lett. **64**, 1943 (1990).
74. M. Copel, M. C. Reuter, E. Kaxiras, *et al.*, Phys. Rev. Lett. **63**, 632 (1989).
75. F. K. Le Goues, M. Copel, and R. Tromp, Phys. Rev. Lett. **63**, 1826 (1989).
76. P. De Padova, R. Larciprete, C. Quaresima, *et al.*, Phys. Rev. Lett. **81**, 2320 (1998).
77. J. M. C. Thornton, A. A. Williams, J. E. Macdonald, *et al.*, J. Vac. Sci. Technol. B **9**, 2146 (1991).
78. R. Cao, X. Yang, J. Terry, *et al.*, Appl. Phys. Lett. **61**, 2347 (1992).
79. S. A. Barnett, H. F. Winfers, C. F. Quate, *et al.*, Surf. Sci. **165**, 303 (1986).
80. M. Richer, J. C. Woicik, J. Nogami, *et al.*, Phys. Rev. Lett. **65**, 3417 (1990).
81. T. M. Grehk, L. S. O. Johansson, S. M. Gray, *et al.*, Phys. Rev. B **52**, 16 593 (1995).
82. T. M. Grehk, C. U. S. Larsson, N. P. Prince, *et al.*, Surf. Sci. **284**, L384 (1993).
83. H. H. Weitering, X. Shi, and S. C. Erwin, Phys. Rev. B **54**, 10 585 (1996).

84. J. J. Pagel, G. Neuhold, H. Haak, *et al.*, J. Vac. Sci. Technol. B **11**, 1439 (1993).
85. T. Okuda, H. Shigeoka, H. Daimon, *et al.*, Surf. Sci. **321**, 105 (1994).
86. J. J. Pagel, G. Neuhold, H. Haak, *et al.*, Phys. Rev. B **52**, 5813 (1995).
87. T. Okuda, H. Daimon, S. Suga, *et al.*, Appl. Surf. Sci. **121–122**, 89 (1997).
88. M.-H. Kang, J.-H. Kang, and S. Jeong, Phys. Rev. B **58**, R13359 (1998).
89. Y.-C. Chao, L. S. O. Johansson, and R. I. G. Uhrberg, Phys. Rev. B **55**, 7198 (1997).
90. P. Soukiassian, J. A. Kubby, P. S. Mangat, *et al.*, Phys. Rev. B **46**, 13 471 (1992).
91. U. A. Effner, D. Badt, J. Binder, *et al.*, Surf. Sci. **277**, 207 (1992).
92. D. M. Riffe, G. K. Wertheim, J. E. Rowe, *et al.*, Phys. Rev. B **45**, 3532 (1992).
93. H. H. Weitering, J. Chen, N. J. Di Nardo, *et al.*, Phys. Rev. B **48**, 8119 (1993).
94. Y.-C. Chao, L. S. O. Johansson, C. J. Karlsson, *et al.*, Phys. Rev. B **52**, 2579 (1995).
95. Y.-C. Chao, L. S. O. Johansson, and R. I. G. Uhrberg, Surf. Sci. **372**, 64 (1997).
96. K. Sakamoto, H. M. Zhang, and R. I. G. Uhrberg, Surf. Rev. Lett. **9**, 1235 (2002).
97. D. S. Lin, T. Miller, and T.-C. Chiang, Phys. Rev. B **44**, 10719 (1991).
98. P. S. Mangat and P. Soukiassian, Phys. Rev. B **52**, 12020 (1995).
99. Y.-C. Chao, L. S. O. Johansson, and R. I. G. Uhrberg, Phys. Rev. B **54**, 5901 (1996).
100. C. L. Park, K. S. An, J. S. Kim, *et al.*, Phys. Rev. B **52**, 8198 (1995).
101. K. S. An, R. J. Park, J. S. Kim, *et al.*, Surf. Sci. **337**, L789 (1995).
102. A. A. Backi, S. C. Erwin, M. S. Tumer, *et al.*, Surf. Sci. **476**, 22 (2001).
103. K. Sakamoto, W. Takeyama, H. M. Zhang, *et al.*, Phys. Rev. B **66**, 165 319 (2002).
104. K. Sakamoto, W. Takeyama, H. M. Zhang, *et al.*, Thin Solid Films **428**, 115 (2003).
105. C.-P. Cheng, I.-H. Hong, and T.-W. Pi, Phys. Rev. B **58**, 4066 (1995).
106. A. L. Wachs, T. Miller, and A. P. Shapiro, Phys. Rev. B **35**, 5514 (1987).
107. M. Herman, J. C. Woicik, A. B. Andrews, *et al.*, Surf. Sci. **290**, L643 (1993).
108. G. Le Lay, V. Yu. Aristov, L. Seehofer, *et al.*, Surf. Sci. **307–309**, 280 (1994).
109. K. Sakamoto, H. Ashima, and H. M. Zhang, Phys. Rev. B **65**, 045305 (2002).
110. T. Okuda, H. Daimon, H. Shigeoka, *et al.*, Electron Spectrosc. Relat. Phenom. **80**, 229 (1996).
111. H. M. Zhang, T. Balasubramanian, and R. I. G. Uhrberg, Phys. Rev. B **65**, 035314 (2001).
112. T. A. Abukawa, F. Hisamata, M. Nakamura, *et al.*, J. Electron Spectrosc. Relat. Phenom. **80**, 233 (1996).
113. H. W. Yeom, T. A. Abukawa, Y. Takakuwa, *et al.*, Phys. Rev. B **54**, 4456 (1996).
114. H. W. Yeom, K. Horikoshi, H. M. Zhang, *et al.*, Phys. Rev. B **65**, 241307 (2002).
115. F. Boscherini, J. J. Joyce, M. W. Ruckman, *et al.*, Phys. Rev. B **35**, 4216 (1987).
116. J. M. Gallego, R. Miranda, S. Molodtsov, *et al.*, Surf. Sci. **239**, 203 (1990).
117. G. Rangelov, P. Augustin, J. Stober, *et al.*, Phys. Rev. B **49**, 7535 (1994).
118. G. Rangelov and Th. Fauster, Surf. Sci. **365**, 403 (1996).
119. G. Rangelov, P. Augustin, Th. Fauster, *et al.*, Surf. Sci. **307–309**, 264 (1994).
120. M. V. Gomoyunova, I. I. Pronin, N. R. Gall’, *et al.*, Pis’ma Zh. Tekh. Fiz. **29** (12), 25 (2003) [Tech. Phys. Lett. **29**, 496 (2003)].
121. M. V. Gomoyunova, I. I. Pronin, N. R. Gall’, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **45** (8), 180 (2003) [Phys. Solid State **45**, 1596 (2003)].
122. B. Kim, K.-J. Kim, and T.-H. Kang, Appl. Surf. Sci. **152**, 44 (1999).
123. K.-J. Kim, T.-H. Kang, K.-W. Kim, *et al.*, Appl. Surf. Sci. **161**, 268 (2000).
124. C. Wigren, J. N. Andersen, R. Nyholm, *et al.*, Phys. Rev. B **48**, 11 014 (1993).
125. W. A. Henle, M. G. Ramsey, F. P. Netzer, *et al.*, Surf. Sci. **254**, 182 (1991).
126. W. A. Henle, M. G. Ramsey, F. P. Netzer, *et al.*, Surf. Sci. **243**, 141 (1991).
127. S. Gokhale, S. Mahemuni, S. V. Deshmukh, *et al.*, Surf. Sci. **237**, 127 (1990).
128. R. Hofmann, W. A. Henle, F. P. Netzer, *et al.*, Phys. Rev. B **46**, 3857 (1992).
129. C. Wigren, J. N. Andersen, R. Nyholm, *et al.*, Phys. Rev. B **47**, 9663 (1993).
130. J. D. Levine, Surf. Sci. **34**, 90 (1973).
131. T. Abukawa and S. Kono, Phys. Rev. B **37**, 9097 (1988).
132. C. A. Papageorgopoulos and M. Kamaratoc, Surf. Sci. **221**, 263 (1989).
133. S. Tanaka, N. Takagi, N. Minami, *et al.*, Phys. Rev. B **42**, 1868 (1990).
134. J. E. Northrup, M. C. Schabel, C. J. Karlsson, *et al.*, Phys. Rev. B **44**, 13799 (1991).
135. B. E. Steele, L. Li, J. L. Stevens, *et al.*, Phys. Rev. B **47**, 9925 (1993).
136. C. Pirri, S. Hong, M. H. Tuiler, *et al.*, Phys. Rev. B **53**, 1368 (1996).
137. W. R. L. Lambrecht, N. E. Christensen, and I. P. Bloch, Phys. Rev. B **36**, 2493 (1987).
138. R. Leckey, J. D. Riley, R. L. Johnson, *et al.*, J. Vac. Sci. Technol. A **6**, 63 (1988).
139. H. L. Meyerheim, U. Döbler, and A. Puschmann, Phys. Rev. B **44**, 5738 (1991).
140. H. Daimon and S. Ino, Surf. Sci. **164**, 320 (1985).
141. M. Tihov, L. Surnev, and M. Kiskinova, Phys. Rev. B **44**, 3222 (1991).
142. A. J. Muskat, A. Rjeb, and D. Roy, Surf. Sci. **302**, L256 (1994).

143. T. Hashizume, M. Katayama, D. R. Jeon, *et al.*, Jpn. J. Appl. Phys. **32**, L1263 (1993).
144. H. H. Weitering, N. J. Di Nardo, R. Perez-Sandoz, *et al.*, Phys. Rev. B **49**, 16 837 (1994).
145. J. Quinn and F. Jona, Surf. Sci. **249**, L307 (1991).
146. M. A. Olmstead, R. I. G. Uhrberg, R. D. Bringans, *et al.*, J. Vac. Sci. Technol. B **4**, 1123 (1986).
147. H. H. Weitering, Surf. Sci. **355**, L271 (1996).
148. W. S. Yang, S. C. Wu, and F. Jona, Surf. Sci. **169**, 383 (1986).
149. W. C. Fan and A. Ignatiev, Phys. Rev. B **41**, 3592 (1990).
150. K. Sakamoto, T. Okuda, H. Nishimoto, *et al.*, Phys. Rev. B **50**, 1725 (1994).
151. K. J. Wan, X. F. Lin, and J. Nogami, Phys. Rev. B **46**, 13635 (1993).
152. D. Jeon, N. Hashizume, T. Sakurais, *et al.*, Phys. Rev. Lett. **69**, 1419 (1992).
153. A. A. Saranin, A. V. Zotov, V. G. Lifshits, *et al.*, Surf. Sci. **426**, 298 (1999).
154. K. S. Pandey, Phys. Rev. Lett. **47**, 1913 (1981).
155. S. T. Erwin and H. Weitering, Phys. Rev. Lett. **81**, 2296 (1998).
156. L. Lottermoser, E. Landemark, D.-M. Smilgies, *et al.*, Phys. Rev. Lett. **80**, 3980 (1998).
157. M.-H. Kang, J.-H. Kang, and S. Leong, Phys. Rev. B **58**, R13359 (1998).
158. T. Sckigushi, F. Shimokoshi, T. Nagao, *et al.*, Surf. Sci. **493**, 148 (2001).
159. G. Lee, S. Hong, H. Kim, *et al.*, Phys. Rev. Lett. **87**, 056104 (2001).
160. O. Kubo, A. A. Saranin, A. V. Zotov, *et al.*, Surf. Sci. **415**, L971 (1998).
161. R. Seiwatz, Surf. Sci. **2**, 472 (1964).
162. O. Bunk, G. Falkenberg, J. H. Zeysing, *et al.*, Phys. Rev. B **59**, 12228 (1999).
163. J. Nakamura, S. Watanabe, and M. Aono, Phys. Rev. B **63**, 193307 (2001).
164. H. W. Yeom, S. Takeda, E. Rotenberg, *et al.*, Phys. Rev. Lett. **82**, 4898 (1999).
165. X. Blasé, A. J. R. De Silva, X. Zhu, *et al.*, Phys. Rev. B **50**, 8102 (1994).
166. E. Landemark, C. J. Karlsson, and R. I. G. Urhberg, Phys. Rev. B **44**, 1950 (1991).
167. K. Hricovini, R. Günther, P. Thiry, *et al.*, Phys. Rev. Lett. **70**, 1992 (1993).
168. M. Rohlfing, P. Krüger, and J. Pollman, Phys. Rev. B **56**, 2191 (1997).
169. T. Engel, Surf. Sci. Rep. **18**, 91 (1993).
170. S. Imata and A. Ischizaka, J. Appl. Phys. **79**, 6653 (1996).
171. G. Hollinger and F. J. Himpsel, Phys. Rev. B **28**, 3651 (1983).
172. M. Tabe, T. T. Chiang, I. Lindau, *et al.*, Phys. Rev. B **34**, 2706 (1986).
173. F. J. Himpsel, F. R. Mc Feely, A. Taleb-Ibrahimi, *et al.*, Phys. Rev. B **38**, 6084 (1988).
174. M. M. B. Holl and F. R. McFeely, Phys. Rev. Lett. **71**, 2441 (1993).
175. M. T. Sieger, D.-A. Luh, T. Miller, *et al.*, Phys. Rev. Lett. **77**, 2758 (1996).
176. D.-A. Luh, T. Miller, and T.-C. Chiang, Phys. Rev. Lett. **79**, 3014 (1997).
177. A. Mascaraque, C. Ottaviani, and M. Capozi, Surf. Sci. **377–379**, 650 (1997).
178. Y. Enta, Y. Miyanishi, H. Irimachi, *et al.*, Phys. Rev. B **57**, 6294 (1998).
179. H. W. Yeom, H. Hamamatsu, T. Ohta, *et al.*, Phys. Rev. B **59**, R10413 (1999).
180. H. W. Yeom and R. Uhrberg, Jpn. J. Appl. Phys. **39**, 4460 (2000).
181. T. Nishimura, Y. Hoshino, H. Namba, *et al.*, Surf. Sci. **461**, 146 (2000).
182. J. H. Oh, H. W. Yeom, Y. Hagimoto, *et al.*, Phys. Rev. B **63**, 205310 (2001).
183. T.-W. Pi, J.-F. Wen, C.-P. Ouyang, *et al.*, Surf. Sci. **478**, L333 (2001).
184. Y. Hoshino, T. Tishimura, T. Nakada, *et al.*, Surf. Sci. **488**, 249 (2001).
185. P. Morgen, T. Jensen, Gundlach, *et al.*, J. Comput. Mater. Sci. **21**, 481 (2001).
186. F. Jolly, F. Rochet, G. Dufour, *et al.*, J. Non-Cryst. Solids **280**, 150 (2001).
187. C. Westphal, S. Dreiner, M. Schurmann, *et al.*, Surf. Rev. Lett. **9**, 735 (2002).
188. J. A. Martin-Gago, E. Roman, M. C. Refolio, *et al.*, Surf. Sci. **424**, 82 (1999).
189. T.-W. Pi, S.-F. Tsai, C.-P. Ouyang, *et al.*, Surf. Sci. **488**, 387 (2001).
190. S. H. Xu, Y. Yang, M. Keeffe, *et al.*, Phys. Rev. B **60**, 11586 (1999).
191. C. Pencey, F. Rochet, G. Dufour, *et al.*, Surf. Sci. **338**, 143 (1995).
192. A. Ourmazd, D. W. Taylor, J. A. Rentschler, *et al.*, Phys. Rev. Lett. **59**, 213 (1987).
193. K.-O. Ng and D. Vanderbilt, Phys. Rev. B **59**, 10132 (1999).
194. A. A. Demkov and O. F. Sankey, Phys. Rev. Lett. **83**, 2038 (1999).
195. Y. Tu and J. Tersoff, Phys. Rev. Lett. **84**, 4393 (2000).
196. R. Buczko, S. J. Pennycook, and S. T. Pantelides, Phys. Rev. Lett. **84**, 943 (2000).
197. H. Kageshima and K. Shiraishi, Phys. Rev. Lett. **81**, 5936 (1998).
198. Y. Miyamoto, Phys. Rev. B **46**, 12 473 (1992).
199. T. Uchiyama and M. Tsukada, Phys. Rev. B **55**, 9356 (1997).
200. K. Kato and T. Uda, Phys. Rev. B **62**, 15978 (2000).
201. L. S. O. Johansson, R. I. G. Uhrberd, R. Lindsay, *et al.*, Phys. Rev. Lett. **74**, 2014 (1995).
202. R. Lindsay, P. L. Wincott, G. Rhornton, *et al.*, Surf. Sci. **398**, 301 (1998).
203. M. Nishijima, J. Yoshinobu, H. Tsuda, *et al.*, Surf. Sci. **192**, 383 (1987).

Translated by V. Isaakyan

Technical Physics, Vol. 49, No. 10, 2004, pp. 1280–1283. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 35–39.
Original Russian Text Copyright © 2004 by Vandakurov.

THEORETICAL
AND MATHEMATICAL PHYSICS

Simple Relationships for Determining the Asymmetry of Solar Rotation

Yu. V. Vandakurov

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

Received March 3, 2004

Abstract—The rate of rotation of the solar atmosphere is determined under the condition that the rate of rotation in one hemisphere (for example, in the northern one) at latitudes of 0, 30, 45, and 60° is known. Three symmetric (about the equator) and two asymmetric rotational modes are taken into account in the calculation, and the reasons why the lowest mode of asymmetric rotation may remain unexcited in the solar convective zone are discussed. A solution with a minimal asymmetry of rotation about the equator turns out to be of practical interest. In this case, the mean rate of solar rotation as a decreasing function of latitude, as well as the parameters of one subsurface high-latitude rapidly rotating flow, are in good agreement with the observations; however, the theory also predicts peaks in the latitude dependence of the rate of rotation. It is assumed that the peaks smooth out due to torsional waves present in the solar convective zone. *© 2004 MAIK "Nauka/Interperiodica".*

INTRODUCTION

At present, the spontaneous occurrence of the complex differential rotation of the solar convective zone with the still faster rotation of solar equatorial layers remains a most challenging question. Recent attempts to reveal a possible effect of the spatial distribution of the turbulent viscosity on the process under study [1, 2] have failed. According to the model proposed by Elliott *et al*. [1], nonuniform rotation generally becomes akin to that observed in the sun when the turbulent viscosity of the medium decreases.

Earlier, we studied the self-setting differential rotation of the stellar convective zone under the assumption of low total dissipation [3]. In this case, the rotation of a star with faster rotation of the equatorial layers does take place; however, the model also predicts the spatial variations of the rate of rotation, which as yet have not been observed.

In Section 1 of this study, we consider a simple solution that relates the general characteristics of the rotation to the rates of rotation at certain ("principal") latitudes. These latitudes are those providing a fairly accurate coincidence of the differences in the rotation frequencies for some rotational modes that are symmetric about the equator and correspond to neighboring latitudes. We are dealing with latitudes of 0, 30, 45, 60, and 90°. For these latitudes (of course, except for the pole), helioseismic observations are available, which simplifies the solution of the entire set of equations. In Section 1, we perform the appropriate calculations and derive simple relationships for the amplitudes of all the rotational modes, including those that are asymmetric about the equator. In general, the symmetry of rotation cannot be disregarded in our equations.

However, the solution mentioned above usually yields an excessively high rotational asymmetry about the equator, which can be reduced by appropriately correcting the rate of rotation at the principal latitudes. As a rule, the dissipation also decreases in this case. Below, we consider only models with a minimal asymmetry of rotation.

The problem being discussed becomes of special importance in view of the recently reported discrepancies in the helioseismic observations [4]. Since the observations were analyzed under the assumption that the solar rotation is symmetric about the equator, the asymmetry of the process might be a cause of these discrepancies. In Sections 1 and 2, we estimate the possible symmetry of rotation using the data cited in [4]. In these sections, we also discuss the problem of slow rotation of the circumpolar solar zones.

Note that some processes yet unexplored introduce a substantial uncertainty into the procedure of comparing the theoretical predictions with the observations. First of all, we mean the amount of the energy spent on the excitation of torsional waves, whose presence in the Sun is supported by numerous observations (see the review in [5]). This issue is considered in Sections 2 and 3.

1. DETERMINATION OF ROTATIONAL MODES

In the case of a differentially rotating medium, the key problem is to determine the amplitudes of individual modes, which specify the rotation of the medium as a whole. If the velocity vector has the azimuth component alone, one can represent the total rate of rotation as an expansion in these components of the vector spherical harmonics with the zero superscript. These harmonics can be expressed through the derivatives of the spherical functions with respect to the polar angle using the expressions presented in [6, Sect. 7.3, Eq. (26); Sect. 5.13]. Taking into consideration only five modes with the least subscripts, we obtain angular velocity Ω in the form

$$\begin{aligned}\Omega = \Omega_0\{u_1 + u_3(7/8)^{1/2}(5\cos^2\vartheta - 1)\\ + u_5[(55)^{1/2}/8](21\cos^4\vartheta - 14\cos^2\vartheta + 1)\\ + 5^{1/2}\cos\vartheta[u_2 + u_4(3/8)^{1/2}(7\cos^2\vartheta - 3)]\},\end{aligned} \tag{1}$$

where $\Omega_0 = \text{const}$; $\vartheta = \pi/2 - \gamma$ is the polar angle; γ is the latitude; and $u_j = u_j(r)$ are the radius-dependent numerical coefficients, which specify the asymmetric (symmetric) rotation of the medium about the equator when j is even (odd).

Note also that, in the case of rigid rotation, u_1 is close to unity and other coefficients u_j are small. At even j, the values of u_j specify the asymmetric rotation of the medium.

Since general equation (1) contains five unknown coefficients $u_j(r)$, we set five conditions on the quantity Ω/Ω_0 at a fixed depth in the convective zone; namely, we assume that this parameter is known at latitudes of 0, 30, 45, 60, and 90° in the northern hemisphere and takes values ν_1, ν_2, ν_3, ν_4, and ν_p, respectively. Our aim is to find these five coefficients; then, we will be able to determine the rate of rotation for any arbitrary angle ϑ. Here, coefficient ν_p is among the known quantities; however, as will be shown below, the feasibility of its determination needs further investigation.

From Eq. (1), one can derive the following expressions for the latitudes listed (principal latitudes):

$$u_1 - 0.935414u_3 + 0.927025u_5 = \nu_1, \tag{2}$$

$$\begin{aligned}u_1 + 1.118034u_2 + 0.233854u_3\\ - 0.855817u_4 - 1.100842u_5 = \nu_2,\end{aligned} \tag{3}$$

$$\begin{aligned}u_1 + 1.581139u_2 + 1.403121u_3\\ + 0.484123u_4 - 0.695269u_5 = \nu_3,\end{aligned} \tag{4}$$

$$\begin{aligned}u_1 + 1.936492u_2 + 2.572389u_3\\ + 2.668172u_4 + 2.143744u_5 = \nu_4,\end{aligned} \tag{5}$$

$$\begin{aligned}u_1 + 2.236068u_2 + 3.741157u_3\\ + 5.477226u_4 + 7.416198u_5 = \nu_p.\end{aligned} \tag{6}$$

Now, subtracting each equation from the preceding one, we obtain

$$\begin{aligned}-1.118034u_2 - 1.169268u_3 + 0.855817u_4\\ + 2.027867u_5 = \nu_1 - \nu_2,\end{aligned} \tag{7}$$

$$\begin{aligned}-0.463105u_2 - 1.169267u_3 - 1.339939u_4\\ - 0.405573u_5 = \nu_2 - \nu_3,\end{aligned} \tag{8}$$

$$\begin{aligned}-0.355353u_2 - 1.169268u_3 - 2.184049u_4\\ - 2.839013u_5 = \nu_3 - \nu_4,\end{aligned} \tag{9}$$

$$\begin{aligned}-0.299576u_2 - 1.169268u_3 - 2.809054u_4\\ - 5.272454u_5 = \nu_4 - \nu_p.\end{aligned} \tag{10}$$

Repeating the subtraction procedure and neglecting small corrections caused by the difference in the seventh decimal place of the second coefficients, we eliminate coefficient u_3:

$$\begin{aligned}-0.654929u_2 + 2.195756u_4 + 2.433440u_5\\ = \nu_1 - 2\nu_2 + \nu_3,\end{aligned} \tag{11}$$

$$\begin{aligned}-0.107752u_2 + 0.844110u_4 + 2.433440u_5\\ = \nu_1 - 2\nu_3 + \nu_4,\end{aligned} \tag{12}$$

$$\begin{aligned}-0.055777u_2 + 0.625005u_4 + 2.433441u_5\\ = \nu_3 - 2\nu_4 + \nu_p.\end{aligned} \tag{13}$$

After subtraction of each equation from the preceding one, we find

$$-0.5472u_2 + 1.3516u_4 = \nu_1 - 3(\nu_2 - \nu_3) - \nu_4, \tag{14}$$

$$-0.0520u_2 + 0.2191u_4 = \nu_2 - 3(\nu_3 - \nu_4) - \nu_p. \tag{15}$$

Thus, with the values of u_j at the principal latitudes known, coefficients u_2 and u_4, which characterize the symmetry of rotation, can be found from Eqs. (14) and (15). Other coefficients u_j are defined by Eqs. (2), (7), and (11). The rotation is symmetric about the equator if the rights of Eqs. (14) and (15) turn to zero, from which it follows that coefficients u_2 and u_4 also vanish.

For the solar rotation, coefficients ν_j with subscripts from 1 to 4 can be determined from the helioseismic data; however, the value of ν_p, which characterizes the rate of rotation at the pole, remains unknown (it may be nonzero if the rotation is asymmetric about the equator). Even if it is assumed that ν_p is known, solving the system of Eqs. (14) and (15) encounters serious difficulties because the coefficients multiplying u_2 are small. Because of this, the solution usually yields an excessively high low-latitude asymmetry, which is almost absent in the helioseismic data. Assuming that

$$u_2 = 0, \tag{16}$$

we find from Eqs. (14) and (15)

$$u_4 = [\nu_1 - 3(\nu_2 - \nu_3) - \nu_4]/1.3516, \tag{17}$$

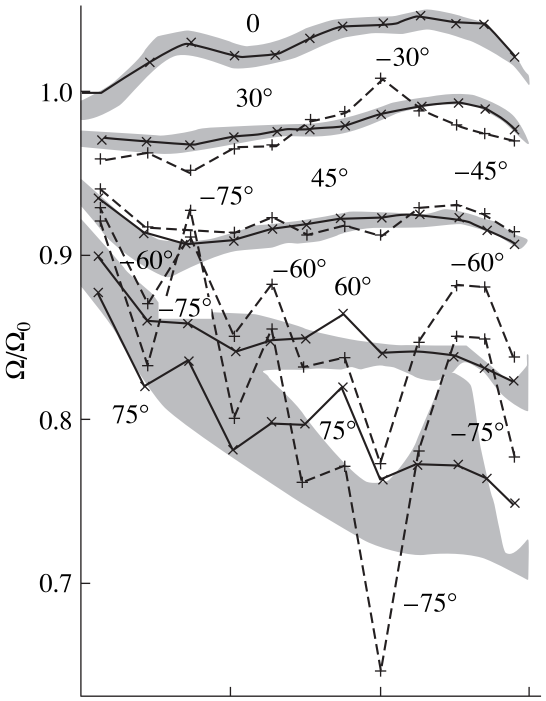
$$\nu_p = \nu_2 - 3(\nu_3 - \nu_4) - 0.2191u_4. \tag{18}$$

If the values of ν_p are the same at various depths of the convective zone, the solution will be unique. Another possibility of satisfying the equations is related to the formation of small but finite-size circumpolar zones where the medium is quiescent, i.e., does not rotate (such a situation may occur, e.g., because of the effluence of the matter). In this case, Eq. (15) should be eliminated from consideration.

A large body of helioseismic data [7, 8] indicates that the rate of solar rotation decreases considerably at high latitudes, which counts in favor of this hypothesis. Anyhow, one may first investigate the rotation of the greater part of the convective zone (see Section 2), leaving aside the determination of ν_p.

2. ROTATION DISTRIBUTION OVER THE SOLAR CONVECTIVE ZONE

Let the solar convective zone consist of 12 spherical layers with their relative radii lying in the range $0.71 \leq r/R_\odot \leq 0.99$. For the northern hemisphere, the initial values of angular velocities at the layer boundaries and at the principal latitudes (except for the pole) are set in accordance with the data presented in [4, (Fig. 1)]. Our goal is to describe the rotation at any depth and in both hemispheres of the convective zone.



Solid curves with crosses and dashed curves with the plus sign show the analytical values of the angular velocity in the northern and southern solar hemispheres, respectively, at the corresponding latitudes (numbers by the curves). The entire convective zone divided into 12 layers across the depth is considered. The shaded zones cover the experimental data [8] with regard to the broadening because of the spread in the observations.

It turns out that small errors (even those within the tolerable limits [4]) involved in the initial data may result in an extremely high spread of the asymmetry of rotation. Therefore, we corrected the initial data for the northern hemisphere in order to minimize rotation asymmetry about the equator in our theoretical model. Sometimes, the corrections even exceeded the admissible errors given in [4].

The results of our calculations are shown in the figure. Here, the solid lines with crosses and the dashed lines with the plus signs indicate the angular velocity in the northern and southern hemispheres, respectively (the data for ±75° are also included). The numbers by the curves indicate latitudes. The curves are normalized to the equatorial angular velocity at the bottom of the convective zone. The shaded regions show the helioseismic data obtained in [4]. The regions at latitudes of ±60° and ±75° are broadened because of the spread in the observations.

It is seen that the angular velocities vary considerably at high latitudes in the southern hemisphere, although the averaged theoretical velocity versus latitude dependences are to some extent similar to those following from the helioseismic observations. It is also of interest that the dashed and solid lines diverge significantly between ±30° and ±45°. This fact indicates that there may be a line to which the dashed and solid curves tend; i.e., the line corresponding to the symmetric rotation. By the way, the solid and dashed curves intersect at $r/R_\odot \approx 0.84$ and ≈ 0.92. Note that the rotation is symmetric throughout the convective zone at a latitude of about ±41°. This value is close to that for the stationary torsional waves observed in the sun (±42°, following [5]).

Thus, the hypothesis that the excitation of the torsional waves assists in smoothing out the abrupt variation of the rate of rotation seems plausible. Here, we mean that, upon formation, a spatial irregularity may be displaced into other layers, so that such an irregularity has no time to form (to be observed) in reality. This effect may explain why the sharp peaks at $r/R_\odot \approx 0.77$ and 0.90 are unobservable.

Of particular interest is the circumpolar rotation in which case the parameter ν_p defined by (18) plays a significant role. Our calculation shows that the value of ν_p varies with depth (ranges from 0.72 to 0.87). Such an appreciable difference in ν_p allows us to adopt (out of the two concepts considered in Section 1) the concept of formation of small circumpolar regions where the medium is quiescent (for example, because of the effluence of the matter). The considerable decrease in the rate of rotation at latitudes higher than 70° was first found in [7, 8].

One may expect that the slow rotation of these circumpolar regions will reduce the effect of smoothing out the rotation nonuniformity through the excitation of torsional waves in these layers. In this context, it seems interesting to consider the helioseismic data [8], which suggest the presence of a subsurface polar flow at latitudes of about 75° that rotates faster than the environment. Its rate of rotation reaches a maximum at $r/R_\odot \approx 0.95$. In the figure, something of the sort is also present in the southern hemisphere, although here the region of fast rotation is somewhat broader (from $r/R_\odot \approx 0.95$ to 0.97). Furthermore, a higher rate region is also observed at a latitude of –60°, which is inconsistent with the data in [8]. It seems likely that it is in this layer that the smoothing effect associated with the torsional waves comes into play.

Note also that, in our model, coefficient u_3 in formula (1) is depth-dependent and ranges between –0.025 and –0.055, while the absolute values of coefficients u_4 and u_5 are comparable to, or noticeably smaller than, the lower limit. It is also worth noting that certain portions of the solid and dashed curves are interchangeable. Other conclusions will be drawn in the following section.

3. DISCUSSION

It is generally accepted that the complex differential solar rotation is the result of viscous turbulent forces arising in the convective zone. However, we derived simple expressions (unrelated to viscous forces) that describe the distribution of the stationary atmospheric rotation. It turns out that in the approximation used, rotation symmetric about the equator takes place only if a number of severe conditions are fulfilled (see Section 1). In the solar convective zone, the asymmetric component of the rate of rotation is apparently present, so that bringing the calculation results into better agreement with the observations requires that an appropriate asymmetry-minimizing procedure be carried out. The fact that the processing of observations under the assumption that the rotation is symmetric about the equator sometimes yields contradictory results [4] also supports the presence of asymmetry in the solar rotation.

The rate of rotation calculated in terms of our model, which assumes a minimal symmetry of rotation, features considerable spatial variations, although averaging gives better agreement with the observation. It is also important that the latitudes where the symmetry of rotation and torsional waves observed in the sun disappear virtually coincide. Thus, the hypothesis that the torsional waves tend to suppress the rotation nonuniformity predicted by the theory deserves attention.

The arguments in favor of the rotation of the circumpolar layers were advanced in Sections 1 and 2. If such an effect actually takes place, the smoothing of the nonuniformities shown in the figure is probably less efficient. Then, the discrepancy between the theoretical predictions and the observations becomes insignificant. In the case of the fast subsurface flow at a latitude of 75° [8], the theory is in good agreement with the observations. However, it is important to note that the theory predicts the existence of the flow in only one hemisphere. Certainly, the point of smoothing efficiency calls for further investigation.

Another issue of interest is how the dissipation changes in the model with minimal asymmetry of rotation. The respective basic equation was considered elsewhere [3, 9]. In essence, one has to estimate quantity d, which is the difference between the inertial force $[\partial v_\varphi/\partial t]/(\Omega_0)^2/r$ and corresponding dimensionless radial component of the viscous turbulent force. In our designations, d is imaginary. If the balance condition is satisfied and the viscous force is absent, d turns to zero.

It is important that, for convective heat transfer to be regular, i.e., for the convective elements to move radially, an axisymmetric toroidal magnetic field must be present in the convective zone. At the bottom of the solar convective zone, this field equals 110 kG [10]. In other layers, the magnetic field is proportional to the square root of the equilibrium density. It appears that d reaches a minimum only if the lower part of the convective zone where $r/R_\odot < 0.81$ is magnetized.

In the model with a high asymmetry of rotation about the equator, the absolute value of d is usually very large, as follows from our calculations. For example, d may vary from $-1000i$ to $+1000i$ at a small change in the convective zone depth. In the case of the minimized model (see the figure), $d = -4.4i$; that is, this model is close to the equilibrium one. There also exists a model with zero d, but it allows for a somewhat higher asymmetry of rotation. Apparently, when determining the value of dissipation, one should take into account the contribution of the torsional waves.

REFERENCES

1. J. R. Elliott, M. S. Miesch, and J. Toomre, Astrophys. J. **533**, 546 (2000).
2. A. S. Brun and J. Toomre, Astrophys. J. **570**, 865 (2002).
3. Yu. V. Vandakurov, Pis'ma Astron. Zh. **28**, 633 (2002) [Astron. Lett. **28**, 560 (2002)].
4. J. Schou, R. Howe, S. Basu, *et al.*, Astrophys. J. **567**, 1234 (2002).
5. S. V. Vorontsov, J. Christensen-Dalsgaard, J. Schou, *et al.*, Science **296**, 101 (2002).
6. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Sci., Singapore, 1988).
7. A. C. Birch and A. G. Kosovichev, Astrophys. J. Lett. **503**, L187 (1998).
8. J. Schou, H. M. Antia, S. Basu, *et al.*, Astrophys. J. **505**, 390 (1998).
9. Yu. V. Vandakurov, Astron. Zh. **76**, 29 (1999) [Astron. Rep. **43**, 24 (1999)].
10. Yu. V. Vandakurov, Zh. Tekh. Fiz. **73** (3), 23 (2003) [Tech. Phys. **48**, 298 (2003)].

Translated by A. Sidorova

Technical Physics, Vol. 49, No. 10, 2004, pp. 1284–1288. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 40–44.
Original Russian Text Copyright © 2004 by Buranov, Gorokhov, Karelin, Repin.

GAS DISCHARGE, PLASMA

Study of the Microscopic Characteristics of a Repetitive Microsecond Barrier Discharge

S. N. Buranov, V. V. Gorokhov, V. I. Karelin, and P. B. Repin

All-Russia Research Institute of Experimental Physics, Russian Federal Nuclear Center,
Sarov, Nizhni Novgorod oblast, 607188 Russia
e-mail: karelin@ntc.vniief.ru

Received December 16, 2003

Abstract—Results are presented from investigations of the electric, optical, and spatial characteristics of a microsecond barrier discharge in atmospheric-pressure air in a uniform electric field. The currents in individual microscopic channels were measured by using sectioned electrodes. It is found that the first series of microscopic discharges is characterized by a relatively low current density (lower than 280 A/cm^2). In the subsequent series, the microscopic discharges have a pronounced channel structure with a high current density. The prints of the channels on the metal anode have the form of microscopic craters 4 μm in diameter and 0.5 μm in depth. *© 2004 MAIK “Nauka/Interperiodica”.*

INTRODUCTION

In recent years, repetitive microsecond barrier discharges (BDs) have gained widespread application in technology, in particular, in ozone therapeutic apparatus [1–4]. At a pulsed power supply, volume (diffuse) discharges ignited in a gas gap are characterized by a high efficiency of ozone production [1]. The structure and microscopic characteristics of BDs supplied with a main-frequency sinusoidal voltage have been studied in detail. There are also publications on BDs with a half-cycle duration of longer than 100 μs. However, the parameters of series of microscopic discharges and the characteristics of microscopic BD channels formed under the action of short voltage pulses are still poorly studied (see, e.g., [1, 5]). This circumstance stimulated our investigations of microsecond BDs. The results of these studies are presented in this paper.

EXPERIMENTAL SETUP AND MEASUREMENT TECHNIQUE

We investigated the characteristics of a BD in atmospheric-pressure air in a discharge gap with a uniform electric field. A block-diagram of the experimental setup is shown in Fig. 1. The discharge was initiated by high-voltage pulsed generator *1* (the design of the generator is described in [6]). The generator operated in either a single-pulse regime or a repetitive regime with a repetition frequency from fractions of 1 Hz to several kHz. Quasi-sinusoidal voltage pulses consisting of two half-waves were formed at electrodes *2*. The amplitude of the first half-wave was up to 9 kV at a voltage growth rate of 4 kV/μs, the duration of the half-wave being 5 μs. The amplitude of the second half-wave did not exceed 2.5 kV.

The discharge gap was formed by ring electrodes *2* with an external diameter 48 mm and an internal diameter of 24 mm. The electrode area was 13.5 cm^2. The electrodes were made of stainless steel and were mounted on insulator *3*. Between the electrodes, plate *5* with a thickness of 1 mm was installed with the help of dielectric inserts *4* at the distance $d = 0.6$ mm from each of the electrodes. Plate *5* was made of KU-1 quartz with the permittivity $\varepsilon = 2.4$ (at a frequency of 500 kHz and temperature of 293 K). The air was input and output through the electrode holes. The gas flow rate was 1 l/min.

To stabilize breakdown, the discharge gap was preionized with a corona discharge ignited between initiating electrodes *6* [5]. The initiating electrodes in the form of arcs were made of nickel foil strips 2 mm in width and 0.1 mm in thickness. The strips were located above the central electrode holes and touched the dielectric surface.

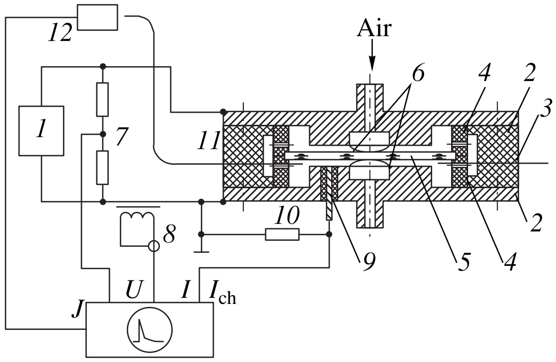


Fig. 1. Block diagram of the experimental setup.

1063-7842/04/4910-1284$26.00 © 2004 MAIK “Nauka/Interperiodica”

In our experiments, we recorded the discharge voltage U, the total discharge current I, the currents I_{ch} through individual microscopic channels, the prints of these channels, the emission intensity J from the discharge in the 0.30- to 0.36-μm wavelength range, and the discharge emission spectrum. Using the waveforms of the currents I and I_{ch}, we calculated the charge Q transferred by a series of microscopic discharges and the charge q transferred by an individual microscopic discharge.

The temporal and amplitude characteristics of the signals were measured with the help of HP 54542A and HP 54720D oscilloscopes with bandwidths of 500 and 2 MHz, respectively. The bandwidth of low-inductive voltage divider *7*, which was used to measure the voltage U, was 500 MHz. Total current I was measured by Rogowski coil *8* with an operating frequency of up to 200 MHz.

The currents I_{ch} in individual microscopic discharges were measured by using sectioned electrodes [7]. The currents were recorded with the help of two current receivers *9* with measuring surfaces 1.5 mm in diameter. The current receivers were mounted in diametrically opposite 2.5-mm-diameter holes in one of the electrodes at a distance of 37 mm from one another. In Fig. 1, only one of them is shown. For the measuring surfaces of the current receivers to be at the same level relative to the electrode, the working surface of the electrode was polished after mounting the current receivers. The ratio of the areas of the current receiver and the electrode was 0.0013; this allowed us to ignore the capacitive component of the recorded current. The current receivers were connected to shunts *10* made of high-frequency resistors with a resistance of 1 Ω and mounted in coaxial return-current electrodes. The shunt bandwidth was not lower than 2 GHz.

The structure of microscopic discharges in a BD was studied by analyzing erosion spots emerging at the measuring surfaces of the current receivers after the passage of the current pulses I_{ch}. The discharge channel prints were photographed through a microscope with a magnification of up to 400. The image scale was determined with the help of diffraction gratings (600 and 1200 groove/mm) that were photographed with the same magnification. The erosion spot depth was evaluated by measuring the length of the crater wall shadow with regard for the angle of electrode illumination. To enhance the contrast of the channel prints, in some experiments, a thin soot layer (<0.25 μm) was deposited on the surfaces of the current receivers and electrodes. After each of the experiments, the electrodes were polished anew.

The timing of a microscopic discharge with respect to the instant of breakdown and its association with the number of a series were performed by using the waveforms of the voltage, total current, and discharge emission. The current receivers were also used to record time delay t_1 between the emergence of microscopic discharges within one series.

The discharge emission was received by 0.6-mm-diameter quartz lightguide *11*. The lightguide aperture was 25° and covered the central region of the discharge gap (11–12% of the total volume of the interelectrode gap), including the measuring surfaces of the current receivers. The discharge emission was detected by SNFT-3 photomultiplier *12* with an operating frequency of up to 80 MHz. In front of the detector, color and neutral filters were installed. The discharge spectrum was recorded by an ISP-30 spectrograph with a slit width of 20–40 μm. In this case, the spectrograph was installed instead of the photomultiplier at the lightguide output. The spectrum was recorded with an exposure time of 4–6 h at a repetition frequency of 1.6 kHz.

EXPERIMENTAL RESULTS

Our experiments showed that a BD was initiated at the front or at the maximum of the voltage pulse (at a voltage of 4.5–9.0 kV) and operated only during the first half-waves of the pulses. Typical waveforms of U, I, J, and I_{ch} for one and several series of microscopic discharges are presented in Figs. 2 and 3, respectively.

It was found that the basic parameters governing the number N of the series of microscopic discharges formed during a supply pulse and their duration were the breakdown voltage U_{br} in the first series and the charge transferred by the first series of microscopic discharges. For example, at $U_{\mathrm{br}} > 8$ kV, the discharge usually consisted of only one series. The full width at half-maximum (FWHM) of the emission pulse correspond-

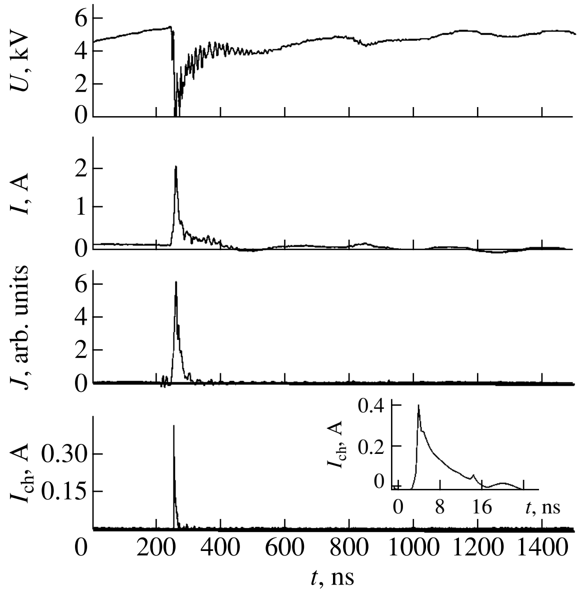


Fig. 2 Waveforms of the signals from one series of microscopic discharges. The current pulse I_{ch} is shown in the inset on a larger scale.

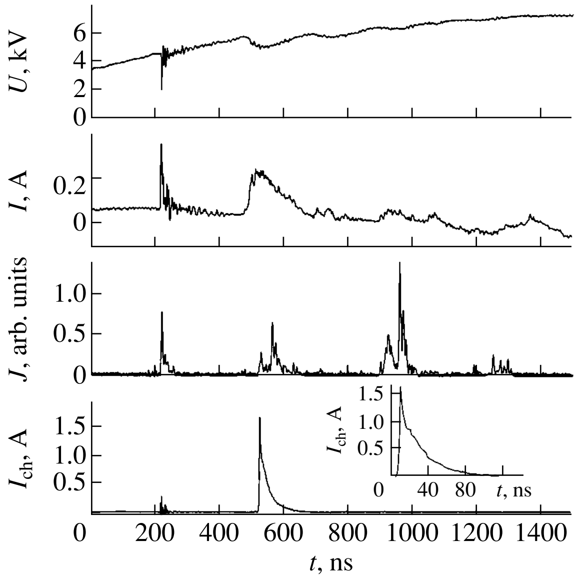


Fig. 3. Waveforms of the signals from four series of microscopic discharges.

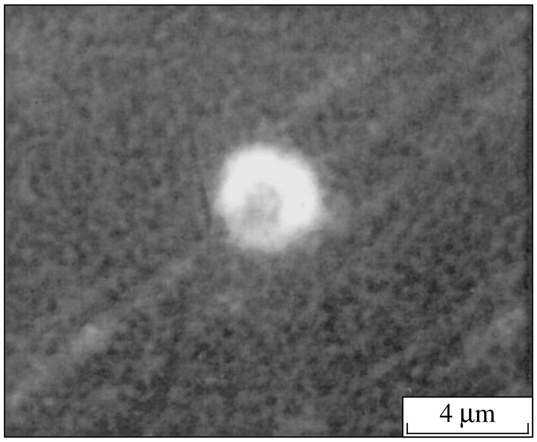


Fig. 4. An erosion spot on the measuring surface of a current receiver. The discharge parameters are $D_{ch} \approx 3.5$ µm, $h \approx 0.5$ µm, $I_{ch} = 1.7$ A, $\tau_{0.5} = 25$ ns, and $U_{br} = 5.5$ kV.

ing to this series was $t_J = 5–15$ ns. At lower values of U_{br}, the voltage increased again to the breakdown one during the supply pulse and, within this time interval, the subsequent series of microscopic discharges were formed. Thus, for $U_{br} < 6$ kV, the number of series was $N = 1–5$, the average number being $N = 4$. In this case, the FWHM of the emission pulse was $t_J = 5–30$ ns for the first series, whereas for the subsequent series, it was $t_J = 50–300$ ns. At high values of U_{br}, the currents in the first series usually exceeded the currents in the subsequent series. The number of series N was found to be inversely proportional to the charge transferred during the first series. Even at moderate values of U_{br} and large values of Q, the discharge sometimes consisted of only one series (Fig. 2).

Our investigations showed that the amplitudes of the I_{ch} pulses, their FWHMs $\tau_{0.5}$, and the dimensions of the channel prints varied in rather wide ranges. When a positive voltage was applied to the current receivers, currents I_{ch} with an amplitude of 5 A and FWHM of $\tau_{0.5} = 0.5–2.0$ ns were recorded at $U_{br} > 8$ kV (one series). In this case (including experiments in which the current receivers and the electrodes were covered with soot), no discharge channel prints were observed.

At $U_{br} = 6–8$ kV, the microscopic discharge currents reached a value of $I_{ch} = 2$ A both in the first and the subsequent series and their FWHMs were $\tau_{0.5} = 1–3$ ns in the first series and $\tau_{0.5} = 7–10$ ns in the subsequent series. The channel prints were observed at $I_{ch} \geq 0.5–1$ A, but only in experiments with current receivers covered by soot and only for microscopic discharges of the subsequent series. The print diameters did not exceed 1 µm. When several prints occurred at the current receiver, the current I_{ch} turned out to be modulated and the number of spikes coincided with the number of prints. The interval between the spikes was 2–5 ns. At $U_{br} < 6$ kV, the microscopic discharge currents I_{ch} in the first series did not exceed 0.5 A, the FWHM of the current pulses being $\tau_{0.5} = 3–7$ ns. In the subsequent series, the currents I_{ch} reach a value of 1.7 A and $\tau_{0.5}$ was 8–25 ns. In this case too, the prints were produced only by microscopic discharges of the subsequent series. When the electrodes were covered with soot, the microscopic discharges sputtered the soot, thereby stripping the electrode surface, or fired the soot into the electrode surface; in this case, prints with diameters of up to 7 µm were observed. When the surfaces of the current receivers and electrodes were not covered with soot, the prints had the form of microscopic craters with diameters of up to $D_{ch} = 4$ µm and depths of up to $h = 0.5$ µm (Fig. 4); inside craters with diameters larger than 2 µm, there was a core.

Our experiments showed that the characteristics of discharges of the first series did not change when the electrode polarity was reversed. It was also found that, in the case of a negative polarity, the discharge currents in the subsequent series were appreciably lower than in the case of a positive polarity. For example, the maximum value of I_{ch} in the first series was $I_{ch} = 5$ A, while in the subsequent series it was $I_{ch} = 0.5$ A and no erosion spots were observed.

Typical calculated values of the charge $Q = \int I(t)dt$ transferred in the first and the subsequent series of microscopic discharges for different values of U_{br} are presented in Table 1. The table also presents the values of the charge $q = \int I_{ch}(t)dt$ transferred by individual microscopic discharges of the corresponding series.

The measurements of t_1 showed that the delay time between two microscopic discharges of the first series recorded with the help of the current receivers varied

Table 1

U_{br}, kV	First series		Subsequent series	
	Q, nC	q, nC	Q, nC	q, nC
8–9	16–40	0.5–10	–	–
6–8	8–20	0.2–5	1–20	0.1–20
5–6	0.4–10	0.05–3	1–50	0.05–45

from zero to the maximum value determined by U_{br}. As U_{br} increased, the scatted in the t_1 values decreased. For example, at $U_{br} < 6$ kV, the maximum delay time was $t_1 = 25$ ns, while at $U_{br} > 8$ kV, the delay time did not exceed 10 ns.

In the subsequent series (in contrast to the first one), the emergence of microscopic discharges was accompanied by a large scatter in t_1 (up to 300 ns). The number of microscopic discharges in these series varied within a wide range and depended on the transferred charge: at $q > 20$ nC, a series could consist of a single microscopic discharge.

In the discharge spectrum, we observed the following bright bands of the second positive system of nitrogen: 405.9 nm (0–3 band), 380.5 nm (0–2 band), 357.7 nm (0–1 band), 337.1 nm (0–0 band), and 315.9 nm (1–0 band), as well as the 391.4-nm band of the first negative system of nitrogen. The lifetimes of the corresponding states does not exceed 1 ns; hence, the shape of the emission pulse must reproduce the current pulse shape. However, the correlation between the I and J pulses was observed only for the first series of microscopic discharges; the ratio between their amplitudes I/J varied from pulse to pulse by no larger than 50%. This means that microscopic discharges of the first series were distributed rather uniformly over the discharge gap. In the subsequent series, the emission from microscopic discharges often did not fall into the lightguide aperture; this caused a discrepancy between the amplitudes of individual pulses in the I and J waveforms (Fig. 3).

DISCUSSION

In contrast to most of the experiments with BDs [1, 4], in our case, the discharge operated at $pd < 200$ Torr cm. The estimated length of the avalanche–streamer transition was larger than d throughout the entire range of the field strength under study ($E/p < 70$ V/(cm Torr)); this is characteristic of the Townsend mechanism for breakdown. At the same time, short breakdown times (<2.5 μs) allowed us to assume that, in our case (in contrast to the classical Townsend discharge), the avalanche stage was governed by photoprocesses [8].

Throughout the entire range of U_{br} under study, microscopic discharges of the first series were almost uniformly distributed over the electrode surface and, probably, were diffuse. The lower estimate of the current density j_{ch} in these discharges may be obtained by assuming that the channel diameter D_{ch} is equal to the diameter D of the current receiver: $j_{ch} = 28–280$ A/cm^2. Table 2 presents the results of processing the waveforms for the maximum value of I_{ch} and two values of U_{br} and estimates of the channel parameters obtained under the same assumption. In Table 2, the following notation is used: U_d is the voltage drop across the gas gap d at the instant of breakdown, calculated with account of the ratio of the gas gap capacitance to the capacitance of the dielectric barrier; τ_f is the duration of the current pulse front; $\tau_{0.1}$ is full width of the current pulse at a level of 0.1; $W_{ch} = \int I_c U dt$ is the energy deposited in the channel; $P = j_{ch}E$ is the specific deposited power; and $n \approx j_{ch}/e\nu$ is the maximum electron density, where e and ν are the electron charge and the electron drift velocity, respectively.

It follows from Table 2 that, in spite of the relatively large amplitudes of I_{ch} and deposited power, the current density, the electron density, and the energy deposited in the channels of the given BD are close to the parameters of microscopic discharges observed in main-frequency BDs and those obtained in numerical calculations [9].

The experimental data allow us to estimate the increase in the temperature ΔT in the channels of the first series: $\Delta T \approx W_{ch}/\rho C_p V_{ch}$, where ρ and C_p are the density and specific heat of air, respectively, and V_{ch} is the channel volume calculated under the assumption that $D_{ch} = D$. Assuming that all the deposited energy is spent on heating and using the data from Table 2, we obtain $\Delta T = 6–35$ K.

It should be noted that a fairly large charge is transferred in the first series at $U_{br} > 8$ kV (Table 1). The large transferred charge leads to a decrease in the voltage drop across the gap; this fact explains the observed

Table 2

U_{br}, kV	U_d, kV	I_{ch}, A	τ_f, ns	$\tau_{0.5}$, ns	$\tau_{0.1}$, ns	E/p, V/(cm Torr)	j_{ch}, A/cm^2	W_{ch}, μJ	P, W/cm^3	q, nC	n, cm^{-3}
9	6.35	5	0.6	1	5.7	70	280	50	1.7×10^7	10	6×10^{13}
6.5	4.6	1.2	2	2.5	5	50	67	8	3×10^6	3	2×10^{13}

dependence of N on the voltage and charge Q in the first series.

In the subsequent series (in contrast to the first series), the discharge has a pronounced channel structure and, as a whole, is close to a spark discharge [8]. Microscopic discharges in the subsequent series are characterized by a longer duration and larger values of q and W_{ch} (up to 100 μJ). We especially note the presence of erosion spots. If we assume that the entire current of a microscopic discharge flows through an anode spot, then, for the erosion spot shown in Fig. 4 ($D_{ch} \approx$ 3.5 μm and I_{ch} = 1.7 A), we obtain the following estimate for the current density: $j_{ch} \approx 1.7 \times 10^7$ A/cm^2. It is evident that this is an overestimated value that does not take into account the current flowing through the spot periphery. At the same time, the presence of erosion spots 0.5 μm in depth at the metal surface with a vaporization temperature of ~3300 K evidences the high local density of energy release. An estimate of the energy that is necessary for the formation of the microscopic crater shown in Fig. 4 ($D_{ch} \approx$ 3.5 μm and $h \approx$ 0.5 μm) gives $W \sim 1$ μJ, the energy deposited in the channel being $W_{ch} \approx 90$ μJ.

We note that the fine structure of BDs was earlier studied mainly by optical methods. According to those measurements, the minimum channel diameters were 0.1–0.3 mm [1, 4]. However, it was repeatedly pointed out in the literature that the measurements of the geometrical parameters of the current channels (the diameters of streamers and sparks and the distributions of the current density) by optical and electron-optical methods are fairly complicated [10]. This is due to the rather low (no better than 0.1 mm) integral resolution of the equipment used (especially when operating with high voltages). On the other hand, the inner channel structure may be invisible against the background of the luminous large-diameter outer shell. The results obtained in this study allow us to suppose that analogous microscopic channels can also be formed in high-frequency BDs.

The reason why the character of the discharge is different in the first and subsequent series is not quite clear. The contraction of a BD is probably determined by its history: the subsequent series are generated in a field that is perturbed by the microscopic discharges of the first series.

CONCLUSIONS

The number of the series of microscopic discharges in a BD is governed by the breakdown voltage and the charge transferred by the first series of microscopic discharges.

The parameters of microscopic discharges in the first and subsequent series are different. Microscopic discharges in the first series appear to be diffuse and are characterized by a relatively low current density and low temperature. In the subsequent series, the BD is contracted, the microscopic discharges have a pronounced channel structure with a high current density, and the diameters of the microscopic channels are appreciably less than those observed earlier in BDs.

Presumably, the channel structure is formed due to the nonuniform distribution of the charges transferred to the dielectric barrier in the first series of microscopic discharges.

REFERENCES

1. V. G. Samoĭlovich, V. I. Gibalov, and K. V. Kozlov, *Physical Chemistry of Barrier Discharge* (Mos. Gos. Univ., Moscow, 1989) [in Russian].
2. S. N. Buranov, V. V. Gorokhov, V. I. Karelin, and P. B. Repin, in *Proceedings of the 2nd All-Russia Scientific–Practical Conference "Ozone in Biology and Medicine"* (Nizhni Novgorod, 1995), p. 103.
3. V. I. Panteleev and A. V. Kosarev, in *Proceedings of the 3rd All-Russia Scientific–Practical Conference "Ozone and Methods of Efferent Therapy in Medicine"* (Nizhni Novgorod, 1998), p. 232.
4. V. V. Lunin, M. P. Popovich, and S. N. Tkachenko, *Physical Chemistry of Ozone* (Mos. Gos. Univ., Moscow, 1998) [in Russian].
5. S. N. Buranov, V. V. Gorokhov, V. I. Karelin, and P. B. Repin, in *Proceedings of the 12th IEEE International Pulsed Power Conference, Monterey, California, 1999*, Vol. 2, pp. 1421–1424.
6. S. N. Buranov, V. V. Gorokhov, V. I. Karelin, and P. B. Repin, Prib. Tekh. Éksp., No. 1, 134 (1999).
7. H. F. A. Verhaart and P. C. T. Laan, J. Appl. Phys. **53**, 1430 (1982).
8. Yu. D. Korolev and G. A. Mesyats, *Field Emission and Explosion in Gas Discharge* (Nauka, Novosibirsk, 1982) [in Russian].
9. V. I. Gibalov and G. Pich, Zh. Fiz. Khim. **68**, 931 (1994).
10. É. M. Bazelyan and Yu. P. Raĭzer, *Spark Discharge* (Mosk. Fiz. Tekh. Inst., Moscow, 1997) [in Russian].

Translated by N. Mende

Technical Physics, Vol. 49, No. 10, 2004, pp. 1289–1295. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 45–51.
Original Russian Text Copyright © 2004 by Kuryshev, Kozlov.

Effect of Convection and Heat Conduction on the Laser Gain in a Nonequilibrium Recombinational Hydrogen Wall Plasma

A. P. Kuryshev and S. A. Kozlov

Voenmekh Ustinov Baltic State Technical University, Pervaya Krasnoarmeiskaya ul.,
St. Petersburg, 190005 Russia

Received February 17, 2004

Abstract—The effect of a conducting body with a large heat capacity on the thermal state of a hydrogen plasma in the boundary surface layer is analyzed. It is shown that, under certain conditions, the energy accumulated in the plasma within the boundary layer can be efficiently transferred to the conducting body; this leads to plasma overcooling. The mathematical model of plasma incorporates mechanisms for convective heat exchange and heat conduction. The possibility is analyzed of the existence of nonequilibrium recombination states with inverse population in the overcooled wall plasma. It is shown that the maximum gain (a few tenths of cm^{-1}) on hydrogen nuclei is achieved at the 3–2 transition for the following initial parameters: the plasma pressure is 1–3 atm, the plasma temperature is 0.5 eV, the tungsten surface temperature is 300 K, and the body radius is 0.5–1.0 m. *© 2004 MAIK “Nauka/Interperiodica”.*

The possibility of creating efficient lasers to be used in thermonuclear fusion, laser-driven thrusters, energy-consuming chemical processes, etc., has been widely discussed in the literature [1, 2]. One of the problems here is to provide the conditions under which a steady-state or pulsed recombining plasma efficiently amplifies radiation at a frequency corresponding to a certain transition of the plasma atoms, molecules, or ions.

Previously, we analyzed the effect of a conducting body with a large heat capacity on the thermal state of a hydrogen plasma in the boundary surface layer within the time-dependent model of heat conduction [3]. It was shown that, under certain conditions, the energy accumulated in the plasma within the boundary layer can be efficiently transferred to the conducting body; this leads to plasma overcooling. The characteristic time scale of these processes is on the order of 10^{-8}–10^{-3} s for an unperturbed plasma temperature of up to 2 eV and a pressure of up to a few atmospheres. It was shown that the overcooled nonequilibrium recombinational plasma that forms in the surface layer can undergo conversion from an ideal into a nonideal state. The possibility was analyzed of the existence of nonequilibrium recombination states with inverse population in overcooled wall plasma. The possibility was also studied of the amplification of radiation in overcooled nonequilibrium recombinational hydrogen wall plasma within time-dependent models of the level-by-level relaxation of hydrogen ions via two-level and multilevel one- and multiphoton transitions. The gain was calculated for radiation from 4–1, 4–2, 4–3, 3–1, and 3–2 transitions of stripped hydrogen nuclei within different models of level-by-level relaxation. It was shown that the maximum gain (a few tenths of cm^{-1}) on hydrogen nuclei can be achieved at the 3–2 transition for the following initial parameters: the plasma pressure is 1–3 atm, the plasma temperature is 0.5 eV, the tungsten surface temperature is 300 K, and the body radius is 0.5–1.0 m.

In this study, the process of convective heat exchange, which, along with heat conduction, occurs in actual physical systems, was incorporated in the mathematical model. Taking into account this process may introduce corrections to the previously obtained results [3].

The problem was investigated using the following model. Under the conditions formulated above, the characteristic time of plasma cooling is much longer than the time of energy exchange between the plasma components. This allows one to use a one-fluid plasma model in solving the problem of heat transfer within the thermal boundary region beyond the double layer. The characteristic spatial scale of the plasma thermal perturbation induced by a body with a large heat capacity is much less than the characteristic size of the body. This makes it possible to use a one-dimensional formulation in solving the problem of heat transfer in plasma (the external problem). To solve the problem of heat transfer inside the body (the internal problem), one can use an analytic solution for a spherical body [4]. The external problem is solved numerically using the well-known mathematical model, namely, the one-dimensional boundary problem of an unsteady temperature field in plasma within the thermal boundary region beyond the double layer with allowance made for heat conduction and convective heat exchange:

$$\frac{\partial\rho}{\partial t}+\frac{\partial}{\partial x}(\rho V_x) = 0,$$

$$\rho\left(\frac{\partial V_x}{\partial t}+V_x\frac{\partial V_x}{\partial x}\right) = -\frac{\partial p}{\partial x},\quad p = \rho RT,$$

$$\rho c_V\left(\frac{\partial T}{\partial t}+V_x\frac{\partial T}{\partial x}\right) = \frac{\partial}{\partial x}\left(\lambda\frac{\partial T}{\partial x}\right)-p\frac{\partial V_x}{\partial x},$$

with the initial conditions

$$T(0,x) = T_0,\quad V_x(0,x) = 0,\quad \rho(0,x) = \rho_0,$$

and the boundary conditions

$$T(t,0) = T_0,\quad \lambda\frac{\partial T}{\partial x}\bigg|_S = q = \sum q_k,$$

$$V_x|_S = 0,\quad \rho(t,0) = \rho_0.$$

The heat exchange between the plasma and the conducting body with a large heat capacity is determined by the contributions from different heat-exchange mechanisms involving electrons, ions, and neutrals. These mechanisms are related to the effect of the electric field, which either forms in the plasma–body interaction (the floating potential) or is produced by an external source. In the latter case, electron and/or ion heat-exchange mechanisms can come into play. After being absorbed, an electron delivers energy on the order of the work function to the body. In an event of surface Auger neutralization [5], an ion delivers energy on the order of the difference between the ionization energy and the work function to the body. Hence, in both cases, the energy is on the order of a few electronvolts. The heat-exchange mechanisms are of quantum nature.

In general, the total heat flux from the plasma to the conducting body depends strongly on the charge of the body, which takes energy from the plasma. Hence, the thermal and charge states of the plasma and the body are closely related to one another. Therefore, one has to solve a complete self-consistent problem. However, within the model of a Maxwellian plasma that is in equilibrium at infinity and under conditions such that the electric field is screened by a thin double layer with a characteristic size on the order of the Debye length, which is much less than the mean free path of the plasma particles, it is possible to separate the electrodynamic and the heat-transfer problems. In calculating energy fluxes in the framework of kinetic theory, one can use known solutions of the electrodynamic problem for the electron and ion distribution functions. The electrons and ions obey Maxwellian–Boltzmann velocity distributions, whereas the neutral particles obey a Maxwellian distribution.

The energy flux densities q_k also depend on the plasma composition. The initial equilibrium composition of hydrogen plasma is described by the equation for the conditional chemical reaction

$$M_T\mathrm{H}^+e^- \longrightarrow M_{\mathrm{H}_2}\mathrm{H}_2+M_\mathrm{H}\mathrm{H}+M_{\mathrm{H}^+}\mathrm{H}^+$$
$$+M_{\mathrm{H}_2^+}\mathrm{H}_2^++M_{\mathrm{H}^-}\mathrm{H}^-+M_e e,$$

where M_T is the number of moles of the H^+e^- plasma “molecules” and M_k is the number of moles of the kth plasma component.

Calculations of the initial equilibrium composition of an unperturbed isothermal plasma for the given plasma parameters (pressure p and temperature T) are based on solving a set of nonlinear algebraic equations for the mole fractions of the plasma components. The set of equations consists of the conservation equations for the number of hydrogen nuclei and electrons, the equations for the chemical equilibrium constants, and the Dalton equation.

For a given plasma composition, the energy flux densities q_k onto the body can be found by the following formulas [3, 4]:

(i) For electrons, we have

$$q_e = J_e^-(2kT_e+e\Phi_b+e\Phi^-),$$

$$J_e^- = \frac{1}{4}n_e v_{T_e}\exp\left(-\frac{e\Phi^+}{kT_e}\right),$$

$$v_{T_e} = \sqrt{\frac{8kT_e}{\pi m_e}},\quad \Phi^{\mp} = \pm\varphi_s\eta(\pm\varphi_s),$$

where J_e^- is the flux density of the plasma electrons onto the body, v_{T_e} is the thermal velocity of plasma electrons, the quantities $\Phi^{\mp}$ take into account a possible change in the electric field polarity, φ_s is the floating potential of the body surface, Φ_b is the work function, $e > 0$ is an elementary charge, and $\eta(x)$ is the Heaviside step function.

(ii) For the ions of the kth plasma component, we have

$$q_k = J_k^-(2kT_k+e\Phi^++e(I_k-\Phi_b)-2kT_s),$$

$$J_k^- = \frac{1}{4}n_k v_{T_k}\exp\left(-\frac{e\Phi^-}{kT_k}\right),\quad v_{T_e} = \sqrt{\frac{8kT_k}{\pi m_k}},$$

where J_k^- is the flux density of the ions of the kth species onto the body, v_{T_k} is the ion thermal velocity, I_k is the ionization energy for the corresponding plasma component, and T_s is the surface temperature of the body that cools the plasma.

(iii) For neutral particles, we have

$$q_k = J_k^-(2kT_k-2kT_s),\quad J_k^- = \frac{1}{4}n_k v_{T_k},$$

where J_k^- is the flux density of the neutral particles of the kth species onto the body and v_{T_k} is the thermal velocity of neutral particles.

The formulas for the energy flux densities q_k were obtained under the assumption that the electrons incident onto the body are completely absorbed and the incident ions lose their charge via Auger neutralization. The neutral particles and neutralized ions undergo diffuse reflection from the body surface.

It is expedient to analyze the above effect of plasma overcooling in the wall region in order to find out whether or not lasing is feasible. It is known [1, 2] that light amplification in the far UV and soft X-ray spectral regions can occur via different mechanisms for creating an inverse population of the working transitions, in particular, via recombinational pumping. For high densities and low temperatures of free electrons, three-body recombination prevails (compared to photorecombination) and high-lying ion states are mainly populated. Collisions with free electrons and spontaneous radiative decays form a level-by-level recombinational flux of the bound electrons into the ground state. The lower the principal quantum number n of a level, the higher the rate of its radiative decay and the lower the rate of its collisional quenching. As a result, inverse population of the levels with sufficiently small quantum numbers $b > a > 1$ can arise. The upper level b is pumped by the recombinational flux, which is mainly sustained by collisions with free electrons, whereas the lower level a is depopulated via spontaneous radiative transitions to lower lying levels, including the ground state.

The main problem in creating an inversely populated medium is that it can be formed only within a narrow range of the main macroscopic parameters—the temperature and density (pressure) of the plasma [1, 2]. Moreover, the range of acceptable temperatures and electron densities is still narrow if a higher gain is required. Gain media with the highest gains,

$$k_{ba} = \frac{\pi^2 c^2 A_{ba}}{\omega_{ba}^2 \Delta\omega_{ba}} \Delta N_{ba},$$

are of practical importance. Here, ω_{ba} is the transition frequency related to the wavelength via the formula

$$\lambda_{ba} = \frac{2\pi c}{\omega_{ba}} = \frac{91.1}{Z^2} \frac{b^2 a^2}{b^2 - a^2} \text{ [nm]},$$

where A_{ba} is the rate of spontaneous radiative decay via the $b \longrightarrow a$ transition; $\Delta\omega_{ba}$ is the effective width of the gain line; $N_{ba} = N_b - N_a^* g_b/g_a$ is the inverse population; N_b and N_a are the populations of the b and a states, respectively; and $g_b = 2b^2$ and $g_a = 2a^2$ are the statistical weights of these states.

When the temperature is so low that the electron-impact excitation from the ground state is insignificant, the gain can be represented as [2]

$$k_{ba} = \alpha k_{ba}^+,$$

where α is the fraction of the recombination centers (hydrogen nuclei) in the ion component and k_{ba}^+ is the gain corresponding to a fully ionized plasma.

The process of the recombinational plasma decay with respect to the H^+ ion component is determined by the continuity equation with a sink,

$$\frac{\partial N}{\partial t} + \frac{\partial}{\partial x}(N_+ V_x) = -\beta_R N_+,$$

$$\beta_R = 5.6 \times 10^{-27} N_+^2 T^{-9/2} \text{ [s}^{-1}\text{]}$$

(where T is in eV, N_+ is in cm^{-3}, and β_R is the recombination coefficient in three-body collisions) and following initial and boundary conditions:

$$N_+(0, x) = N_+^0, \quad N_+(t, 0) = N_+^0.$$

The kinetics of the level-by-level relaxation of electrons in a hydrogen atom is determined by the Kolmogorov set of equations for the populations within a multilevel time-dependent model with allowance made for multiphoton transitions:

$$\frac{dN_4}{dt} = \beta_R N_+ - K_4 N_4,$$

$$\frac{dN_3}{dt} = K_{43} N_4 - K_3 N_3, \quad \frac{dN_2}{dt} = K_{43} N_4 + K_{32} N_3 - K_2 N_2,$$

$$\frac{dN_1}{dt} = K_{41} N_4 + K_{31} N_3 + K_{21} N_2,$$

$$\sum_{i=1}^{4} N_i(t) + N_+(t, R) = N_+^0,$$

$$N_i(0) = 0.$$

Figure 1 shows a schematic of the transitions corresponding to the above set of equations for the popula-

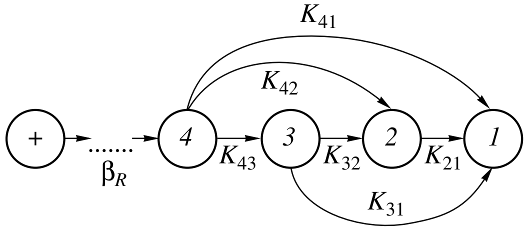


Fig. 1. Schematic of the transitions in a multilevel model with multiphoton transitions.

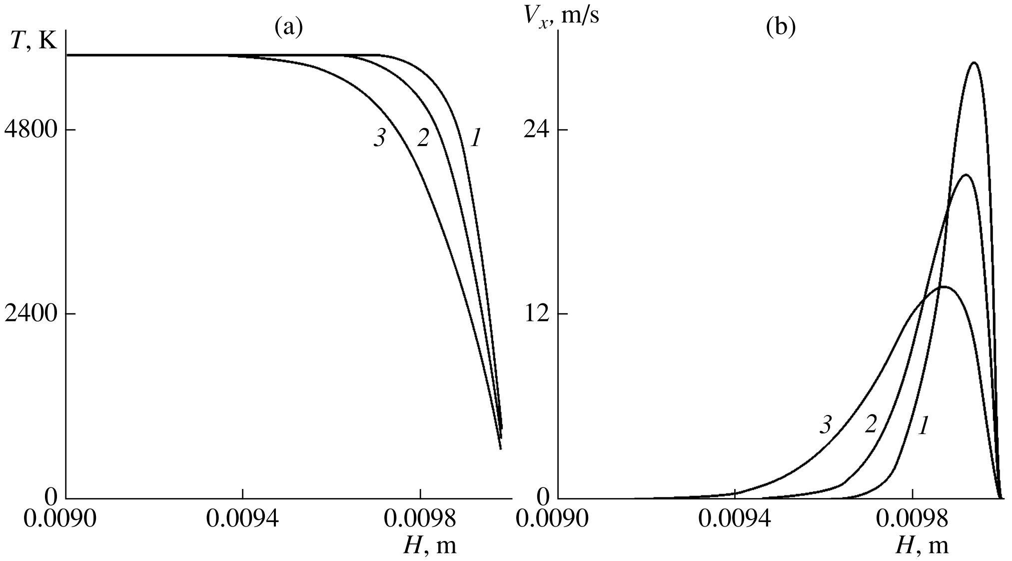


Fig. 2. Profiles of (a) plasma temperature T and (b) plasma flow velocity V_x across the boundary level at times t = (*1*) 3×10^{-7}, (*2*) 6×10^{-7}, and (*3*) 1.5×10^{-6} s for p = 2 atm, T = 0.5 eV, T_s = 300 K, and R = 0.5 m. The thickness of the boundary level is 0.01 m.

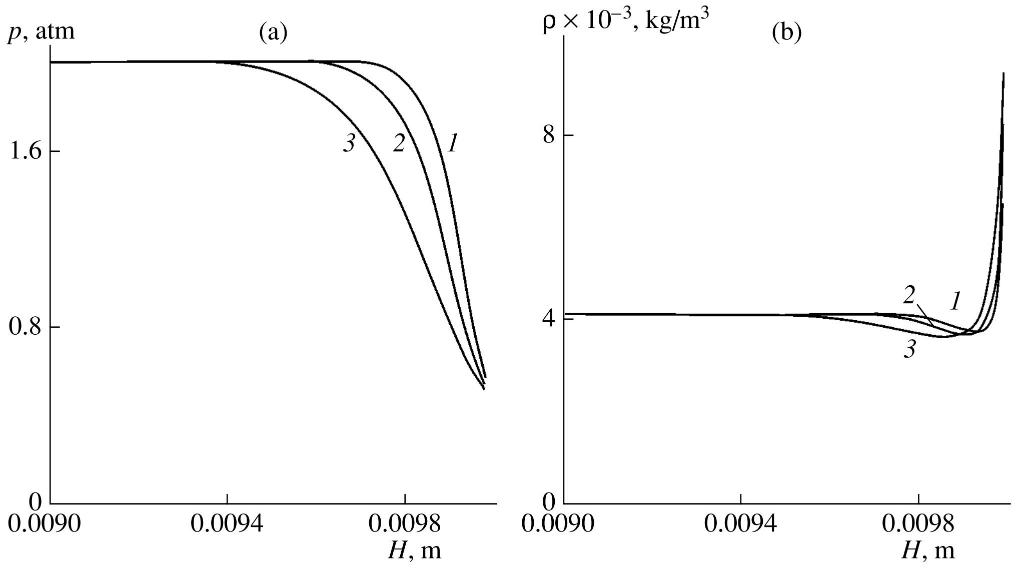


Fig. 3. Profiles of (a) plasma pressure p and (b) plasma density ρ across the boundary level at the same times and conditions as in Fig. 2.

tions (here, n = 1, 2, 3, 4 is the number of the energy level of a hydrogen atom).

To determine the number of hydrogen atom states, we used data on the position of the sink "bottleneck" [1]. The transition intensities K_{ba} are presented in [2].

Based on the above mathematical model, we analyzed the possibility of lasing in a recombining hydrogen plasma with allowance made for heat conduction and convective heat exchange. In numerical calculations, the above set of equations was represented in a divergent form and was solved by an explicit Lax scheme.

The gain was calculated with allowance made for the Doppler and Holtsmark broadenings of the spectral line. The computation results are shown in Figs. 2–9. The profiles of the main plasma macroparameters (temperature, flow velocity, pressure, and density) evolving from an initial state determined by the conditions of an isobaric–isothermal equilibrium with the ambient medium are shown in Figs. 2 and 3. Figures 4–6 show the profiles of the density of atomic hydrogen ions in

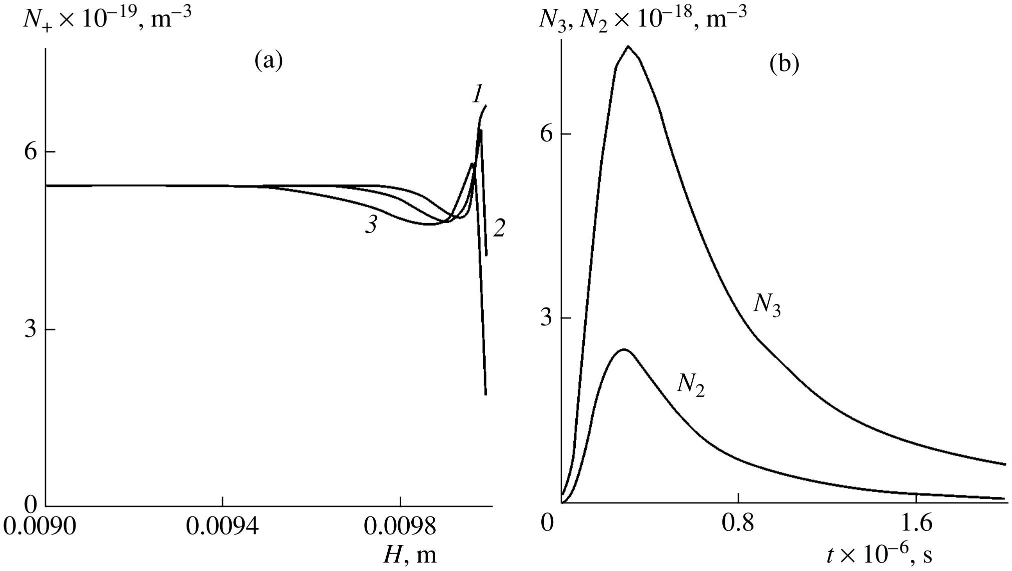


Fig. 4. (a) Profiles of the density N_+ of hydrogen nuclei across the boundary level at the same times as in Fig. 2 and (b) the time evolution of the N_3 and N_2 populations in the boundary layer. The conditions are the same as in Fig. 2.

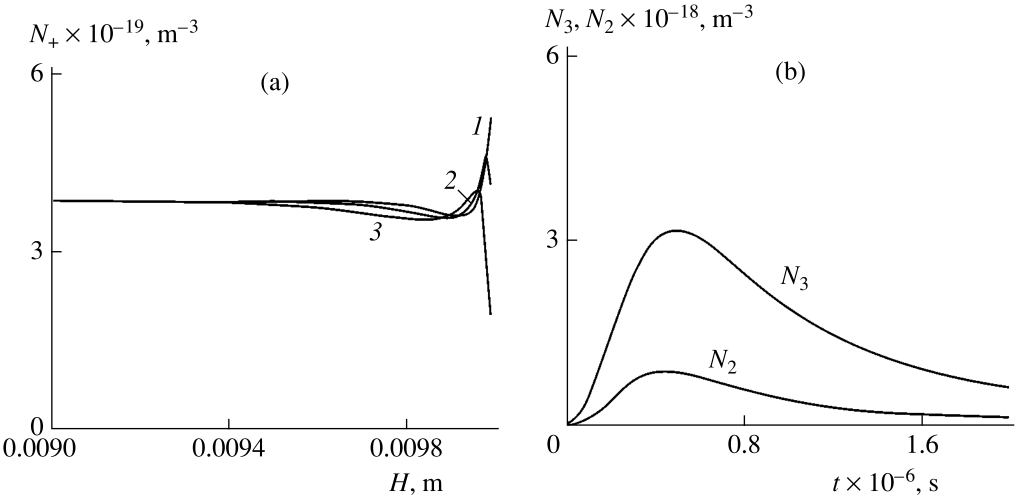


Fig. 5. The same as in Fig. 4, but for $p = 1$ atm.

the boundary level (it is this density that determines the process of the level-by-level relaxation of electrons in atomic hydrogen) and the time evolution of the populations of the 3–2 working levels just at the plasma–body interface for different initial pressures. The time evolution of the gain k_{ba}^{+} at the 3–2 transition of a hydrogen nucleus with allowance for the Doppler and Holtsmark broadenings of the emission line is shown in Figs. 7 and 8 for two different models of level-by-level relaxation. For comparison, Fig. 9 shows this gain calculated without making allowance for convection.

As was mentioned above, lasing is feasible only within a narrow range of the main macroscopic plasma parameters, namely, the temperature and density (pressure). The computation results show that the maximum gain is achieved within the following range of the initial parameters: the plasma temperature is ~0.5 eV, the pressure is 1–3 atm, and the radius of the conducting body that cools the plasma is 0.5–1.0 m. Within these parameter ranges, the plasma is a weakly ionized medium and the energy exchange between the plasma and the body is determined by the neutral (atomic hydrogen) component. The energy flux density is on the order of 10 MW/m^2. Under such conditions, the initial equilibrium plasma composition can be calculated by the Saha formula.

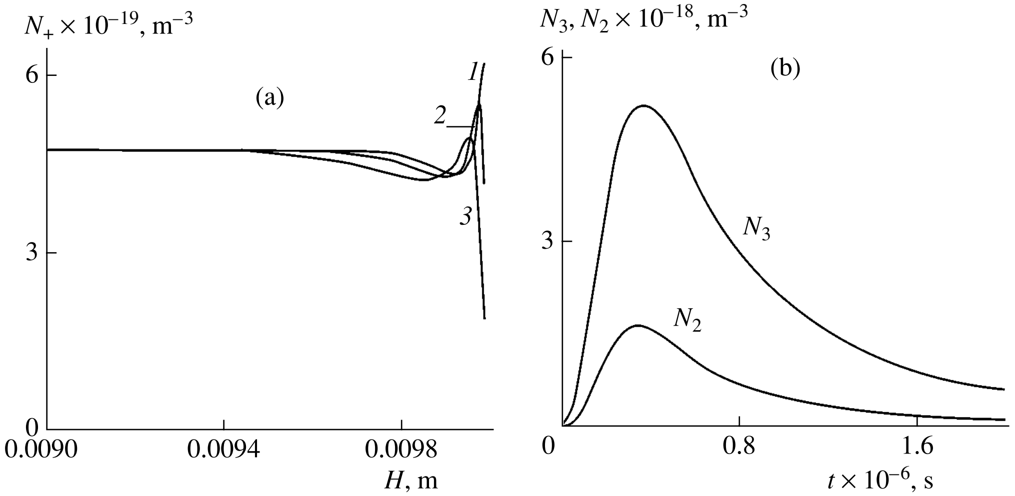


Fig. 6. The same as in Fig. 4, but for p = 1.5 atm.

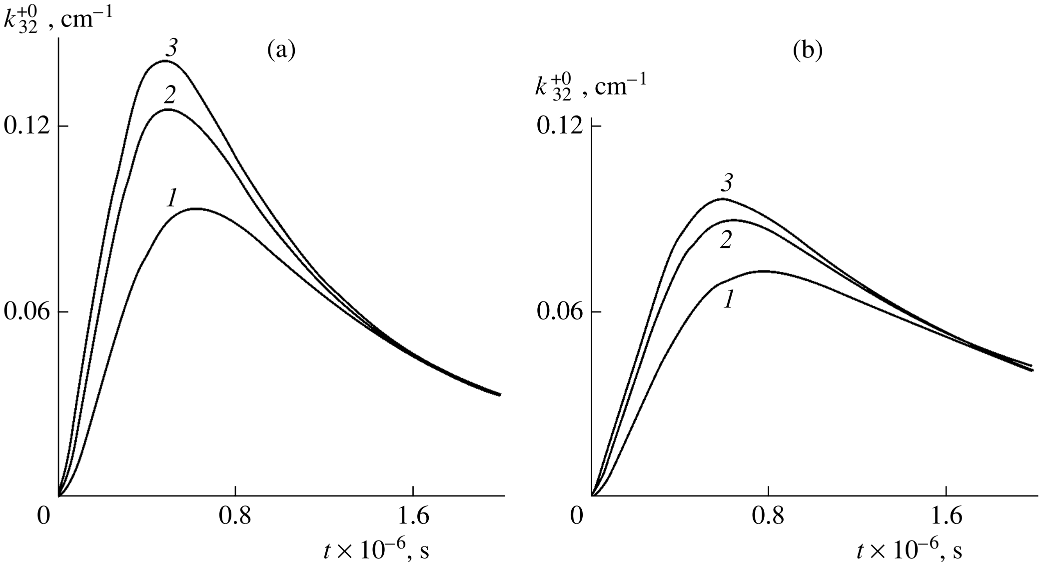


Fig. 7. Time evolution of the gain k_{32}^{+0} calculated by the two-level time-dependent model of level-by-level relaxation with allowance made for the (a) Doppler and (b) the Holtsmark broadening for p = (*1*) 1, (*2*) 1.5, and (*3*) 2 atm; T = 0.5 eV; T_s = 300 K; and R = 0.5 m. The thickness of the boundary level is 0.01 m.

The results obtained cam be summarized as follows:

The lasing process is pulsed in character. Lasing terminates due to the intensification of the recombinational plasma decay when the characteristic recombinational length becomes less than the thickness of the boundary level. The characteristic time of this process is about 1 μs. The highest gain at the 3–2 transitions of hydrogen nuclei is on the order of 0.1 cm^{-1}.

With allowance made for convection, the gain increases approximately twice. Convection mainly determines the transfer of H$^+$ ions to the body and only slightly affects the thermal state of the plasma in the boundary level. In solving the problem, one can use the energy equation in the form of a time-dependent equation for heat transfer.

On the characteristic time scale of the problem (~1 μs), the body that cools the plasma is heated insignificantly because the density of the energy flux transferred from the plasma is rather low (~10 MW/m^2). Thus, the problems of the thermal states of the plasma and the body can be solved separately.

The gains calculated within the two-level and multilevel models of level-by-level relaxation are close to each other. The former gives a somewhat higher gain;

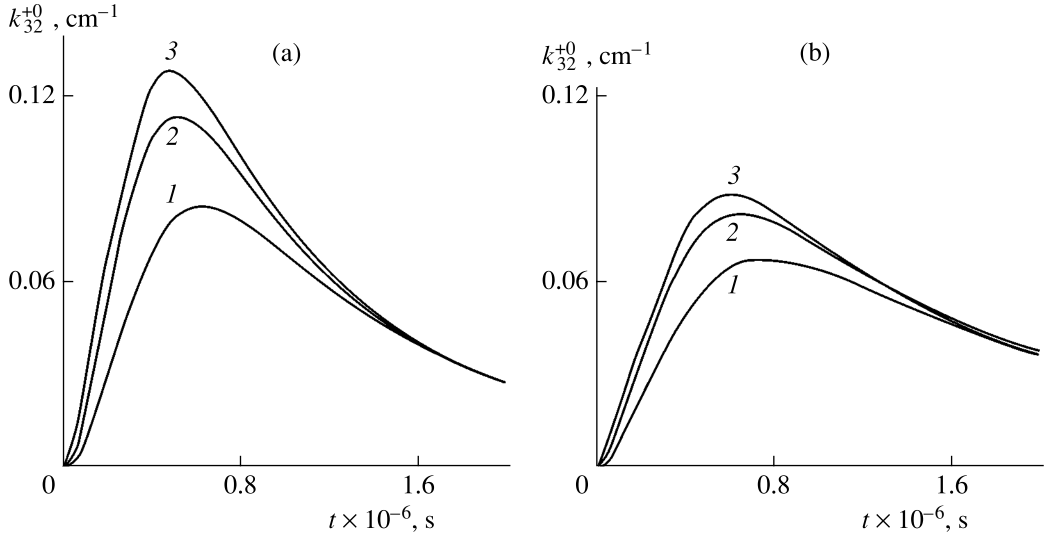


Fig. 8. The same as in Fig. 7, but for the multilevel time-dependent model of level-by-level relaxation.

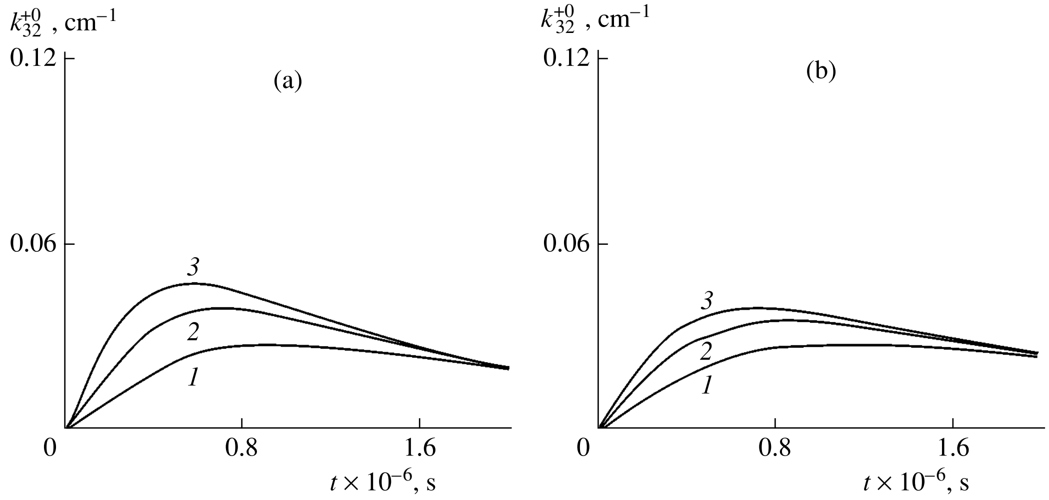


Fig. 9. The same as in Fig. 7, but for the multilevel time-dependent model of level-by-level relaxation without making allowance for convection.

this is quite natural, taking into account the physical content of the models.

REFERENCES

1. L. I. Gudzenko and S. I. Yakovlenko, *Plasma Lasers* (Atomizdat, Moscow, 1978) [in Russian].
2. A. V. Borovskiĭ, S. A. Zapryagaev, N. L. Manakov, and O. V. Zatsarinnyĭ, *Multiply Charged Ion Plasma* (Nauka, Moscow, 1995) [in Russian].
3. A. P. Kuryshev and V. V. Sakhin, in *Modern Problems of Nonequilibrium Gas Dynamics and Thermodynamics: Collection of Articles*, Ed. by V. N. Uskov (Balt. Gos. Tekhnol. Univ., St. Petersburg, 2002) [in Russian].
4. A. P. Kuryshev and V. V. Sakhin, Zh. Tekh. Fiz. **71** (11), 106 (2001) [Tech. Phys. **46**, 1453 (2001)].
5. Y.-H. Ohtsuki, *Charged Beam Interaction with Solids* (Taylor & Francis, London, 1983; Mir, Moscow, 1985).

Translated by N. Ustinovskiĭ

Technical Physics, Vol. 49, No. 10, 2004, pp. 1296–1300. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 52–56.
Original Russian Text Copyright © 2004 by Barannikova, Danilov, Zuev.

Plastic Strain Localization in Fe–3%Si Single Crystals and Polycrystals under Tension

S. A. Barannikova, V. I. Danilov, and L. B. Zuev
Institute of Strength Physics and Materials Science, Siberian Division,
Russian Academy of Sciences, Akademicheskiĭ pr. 2/1, Tomsk, 634021 Russia

Received January 23, 2004

Abstract—Plastic flow localization in siliceous iron single crystals and polycrystals under identical stretching conditions is considered. The localization patterns are analyzed at the stages of linear and parabolic deformation hardening, as well as at the stages of necking and plastic fracture. The localization patterns observed in the alloy in the single-crystalline and polycrystalline states are compared. *© 2004 MAIK “Nauka/Interperiodica”.*

INTRODUCTION

Plastic strain macrolocalization remains a major problem related to plastic flow. Systematic efforts in this field, which date back to the late 1980s, have led to particular success. For example, the gradient theory of plasticity [1] treats the localization as a result of the flow instability due to a negative deformation hardening coefficient at a number of the process stages. The dislocation mechanisms of strain localization and plastic flow instability are described at length in [2, 3]. In [4, 5], the macrolocalization was experimentally studied by two-beam speckle interferometry [6, 7], which makes it possible to record the space–time distributions of the local components of the plastic deformation tensor in a specimen under load. It was shown that this phenomenon may be classified as a sequence of self-organization processes in open systems. The fact that the macrolocalization shows up at all stages of plastic deformation from the yield stress to fracture is of special importance. According to the present-day categorization [8], the evolving localization patterns are viewed as various types of autowaves and dissipative structures [9].

Although the experimental data concerning the localization have been obtained for at least ten metals and alloys [8], it still remains unclear whether the macrolocalization-governing laws are of a general character. While the regular distributions of localized strains were first observed in α-Fe as early as in the late 1980s [8], subsequent studies were performed largely on fcc materials. Therefore, a demand arose for verifying the basic laws that govern the evolution of the localization patterns in bcc materials. In this respect, Fe–3%Si alloy (siliceous iron), the model material for plastic flow investigation (see, e.g., [10, 11]), seems to be the most appropriate. It is also of interest to contrast the localization patterns in the single-crystalline and polycrystalline states of this material, since grains boundaries and their condition are known to considerably influence plastic flow [12].

MATERIAL AND METHOD OF INVESTIGATION

The experiments were carried out on single-crystalline and polycrystalline specimens made of Fe–3%Si alloy. The single crystals were grown by the Bridgman method in an inert gas. The specimens, in the form of double blades with a working area measuring 28 × 5 × 1.5 mm, were cut on an electrical discharge machine from a single-crystal ingot so that their longer axes were aligned with the [143] direction and the working face had indices (168). Prior to the tests, the specimens were homogenized at 1300 K for 16 h.

Polycrystalline specimens with a working area measuring 50 × 10 mm were punched out from 0.3-mm-thick transformer steel sheets and then subjected to stress relieving in a vacuum at 1373 K. The mean grain size was 4.5 ± 3 mm. The axis of tension was aligned with the rolling direction.

Mechanical tests were carried out following the simple uniaxial tension scheme with an Instron-1185 testing machine at 300 K. The strain rate was 1.2×10^{-4} s^{-1} for the single-crystalline specimens and 6.67×10^{-5} s^{-1} for the polycrystals. Such a difference is due to the fact that simultaneous recording of flow curves and displacement vector fields (see below) requires that the speed of the mobile specimen grips be the same in both cases (the lengths of the single-crystalline and polycrystalline specimens were different in our experiments).

The fields of the displacement vectors $r(x, y)$ of points on the working surface were successively recorded from the yield stress to fracture using the method of speckle interferometry [6, 7, 13].[1] The increase in total strain $\Delta\varepsilon_{tot}$ upon recording each of the speckle images was 2×10^{-3}. Then, the distributions of the longitudinal (ε_{xx}), transverse (ε_{yy}), shear (ε_{xy}), and

---

[1] The x axis coincides with the tension direction, and the y axis lies in the specimen plane.

rotational (ω_z) components of the plastic strain tensor $\beta_{ij} = \nabla r(x, y)$ over all points on the surface of observation were calculated and plotted by numerically differentiating the field of $r(x, y)$ with respect to coordinates x and y. In most cases, the behavior of the component ε_{xx} (the local elongation along the axis of tension) of tensor β_{ij}, was analyzed during stretching. This component carries the most information and is defined as $\varepsilon_{xx} = \partial u/\partial x$, where u is the component of r in the x direction. We considered the space distribution of ε_{xx} over the specimen, $\varepsilon_{xx}(x, y)$, for a given time instant (or the space distribution of the strain, since $\varepsilon \sim t$ for constant-rate stretching) and space–time distributions, $\varepsilon_{xx}(x, y)$, along the specimen axis for a sequence of time instants.

EXPERIMENTAL RESULTS

The calculation shows that, for the given orientation of the single crystals, slip is the most pronounced in the system (110)[$\bar{1}11$] with a Schmid factor $m_1 = 0.47$ and in the systems (101)[$\bar{1}11$] and (1$\bar{1}$0)[$\bar{1}\bar{1}1$] with $m_2 = m_3 = 0.38$. The metallographic examination carried out immediately after the onset of plastic flow confirmed this statement.

The polycrystalline sheets were highly textured along the easiest magnetization axis [001]. Under these conditions, the strain at the beginning of the plastic flow ($\varepsilon \leq 1.5 \times 10^{-2}$) was localized in the slip bands. One or two slip systems acted in each of the grains. Slip traces in adjacent grains differed substantially, which is typical of grain boundaries of a general type.

Plastic flow in single crystals and polycrystals of Fe–3%Si alloy has distinct stages. Typical stress–strain curves $\sigma(\varepsilon)$ for such specimens are demonstrated in Figs. 1 and 2. Specifically, stages of linear and parabolic deformation hardening can be distinguished. In the single-crystalline specimens, the easy slip stage is absent. After the extended elasticity–plasticity transition, the stage of linear hardening starts at $\varepsilon_{\text{tot}} = 10^{-2}$ with a coefficient $\theta = 900$ MPa. This is clearly seen from the dependence of the deformation hardening coefficient $\theta = d\sigma/d\varepsilon$ on the total strain ε_{tot} (Fig. 1, curve θ). The stage of linear hardening lasts $\approx 2.5 \times 10^{-2}$. Then, the hardening coefficient varies in such a way that the hardening exponent $n < 1$ in the relationship $\sigma \sim \varepsilon^n$, which describes the flow curve at the parabolic stage, drops stepwise from $n = 0.6$ to 0.4, thereby separating this stage into two regions. The deformation process ends in necking with subsequent plastic fracture of the specimen.

The flow curves for the polycrystalline specimens (Fig. 2) are characterized by a higher coefficient of linear deformation hardening ($\theta = 1750$ MPa). To more accurately separate the stages in the flow curves in this case, the dependences $\sigma(\varepsilon)$ were plotted in the coordinates of true strains and stresses: $e = \ln(1 + \varepsilon)$ and $s = \sigma(1 + \varepsilon)$ [14]. Such a procedure allowed us to subdivide the parabolic stage for the polycrystals into two regions with $n = 0.5$ and 0.4.

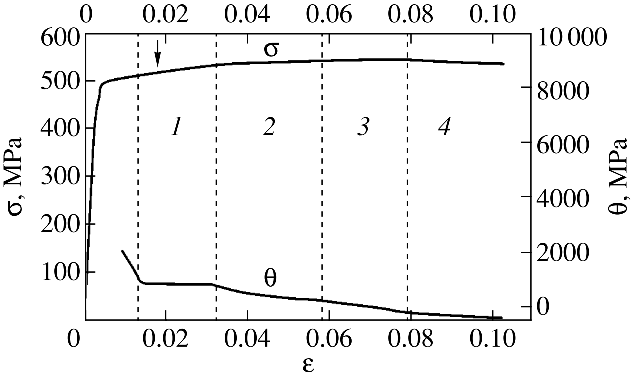


Fig. 1. Flow curve (σ) for the single-crystalline specimen of Fe–3%Si alloy and the dependences of the deformation hardening coefficient on the strain (θ): *1*, linear stage; *2*, parabolic stage I; *3*, parabolic stage II; and *4*, fracture.

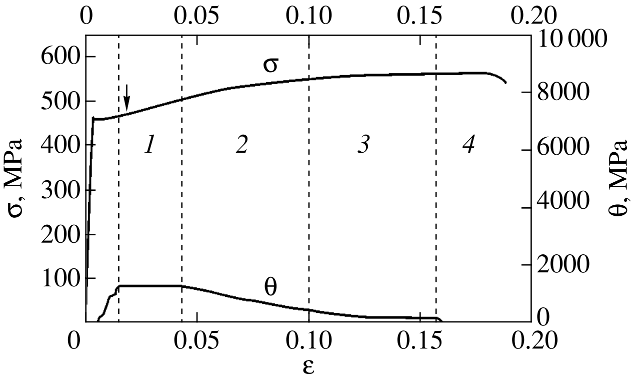


Fig. 2. The same as in Fig. 1 for the polycrystalline specimen.

The distributions of local elongations ε_{xx} over the specimens suggest that the strains in both the single crystals and polycrystals are distributed nonuniformly throughout the deformation process. From Figs. 3 and 4, which demonstrate the typical distributions of elongation ε_{xx}, shear component ε_{xy}, and rotational component ω_z, it follows that the plastic strain is localized in particular zones of the specimen, while other regions of the material remain virtually unstrained at the given strain rate. It is easy to see that the distributions of the components for the single crystals and polycrystals differ only slightly in local strain amplitudes. Straightforward quantitative estimates show that the increase in the plastic strain is almost completely concentrated in the localization zones. If N is the number of strain maxima with a magnitude $\varepsilon_{xx}^{\text{mag}}$ in a specimen of length L and l is the size of the strain localization zone along the x direction, the mean elongation within this zone is propor-

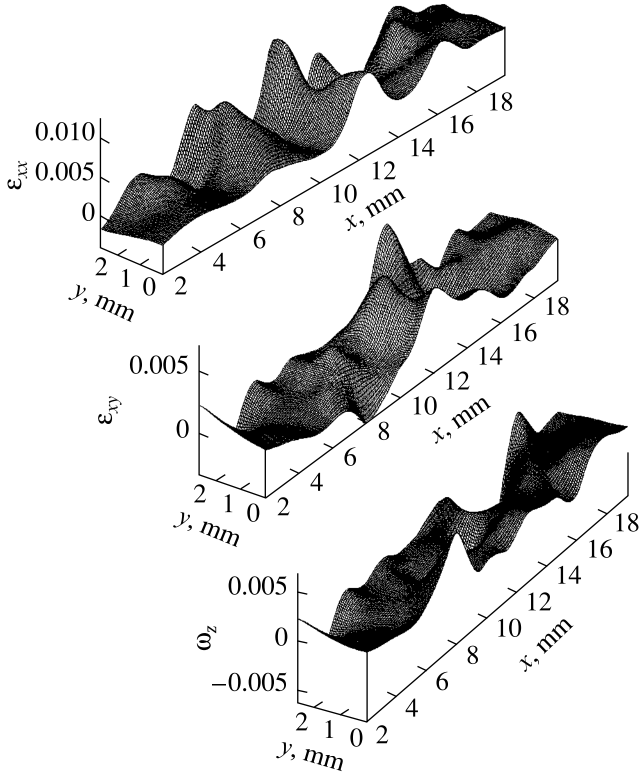


Fig. 3. Distribution of the plastic strain tensor components in the single-crystalline specimen for the total strain shown by the arrow in Fig. 1.

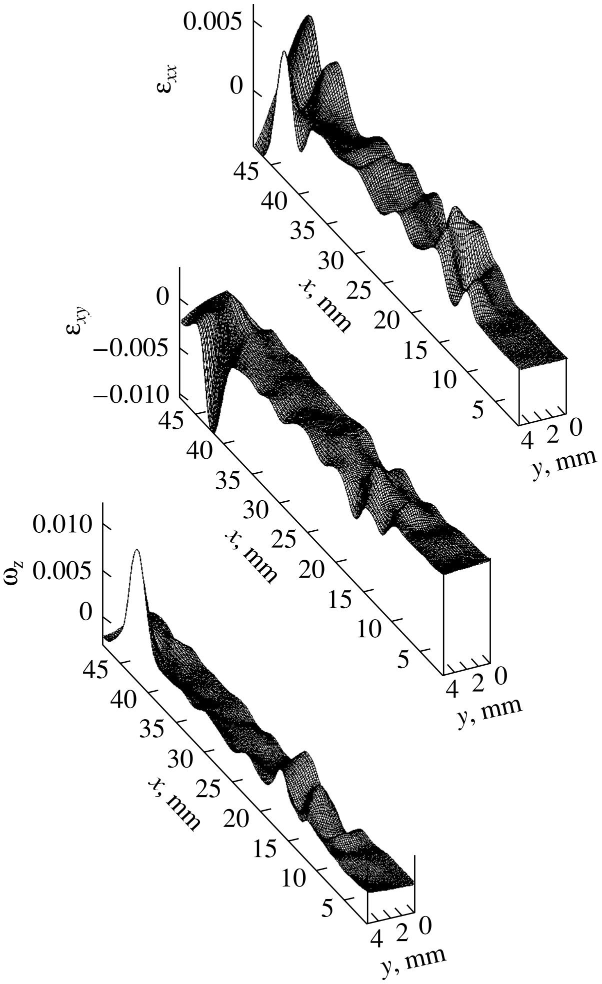


Fig. 4. Distribution of the plastic strain tensor components in the polycrystalline specimen for the total strain shown by the arrow in Fig. 2.

tional to $\sim\langle\varepsilon_{xx}\rangle l$ and the total elongation of the specimen is given by

$$\delta L = \sum_N \varepsilon_{xx}^{\mathrm{mag}} l \approx N\langle\varepsilon_{xx}\rangle l. \quad (1)$$

From Fig. 3 (polycrystal), it follows that $N = 5$ and $l \approx 3.8$ mm. The mean strain in the strained zone is $\langle\varepsilon_{xx}\rangle \approx \varepsilon_{xx}^{\mathrm{mag}}/2 \approx 4.8 \times 10^{-3}$. Accordingly, $\delta L \approx 0.09$ mm and $\Delta\varepsilon_{\mathrm{tot}} = \delta L/L \approx 1.8 \times 10^{-3}$. For the single crystal (Fig. 4), $N = 4$, $l \approx 2.5$ mm, and $\langle\varepsilon_{xx}\rangle \approx 5 \times 10^{-3}$, and the same estimation yields $\delta L \approx 0.05$ mm and $\Delta\varepsilon_{\mathrm{tot}} \approx 1.7 \times 10^{-3}$. Both values estimated agree, in general, with the above increase in the total strain at each step of recording the speckle interferograms, 2×10^{-3}, but are somewhat lower than this value because of the plastic deformation of the rest of the material. Thus, the plastic strain under stretching is localized in several rather thin ($l \ll L$) layers of the specimen.

The distributions of component ε_{xx} suggest that the process of macroscopic strain localization at the linear hardening stage is spatially and temporally ordered, i.e., produces a typical wave pattern, which can be characterized by wavelength and wave propagation velocity. From the distributions of component ε_{xx} in the single-crystalline and polycrystalline specimens (Figs. 3, 4), it follows that the spatial period of the localization zones (wavelength) is 3–4 mm for the single crystal and 5–6 mm for the polycrystal. The propagation rate of the localization autowave, V_{aw}, that was determined with the technique described in [8] was found to be 5.8×10^{-5} m/s in the single crystals and 1.7×10^{-5} m/s in the polycrystalline specimens.

This discrepancy is consistent with the autowave concept of evolution of plastic flow localization at the stage of linear gardening, which was proposed for fcc materials [5, 8]. The rates of localization zone propagation mentioned above are well fitted by the relationship between the velocity of phase autowaves of local plastic strain and the hardening coefficient at the linear stage: $V_{\mathrm{aw}} = \Xi/\theta^*$. Here, $\Xi = 6.33 \times 10^{-7}$ m/s is the constant obtained by generalizing the early data for a variety of metals and alloys and $\theta^* = \theta/G$ is the deformation hardening coefficient normalized to shear modulus G [8]. Since $\theta^*_{\mathrm{single}} \approx 1.1 \times 10^{-2}$ and $\theta^*_{\mathrm{poly}} \approx 2.1 \times 10^{-2}$, the calculated values of the wave propagation velocity is

$\approx 5.75 \times 10^{-5}$ m/s in the single crystals and $\approx 3.0 \times 10^{-5}$ m/s in the polycrystals. Both values are in good agreement with those given above (especially for the single crystals). In the single crystals, change over from the linear to parabolic hardening is preceded by a short region of the chaotic distributions of ε_{xx} over the specimen. In this case, the previous localization pattern "breaks" and a new one spontaneously appears from the chaos. The presence of such transition regions with chaotic distribution of small plasticity-localization zones was observed by Zuev *et al.* in deformed Al polycrystals [15].

As was mentioned previously, the parabolic stage of deformation hardening in the single-crystalline and polycrystalline specimens consists of into two portions with different *n*. In the first portion ($n \approx 0.6$ for the single crystals and $n \approx 0.5$ for the polycrystals), the strain localization zones are stationary. In this case, the macrolocalization patterns in the single crystals represent a set of stationary equidistant strain zones with roughly the same magnitude. In the second portion, when *n* decreases to 0.4, the zones become mobile in both the single crystals and polycrystals and the strain magnitudes change so that one of the ε_{xx} maxima gradually grows, while local strain increments in other zones remain the same or decrease slightly. A similar strain redistribution pattern was observed in polycrystalline Zr–Nb alloy at the prefracture stage [16].

DISCUSSION

The localized strained zones behave in the most intriguing way at he stage of parabolic hardening at $n < 0.5$, where they move and tend to coalesce. The motion of the zones at this stage is mutually coordinated so that their velocity is the greater, the farther they are from the site of fracture at the beginning of straining (Fig. 5 for the polycrystal). Since the velocities of the zones are "matched" to each pother, all strained zones come to the site of fracture simultaneously. Extrapolating the straight lines $X(t)$ in the kinetic diagram (Fig. 5) until they meet, one can find a pole with the fracture coordinates $X_f = 40 \pm 3$ mm and $t_f = 2520 \pm 90$ s. The former fairly accurately determines the site of fracture, while the latter specifies the point in time at which the specimen becomes discontinuous (the discontinuity somewhat lags behind the onset of cracking). At this stage of the process, the position of each of the strained zones versus time is clearly given by

$$X = X_f + V_0(t - t_f), \quad (2)$$

where the sign of V_0 depends on the initial position of a specific plastic-flow zone.

In other words, the events taking place at the site of fracture do not prevent the development of plastic deformation in the rest of the specimen but force the mobile localized-flow zones to associate at the site where a crack originates. The same, yet weaker, tendency is also observed in the single-crystalline specimen (Fig. 6). Nevertheless, in this case, too, there exists a pole with the coordinates $X_f = 3 \pm 3$ mm and $t_f = 890 \pm 60$ s, which specify the site and time of cracking.

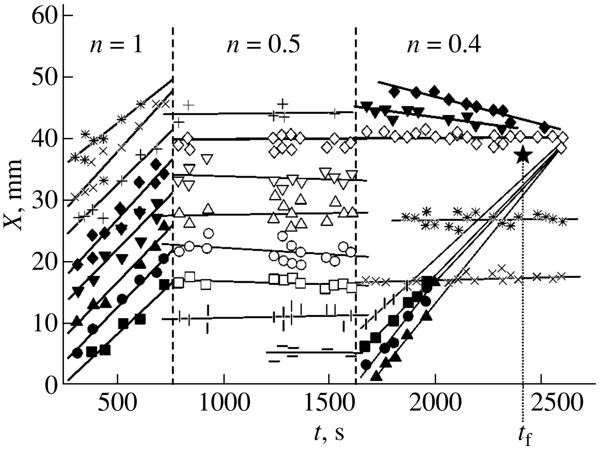


Fig. 5. Evolution of the macrolocalization patterns during the plastic deformation of the polycrystal (★ indicates the site and time of cracking). The symbols and lines show that the localization zones move at the stage of linear hardening ($n = 1$), are stationary at the parabolic stage ($n = 0.5$), and tend to associate near the site of fracture (for $n < 0.5$).

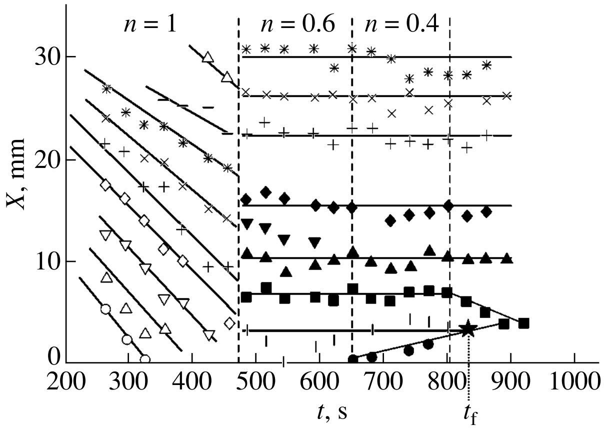


Fig. 6. Evolution of the macrolocalization patterns in the polycrystal.

As follows from Figs. 5 and 6, some of the localized-flow zones that appear at the parabolic stage with $n \approx 0.5$ (well before fracture) remain immobile during loading at any $n \leq 0.5$. This is observed in both the single crystals and polycrystals. In Figs. 5 and 6, their positions are shown by the thick horizontal lines. The distributions of the plastic strain tensor components over the specimens (these distributions are similar to those represented in Figs. 3 and 4) demonstrate that the strain magnitude in these zones is the highest from the time of occurrence. It is near these strained zones that the poles of the curves $X(t)$ for the single crystals and polycrystals arise and plastic cracks originate. A similar effect was noted earlier [16].

The mechanism that provides the synchronous motion of the localized plasticity zones at the final stage of the process may be the well-known phenomenon of acoustic emission from the fracture zone. Elastic pulses of acoustic emission may initiate the development of plasticity and cause the localized plasticity zones to move. Such a mechanism was suggested in [15] to explain the existence of macroscopic distances between the localized plastic-strain zones (the localized strain wavelength). The effect of acoustic emission may be sufficiently pronounced for the plastic flow kinetics to be changed, as indirectly indicated by the formation of the so-called Wallner lines at fracture [17]. These lines are associated with the curvature of the propagating crack path due to ultrasonic pulses of acoustic emission.

CONCLUSIONS

Thus, it is established that flow curves $\sigma(\varepsilon)$ for single-crystalline and polycrystalline specimens of Fe–3%Si bcc alloy have the linear ($\sigma \sim \varepsilon$) and parabolic ($\sigma \sim \varepsilon^n$) stages of deformation hardening. The latter, in turn, consists of two portions with different exponent $n < 1$. From the character of strain localization in this alloy, one can draw the following conclusions.

The laws of plastic strain macrolocalization that were previously derived for the stages of linear hardening (the wave nature of straining) and parabolic hardening (a stationary set of the localization zones at $n \approx 0.5$) in fcc and hcp single crystals and polycrystals remain in full measure valid for single crystals and polycrystals of Fe–3%Si bcc alloy.

For a parabolicity exponent $n < 0.5$, the plasticity localization zones become mobile and consistently move, meeting at the site of crack origination at the time of fracture.

These findings seem to be useful in estimating the ultimate states of plastic materials. Also, they can be used for predicting the limiting plasticity of metals, locating the site of fracture, and determining the time to fracture (lifetime) during mechanical tests [18]. This can be done by extrapolating the experimentally found temporal dependences of the coordinates of the localized plasticity zones, $X(t)$, until they meet. In this case, the space and time coordinates of the point of intersection define the site and time of fracture with a reasonable accuracy.

ACKNOWLEDGMENTS

S.A. Barannikova thanks the Ministry of Education of the Russian Federation for its support (grant no. PD 02-1.2-63, 2002).

REFERENCES

1. E. C. Aifantis, Int. J. Non-Linear Mech. **31**, 797 (1996).
2. G. A. Malygin, Fiz. Tverd. Tela (St. Petersburg) **37**, 3 (1995) [Phys. Solid State **37**, 1 (1995)].
3. G. A. Malygin, Usp. Fiz. Nauk **169**, 979 (1999) [Phys. Usp. **42**, 887 (1999)].
4. L. B. Zuev, V. I. Danilov, and V. V. Gorbatenko, Zh. Tekh. Fiz. **65** (5), 91 (1995) [Tech. Phys. **40**, 456 (1995)].
5. L. B. Zuev, Ann. Phys. **10**, 965 (2001).
6. L. B. Zuev, V. I. Danilov, and N. M. Mnikh, Zavod. Lab. **56**, 90 (1990).
7. L. B. Zuev, S. N. Polyakov, and V. V. Gorbatenko, Proc. SPIE **4900**, 1197 (2002).
8. L. B. Zuev, V. I. Danilov, and B. S. Semukhin, Usp. Fiz. Met. **3**, 237 (2002).
9. V. A. Vasil'ev, Yu. M. Romanovskiĭ, and V. G. Yakhno, *Self-Sustained Processes* (Nauka, Moscow, 1987) [in Russian].
10. B. Šesták and Z. Arnold, Phys. Status Solidi A **23**, 155 (1974).
11. B. Šesták and V. Novák, Phys. Status Solidi A **23**, 703 (1974).
12. P. Lejček, S. Hofman, and A. Krajnikov, Mater. Sci. Eng., A **234–236**, 283 (1997).
13. I. S. Klimenko, *Holography of In-Focus Images and Speckle-Interferometry* (Nauka, Moscow, 1985) [in Russian].
14. *Handbook of Experimental Mechanics*, Ed. by A. S. Kobayashi (Prentice-Hall, Englewood Cliffs, 1987; Mir, Moscow, 1990), Vol. 1, pp. 11–53.
15. L. B. Zuev, B. S. Semukhin, and N. V. Zarikovskaya, Zh. Tekh. Fiz. **71** (5), 57 (2001) [Tech. Phys. **46**, 563 (2001)].
16. T. M. Poletika, V. I. Danilov, G. N. Narimanova, *et al.*, Zh. Tekh. Fiz. **72** (9), 57 (2002) [Tech. Phys. **47**, 1125 (2002)].
17. *High Speed Physics*, Hrsg. von K. Vollrath and G. Thomer (Springer, Vienna, 1967; Mir, Moscow, 1971), Vol. 2, pp. 5–68.
18. B. Dhillon and C. Singh, *Engineering Reliability* (Wiley, New York, 1981; Mir, Moscow, 1984).

Translated by V. Isaakyan

Technical Physics, Vol. 49, No. 10, 2004, pp. 1301–1307. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 57–63.
Original Russian Text Copyright © 2004 by Malygin, Khusainov.

SOLIDS

Stability of the Mechanical Behavior of an Arched TiNi Strip under the Conditions of the Constrained Shape Memory Effect

G. A. Malygin* and M. A. Khusainov**
** Ioffe Physicotechnical Institute, Russian Academy of Sciences,*
Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia
e-mail: malygin.ga@mail.ioffe.ru
*** Yaroslav Mudryĭ State University, Sankt-Peterburgskaya ul. 41, Novgorod, 173003 Russia*
e-mail: vestnik@novsu.ac.ru

Received March 11, 2004

Abstract—Unstable form changing (click) of an arched TiNi strip annealed at 773 K is investigated. After the annealing, the strip is placed into the grips of a deforming machine and is transferred to the martensitic state by bending in the direction opposite to its initial bend. Subsequent heating of the strip, which carries it to the austenitic state, is accompanied by a click under the conditions of constrained form changing. Analysis of the martensitic deformation in terms of the theory of diffuse martensitic transformations makes it possible to establish the conditions under which the form changing process becomes unstable. © *2004 MAIK “Nauka/Interperiodica”.*

INTRODUCTION

Titanium nikelide (TiNi) is today viewed as the most promising material for temperature sensors in electromechanical and microelectromechanical systems [1, 2]. This is due to the fact that, compared to other materials, TiNi offers a more beneficial relationship between the energy delivered to the sensor and its mechanical displacement. The so-called bidirectional shape memory effect [3–5] extends the functionality of the TiNi sensors. This effect arises in TiNi strips and plates (provided that the amount of nickel exceeds that in the equiatomic composition) annealed under bending at 700–800 K.

Bending during the annealing leads to different orientations of matrix-coherent disk-shaped Ti_3Ni_4 intermetallic precipitates in the expanded and compressed layers of strips and plates [4, 6] and generates internal stresses of opposite sign in these layers. Subsequent stress relaxation during the direct, $B2 \longrightarrow R \longrightarrow B19'$, and reverse, $B19' \longrightarrow R \longrightarrow B2$, two-stage martensitic transformations (the formation and decomposition of the R and M martensite modifications, respectively) is accompanied by a temperature-reversible change of sign of curvature in the strip or plate. This is the manifestation of the bidirectional shape memory effect (SME). The quantitative theory of the bidirectional SME based on the concept of diffuse martensitic transformations [7, 8] is developed in [9].

The functionality of titanium nickelide is extended further when the bidirectional SME occurs under constrained conditions. It was established [10, 11] that mechanical restrictions placed on the displacement of the ends of an arched TiNi strip make the SME unstable and cause the click phenomenon in the strip in a narrow temperature range as the temperature increases. Note that this phenomenon also takes place in TiNi disks annealed in the curved state when their edges are free to move [12]. One may suppose that the effect of unstable martensitic deformation is caused in this case by the anisotropic and nonuniform distribution of Ti_3Ni_4 particles across the thickness and along the radius of the plate.

The aim of this study is to analyze the mechanical stability of an arched strip under constrained conditions [10, 11] in terms of the recently developed theory of bidirectional SME [9, 13] and to find the factors responsible for the click when the temperature is varied.

EXPERIMENTAL CONDITIONS AND RESULTS

The TiNi strips (the nickel content was 50.5 at. %) of thickness $2h = 0.4–0.5$ mm, width $b = 6–8$ mm, and length $2l = 19–21$ mm were mounted in a mandrel with a given radius of curvature R_0 and a maximum bending deflection $W_0 = R_0[1 - \sqrt{1-(L_0/R_0)^2}]$, where $L_0 = R_0\sin(l/R_0)$ is half the chord between the ends of the bent strip (Fig. 1a). The fixed strip was annealed at a temperature of 773 K for 30 min in order to give rise to the bidirectional shape memory effect. Further, the strip was rolled to $\varepsilon_\Sigma \approx 35\%$ with intermediate annealings and finally annealed at 693 K for 1.5 h. These procedures were aimed at improving the reversibility of the martensitic transformation. As a result, the starting and final temperatures of the direct and reverse martensitic transformations (M_s, M_f and A_s, A_f, respectively) were

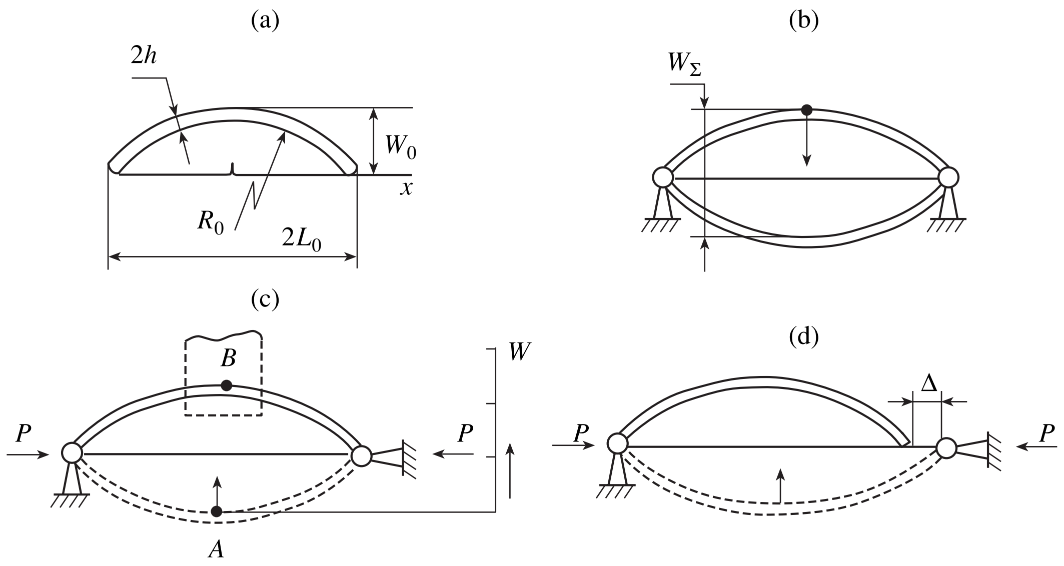


Fig. 1. Stages of the form changing process in the arched strip: (a) setting the initial shape, (b) bending by applying force F in martensite, (c) recovery of the shape upon heating, and (d) recovery of the shape in the presence of gap Δ.

found to be $M_s = 307$ K, $M_f = 279$ K, $A_s = 297$ K, and $A_f = 325$ K.

Next, the arched strip was placed into a deforming machine with hinged grips (Fig. 1b) and bent by applying force F in the direction opposite to the initial bend at 293 K. The amount of the deflection in the opposite direction was equal to $W_\Sigma/2$. In this case, the material of the strip is transferred to the martensitic state. Now, if the strip is heated with force F removed, it recovers the initial state, giving forth a clap (click) (Fig. 1c). The clap is heard when the strip's ends rest on the stationary hinged grips, and compressive force P appears at the ends upon straightening. This force additionally bends the strip, making its strained state unstable [13]. If one or both ends are free, the transition to the austenitic state and the form changing process occur steadily. Therefore, to stabilize these processes when the strip is in the hinged grips, a sufficiently wide gap Δ must be provided in order that one or both ends of the strip are free to move (Fig. 1d). Varying the width of the gap, one can control the constraint.

Under the conditions of constrained form changing, the functionality of the arched strip expands. Figure 2 shows the dependence of static reaction force Q_r that is exerted by obstacle B (dynamometer) placed at the center of the strip on the total bending deflection $W = AB$ of the central part of the strip during heating. When the form changing process is constrained, the phenomenon of clap is of dynamic character and the strip acquires a kinetic energy in the course of form changing. If there is an obstacle in the path of the strip, the former, when meeting the strip, undergoes an impact, with the impact force depending on the strip–obstacle distance. Figure 2 demonstrates the experimental results on determining impact force Q_i of the strip under the conditions of the constrained SME. It is seen that, unlike static reaction force Q_r, the dynamic force is maximal for intermediate values of the deflection, when the kinetic energy of the strip reaches a maximum. Thus, the arched TiNi strip can operate not only as a static pusher (mechanical drive or sensor), but also as a dynamic pulsed loading unit operating in a narrow temperature range, provided that the motion of its ends is limited.

In what follows, we analyze the mechanical behavior of the arched TiNi strip under the conditions of the

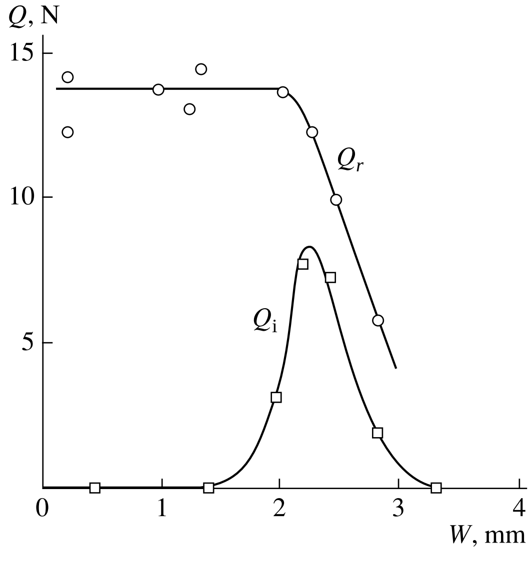


Fig. 2. Reaction force Q_r and impact force Q_i with which the strip strikes the obstacle vs. the free bend of the strip during heating.

constrained SME. The aim is to find the geometry of the strip that favors the effect of click.

BASIC RELATIONSHIPS

With allowance for the internal bending moment (which is due to the anisotropic Ti_3Ni_4 particle distribution over the thickness of the strip) and external bending moment (due to compressive force P at the ends of the constrained strip), the total curvature R^{-1} of the strip (R is the radius of curvature) is given by

$$R^{-1}(x, T) = R_0^{-1} + R_e^{-1} + R_\mathrm{m}^{-1}(x, T) + R_P^{-1}(x, T). \quad (1)$$

Here, R_0^{-1} is the curvature of the strip upon annealing (Fig. 1a); $R_e^{-1} = -(3/4h)|\varepsilon_0|$ and $R_\mathrm{m}^{-1} = -(3/2h)\tilde{\tilde{\varepsilon}}_\mathrm{m}(x, t)$ are the changes in the curvature due to the elastic, ε_0 [9], and martensitic, $\tilde{\tilde{\varepsilon}}_\mathrm{m}(x, T)$ [13], strains arising in the strip after the relaxation of elastic bending stresses, respectively; T is the temperature; and x the coordinate measured from the center of the spanning chord (Fig. 1a). The last term on the right of Eq. (1), $R_P^{-1}(x, T) = -M(x, T)/EJ$, is the elastic contribution to the curvature of the strip because of the constrained motion of its ends. Here, $M(x, T) = W_0(x, T)P(T)$ is the bending moment and $W_0(x, T)$ is the bending deflection of the strip in the case of the unconstrained SME. We have

$$W_0(x, T) = R_0(T)\left[\sqrt{1-\left(\frac{x}{R_0(T)}\right)^2} - \sqrt{1-\left(\frac{L_0(T)}{R_0(T)}\right)^2}\right],$$

$$R_0(T) = R_0\left[1 - \frac{3R_0}{2h}\left(\frac{1}{2}|\varepsilon_0| + \bar{\bar{\varepsilon}}_\mathrm{m}(T)\right)\right]^{-1}, \quad (2)$$

$$L_0(T) = R_0(T)\sin\frac{l}{R_0(T)}.$$

In (2) and above, $R_0(T)$ and $L_0(T)$ are the radius of curvature and half the distance between the strip's ends when the temperature varies under the conditions of the unconstrained SME, respectively [9, 13]; $P(T) = EA\varepsilon(T)$ is the longitudinal compressive force at the strip's ends (Fig. 1c); E is the modulus of elasticity; $A = 2hb$ is the cross-sectional area of the strip; $\varepsilon(T) = [l_0 - L_0(T)]/l$ is the longitudinal compressive strain in the strip; $2l_0$ is the free spacing between the grips, which is related to gap Δ (Fig. 1d); and $J = b(2h)^3/12$ is the moment of inertia of the strip's cross section.

The averaged martensitic deformations $\bar{\bar{\varepsilon}}_\mathrm{m}(T)$ and $\tilde{\tilde{\varepsilon}}_\mathrm{m}(x, T)$ (under the conditions of the unconstrained and constrained SMEs, respectively) are determined when the bending moments applied to the strip are in equilibrium [9, 13]:

$$\bar{\bar{\varepsilon}}_\mathrm{m}(T) = \frac{1}{h^2}\int_{-h}^{h}\varepsilon_p(y, T)y\,dy,$$
$$\tilde{\tilde{\varepsilon}}_\mathrm{m}(x, T) = \frac{1}{h^2}\int_{-h}^{h}\varepsilon_p(x, y, T)y\,dy. \quad (3)$$

That is, they depend on the temperature and on the distributions of martensitic strain ε_p and stresses σ'_0 across the thickness of the strip (coordinate y) and along its length (coordinate x):

$$\begin{aligned}\varepsilon_p(x, y, T) &= [\varepsilon_R\varphi_R(\sigma'_0(x, y, T), T)\\ &+ \varepsilon_M\varphi_M(\sigma'_0(x, y, T), T)]\,\mathrm{sgn}[\sigma'_0(x, y, T)],\end{aligned} \quad (4)$$

where φ_R and φ_M are the volume fractions of R and M martensites, respectively; $\varepsilon_R = m_R\xi_R$; $\varepsilon_M = m_M\xi_M$; ξ_R and ξ_M are the shear strains of the lattice when it configures into the R and M modifications, respectively; and m_R and m_M are the orientation factors.[1]

The expressions for volume fractions φ_R and φ_V (according to the theory of diffuse martensitic transformations) are given in [13]. Their values depend on the temperature and bending stress:

$$\begin{aligned}&\sigma'_0(\bar{x}, \bar{y}, T)\\ &= E\left[\frac{1}{2}|\varepsilon_0|\left(\frac{3}{2}\bar{y} \pm 1\right) + 3\frac{W_0(\bar{x}, T)}{h}|\varepsilon(T)|\bar{y}\right],\end{aligned} \quad (5)$$

where $\bar{y} = y/h$ and $\bar{x} = x/L_0$.

To determine the shape of the strip when the temperature varies, we will make use of an equation relating the radius of curvature of the strip and its bending deflection $W(x, T)$ [14, 15]:

$$R^{-1}(x, T) = -\frac{W''}{(1 + W'^2)^{3/2}}, \quad (6)$$

where $W' = dW/dx$ and $W'' = dW'/dx$. The minus sign means that the curvature is positive when the strip is convex.

Since the angle of rotation $\Omega(x, T)$ of the strip's cross sections is $\Omega(x, T) = W'(x, T)$, the single integra-

---

[1] In [13], it was assumed that strains ε_R and ε_M depend on the volume concentration of Ti_3Ni_4 particles in the strip. Since this assumption was shown to be not quite correct, strains ε_R and ε_M in (4) depend only on the lattice strains and orientation factors.

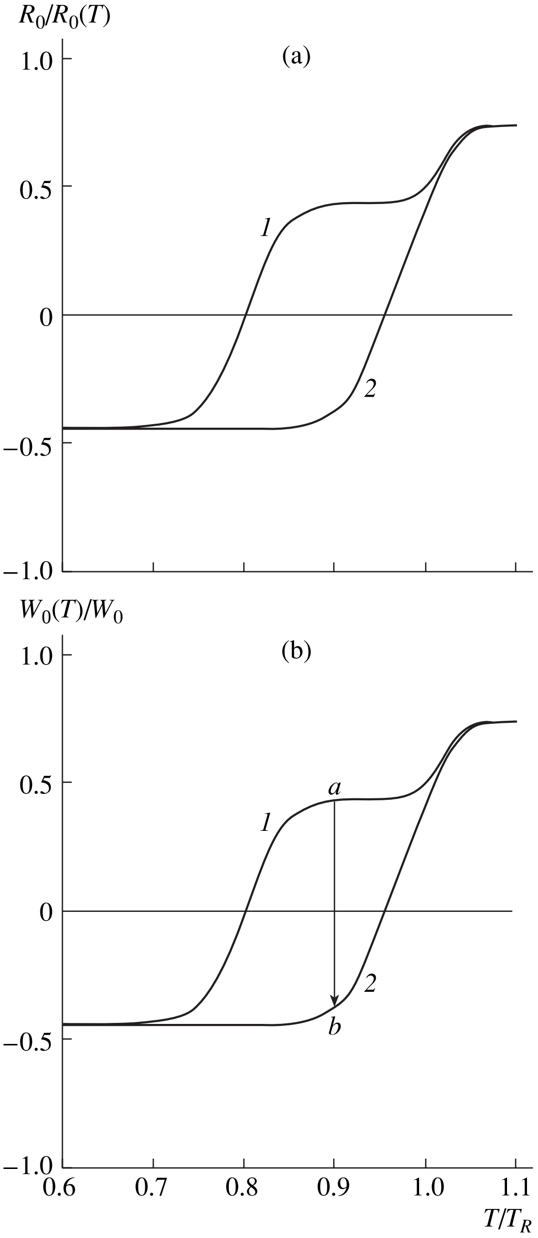


Fig. 3. Variation of the (a) curvature and (b) bending deflection of the arched strip with (*1*) decreasing and (*2*) rising the temperature under the conditions of the unconstrained SME.

tion of Eq. (6) yields

$$\frac{\Omega}{(1+\Omega^2)^{1/2}} = \omega(x, T),$$
$$\omega(x, T) = -\int_0^x \frac{dx}{R(x, T)}. \tag{7}$$

The rotation of the cross sections meets the edge condition $\Omega(x, T) = 0$ when $x = 0$. From the first relationship of (7), it follows that $\Omega(x, T) \approx \omega(x, T)$ if the rotation of the cross sections is small ($\Omega \ll 1$). In the general case, $\Omega = \omega/\sqrt{1-\omega^2}$; consequently, the sag of the strip as a function of coordinate x and temperature T is given by the integral

$$W(x, T) = \int_{-L_0(T)}^{x} \frac{\omega(x, T)dx}{\sqrt{1-\omega^2(x, T)}}. \tag{8}$$

Sag (8) meets the boundary condition $W = 0$ at the strip's ends $x = \pm L_0$. For $\omega(x, T) = -x/R_0(T)$, we can derive relationship (2) for the bending deflection under the conditions of the unconstrained SME from formulas (7) and (8).

UNCONSTRAINED SHAPE MEMORY EFFECT

In the quantitative calculations that follow, it is convenient to write the curvature of the strip (Eq. (1)) in an expanded and reduced form:

$$\frac{R_0}{R(x, T)} = 1 - \frac{3R_0}{2h}\left(\frac{1}{2}|\varepsilon_0| + \tilde{\tilde{\varepsilon}}_{\mathrm{m}}(\bar{x}, T)\right) + 3\left(\frac{R_0}{h}\right)^2 \frac{W_0(\bar{x}, T)}{R_0}|\varepsilon(T)|. \tag{9}$$

Figure 3a shows (according to (9)) the temperature dependence of the curvature of the strip ($l = 10$ mm, $b = 7$ mm, $h = 0.25$ mm, $R_0 = 29.5$ mm, $R_0/h = 118$, and $\varepsilon_0 = 3 \times 10^{-3}$; for the other parameters, see [9, 13]) for the unconstrained SME when $\varepsilon(T) = 0$ in (9) and martensitic strain $\bar{\bar{\varepsilon}}(T)$ is uniformly distributed along the strip's length. As the temperature declines (curve *1*), the curvature of the strip varies with temperature stepwise because of the two-step martensitic transformation (the sequential formation of the *R* and *M* martensite modifications). On heating (curve *2*), the temperature ranges where the *R* and *M* modifications disappear virtually join [9] and the recovery of the initial curvature occurs through the one-step $B19' \longrightarrow B2$ martensitic transformation.

Figure 3b demonstrates the temperature dependence of the sag at the central part ($x = 0$) of the strip (according to (7) and (8)), which corresponds to its curvature shown in Fig. 3a. The complete shape of the strip at different temperatures is depicted in Fig. 4, which shows its shapes after the annealing (dashed curve) and after the elastic relaxation of the internal anisotropic stresses (curve *1*). Curves *2* and *3* in Fig. 4 illustrate the bidirectional SME at $T = 0.9T_R$, i.e., the reversible change of the sign of curvature and of the bending deflection during the direct (curve *2*) and reverse (curve *3*) martensitic transformations (T_R is the characteristic temperature of the *R* transformation [9]). Under the given conditions and parameters of the strip, the maximum sag at its central part is ±0.8 mm. The transition of the strip to the fully martensitic state and the change of sign of the sag (point *b*) by applying force F (Fig. 1b) to the strip at a temperature of $0.9T_R$ (point *a*) are shown by arrow *ab* in Fig. 3b.

From comparison of the curves in Figs. 3a and 3b, it follows that, in the reduced coordinates, the bending deflection $W_0(0, T) = W_0(T)$ at the central part of the strip varies with temperature in the same way as the reduced curvature and that these values quantitatively coincide. This coincidence stems from the fact that (i) the deflection of the strip is smaller than its length ($\approx 0.5W_0/l \approx 0.88$) and (ii) the curvature R_0^{-1} of the strip is also smaller than its length both in the initial (after the annealing) state ($l/R_0 \approx 0.34$) and during the subsequent temperature variations ($R^{-1}(T) < R_0^{-1}$). Indeed, in the case of the unconstrained SME, it follows from (7) that $\omega_0(x, T) = -x/R_0(T)$. Next, since $|\omega_0(x, T)| \ll 1$, we have the following relationship for the reduced bending deflection $W_0(T)/W_0$ of the strip (according to (8)):

$$\frac{W_0(T)}{W_0} \approx \frac{L_0^2(T)}{2W_0R_0(T)} = \frac{R_0(T)}{4R_0}\frac{\left[\sin\left(\frac{l}{R_0(T)}\right)\right]^2}{\left[\sin\left(\frac{l}{2R_{0)}}\right)\right]^2} \approx \frac{R_0}{R_0(T)}. \qquad (10)$$

The second ratio on the right of (10) is valid if $l/2R_0 \ll 1$ and $l/R_0(T) \ll 1$ and confirms the statement that the curvature and deflection of the strip vary with temperature in a similar manner. Hence, under the above conditions, the bending deflection at the central part of the strip can be calculated from its curvature (see (9)).

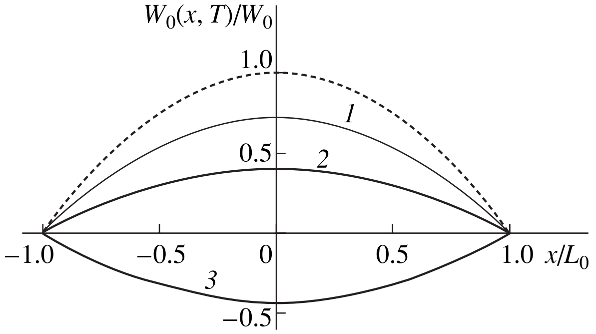


Fig. 4. Shape of the strip at different temperatures: dashed curve, after annealing; (*1*) after the elastic relaxation of internal stresses ($T = 1.1T_R$); and (*2*) and (*3*) after the martensitic relaxation of internal stresses as the temperature declines and rises, respectively ($T = 0.9T_R$).

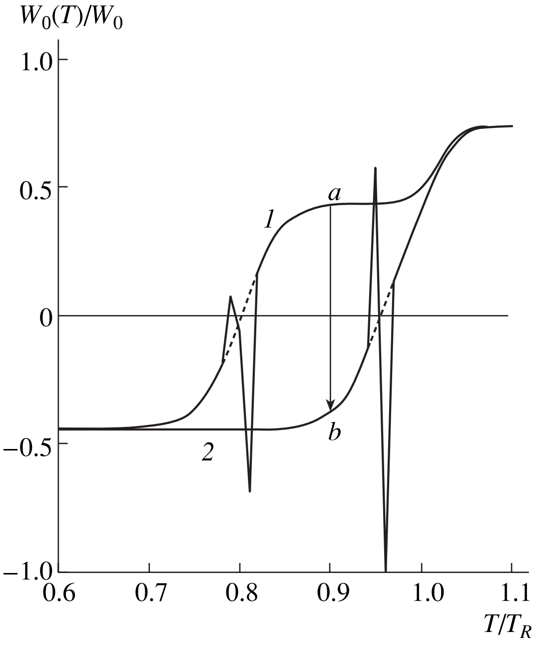


Fig. 5. Variation of the maximum bending deflection of the stripe under the constrained SME conditions during (*1*) direct and (*2*) reverse martensitic transformations.

CONSTRAINED SHAPE MEMORY EFFECT

Figure 5 shows the temperature dependence of the deflection at the central part when the temperature declines (curve *1*) and rises (curve *2*) under the conditions of the constrained SME; that is, $\varepsilon(T) \neq 0$ in (9) and martensitic strain $\tilde{\tilde{\varepsilon}}_{\mathrm{m}}(x, T)$ is nonuniformly distributed along the strip's length. Contrary to the unconstrained SME (dashed curve), the process of form changing during heating here is unstable: the strip exhibits severe vibrations, tending to straighten out ($W(T) \longrightarrow 0$), and the motion of its ends in the grips is limited. In the case shown in Fig. 5, the maximum compressive strain (constraint) of the strip is $\varepsilon(T_1) = \varepsilon(T_2) = (l_0 - L_0(T_{1,2})/l = -2.9 \times 10^{-4}$, where T_1 and T_2 are the temperatures at which the strip completely straightens out $L_0(T_{1,2}) = l$ as the temperature declines and rises, respectively. The free grip spacing is given by the expression $2l_0 = 2L_0(T_a) + \Delta$, where T_a is the temperature at which the strip is mounted in the grips and Δ is the total gap between the strip's ends and grips at this temperature (Fig. 1d). For the maximal compressive strain $\varepsilon(T_2)$ of the strip (see above) and the temperature at which the strip is mounted in the grips, $T_a = 1.1T_R$, gap Δ equals 0.2 mm and the compressive (bending) force at the strip's ends (Fig. 1c) is $P = EA\varepsilon(T_2) \approx 100N$, where $E = 100$ GPa. When the gap is narrow or is absent at all, the instability of the form changing process builds up, while for wide gaps, it weakens or even completely disappears. The process becomes unstable when the gap is smaller than the critical value, $\Delta < \Delta_{\mathrm{c}}$, where

$$\Delta_{\mathrm{c}} = 2L_0(T_{1,2}) - 2L_0(T_a) \approx 2[l - L_0(T_a)]. \qquad (11)$$

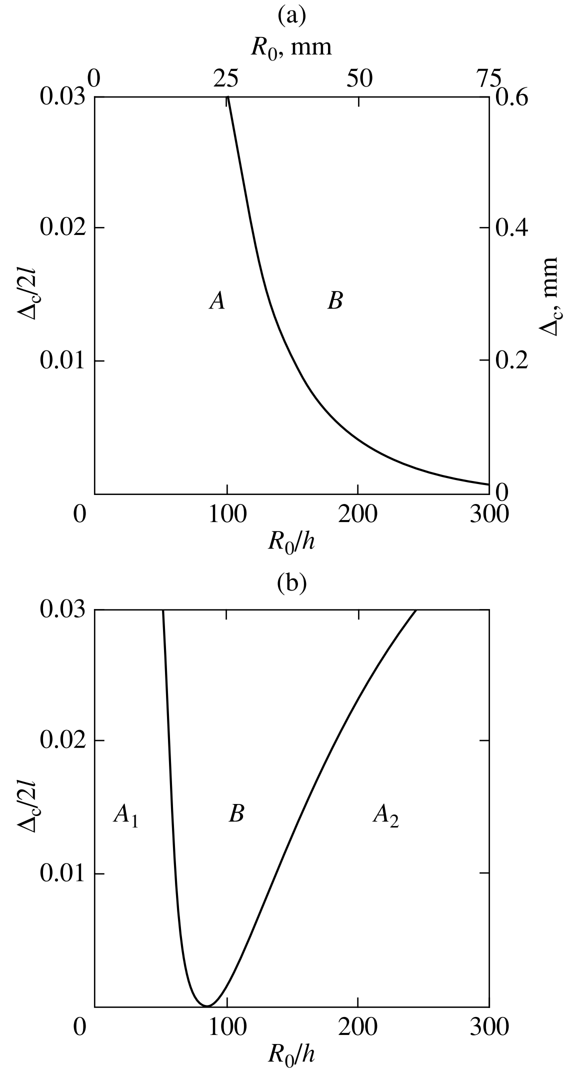


Fig. 6. Critical gap Δ_c at which the clicking effect arises vs. the radius of initial curvature R_0 of the strip. Strip mounting temperature T_a equals (a) $1.1T_R$ (a) and (b) $0.9T_R$.

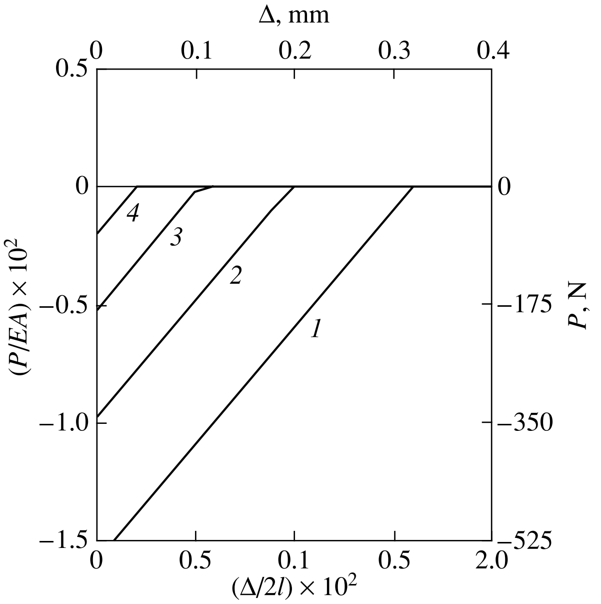


Fig. 7. Compressive force P at the ends of the strip under the conditions of the constrained SME vs. gap Δ. The initial radius of curvature of the strip R_0/h = (*1*) 100, (*2*) 120, (*3*) 150, and (*4*) 200.

Figures 6a and 6b demonstrate (according to (2) and (11)) how critical gap Δ_c varies with the initial radius of curvature R_0 of the strip in the reduced coordinates $\Delta_c/2l - R_0/h$ for $l/h = 40$. The temperatures at which the strip was mounted in the grips were taken to be $T_a = 1.1T_R$ and $0.9T_R$, respectively. In Fig. 6a, the domains of unstable and stable form changing are marked by letters *A* and *B*, respectively. The upper and right-hand scales in Fig. 6a show the values of R_0 and Δ_c for a strip $2l$ = 20 mm long and $2h$ = 0.5 mm thick. For initial radii of curvature of the strip R_0 ranging from 30 to 50 mm (R_0/h = 120–200) [10], clicking was heard when the gap varied between 0.1 and 0.5 mm. As is seen from Fig. 6a, this is in good agreement with the theory. From Fig. 6b, it follows that, if the temperature of mounting the strip into the grips is lower than the characteristic temperature T_R of the onset of martensitic transformation, there exist two ranges of parameters where the form changing during heating is unstable, $55 < R_0/h < 90$ (domain A_1) and $R_0/h > 90$ (domain A_2), with the domain *B* of stable form changing in between.

The elastic compressive strain of the strip for the constrained SME is given by the formula

$$\varepsilon(T) = [L_0(T_a) - L_0(T) + \Delta/2]l^{-1}. \tag{12}$$

The strain depends on the running temperature T of the strip, temperature T_a at which the strip is mounted into the grips, the initial radius of curvature R_0 of the strip, and gap Δ. As was mentioned above, strain ε reaches a maximum

$$\varepsilon(T_1) = \varepsilon(T_2) = [L_0(T_a) - l + \Delta/2]l^{-1} \tag{13}$$

at temperatures T_1 and T_2. At these temperatures, the strip is completely straightened out ($L_0(T_{1,2}) = l$) as the temperature declines and rises, respectively. Compressive force P acting on the ends of the strip, $P = EA\varepsilon(T_{1,2})$, is also maximal at these temperatures. Figure 7 plots the maximum compressive force during heating $P/EA = \varepsilon(T_2)$ versus relative gap $\Delta/2l$ for four values of the initial radius of curvature R_0/h at $T_a = 1.1T_R$. As the gap and radius of curvature increase, the compressive force and compressive strain at the strip's ends decrease and vanish when Δ and R_0 take the critical values (Fig. 6a). For $R_0 > 50$ mm ($R_0/h > 200$), the critical gap at which the effect of clicking is absent becomes less than 0.02 mm. The dimensional scales in Fig. 7 show how the compressive force varies with Δ.

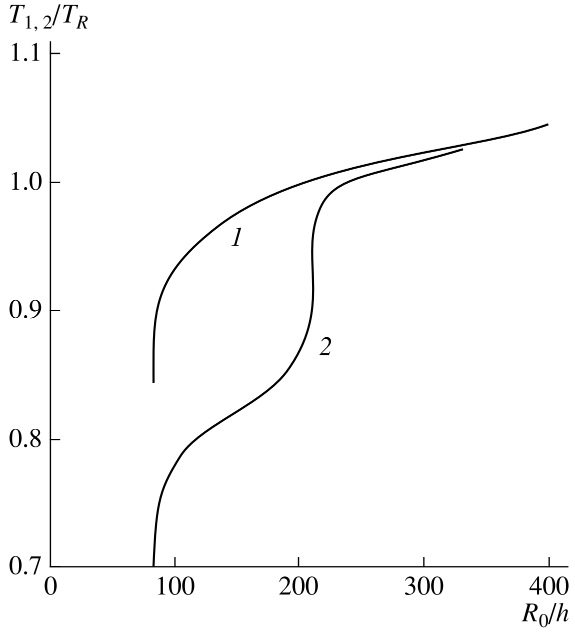


Fig. 8. Critical temperatures T_1 and T_2 at which the form changing process becomes unstable vs. the radius of initial curvature R_0 for (*1*) direct and (*2*) reverse martensitic transformations.

Compressive forces in the range $P = 10\text{–}120$ N (for gap widths of 0.1–0.5 mm and radii of curvature of 30–50 mm) [10, 11] correspond to the maximum compression strain in the range $(0.3\text{–}3.0) \times 10^{-4}$.

Critical temperatures T_1 and T_2, at which the deformation of the strip is unstable, are determined from the zero-curvature conditions that set in when the temperature declines and rises, respectively:[2]

$$\frac{R_0}{R_0(T_{1,2})} = 1 - \frac{3R_0}{2h}\left(\frac{1}{2}|\varepsilon_0| + \bar{\bar{\varepsilon}}_{\mathrm{m}}(T_{1,2})\right) = 0. \qquad (14)$$

Figure 8 plots the critical temperatures against the initial radius of curvature in dimensionless coordinates. For small radii ($R_0/h < 100$ or $R_0 < 25$ mm), the critical temperatures sharply drop. For large radii ($R_0/h > 200$ or $R_0 > 50$ mm), these temperatures are nearly the same.

Thus, our investigation into the unstable form changing of an arched TiNi strip under the conditions of the constrained shape memory effect demonstrates that the clicking effect depends on a number of factors and relationships between them. Among these factors are the geometry of the strip, its curvature specified by annealing, and the temperature at which the strip is mounted into the grips. Another group includes structural factors, such as the parameters of martensitic transformation and elastic internal stresses (strains ε_0), which depend on the Ti_3Ni_4 particle concentration in the strip, i.e., on the temperature and duration of annealing of the bent strip.

REFERENCES

1. P. Krulevitch, A. P. Lee, P. B. Ramsey, *et al.*, J. Microelectromech. Syst. **5**, 270 (1996).
2. J. L. Seguin, M. Bendahan, A. Isalgue, *et al.*, Sens. Actuators **74**, 65 (1999).
3. N. Nishida and T. Honma, Scr. Metall. **18**, 1293 (1984).
4. T. Honma, in *Shape Memory Alloy-86*, Ed. by Ch. Youyi, T. Y. Hsu, and T. Ko (Academic, Guilin, 1986), pp. 83–88.
5. H. Takagi, K. Okano, S. Juodkazis, *et al.*, Adv. Eng. Mater. **5**, 732 (2003).
6. D. Y. Li and L. Q. Chen, Acta Mater. **45**, 471 (1997).
7. G. A. Malygin, Zh. Tekh. Fiz. **66** (9), 112 (1996) [Tech. Phys. **41**, 1145 (1996)].
8. G. A. Malygin, Usp. Fiz. Nauk **171**, 187 (2001) [Phys. Usp. **44**, 173 (2001)].
9. G. A. Malygin, Fiz. Tverd. Tela (St. Petersburg) **45**, 1700 (2003) [Phys. Solid State **45**, 1784 (2003)].
10. M. A. Khusainov, Vestn. Novgor. Gos. Univ., No. 10, 34 (1998).
11. M. A. Khusainov and V. N. Belyakov, in *Proceedings of the 3rd International Scientific Likhachev Seminar Modern Problems of Strength "Topical Problems of Strength," Staraya Russa, 1997*, Vol. 2, Part 1, pp. 139–142.
12. M. A. Khusainov, Zh. Tekh. Fiz. **67** (6), 118 (1997) [Tech. Phys. **42**, 692 (1997)].
13. G. A. Malygin, Fiz. Tverd. Tela (St. Petersburg) **45**, 2233 (2003) [Phys. Solid State **45**, 2342 (2003)].
14. V. I. Feodos'ev, *Strength of Materials* (Nauka, Moscow, 1972) [in Russian].
15. A. S. Vol'mir, *Stability of Strained Systems* (Nauka, Moscow, 1967) [in Russian].

Translated by Yu. Vishnyakov

---

[2] Equation (14) yields temperatures T_1 and T_2 in the zero-order approximation when the constrained strain of the strip is calculated by the iteration method [13].

Technical Physics, Vol. 49, No. 10, 2004, pp. 1308–1312. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 64–68.
Original Russian Text Copyright © 2004 by Korolev, Volkov.

SOLIDS

Nanoprecipitation Resulting from the Decomposition of Supersaturated Solid Solutions in the Tracks of Swift Heavy Ions

D. N. Korolev and A. E. Volkov
Russian Research Center Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia
e-mail: deck@dni.polyn.kiae.su

Received March 23, 2004

Abstract—The electron subsystem of a material is strongly excited when swift heavy ions pass through the material. The subsequent relaxation of this excitation results in considerable short-term ($<10^{-9}$ s) heating of the material in the nanometer vicinity of the projectile trajectory. Nanoprecipitation stimulated by such thermal spikes in supersaturated solid solutions is studied. Nanoprecipitates are shown to form when the temperature in the track reaches a point where the characteristic time of precipitation becomes shorter than the time of cooling of the track. The region of most efficient precipitation may be offset from the track axis. The initial cylindrical nonuniformity of the spatial density of nucleating clusters may cause the formation of nanodimensional tubular heterostructures extended along the trajectory of the heavy ions. The parameters of the system that are the most favorable to the tubular mode of precipitation are found. *© 2004 MAIK “Nauka/Interperiodica”.*

INTRODUCTION

An appreciable part of the energy of swift heavy ions (with a mass $M > 50m_p$, where m_p is the mass of a proton, and energy $E > 1$ MeV/nucleon) passing through a material is spent on the excitation of its electron subsystem in the nanodimensional region near the trajectory of the projectile. Recent experiments have shown that the relaxation of such high-energy electron excitations causes nanoprecipitation in supersaturated solid solutions of metals [1, 2], insulators [3], and semiconductors [4]. The amount of this effect correlates with electron losses of the swift heavy ion (SHI) energy and opens up new possibilities of controlling the precipitation kinetics over a small range (about several nanometers). Specifically, this effect may be used to advantage for producing arrays of nanodimensional heterostructures (quantum dots and nanotubes) involved in multilayer systems with a single technological process [4].

Energy transfer from the electrons excited to the atomic subsystem results in local heating of the material near the track. In this case, the temperature in the SHI tracks may rise to the point where it stimulates phase transitions in the material or in fine impurity inclusions present in it [5–8]. In this work, we study the effect of such thermal spikes [9–13] on nanoprecipitation in the vicinity of the SHI trajectory.

The basic idea of this article is that the temperature dependence of the precipitation time has a sharp minimum at some characteristic temperature that depends on the parameters of the system [14]. Impurities may precipitate if the temperature in the track reaches the point where the precipitation time becomes shorter than the time of cooling of the track.

The region that is the most favorable to the precipitation may be offset from the track axis. In this case, the precipitate spatial distribution is cylindrically symmetric with an off-axis maximum; i.e., the spatial configuration of precipitates appears to be tubular.

Here, we analyze the temperature dependence of the precipitation time and also present the parameters of the matrix, impurity atoms, and heavy projectiles, as well as the irradiation conditions, that facilitate impurity precipitation in the SHI track. In addition, we estimate the parameters ensuring the nucleation of tubular nanodimensional clusters along the SHI trajectory.

MODEL

Assume that, before irradiation, the concentration of impurity atoms in the matrix exceeds the value corresponding to their solubility limit at an irradiation temperature (supersaturated solid solution).

Let a portion of the energy of the projectiles go into local heating of the material in the SHI track. Then, we assume that the initial temperature in the track does not exceed the melting point; otherwise, we will describe the system, starting from the time when the temperature in the track decreases below the melting point. Finally, the size of the resulting clusters is assumed to be smaller than the thermal spike diameter, so that the influence of temperature gradients on the precipitation kinetics can be neglected.

1063-7842/04/4910-1308$26.00 © 2004 MAIK “Nauka/Interperiodica”

The effect of temperature on the precipitation in the track heated depends on two competitive processes. On the one hand, the diffusion mobility of impurity atoms increases with temperature, thereby rising the frequency of their collisions and, hence, the precipitation rate. On the other hand, a rise in temperature raises the solubility limit of an impurity in the matrix. Accordingly, the supersaturation of the solid solution decreases (with the concentration remaining unchanged), the thermodynamic driving force of the phase transition diminishes, and the precipitation rate drops. Because of competition between these tendencies, a temperature interval generally arises where the precipitation rate reaches a maximum [14]. Thus, for a certain set of the system's parameters, a rise in the temperature of the SHI track may considerably cut the characteristic precipitation time. Evidently, precipitation will actually take place if the precipitation time is shorter than the time of track cooling.

The cooling time depends on the heat conduction mechanism and is estimated from the expression

$$t_T \approx \frac{R_T^2}{\chi_i}, \tag{1}$$

where R_T is the initial radius of the region heated and χ_i is the thermal diffusivity of the material.

As follows from the experimental data and numerical estimations, the radius of the thermal spike ranges from several nanometers to several tens of nanometers [12, 15, 16]. The shortest cooling time can be found by taking the thermal diffusivity of metals ($\chi_i = 10^{-3}$–10^{-2} cm^2/s):

$$t_T = 10^{-9}\text{–}10^{-10}\ \text{s}. \tag{2}$$

At a high concentration, even two impurity atoms may form a stable nucleus. Structural inhomogeneities in the matrix may also favor impurity precipitation. However, we will seek the upper estimate of the nucleation time and consider the case of slowest homogeneous nucleation, which occurs at low supersaturations and high temperatures. Such a process can be described as the overcoming of an energy barrier whose height depends on the supersaturation of the solution.

Taking into account the nanometer size of the precipitates and a very short (on the order of nanoseconds) time of track cooling, we may assume that the forming nuclei introduce insignificant, if any, perturbations into the spatial distribution of impurity atoms in the track. In this case, the critical barrier for precipitation, φ_c, takes the form [14, 17, 18]

$$\varphi_c = \frac{16\pi}{3}\frac{\Omega^2\gamma^3}{\ln^2(c_i/C)T(T_0 - T)^2}. \tag{3}$$

Here, $T_0 = \Psi/\ln(c_i/C)$ (temperature $T < T_0$ is measured in energy units); c_i and C are the atomic concentrations of impurity atoms in the precipitates and matrix, respectively; Ψ is the Gibbs energy of impurity atom dissolution in the matrix; Ω is the volume occupied by an impurity atom in the precipitate; and γ is the surface tension coefficient of the nuclei. The precipitates are assumed to be spherical.

The characteristic time of precipitation t_p is defined as the average time it takes for nuclei of the new phase to grow to the critical size (n_c) corresponding to the precipitation barrier [19]:

$$t_p(n_{\min} \longrightarrow n_c) \simeq \left(\frac{2}{\pi}Z\right)^{-1/2}\frac{\exp[\varphi(n_c)]}{2\pi\nu\exp(-U/T)}, \tag{4}$$

where ν is the oscillation frequency of impurity atoms, U is the migration barrier for impurity atoms in the matrix, and

$$Z = \left|\frac{d^2\varphi}{dy^2}\right|_{y = n_c} = \frac{\ln^4(c_i/C)}{32\pi\Omega^2\gamma^3}\frac{(T_0 - T)^4}{T} \tag{5}$$

is the Zel'dovich factor.

PRECIPITATION CONDITIONS IN THE TRACKS

For Eq. (4), it follows that precipitation time $t_p(T)$ has a deep minimum at a temperature that substantially depends on parameters Ψ, γ, and U, which characterize the properties of impurities and precipitates in the matrix (Fig. 1). For typical values of the solubility limit (Ψ = 0.5–2.0 eV, solutions of point defects or impurity atoms), surface tension coefficient (γ varies from 100 erg/cm^2 for impurity precipitates to 1000 erg/cm^2 for voids and bubbles), and migration barrier (U = 0.4–1.5 eV) [18, 20, 21], the following results were obtained. An increase in Ψ increases the driving force

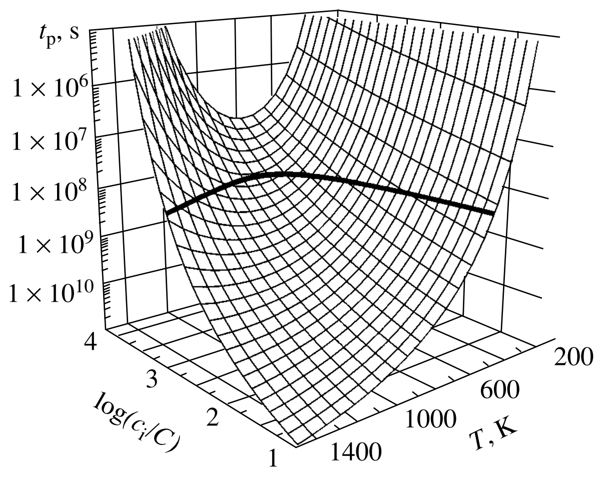


Fig. 1. Precipitation time t_p as a function of temperature T in the track and supersaturation $\log(c_i/C)$ (U = 0.4 eV, γ = 50 erg/cm^2, Ψ = 2 eV). The curve on the surface corresponds to precipitation time $t_p = 10^{-9}$ s.

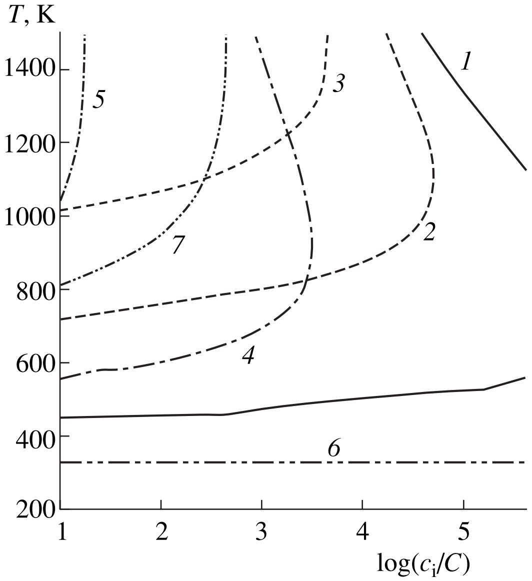


Fig. 2. Precipitation time isolines bounding from the right the region where the precipitation time does not exceed 10^{-9} s. Curves *1–3* drawn for $\gamma = 500$ erg/cm^2 and $\Psi = 2$ eV show the dependence of the precipitation time t_p on migration barrier height U ($U = 0.4$, 0.7, and 1.0 eV, respectively); curves *1, 4,* and *5* ($\gamma = 500$ erg/cm^2, $U = 0.4$ eV) show the dependence of the precipitation time on the solubility limit of the impurity ($\Psi = 2$, 1.5, and 1 eV, respectively); and curves *1, 6,* and *7* ($U = 0.4$ eV, $\Psi = 2$ eV) demonstrate the dependence of the precipitation time on the surface tension coefficient ($\gamma = 500$, 200, and 800 erg/cm^2, respectively).

of the phase transition and decreases the characteristic precipitation time (curves *1, 4, 5* in Fig. 2; here, Ψ varies from 1.0 to 2.5 eV). For high Ψ, the precipitation time does not exceed 1 ns over an extended region in the plane (T, $\log(c_i/C)$). Moreover, for the supersaturation reached in the experiments ($\log(c_i/C) \lesssim 3$) [1–3], precipitation time t_p becomes shorter than 100 ps ($t_p \lesssim 10^{-10}$ s).

The dependence of precipitation time t_p on surface tension γ is shown in Fig. 2 (curves *1, 6, 7*; here, γ varies from 200 to 800 erg/cm^2). At high concentrations of impurity atoms in the matrix ($C \sim 1\%$), the precipitation time does not exceed 1 ns even if the surface tension coefficient is above 800 erg/cm^2. Such high values correspond, for example, to the formation of gas bubbles or vacancy pores in the tracks due to the decomposition of supersaturated solutions of gases and point defects [20].

The diffusion barrier height also has a significant effect on the precipitation kinetics. As is seen from Fig. 2 (curves *1–3*), low diffusion barriers facilitate impurity precipitation in the SHI track [3, 15, 16].

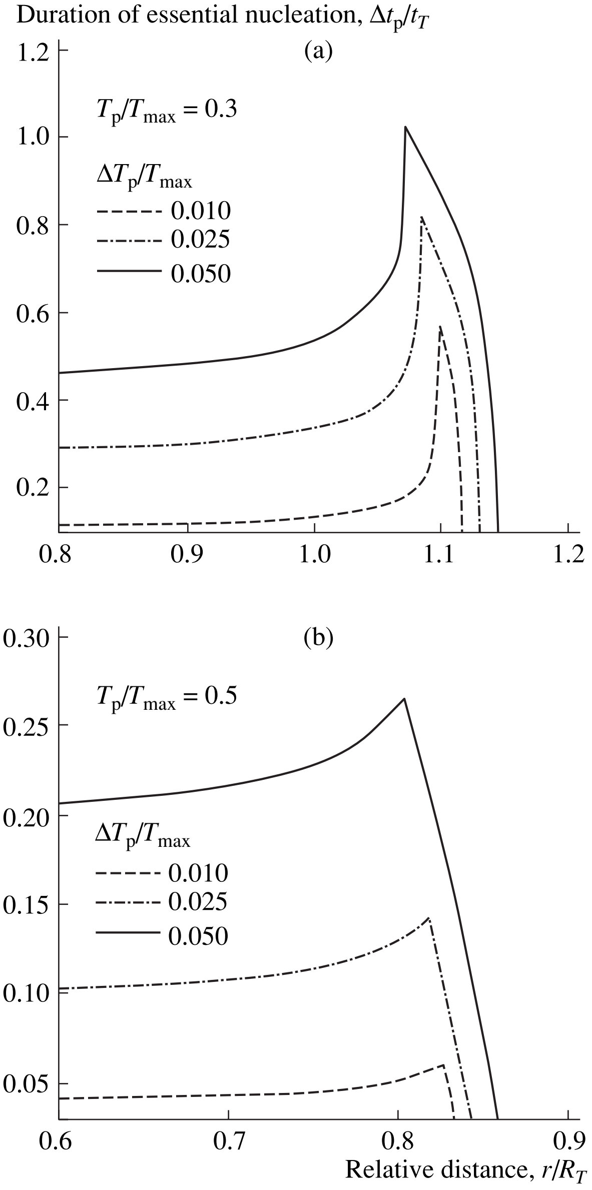


Fig. 3. Duration $\Delta t_p(r)$ of the temperature interval vs. distance r from the track axis. T_{max} is the initial temperature in the track axis. (a) Sharp peak at $\Delta T_p/T_{max} = 0.3$ and (b) step dependence at $\Delta T_p/T_{max} = 0.5$.

ON THE TUBULAR PRECIPITATION IN THE SHI TRACK

At the time when the thermal spike appears, the axial region of the track has the highest temperature. Heat transfer warms the peripheral regions of the track, and the temperature of the axial region declines. When the temperature appropriate for the precipitation persists at the periphery longer than at the center, tubular precipitation may occur. In this case, the concentration of the forming clusters has a cylindrically symmetric distribution with an off-axis maximum. Subsequently, such a spatial distribution may evolve into nanodimen-

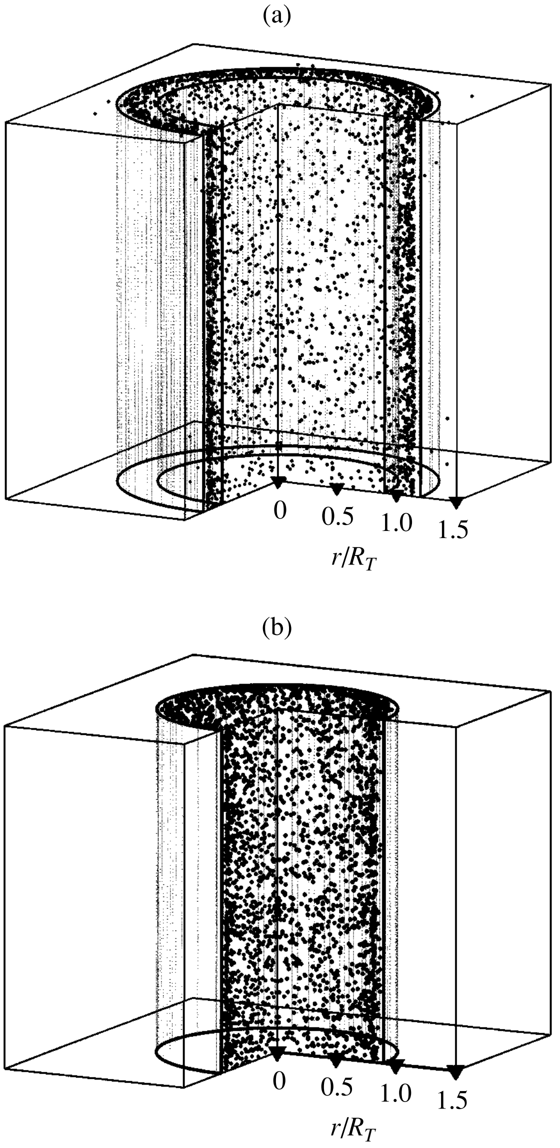


Fig. 4. Distribution of precipitates in the SHI track for the (a) tubular and (b) cylindrical precipitation.

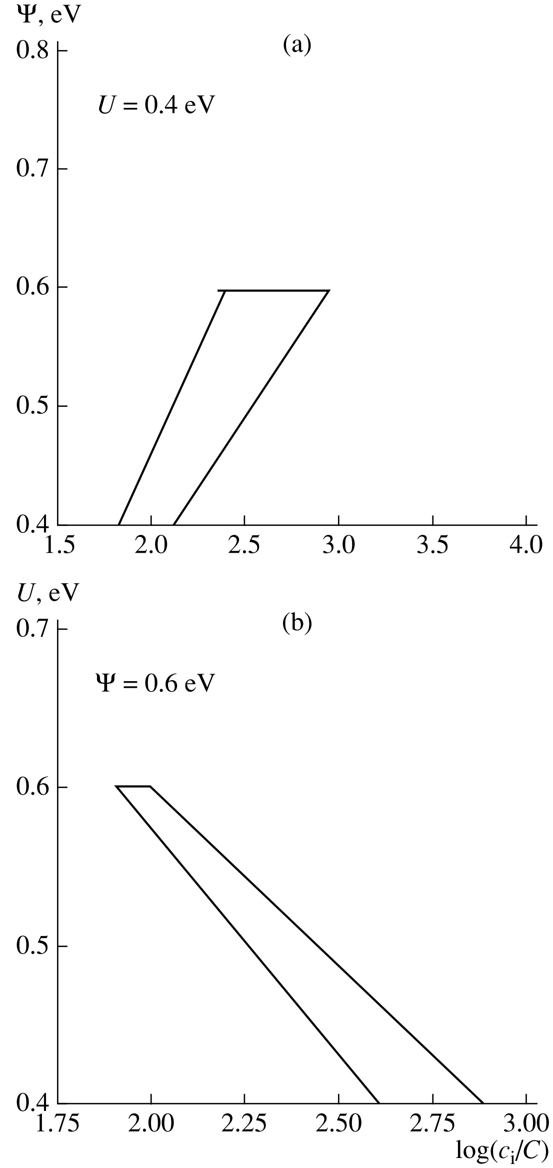


Fig. 5. The solid line bounds the domain of parameters that is favorable for the tubular precipitation in the SHI track.

sional tubular heterostructures extended along the SHI trajectory.

To find the conditions necessary for tubular precipitation, let us see how the duration of the temperature interval where the precipitation is the most efficient depends on distance r from the track axis. The duration of this temperature interval ΔT_p centered at $T_p \le T_{max}$, so that $|T - T_p| \le \Delta T_p/2$ (T_{max} is the initial temperature at the track axis), versus relative distance r/R_T is shown in Fig. 3. The corresponding distribution of precipitates in the track is shown in Fig. 4.

As follows from Fig. 3a, if precipitation time t_p is minimal at temperature T_p that does not exceed $0.3T_{max}$, the time of efficient precipitation in the peripheral region is one order of magnitude longer than that at the track axis. Furthermore, if the maximal temperature in the track under the given irradiation conditions is close to the melting point of the material ($T_{max} \approx T_m$, where T_m = 1500–2000 K is the melting point), the tubular precipitation is possible at a low solution energy (Ψ) and high mobility of impurity atoms (U) when the characteristic temperature of precipitation T_p lies between 500 to 650 K (Figs. 5a, 5b).

DISCUSSION

Consider the Fe–Cu system (the solid solution of copper in iron) as an example. In the experiments conducted in [1, 2], the copper concentration in the iron was 1–2% ($\log(c_i/C) \approx 2$). The bonding energy E_b^{Cu-pr} of copper atoms in the matrix with copper precipitates and barrier $E_m^{v,Cu}$ for copper atom migration via the vacancy mechanism were estimated as E_b^{Cu-pr} = 0.5–0.8 eV and $E_m^{v,Cu}$ = 0.4 eV [21]. These values of the parameters are favorable for precipitation in the SHI track if the thermal spike exceeds 500–600 K (see above). Moreover, these values are very close to those at which the tubular precipitation may take place provided that the temperature in the track is close to the melting point of the matrix.

ACKNOWLEDGMENTS

This work was supported by grant no. 15 (2003) for Junior Scientists from the Russian Research Center Kurchatov Institute and by grant no. NSh-1795.2003.2 of the Scientific School.

REFERENCES

1. A. Barbu, P. Pareige, and V. Jacquet, Nucl. Instrum. Methods Phys. Res. B **146**, 278 (1998).
2. A. Iwase, T. Hasegawa, Y. Chimi, *et al.*, Nucl. Instrum. Methods Phys. Res. B **195**, 309 (2002).
3. E. Valentin, H. Bernas, C. Ricolleau, and F. Creuzet, Phys. Rev. Lett. **86**, 99 (2001).
4. P. I. Gaiduk, A. Nylandsen-Larsen, C. Trautmann, and M. Toulemonde, Phys. Rev. B **66**, 045316 (2002).
5. R. E. Voskoboinikov, A. E. Volkov, and V. A. Borodin, Radiat. Eff. Defects Solids **152**, 181 (2000).
6. H. Dammak, A. Dunlop, and D. Lesueur, Philos. Mag. A **79**, 147 (1999).
7. A. Audouard, A. Dunlop, D. Lesueur, *et al.*, Eur. Phys. J.: Appl. Phys. **3**, 149 (1998).
8. J.-M. Castantini, F. Brisard, M. Toulemonde, and F. Studer, Nucl. Instrum. Methods Phys. Res. B **122**, 514 (1997).
9. M. I. Kaganov, I. M. Lifshitz, and L. V. Tanatarov, J. Nucl. Energy, Part A: Reactor Sci. **12**, 69 (1960); Zh. Éksp. Teor. Fiz. **30**, 232 (1956) [Sov. Phys. JETP **3**, 216 (1956)].
10. I. A. Baranov, Yu. V. Martynenko, S. O. Tchepelevich, and Yu. N. Yavlinskiy, Usp. Fiz. Nauk **156**, 178 (1998) [Sov. Phys. Usp. **31**, 1015 (1988)].
11. M. Toulemonde, C. Dufour, and E. Paumier, Phys. Rev. B **46**, 14 362 (1992).
12. A. E. Volkov and V. A. Borodin, Nucl. Instrum. Methods Phys. Res. B **146**, 137 (1998).
13. G. Schiwietz, G. Xiao, P. L. Grande, *et al.*, Nucl. Instrum. Methods Phys. Res. B **146**, 131 (1998).
14. J. Frenkel, *Kinetic Theory of Liquids* (Nauka, Leningrad, 1945; Clarendon Press, Oxford, 1946).
15. A. I. Ryazanov, A. E. Volkov, and S. Klaumünzer, Phys. Rev. B **51**, 12107 (1995).
16. A. I. Ryazanov, H. Trinkaus, and A. E. Volkov, Phys. Rev. Lett. **84**, 919 (2000).
17. Ya. B. Zel'dovich, Acta Physicochim. URSS **18**, 1 (1943); Zh. Éksp. Teor. Fiz. **12**, 525 (1992).
18. A. E. Volkov, M.-O. Ruault, H. Bernas, and V. A. Borodin, Nucl. Instrum. Methods Phys. Res. B **178**, 327 (2001).
19. C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*, 2nd ed. (Springer-Verlag, Berlin, 1985; Mir, Moscow, 1986), Vol. 13, Chaps. 3, 5.
20. A. E. Volkov and A. I. Ryazanov, J. Nucl. Mater. **273**, 155 (1999).
21. Yu. N. Osetsky and A. Serra, Philos. Mag. A **73**, 249 (1996).

Translated by E. Yablonskaya

Technical Physics, Vol. 49, No. 10, 2004, pp. 1313–1318. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 69–74.
Original Russian Text Copyright © 2004 by Bulgakov, Kononenko.

SOLID-STATE ELECTRONICS

Dispersion Properties of Cyclotron Waves in Periodic Semiconductor–Insulator Structures

A. A. Bulgakov and V. K. Kononenko

Usikov Institute of Radiophysics and Electronics, National Academy of Sciences of Ukraine,
Kharkov, 61085 Ukraine
e-mail: bulgakov@ire.kharkov.ua

Received October 22, 2003

Abstract—The band spectrum of cyclotron waves propagating in a periodic layered semiconductor–insulator structure at an angle to an external magnetic field that is applied perpendicularly to the layers is calculated for two relationships between the characteristic frequencies of the semiconductor: $\omega_H > \omega_P$ and $\omega_H < \omega_P$. The wave field distributions across the layers and over the period of the structure are analyzed. In both spectra, transmission bands arise when the conditions for dimensional resonance across the semiconductor layer are fulfilled. The graphic solution of the dispersion relation demonstrates that the cyclotron wave spectrum can be subdivided into two spectra of normal waves according to the Bloch wavenumbers of the periodic structure. The cases where the band spectra complement each other or overlap are considered. *© 2004 MAIK “Nauka/Interperiodica”.*

INTRODUCTION

Previously [1], we considered the band spectrum of electromagnetic waves propagating in a layered periodic semiconductor structure at an angle to an external magnetic field that is applied perpendicularly to the layers (along the periodicity axis). Our statement of the problem differs from those both in [2], where the periodicity axis, the direction of the electromagnetic wave, and the external magnetic field are aligned, and in [3–5], where the magnetic field is normal to the periodicity axis and to the plane of wave propagation. In our statement, the Maxwell equations for a gyrotropic medium are not split into two equations for the field components of independent polarizations and normal waves in such a medium are elliptically polarized waves having six field components each [6]. It was also shown [1] that the band spectrum includes a specific type of electromagnetic waves, the so-called cyclotron waves, which introduce numerous transmission bands into the spectrum. In this work, we study the band structure of the cyclotron wave spectrum and also the wave field distributions across the layers and over the period of the structure.

MATHEMATICAL BACKGROUND

Let us consider an infinite stack of alternating semiconductor and insulating layers (with thicknesses d_1 and d_2, respectively) in the presence of a permanent magnetic field directed perpendicularly to the layers (the z axis). A dispersion relation that describes the propagation of waves in this structure can be found by the transformation matrix method [7] in the form [1]

$$\xi^4 + B_3\xi^3 + B_2\xi^2 + B_1\xi + 1 = 0, \qquad (1)$$

where $\xi = \exp(i\bar{k}d)$, $\bar{k}$ is the Bloch wavenumber, and $d = d_1 + d_2$ is the period of the structure. Coefficients B_1, B_2, and B_3 are expressed in terms of the elements of the one-period transformation matrix, which relates the fields at the extremities of the period of the structure.

It was shown numerically [1] that, without regard for dissipation, dispersion relation (1) can be represented in the form

$$\cos\bar{k}_1 d = -\frac{a_1}{2}, \qquad (2)$$

$$\cos\bar{k}_2 d = -\frac{a_2}{2}. \qquad (3)$$

Equation (1) and expressions (2) and (3) are related as

$$a_{1,2} = \frac{B_1}{2} \pm \sqrt{\left(\frac{B_2}{2}\right)^2 + 2 - B_2}. \qquad (4)$$

Expression (4) is found as a solution of the quadratic equation; hence, coefficients a_1 and a_2 may take either real or complex conjugate values.

The fact that the dispersion relation can be represented in the form of Eqs. (2) and (3) means that two wave spectra exist, each being characterized by the respective Bloch wavenumber $\bar{k}_1$ or $\bar{k}_2$. Physically, these spectra correspond to two normal (generally, elliptically polarized) waves propagating in any anisotropic medium [6]. In the transmission bands of these spectra, the Bloch wavenumbers are real and must meet the conditions $|a_{1,2}| \le 2$, where a_1 and a_2 are real numbers. In any other case, $\bar{k}_1$ and $\bar{k}_2$ are complex numbers and characterize the opacity bands. Since Eqs. (2) and (3) are mutually independent, the corresponding spectra of the normal waves may complement each other or overlap.

To gain an insight into the physical processes that take place in the presence of two band spectra, one must know the specific features of their formation and the distribution of the electromagnetic fields in each of the layers. Accordingly, independent expressions for the field components in the material of the layers must be derived.

A method for determining the wave fields in gyrotropic media is presented in [8]. In this method, the field components are expressed via differential operators and scalar function $\Psi(x, y, z)$, which defines the z dependence of the field components in the direction of permanent magnetic field B_0. For the configuration described above, this function can be written as a sum of the eigenfunctions of the wave equation:

$$\Psi = (A_1 \cos k_{1z} z + A_2 \sin k_{z1} z + A_3 \cos k_{z2} z + A_4 \sin k_{z2} z) \exp[i(k_x x + k_y y)]. \tag{5}$$

Here, k_{z1} and k_{z2} are the wavenumbers describing the transverse distribution of the field for two normal waves in the semiconductor layers with the permittivity tensor [9]

$$\hat{\varepsilon} = \begin{vmatrix} \varepsilon_1 & i\varepsilon_2 & 0 \\ -i\varepsilon_2 & \varepsilon_1 & 0 \\ 0 & 0 & \varepsilon_3 \end{vmatrix}, \tag{6}$$

where

$$\varepsilon_1 = \varepsilon_L \left[1 + \frac{\omega_p^2}{\omega_H^2 - \omega^2}\right], \quad \varepsilon_2 = \frac{\varepsilon_L \omega_p^2 \omega_H}{\omega(\omega_H^2 - \omega^2)},$$

$$\varepsilon_3 = \varepsilon_L \left[1 - \frac{\omega_p^2}{\omega^2}\right],$$

ε_L is the permittivity of the semiconductor lattice, ω_p is the plasma frequency, ω_H is the cyclotron frequency, and ω is the frequency of the electromagnetic wave.

The transverse wavenumbers can be expressed as

$$k_{z1,2}^2 = \frac{1}{2}\left[2\varepsilon_1 k_0^2 - \left(1 + \frac{\varepsilon_1}{\varepsilon_3}\right)k_{xy}^2\right] \pm \sqrt{\frac{1}{4}\left[2\varepsilon_1 k_0^2 - \left(1 + \frac{\varepsilon_1}{\varepsilon_3}\right)k_{xy}^2\right]^2 - \varepsilon_1\left[\varepsilon_V k_0^4 + \frac{k_{xy}^2}{\varepsilon_3} - \left(1 + \frac{\varepsilon_V}{\varepsilon_3}\right)k_{xy}^2 k_0^2\right]}, \tag{7}$$

where $k_0 = \omega/c$, $\varepsilon_V = \varepsilon_1 - (\varepsilon_2^2/\varepsilon_1)$ is the Voigt permittivity, and $k_{xy} = \sqrt{k_x^2 + k_y^2}$.

After substituting function $\Psi(x, y, z)$ into the corresponding expressions for the fields [8, 9], we obtain relationships for the transverse (relative to the magnetic field direction or to the periodicity axis) field components in the semiconductor layers:

$$E_{x,y} = C_{1,3} A_1 \cos k_{z1} z + C_{1,3} A_2 \sin k_{z1} z + C_{2,4} A_3 \cos k_{z2} z + C_{2,4} A_4 \sin k_{z2} z, \tag{8a}$$

$$H_{x,y} = -D_{1,3} A_1 \sin k_{z1} z + D_{1,3} A_2 \cos k_{z1} z - D_{2,4} A_3 \sin k_{z2} z + D_{2,4} A_4 \cos k_{z2} z, \tag{8b}$$

where $C_{1,2} = (ik_y Q_{1,2} - \varepsilon_2 k_x P)S$, $D_{1,2} = k_{z1,2}(k_x/k_0 Q_{1,2} - i\varepsilon_2 k_0 k_y)S$, $C_{3,4} = (-ik_x Q_{1,2} - \varepsilon_2 k_y P)S$, $D_{3,4} = k_{z1,2}(k_y/k_0 Q_{1,2} + i\varepsilon_2 k_0 k_x)S$, $P = k_0^2 - k_{xy}^2/\varepsilon_3$, $Q_{1,2} = \varepsilon_1 P - k_{z1,2}^2$, and $S = \exp[i(k_x x + k_y y)]$.

In the dielectric medium, the expressions for the transverse components can be derived from the Maxwell equations. At arbitrary point z of the insulator, the field components of a plane wave $\exp[-i(\omega t - \mathbf{kr})]$, when expressed in terms of the fields at point $z = 0$, take the form

$$E_{x,y}(z) = F_1 E_{x,y}(0) + \frac{F_{2,4}}{\varepsilon} H_x(0) - \frac{F_{3,2}}{\varepsilon} H_y(0), \tag{9a}$$

$$H_{x,y}(z) = -F_{2,4} E_x(0) + F_{3,2} E_y(0) + F_1 H_{x,y}(0). \tag{9b}$$

Here,

$$F_1 = \cos k_z z, \quad F_2 = i\frac{k_x k_y}{k_0 k_z} \sin k_z z,$$

$$F_3 = i\frac{(k_x k_y)^2 + \varepsilon(k_0 k_z)^2}{k_0 k_z (k_y^2 - \varepsilon k_0^2)} \sin k_z z,$$

$$F_4 = i\frac{k_y^2 - \varepsilon k_0^2}{k_0 k_z} \sin k_z z,$$

$k_z = \sqrt{\varepsilon k_0^2 - k_{xy}^2}$, and ε is the permittivity of the insulator.

Thus, we obtained two systems of equations ((8) and (9)) consisting of the independent expressions for the field components in the semiconductor and insulator that are transverse to the magnetic field. Either of the systems contains four unknown coefficients: $E_x(0)$, $E_y(0)$, $H_x(0)$, and $H_y(0)$, and A_1–A_4, respectively. In order to find them, we make use of (i) four boundary conditions for the field components at the interfaces:

$$L_i^{(s)}\Big|_{z = d_1} = L_i^{(d)}\Big|_{z = d_1}, \tag{10}$$

where $L_i^{(s)}$ and $L_i^{(d)}$ ($i = x, y$) are the field components in the semiconductor and insulator, respectively, that are parallel to the boundary of the layers, and (ii) four periodicity conditions (Floquet theorem) at the extremities of the period:

$$L_i^{(s)}\Big|_{z = 0}\exp(-i\bar{k}d) = L_i^{(d)}\Big|_{z = d}. \tag{11}$$

The resulting system of eight equations is linear and homogeneous, so that its coefficients can be expressed through any of them. Thus, the problem of determining the transverse field components in the layers is reduced to solving the system of seven linear inhomogeneous equations.

BAND STRUCTURE OF THE SPECTRUM AND FIELD DISTRIBUTION

We numerically analyzed dispersion relations (2) and (3) for two sets of parameters where the characteristic frequencies of the semiconductor met the inequalities $\omega_H < \omega_p$ ($B_0 = 0.05T$) and $\omega_H > \omega_p$ ($B_0 = 0.23T$), respectively. The semiconductor was taken to be n-InSb with an electron concentration of 3×10^{14} cm^{-3}, $\varepsilon_1 = 17.8$, and $d_1 = 0.01$ cm; for the insulator, $\varepsilon_2 = 2.0$ and $d_2 = 0.03$ cm.

A part of the band spectrum of cyclotron waves for the first case is shown in Fig. 1. In this domain, the expressions for k_{z1} and k_{z2} in the semiconductor take the form

$$k_{z1} \approx \pm\sqrt{\left|\frac{\varepsilon_1}{\varepsilon_3}\right|}k_{xy}, \tag{12}$$

$$k_{z2} \approx \pm ik_{xy}, \tag{13}$$

when $\omega \le \omega_H$ and $k_{xy} \gg k_0$. That is, with an increase in $k_{xy}d$, the transverse wavenumber is imaginary for one of the partial modes in the semiconductor and real for the other. This feature also specifies the type of these modes: one is of the surface type (the field exponentially drops from the interface toward the bulk of the layer), while the other is a bulk (waveguide) mode. For the latter, the condition for dimensional resonance is met in the transmission bands: $k_{z1}d_1 = p\pi$, where p is an

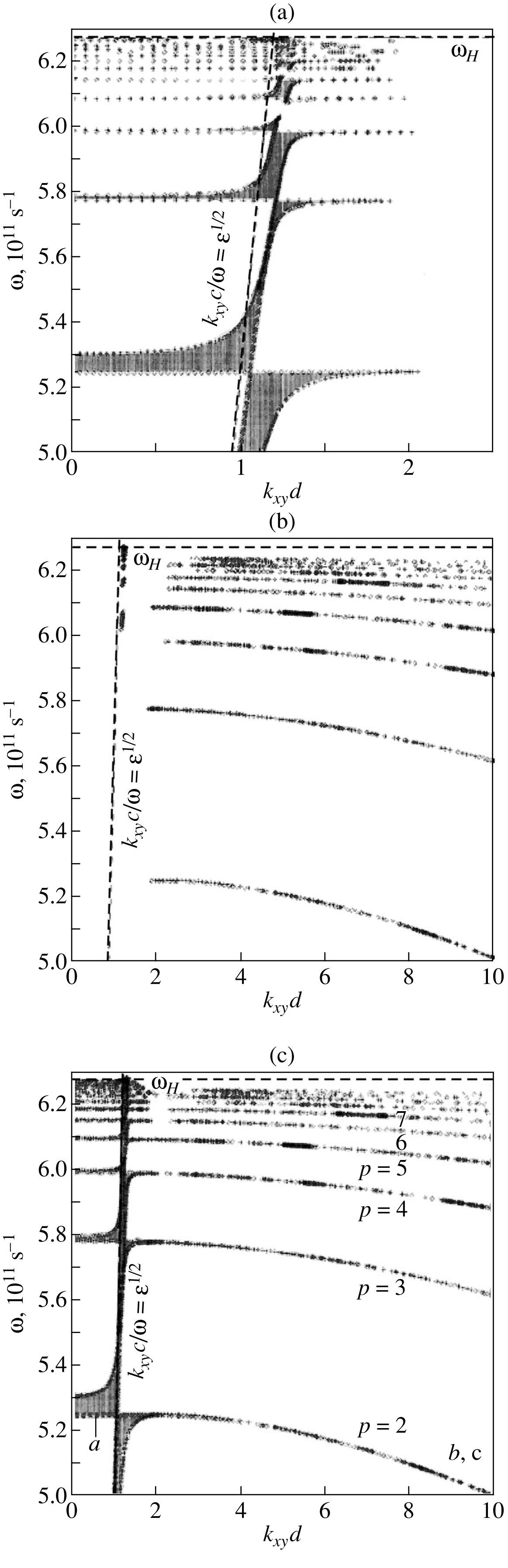


Fig. 1. Band spectra of cyclotron waves for Bloch wavenumbers (a) $\bar{k}_1$ and (b) $\bar{k}_2$ and (c) the resulting spectrum. $\omega_H < \omega_p$. The transmission bands are shaded. The order of half-wave resonance in the semiconductor layer is $p = 2, 3, 4, \ldots$.

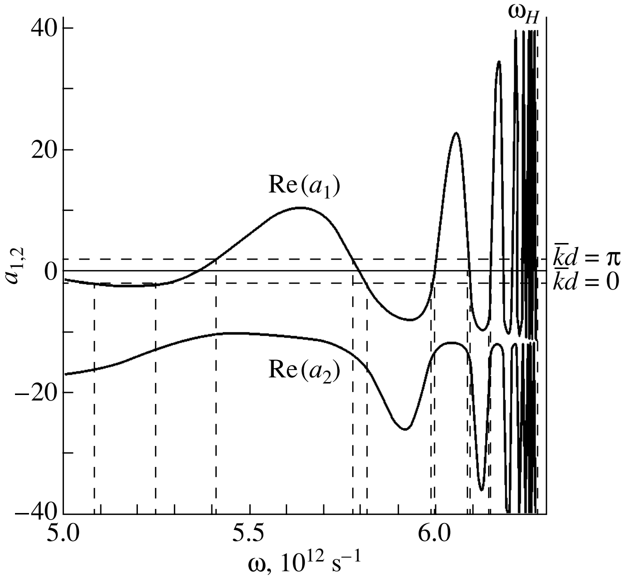


Fig. 2. Graphic solution of dispersion relations (1) and (2) at $k_{xy}d = 1$ for $\omega_H < \omega_p$. The vertical dashed lines indicate the boundaries of the transmission and opacity bands.

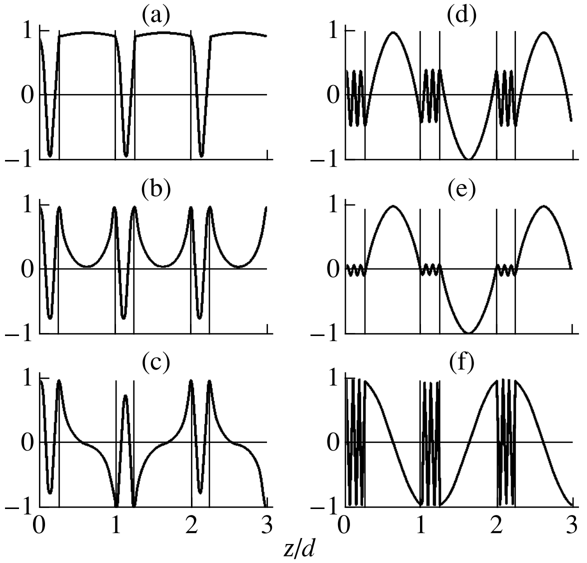


Fig. 3. Distribution of the E_x component real part across the layers over three periods of the structure. (a–c) The fields at the respective points of the spectrum in Fig. 1c and (d–f) the same for the spectrum in Fig. 5. The fields are normalized to the amplitude values.

integer showing the number of half-waves that are accommodated across the semiconductor layer.

Figure 2 shows the band spectrum in the form of alternating transmission and opacity bands. The spectrum is seen to thicken as the bands approach the cyclotron frequency. The plot represents a graphic solution of dispersion relations (2) and (3) and defines the position of the transmission bands corresponding to the range $-2 \le \mathrm{Re}(a_{1,2}) \le 2$ at $\mathrm{Im}(a_{1,2}) = 0$ in formula (4). It is obvious that this part of the spectrum ($k_{xy}d = 1$ in Fig. 1a) involves only solutions to Eq. (2), since the values of Re(a_2) in Eq. (3) fall beyond the given interval and lie below the transmission band with the boundaries $\bar{k}d = 0$ and $\bar{k}d = \pi$ in Fig. 2. In contrast, the formation of the spectrum in the interval $k_{xy}d > 2$ (Fig. 1b) is related to only solutions to Eq. (3). This example clearly illustrates that the dispersion curves for $\bar{k}_1$ and $\bar{k}_2$ complement the corresponding spectra to the resulting energy spectrum of the periodic structure (Fig. 1c) if the characteristic frequencies are appropriately selected.

It is noteworthy that the position of cyclotron waves in the resulting band spectrum of the periodic structure [1] is such that the phase velocity of the waves in the transmission bands is the least, as compared with other spectral regions. As a result, the transmission bands may extend far into the high-k_{xy} range, since one of the transverse wavenumbers in the semiconductor remains real with an increase in $k_{xy}d$. However, taking into account dissipative losses in the semiconductor, which limit the maximal wavenumber k_{xy} and, accordingly, the minimal phase velocity, may change the transmission band. Therefore, it is important to estimate the effect of dissipation on the band spectrum of cyclotron waves.

The field distribution across the layers over three periods of the structure in different regions of the spectrum is shown in Fig. 3. The distribution represents an interference pattern produced by the superposition of the partial modes and their multiple reflections from the boundaries of the structure. For example, point *a* of the spectrum in Fig. 1c meets the condition of dimensional resonance: two half-waves are accommodated across the semiconductor layer, which is embodied in the field distribution shown in Fig. 3a. The wave in the dielectric layer is also of the bulk type (k_z is real). The field distribution is symmetric about the center of the layer. Since point *a* lies at the boundary $\bar{k}_1 d = 0$, the field pattern repeats each alternate period of the structure.

At points *b* and *c*, the wave in the dielectric layer is of the surface type (k_z is imaginary) and the field exponentially drops inward to the layer. The pattern repeats each alternate period of the structure at the boundary $\bar{k}_2 d = 0$ (Fig. 3b) and in every two periods at the boundary $\bar{k}_2 d = \pi$ (Fig. 3c).

The relationship $\omega_H > \omega_p$ between the characteristic frequencies of the semiconductor differs from the previous one in that both transverse wavenumbers $k_{z1,2}$ in the semiconductor layers are real. Part of the related band spectrum ($k_{xy}d = 2.6$) is illustrated in Fig. 4.

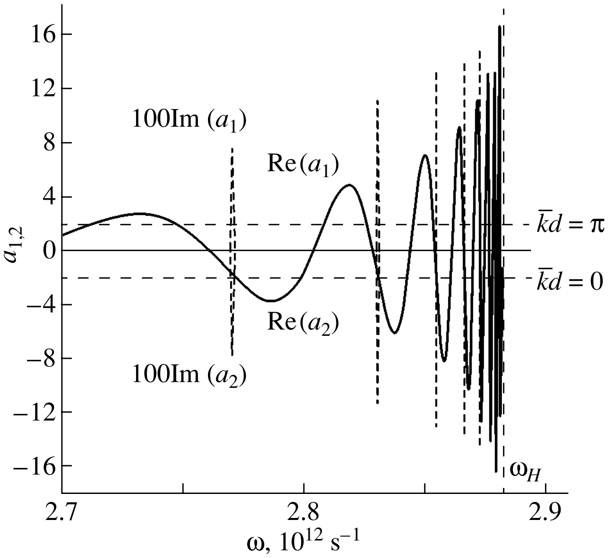


Fig. 4. Graphic solution of dispersion relations (1) and (2) at $k_{xy}d = 2.6$. $\omega_H > \omega_p$.

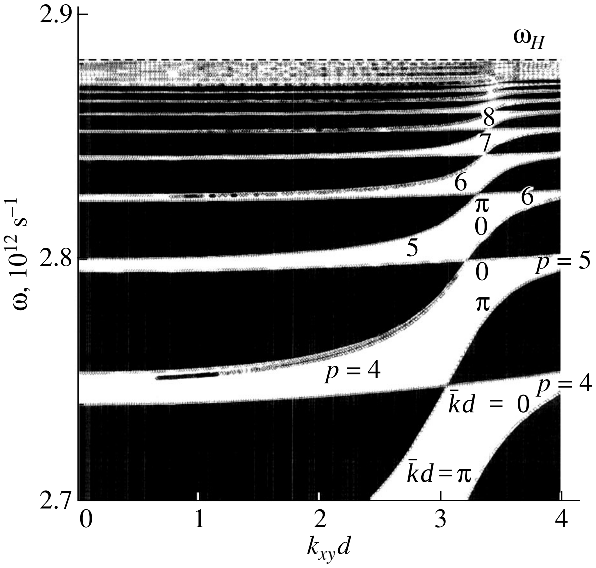


Fig. 5. Band spectrum of cyclotron waves that corresponds to Bloch wavenumber $\bar{k}_1$. $\omega_H > \omega_p$. The opacity bands (shown dark) are bounded by the curves corresponding to the same values of $\bar{k}\,d$.

It is seen that the solutions to both dispersion relations ((2) and (3)) fall into the interval corresponding to the transmission bands. This spectrum is difficult to analyze, since the dependences obtained for a_1 and a_2 outline the bands in the entire spectrum rather than separate the bands for the two normal waves. This stems from the fact that the values of a_1 and a_2 calculated by formula (4) are always taken as positive and negative, respectively. Therefore, the upper sections of the curves in Fig. 4 (fine lines) formally refer to one spectrum and the lower sections to the other.

However, if one takes into consideration that $|k_{z1}| \gg |k_{z2}|$ and k_{z2} depends on frequency only slightly in this spectral range, it becomes evident that the dependence mentioned above must be a rapidly oscillating function for the first normal mode and a nearly linear function for the second. Such behavior is due to the trigonometric functions of arguments $k_{z1}d_1$ and $k_{z2}d_1$ that enter into the dispersion relations. As a result, if k_{z1} is large, even its small changes give rise to numerous transmission and opacity bands for the first normal mode. For the second normal mode, k_{z2} and its variation are small, so that the dispersion relation remains within the transmission band. It is such an approach (unlike the formal one) to selecting the sections of the curves shown in Fig. 4 that provides a correspondence to the band spectra of the two normal modes characterized by Bloch wavenumbers k_1 and k_2 in the periodic structure.

From the mathematical point of view, the entire spectrum can be resolved into normal mode spectra by changing the sign of the radical in (4) each time the dependence passes through the region where the determinant turns to zero, i. e., where $a_1 = a_2$.

Another feature of this spectrum is the presence of the regions in Fig. 4 that are associated with the complex conjugate values of a_1 and a_2 (see (4)) . These regions correlate with some of the points where the dependences $\mathrm{Re}(a_{1,2}) = f(\omega)$ meet and appear as successive peaks that are symmetric about the abscissa axis. In Fig. 4, they are shown by the dashed lines and correspond to the values $100\mathrm{Im}(a_{1,2})$. Hence, one may expect that extra (relatively narrow and similarly located) opacity bands will appear in the band spectra of both normal modes.

Thus, the spectrum of cyclotron waves in the periodic structure can be subdivided into two spectra of normal modes with Bloch wavenumbers $\bar{k}_1$ and $\bar{k}_2$ (see Figs. 5 and 6, respectively).

The spectrum corresponding to $\bar{k}_1$ consists of alternating transmission and opacity bands, whose density along the frequency axis increases as the cyclotron resonance frequency is approached. The condition of half-wave dimensional resonance across the semiconductor layer is met in each of the transmission bands. The order of resonance grows with frequency, differing by unity in each successive band. Within the transmission band, the order of resonance varies in the same way because of its bend and the related frequency shift at its edges.

The above features of the band spectrum are also reflected in the field distributions across the layers (Figs. 3d–3f). These distributions correspond to three spectrum points in Fig. 5 lying within the same transmission band. In this case, the order of dimensional res-

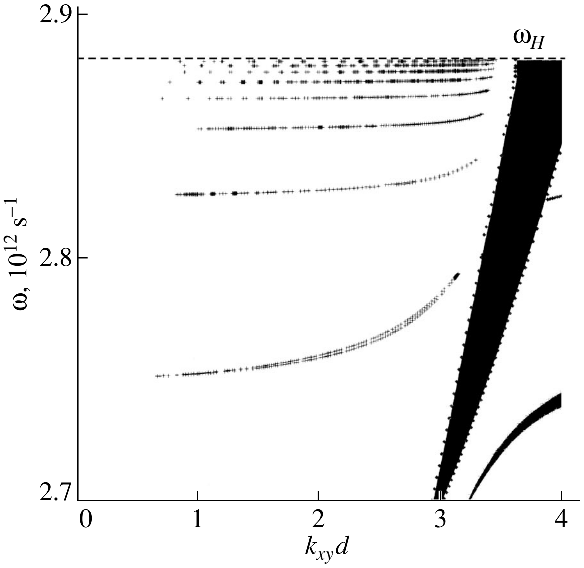


Fig. 6. Band spectrum of cyclotron waves corresponding to Bloch wavenumber $\bar{k}_2$. $\omega_H > \omega_p$. The opacity bands are dark.

onance in the semiconductor changes from $p = 5$ (Fig. 3d) to $p = 6$ (Fig. 3f). At the intermediate point (Fig. 3e), 5.25 half-waves are accommodated across the layer and the field oscillation amplitude is minimal. Since all the three points are chosen at the boundary $\bar{k}d = \pi$, the field pattern repeats in every two periods of the structure.

The spectrum of the normal mode with $\bar{k}_2$ (Fig. 6) is essentially a wide transmission region including a sequence of opacity bands. This sequence involves the bands associated with the complex conjugate solutions in (4) and a broadening band corresponding to the real solutions with $|a_2| > 2$. Since $k_{z2}d_1 \approx \pi$ for this normal mode, the condition of half-wave resonance in the semiconductor layer is met throughout in the transmission region. Note that the field distribution shown in Fig. 3 is a superposition of the fields of the normal modes and depends, in particular, on the relationship between their amplitudes.

Thus, if the characteristic frequencies of the semiconductor meet the second inequality, the dispersion relations for $\bar{k}_1$ and $\bar{k}_2$, as well as the corresponding spectra, overlap.

CONCLUSIONS

We analyzed the band spectrum of cyclotron waves propagating in a periodic layered semiconductor–dielectric structure at an angle to an external magnetic field that is applied perpendicularly to the layers.

The spectra for two relationships between the characteristic frequencies of the semiconductor, $\omega_H < \omega_p$ and $\omega_H > \omega_p$, were studied. The field distributions across the layers and over the period of the structure were calculated. It was established that the transmission bands in both spectra form when the condition for dimensional resonance in the semiconductor layer is met. Such a structure of the transmission bands may find use, e.g., in magnetically tuned multichannel filters.

By graphically solving the dispersion relation, we demonstrated that the spectrum of cyclotron waves can be subdivided into two spectra of normal modes that correspond to the Bloch wavenumbers in the periodic structure. The cases where the transmission bands complement each other or overlap were considered.

REFERENCES

1. A. A. Bulgakov and V. K. Kononenko, Zh. Tekh. Fiz. **73** (11), 15 (2003) [Tech. Phys. **48**, 1372 (2003)].
2. A. C. Baynham and A. D. Boardman, J. Phys. C **2**, 619 (1969).
3. A. A. Bulgakov and O. V. Shramkova, Radiotekh. Élektron. (Moscow) **46**, 236 (2001).
4. J. J. Brion, R. F. Wallis, A. Hardstein, and E. Burstein, Phys. Rev. Lett. **28**, 1455 (1972).
5. R. F. Wallis, J. J. Brion, E. Burstein, and A. Hartstein, Phys. Rev. B **9**, 3424 (1974).
6. V. L. Ginzburg, *The Propagation of Electromagnetic Waves in Plasmas,* 2nd ed. (Nauka, Moscow, 1967; Pergamon, Oxford, 1970).
7. F. G. Bass, A. A. Bulgakov, and A. P. Tetervov, *High-Frequency Properties of Semiconductors with Superlattices* (Nauka, Moscow, 1989) [in Russian].
8. A. G. Gurevich, *Microwave Ferrites* (Fizmatgiz, Moscow, 1960) [in Russian].
9. A. I. Akhiezer, I. A. Akhiezer, and R. V. Polovin, *Plasma Electrodynamics,* Ed. by A. I. Akhiezer (Nauka, Moscow, 1974; Pergamon, Oxford, 1975).

Translated by A. Sidorova

Technical Physics, Vol. 49, No. 10, 2004, pp. 1319–1324. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 75–79.
Original Russian Text Copyright © 2004 by Primak.

OPTICS, QUANTUM ELECTRONICS

Sensitivity of Prismatic Integrated Optical Sensors under Nearly Critical Conditions

I. U. Primak

Institute of Applied Optics, National Academy of Sciences of Belarus, Mogilev, 212793 Belarus
e-mail: ipo@physics.belpak.mogilev.by

Received June 17, 2003

Abstract—The sensitivity of integrated optical sensors based on prismatic excitation of planar waveguides is analyzed. It is shown that the sensitivity peak is attained when the waveguide modes are excited under nearly critical conditions. *© 2004 MAIK "Nauka/Interperiodica".*

INTRODUCTION

A considerable number of integrated optical sensors for parameters of medium have been developed [1–6]; however, many aspects of the optimization of such devices remain unclear. This study is devoted to analysis of sensors in which the power of a light beam reflected from a prismatic device for exciting planar optical waveguides is the quantity being detected [2–4]. The possibility for optimizing such sensors was considered in [4], where the mode excitation under far-from-critical conditions was analyzed. However, the results of investigation of waveguide modes for thicknesses close to critical [5, 6], on the one hand, and the reflection of beams from the medium being studied in the vicinity of critical angles, on the other hand [7, 8], lead to the conclusion that the sensitivity of the sensors might have a peak in the vicinity of the critical conditions. We will derive an analytic expression for the sensitivity of sensors and will solve the problem on its maximization at nearly critical conditions. It will be shown that a transition to these conditions makes it possible to improve the sensitivity of sensors by several orders of magnitude.

INTENSITY OF THE REFLECTED BEAM

To study the sensitivity of sensors, we will derive an analytic expression for the coefficient of beam energy reflection R from the base of a prismatic coupling device. A schematic diagram for the devices studied here is shown in Fig. 1. It includes a prism with a film structure at the base, consisting of a buffer layer and a waveguide with thicknesses g and d. The waveguide is in contact with the medium being studied. The prism, the buffer layer, the waveguide, and the medium have complex permittivities $\varepsilon_p = \varepsilon'_p$, $\varepsilon_g = \varepsilon'_g + i\varepsilon''_g$, $\varepsilon_w = \varepsilon'_w + i\varepsilon''_w$, and $\varepsilon_c = \varepsilon'_c + i\varepsilon''_c$ (here and below, complex quantities are written in the form $Z = Z' + iZ''$, where $Z' = \mathrm{Re}Z$ and $Z'' = \mathrm{Im}Z$). We will assume in the further analysis that $\varepsilon'_g < \varepsilon'_c$. This condition, which is most interesting for practical applications [5, 6], can be satisfied by using a metallic buffer layer. The waveguide is excited by a coherent light beam with TE or TM polarization. The power of the reflected beam is detected by a photodetector. We will assume that the field of the exciting beam depends on time as $\exp(i\omega t)$ and the field distribution at the prism base is described by the law $\psi = \psi_0(xw_0^{-1}, zw^{-1}\sin\alpha)\exp(-i\beta_0 z)$, where the quantity ψ has the meaning of the electric field component E_x for TE polarized waves or the magnetic field component H_x in the case of TM polarization; w_0 is the beam radius in the medium surrounding the prism; $w = w_0 N$, where N is the coefficient accounting for the beam refraction at the lateral face of the prism; α is the angle between the beam axis and the prism base; and $\beta_0 = k_0\sqrt{\varepsilon_p}\cos\alpha$, $k_0 = 2\pi\lambda_0^{-1}$ being the wave number of vacuum.

According to [9], the energy reflection coefficient for the prism base can be expressed in terms of the reflection coefficient $r(k_x, \beta)$ for a plane wave and the Fourier transform $a(k_x, \beta)$ of the incident beam via the

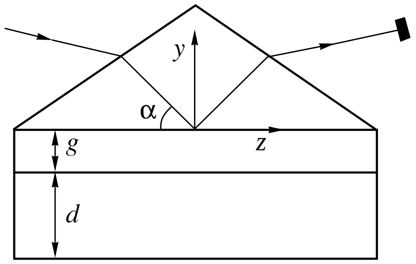


Fig. 1. Schematic diagram of an prismatic integrated optical sensor.

1063-7842/04/4910-1319$26.00 © 2004 MAIK "Nauka/Interperiodica"

relation

$$R = A\int\int_{-\infty}^{\infty} dk_x d\beta |ra|^2 \sqrt{k_0^2\varepsilon_P - \beta^2}$$
$$\times\left(\int\int_{-\infty}^{\infty} dk_x d\beta |a|^2 \sqrt{k_0^2\varepsilon_P - \beta^2}\right)^{-1}, \tag{1}$$

where A is the coefficient taking into account the beam reflection from the lateral faces of the prism and

$$a = \int\int_{-\infty}^{\infty} d\xi d\zeta \psi_0(\xi, \zeta)$$
$$\times \exp[ik_x\xi w_0 + i(\beta - \beta_0)\zeta w/\sin\alpha],$$
$$\zeta = xw_0^{-1}, \quad \zeta = zw^{-1}\sin\alpha.$$

A rigorous expression for function $r(k_x, \beta)$, which is applicable in the case of an arbitrary layered structure, is given in [10]. However, this expression can hardly be used to analyze the sensor sensitivity in view of its complexity. An approximate analytic expression for function $r(k_x, \beta)$ proposed in [10] was derived for conditions of resonant mode excitation. However, this expression becomes inapplicable when the waveguide mode is excited under nearly critical conditions, i.e., when the real part h' of the mode propagation constant tends to $k_0\sqrt{\varepsilon_c'}$. For this reason, we consider another approximation, which is applicable both in the vicinity of and far from critical conditions. This approximation is given by

$$r = (1-\delta)(1+\delta)^{-1} - 4\delta\Delta\bar{k}_{yc}[(1-\delta)^2(\nu - \bar{k}_{yc})]^{-1}, \tag{2}$$

where

$$\nu(k_x, \beta) = \sqrt{k_0^2\varepsilon_c - k_x^2 - \beta^2} \quad (\nu' \geq 0),$$
$$\Delta\bar{k}_{yc} = \bar{k}_{yc} - k_{yc}$$
$$= ik_{yg}\left(\frac{1-\delta}{1+\delta}\right)\frac{\exp(-2ik_{yg}g)A_0(-g+0)}{k_{yc}\int_{-\infty}^{\infty} A_0 dy}, \tag{3}$$

$$A_0(y) = Y^2(y)\varepsilon^{-T}, \quad \delta = \frac{k_{yg}}{k_{yp}}\left(\frac{\varepsilon_p}{\varepsilon_g}\right)^T,$$
$$k_{yp,\,g} = \sqrt{k_0^2(\varepsilon_{p,\,g} - \varepsilon_c) + k_{yc}^2} \quad (k_{yg}'' < 0),$$

k_{yc} and $Y(y)$ are the eigenvalue and the eigenfunction of the spectral problem

$$\varepsilon^T\frac{\partial}{\partial y}\left[\varepsilon^{-T}\frac{\partial}{\partial y}Y\right] + [k_0^2(\varepsilon - \varepsilon_c) + k_{yc}^2]Y = 0, \tag{4}$$

and $\bar{k}_{yc}$ is the transverse constant of propagation of the leaky mode of the waveguide structure loaded by the prism in the medium studied; $T = 0$ for TE waves and $T = 1$ for TM waves; the complex function $\varepsilon = \varepsilon(y)$ describes the permittivity of the waveguide structure without a prism and assumes values of ε_g for $y > -g$, ε_w for $-g - d < y < -g$, and ε_c for $y < -g - d$.

Approximation (2) was obtained according to the scheme analogous to that described in [10], but using the Taylor expansion of function $\Phi(\nu)$ in the vicinity of point $\nu = k_{yc}$, where $\Phi(k_{yc}) = 0$ is the dispersion equation for the waveguide mode being excited. The passage from variable β in the above-mentioned expansion [10] to variable ν allows us to remove the problem associated with the closeness of the branching point of function $\nu(0, \beta) = \sqrt{k_0^2\varepsilon_c - \beta^2}$ to the root of the dispersion equation under nearly critical conditions and, hence, to radically improve the accuracy of the approximation.

It should be noted that, applying the perturbation method [11] to Eq. (4), we can derive a relation important for the further analysis,

$$k_{yc} = \eta - \left[k_0^2\int_{-\infty}^{\infty}(\varepsilon_c''A_1 - \varepsilon''A_2)dy\right]\left(2i\eta\int_{-\infty}^{\infty}A_1 dy\right)^{-1}, \tag{5}$$

where $A_1 = Y_0^2\varepsilon'^{-T}$, $A_2 = \varepsilon'^{-T}(A_1[\varepsilon_c'^{T} - T(\eta k_0^{-1})^2] + T(k_0^2\varepsilon')^{-1}[\nabla_y Y_0]^2)$; η and $Y_0(y)$ are the eigenvalue and the eigenfunction of problem (4) for $\varepsilon_g'' = 0$, $\varepsilon_w'' = 0$, and $\varepsilon_c'' = 0$.

SENSOR SENSITIVITY

To apply the results obtained in the previous section to analysis of the sensitivity of the sensors studied here, we take into account the fact that these devices register the increment ΔR in the reflection coefficient due to a change in the parameters of the medium being studied. Since these changes are manifested in the permittivity increment $\Delta\varepsilon_c = \Delta\varepsilon_c' + i\Delta\varepsilon_c''$, we have

$$\Delta R = R(\varepsilon_c + \Delta\varepsilon_c) - R(\varepsilon_c); \tag{6}$$

in this case, sensitivity S is defined as the derivative

$$S = \partial R/\partial p, \tag{7}$$

where p is a parameter characterizing the action inducing the change in the medium permittivity (e.g., p may have the meaning of the impurity concentration in the medium) [3, 4].

To derive an explicit expression for S, we use the familiar fact that variations of quantities δ, k_{yp}, and Δk_{yc} associated with increment $\Delta\varepsilon_c$ are negligibly small as compared to variations of parameters ν and k_{yc} [4]. Suppose that the sensor is excited by a light beam that corresponds to narrow effective ranges of variation of the wave vector components of the plane waves constituting the beam: $\Delta k_x = |k_x| \sim w_0^{-1}$ and $\Delta\beta = |\beta - \beta_0| \sim w^{-1}\sin\alpha$, where $(k_0 w_0)^{-1} \ll 1$. We also assume that the beam is focused on the coupling prism base; i.e., $\psi_0^* = \psi_0 \exp(i\varphi)$, where the asterisk indicates complex conjugation and φ is a certain real constant. Using relations (1), (2), (6), and (7), we derive the expressions accurate to quantities on the order of $(\Delta k_x/k_0)^4$ and $(\Delta\beta/k_0)^4$,

$$S = A\left|(1-\delta)(1+\delta)^{-1}\right|^2 K|L|F, \tag{8}$$

where

$$K = (k'_{yc} - \nu'_0)^{-1}\frac{\partial\varepsilon''_c}{\partial p}, \quad L = \frac{k_0^2}{2\nu_0} - \frac{\partial k_{yc}}{\partial\varepsilon_c},$$

$$\begin{aligned} F = {} & 4u(1+u)(1+v^2)^{-2}[D_1(v^2 - 2u - 1) \\ & - 2D_2 v(1+u) + GH_1^2 + BH_1H_2 + CH_2^2], \end{aligned} \tag{9}$$

$$\nu_0 = \nu(0, \beta_0),$$

$$\frac{\partial k_{yc}}{\partial\varepsilon_c} = \left[k_0^2 \int_{-d-g}^{\infty} A_0 dy - iTk_{yc}\varepsilon_c^{-1}A_0(-d-g-0)\right] \times \left[2k_{yc}\int_{-\infty}^{\infty} A_0 dy\right]^{-1},$$

$$H_1 = \frac{\beta_0 \sin\alpha}{wp_1|\nu_0|}, \quad H_2 = \frac{k_0 \sin\alpha}{w|\nu_0|^2},$$

$$D_1 = \frac{1 - DL'(L'')^{-1}}{\sqrt{1 + [L'(L'')^{-1}]^2}}, \quad D_2 = \frac{D + L'(L'')^{-1}}{\sqrt{1 + [L'(L'')^{-1}]^2}},$$

$$D = \frac{\partial\varepsilon_c'}{\partial p}\left(\frac{\partial\varepsilon''_c}{\partial p}\right)^{-1},$$

$$v = p_1^{-1}(\nu''_0 - \bar{k}''_{yc}), \quad u = p_1^{-1}\mathrm{Re}[2\delta\Delta\bar{k}_{yc}(1-\delta)^{-2}],$$

$$p_1 = \bar{k}'_{yc} - \nu'_0,$$

$$\begin{aligned} G = {} & -P_z(1+v^2)^{-1}\mathrm{Im}\{4iu(D_1 + vD_2) \\ & \times [1 + (\nu_0^*|\nu_0|^{-1})^2] + (\nu_0^*|\nu_0|^{-1})^2(1+v^2)^{-1} \\ & \times [(D_2 + iD_1)(3v^4 + 3 - 18v^2 + 4u - 20uv^2) \\ & + 4v(1-v^2)[3(1+u)(iD_2 - D_1) - 2uD_1]\}, \\ B = {} & \mathrm{Im}\{(1+v^2)^{-1} \\ & \times [(1+u)(1 - 3v^2[\Phi_1 D_2 + \Phi_2 D + i(\Phi_1 D_1 + \Phi_2)] \\ & - [\Phi_1 D_1 + \Phi_2 - i(\Phi_1 D_2 + \Phi_2 D)](3 + 4u - v^2)] \end{aligned} \tag{10}$$

$$+ 2iu\Phi_1 D_1 + u(i + D)\Phi_2(1 + (|\nu_0|(\nu_0^*)^{-1})^2)\},$$

$$C = \mathrm{Im}\{\Phi_3[(1 - iD)(v^2 - 2u - 1) - 2v(1+u)(i + D)]\},$$

$$\Phi_1 = k_0\nu_0^*(\beta_0|\nu_0|)^{-1}(P_z\varepsilon'_c(\nu_0^*|\nu_0|^{-1})^2 + P_x|\nu_0 k_0^{-1}|^2),$$

$$\Phi_2 = -ik_0\beta_0|\nu_0 L|^{-1}(\nu_0^*|\nu_0|^{-1})^4 P_z,$$

$$\Phi_3 = 0.5k_0^2\nu_0^*|\nu_0^2 L|^{-1} \times (1.5\varepsilon'_c(\nu_0^*|\nu_0|^{-1})^4 P_z + (0.5P_x - P_z)(\nu_0^* k_0^{-1})^2),$$

$$P_x = \int\int_{-\infty}^{\infty}\left|\frac{\partial\psi_0}{\partial\xi}\right|^2 d\xi d\zeta\left((\sin\alpha/N)^2\int\int_{-\infty}^{\infty}|\psi_0|^2 d\xi d\zeta\right)^{-1},$$

$$P_z = \int\int_{-\infty}^{\infty}\left|\frac{\partial\psi_0}{\partial\zeta}\right|^2 d\xi d\zeta\left(\int\int_{-\infty}^{\infty}|\psi_0|^2 d\xi d\zeta\right)^{-1}.$$

In deriving expression (9), we omitted the terms proportional to $p_1^{-1}\,\mathrm{Im}[2\delta\Delta\bar{k}_{yc}(1-\delta)^{-2}]$, since weakly attenuating modes, which are of the main interest, satisfy the inequality $\mathrm{Im}|k_{yc}\int_{-\infty}^{\infty} A_0\,dy| \gg \mathrm{Re}|k_{yc}\int_{-\infty}^{\infty} A_0\,dy|$ [10]; consequently, $|p_1^{-1}\,\mathrm{Im}[2\delta\Delta\bar{k}_{yc}(1-\delta)^{-2}]| \ll |u|$.

The problem of determining the maxima of function $|S(w, \alpha, g, d)|$ for given permittivities ε_p, ε_g, ε_w, and ε_c is of practical importance. To solve this problem, we take into account the fact that the maximal sensitivity should be expected under nearly critical conditions, i.e., for $d \longrightarrow d_c$, where d_c is the critical thickness of the waveguide. In this case, coefficients L and K have the form $L = 0.5k_0^2\nu_0^{-1}$ and $K = (k'_{yc})^{-1}\partial\varepsilon''_c/\partial p$. Analysis of formula (5) for $d \longrightarrow d_c$ in the case when the mode attenuation in the waveguide structure is determined by absorption in the buffer layer, which means that the inequality [11]

$$\left|\int_{-d-g}^{-g}(\varepsilon''_c A_1 - \varepsilon''_w A_2)dy\right| \ll \left|\varepsilon''_g\int_{-g}^{\infty} A_2 dy\right| \tag{11}$$

is satisfied, shows that we can disregard the dependence of k'_{yc} on d. This allows us to reduce the search for the maxima of function $|S(w, \alpha, g, d)|$ to the solution of the system of equations $\partial|L|/\partial\nu''_0 = 0$, $\partial F/\partial H_1 = 0$, $\partial F/\partial H_2 = 0$, $\partial F/\partial u = 0$, and $\partial F/\partial v = 0$ in parameters ν''_0, H_1, H_2, u, and v. This system has the analytic solution $\nu''_0 = \nu''_{0\mathrm{opt}}$, $H_1 = H_{1\,\mathrm{opt}}$, $H_2 = H_{2\,\mathrm{opt}}$, $u = u_{\mathrm{opt}}^{(s)}$, $v = v_{\mathrm{opt}}^{(s)}$, and $F = F_{\mathrm{opt}}$, where

$$\nu''_{0\mathrm{opt}} = -k_0\sqrt{0.5|\varepsilon''_c|}, \tag{12}$$

$$H_{1\,\mathrm{opt}} = H_{2\,\mathrm{opt}} = 0, \tag{13}$$

$$F_{\text{opt}} = \pm\frac{2}{3}\sqrt{\frac{1+D^2}{3}}, \quad u_{\text{opt}}^{(s)} = -\frac{3F_{\text{opt}}}{9F_{\text{opt}}+2D_1}, \quad v_{\text{opt}}^{(s)} = \frac{2D_2}{9F_{\text{opt}}+2D_1}, \quad s = \operatorname{sgn}(F_{\text{opt}}). \tag{14}$$

It follows from relations (12), as well as relations

$$\beta_0 = k_0\sqrt{\varepsilon_c - [0.5k_0\varepsilon_c''(\nu_{0\text{opt}}'')^{-1} + i\nu_{0\text{opt}}''k_0^{-1}]^2} = k_0\sqrt{\varepsilon_c'},$$

$$\alpha = \arccos[\beta_0/(k_0^2\varepsilon_p)^{0.5}],$$

that the optimal angle of incidence of the exciting beam is equal to the critical value $\alpha_k = \arccos[(\varepsilon_c'/\varepsilon_p')^{0.5}]$. Conditions (13) are satisfied in the limit $w \longrightarrow \infty$. This means that the maximum of quantity $|S|$ takes place when waveguides are excited by a plane wave. Expressions (3), (10), and (14) lead to the existence of two sets of optimal thicknesses of the waveguide and buffer layers: $d = d_{\text{opt}}^{(s)}$ and $g = g_{\text{opt}}^{(s)}$. Using expressions (10) and (14), we can write the equation for determining $d_{\text{opt}}^{(s)}$,

$$k_{yc}'' - \nu_{0\text{opt}}'' + (k_{yc}' - 0.5k_0^2\varepsilon_c''(\nu_{0\text{opt}}'')^{-1}) \times [v_{\text{opt}}^{(s)} + u_{\text{opt}}^{(s)}](1 + u_{\text{opt}}^{(s)})^{-1} = 0. \tag{15}$$

Here, $k_{yc}' = k_{yc}'(d_{\text{opt}}^{(s)})$, $k_{yc}'' = k_{yc}''(d_{\text{opt}}^{(s)})$, and q = Im$[0.5\delta^{-1}(1 + \delta^2)]$. Having determined $d_{\text{opt}}^{(s)}$, we can find the value of $g_{\text{opt}}^{(s)}$ using an expression derived from Eqs. (3), (10), and (14),

$$g_{\text{opt}}^{(s)} = 0.5|k_{yg}|^{-1} \times \ln\left|\frac{0.5(1-\delta)^2 k_{yc}u^{(s)}(k_{yc}' - 0.5k_0^2\varepsilon_c''(\nu_{0\text{opt}}'')^{-1})\int_{-\infty}^{\infty} A_0\,dy}{k_{yg}\delta(1+u^{(s)})A_0(-g+0)}\right|. \tag{16}$$

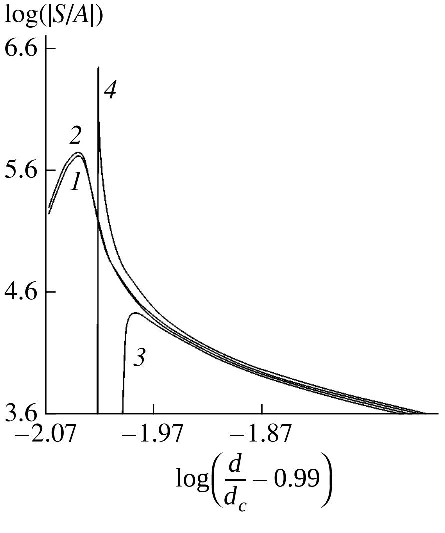


Fig. 2. Dependence of the sensor sensitivity on the waveguide thickness.

It should be noted that, in the absence of absorption in the waveguide structure ($\varepsilon_g'' = 0$, $\varepsilon_w'' = 0$, $\varepsilon_c'' = 0$), the root of Eq. (15) is the critical waveguide width d_c for the mode being excited. In this case, the sensitivity is independent of the number m ($m = 0, 1, \ldots$) of the given mode. This is due to the fact that the dependence of quantity K on d can be neglected (see above) under conditions (11) and for $d \longrightarrow d_c$. An exception from this rule exists for the principal mode of the TM polarization, which degenerates to the plasmon mode of the interface $y = -g$ for $d \longrightarrow 0$ in the case of a metallic buffer layer.

To test approximation (8) and the validity of the assumption concerning the maximization of sensitivity in the vicinity of critical conditions, calculations were made for the sensitivity of a specific sensor as a function of the waveguide thickness. The results of these calculations are shown in Fig. 2 in the case of variations of absorption of the aqueous medium ($\Delta\varepsilon_c' = 0$, $\Delta\varepsilon_c'' \neq 0$, $D = 0$) for $\varepsilon_c = 1.774224 - i3.73 \times 10^{-8}$, $\varepsilon_w = 2.295225 - i3.03 \times 10^{-5}$, $\varepsilon_g = -18 - i0.47004$, and $\varepsilon_p = 3.055154$. We consider the TE-polarization mode ($m = 1$). Exact calculations of sensitivity

$$S = \frac{\partial R}{\partial \varepsilon_c''} \tag{17}$$

(the results of these calculations are represented by curve *1* in Fig. 2) were performed using numerical differentiation of function $R(\varepsilon_c)$, where the value of $R = |r|^2$ was calculated using recurrence relations [12]. In these calculations, we used the values of thicknesses d and g determined after the numerical solution of Eq. (15) and on the basis of expression (16), respectively. For a given value of angle of incidence α, we assumed that $\nu_{0\text{opt}}'' = \nu_0''$, $d = d_{\text{opt}}^{(1)}$, and $g = g_{\text{opt}}^{(1)}$ used the values of $u_{\text{opt}}^{(1)}$ and $v_{\text{opt}}^{(1)}$ determined from Eq. (14) for $s = 1$. The values of quantities k_{yc} were calculated by solving Eq. (4) by the method of contour integration [12]. As a result of successive variation of angle α in the vicinity of critical value α_k and application of the above-described procedures for determining the values of d, g, and S, we plotted the dependence $S(d)$, where $d = d(\alpha)$. Using the obtained dependence $d = d(\alpha)$ and expressions (8) and (14), we calculated approximately the values of S (the results of these calculations are represented by curve *2* in Fig. 2). These results confirm the correctness of expression (8). It follows from these results that the maximal value of $|S/A| = 5.8 \times 10^5$ is attained for $d = 6.303 k_0^{-1}$ in the vicinity of the critical waveguide thickness $d_c = 6.308 k_0^{-1}$. At the same time, the analytic description of sensitivity obtained in [4] is ineffective for $d \longrightarrow d_c$. This follows from the results of exact and approximate calculations of values of S using expressions (2) and (3) from [4]. These results are represented by curves *3* and *4* in Fig. 2.

Let us now consider the choice of optimal polarization of the exciting beam This polarization should be chosen from the maximization condition for quantity K, since other parameters in expression (8) for $\nu_0'' = \nu_{0\mathrm{opt}}''$, $H_1 = H_{1\,\mathrm{opt}}$, $H_2 = H_{2\,\mathrm{opt}}$, $u = u_{\mathrm{opt}}^{(s)}$, $v = v_{\mathrm{opt}}^{(s)}$, and $F = F_{\mathrm{opt}}$ are independent of polarization. In this connection, we consider the ratio $K|_{T=0}(K|_{T=1})^{-1}$, which has the following form for $d \longrightarrow d_c$ on account of Eqs. (5) and (11):

$$K|_{T=0}(K|_{T=1})^{-1} = \varepsilon_c'(2\varepsilon_c' - \varepsilon_g')(\varepsilon_c'(\varepsilon_w' + \varepsilon_g') - \varepsilon_g'\varepsilon_w')^{-1}. \tag{18}$$

It follows that, for $\varepsilon_c' > 0.5\varepsilon_w'$, it is preferable to use an exciting beam with the TE polarization. For example, for the above values of permittivities ε_c, ε_w, and ε_g, this condition is satisfied; in accordance with expression (18), $K|_{T=0}(K|_{T=1})^{-1} = 2.8$ in this case.

Let us now estimate the effect of the beam boundedness on sensitivity. As follows from expressions (8) and (9), the dependence of the sensitivity on the beam diameter for $|H_1 H_2^{-1}| \ll 1$ (which is valid for $\alpha \longrightarrow \alpha_k$) is described by function $F(H_2)$, where $H_2 = H_2(w)$. In accordance with (13), the peak of $|F(H_2)|$ is attained for $H_2 = 0$, i.e., in the limit $w \longrightarrow \infty$. This fact is confirmed by the calculations of function $F(H_2)$ for a sensor whose parameters satisfy expressions (12) and (14). For a Gaussian beam ($\psi_0(\xi, \zeta) = \exp(-\xi^2 - \zeta^2)$) and for the above parameters D, ε_c, ε_w, ε_g, ε_p, T, and m, we calculated the values of $F = S(A|(1 - \delta)(1 + \delta)^{-1}|^2 K|L|)^{-1}$ in accordance with expressions (1) and (17). In accordance with expressions (12) and (14)–(16), the parameters of the exciting beam, prismatic coupling device, and waveguide were found to be $\beta_0 = 1.332 k_0$, $g_{\mathrm{opt}}^{(1)} = 0.6067 k_0^{-1}$, $d_{\mathrm{opt}}^{(1)} = 6.303 k_0^{-1}$, and $g_{\mathrm{opt}}^{(-1)} = 0.4468 k_0^{-1}$, $d_{\mathrm{opt}}^{(-1)} = 6.304 k_0^{-1}$. The results of calculation of the sensitivity of sensors with such parameters are shown by curves *1* and *2*. It can be seen from the figure that, for $H_2 \longrightarrow 0$, quantities F tend to constant values close to F_{opt} (the values of F_{opt} are shown by dashed lines in Fig. 3). The difference of F_{opt} from these constant values is due to the approximate nature of expression (9), which was derived disregarding the terms proportional to $p_1^{-1}\,\mathrm{Im}[2\delta\Delta\bar{k}_{yc}(1 - \delta)^{-2}]$. With increasing H_2 (decreasing w), the values of F deviate from the limiting values $F(0)$. In particular, the values of S/A vary from the limiting values -5.8×10^5 and 6.0×10^5 (for $w \longrightarrow \infty$) to -75.3 and 5123.7, respectively, for $w = 10^4 k_0^{-1}$.

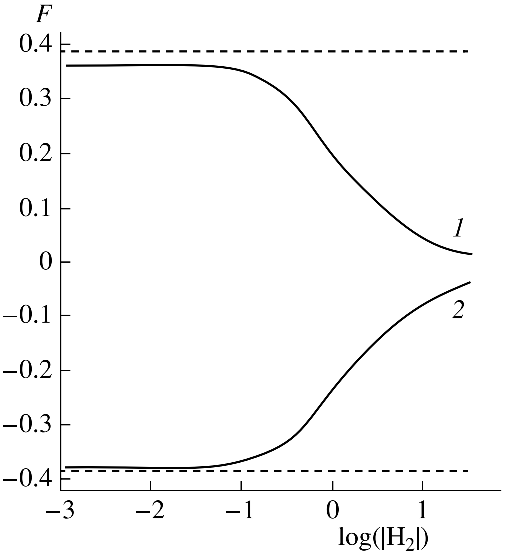


Fig. 3. Dependence of parameter F on H_2 for $d \longrightarrow d_c$.

Finally, it should be observed that a comparison of the given sensor with others [3, 4, 7, 8] confirms its high sensitivity. For example, the maximal sensitivity of a sensor operating on the basis of the surface plasmon resonance [2, 3] for the same values of parameters D, ε_p, ε_g, ε_c, and $w = 10^4 k_0^{-1}$ is lower by a factor of 64 than the sensitivity of our sensor (in the limit $w_0 \longrightarrow \infty$, the difference in the sensitivities reaches 7.1×10^3).

CONCLUSIONS

We have studied integrated optical sensors in which the power of a light beam reflected from a prismatic device for exciting planer waveguides is detected. The analytic solution of the problem on the maximization of the sensitivity of such sensors is obtained in general form. The solution makes it possible to determine the parameters of the exciting beam as well as of the buffer and waveguide layers for which the maximal intensity can be attained. It is shown that these maxima are attained when the waveguide mode is excited under nearly critical conditions. It is demonstrated that the sensor sensitivity substantially depends of the exciting beam radius w_0 and attains its maximal value in the limit $w_0 \longrightarrow \infty$. The maximal sensitivity of the sensor studied to the change in the absorption by the aqueous medium exceeds the maximal sensitivity of the sensor operating on the basis of surface plasmon resonance by a factor of 7.1×10^3.

ACKNOWLEDGMENTS

The author thanks A.B. Sotskiĭ, A.A. Romanenko, V.A. Karpenko, and E.A. Ermolaev for fruitful discussions and their interest in this research.

REFERENCES

1. W. Lukosz, Sens. Actuators B **29**, 37 (1955).
2. A. K. Nikitin, A. A. Tishchenko, and A. I. Chernyaĭ, Zarubezhn. Radioélektronika, No. 10, 14 (1990).
3. J. Villatoro and A. Garcia-Valenzuela, Appl. Opt. **38**, 4837 (1999).
4. I. U. Primak, A. B. Sotskiĭ, and A. V. Khomchenko, Pis’ma Zh. Tekh. Fiz. **23** (13), 46 (1997) [Tech. Phys. Lett. **23**, 515 (1997)].
5. K. Tiefenthaler and W. Lukosz, J. Opt. Soc. Am. B **6**, 209 (1989).
6. O. Parriaux and G. J. Veldhuis, J. Lightwave Technol. **16**, 573 (1998).
7. B. V. Ioffe, *Refractometry Methods in Chemistry* (Khimiya, Leningrad, 1983) [in Russian].
8. M. C. Peña-Gomar and A. Garcia-Valenzuela, Appl. Opt. **39**, 5131 (2000).
9. A. A. Romanenko, A. B. Sotskiĭ, and A. V. Khomchenko, Opt. Akust., No. 4, 71 (1996).
10. A. A. Romanenko and A. B. Sotskiĭ, Zh. Tekh. Fiz. **68** (4), 86 (1998) [Tech. Phys. **43**, 427 (1998)].
11. M. Adams, *An Introduction to Optical Waveguides* (Wiley, New York, 1981; Mir, Moscow, 1984).
12. A. A. Romanenko, A. B. Sotskiĭ, and A. V. Khomchenko, Preprint No. 649, IF AN BSSR (Belarus Institute of Physics, Belarussian Academy of Sciences, Minsk, 1991).

Translated by N. Wadhwa

Technical Physics, Vol. 49, No. 10, 2004, pp. 1325–1328. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 80–83.
Original Russian Text Copyright © 2004 by Rumyantsev, Pashkevich.

OPTICS, QUANTUM ELECTRONICS

Light Scattering by the Apical Oxygen Sublattice in a Thin Layer of YBaCuO Crystal

V. V. Rumyantsev and Yu. G. Pashkevich
Galkin Institute Physics and Technology, National Academy of Sciences of Ukraine, Donetsk, 83114 Ukraine
e-mail: rumyants@host.dipt.donetsk.ua

Received June 27, 2003

Abstract—Peculiarities of scattering of a TM-polarized light wave by the apical oxygen sublattice, which are associated with dispersion of a phonon *n*-polariton localized in an ultrathin YBaCuO layer, are studied. The difference between the angle of scattering of the luminous flux from the angle of reflection is estimated for the maximum of the Raman frequency shift. It is shown that the amplitude of *g*-oscillations of bridge oxygen ions in the vicinity of the resonance frequency increases sharply; consequently, it becomes possible to observe additional (in respect to bulk) scattering of coherent electromagnetic waves at the Stokes and anti-Stokes frequencies. *© 2004 MAIK “Nauka/Interperiodica”.*

INTRODUCTION

Interest in studying ultrathin high-temperature superconducting (HTSC) objects, in particular, 1 : 2 : 3 compounds [1, 2], has increased due to the promising application of epitaxial nanocrystalline films and layered structures with preset characteristics in solid-state electronics. Optical methods, including those based on Raman scattering of light and Raman scattering at surface polaritons, are highly informative in the analysis of microscopic mechanisms of superconductivity required for controlling technological processes and for determining the defectiveness of the structure of such quasi-two-dimensional crystals.

Raman scattering spectra and the IR absorption spectra of compounds of the YBaCuO type have been measured extensively [3–8] and detailed analysis of the vibrational spectrum of these crystals has been carried out. The application of vibrational spectroscopy has made it possible to identify phase transformations and to study the mechanisms of structural instabilities in bulk samples [9–11]. At the same time, a detailed analysis of Raman spectra of real crystals leads to the problem of the role of boundaries (outer surfaces, interfaces, domain walls, etc.) in light scattering. It has also been noted that the oxygen deficiency and the crystal structure near the surface and in the bulk of a YBaCuO single crystal differ substantially. This complicates the interpretation of experimental results on electron tunneling and photoemission as well as the selection of the technique for determining the oxygen deficit index (in view of the small coherence length ξ of HTSC materials of the YBaCuO type, experimental data are often determined by a surface layer of thickness d on the order of ξ [12]).

The above arguments support the importance of the problem of light scattering by a thin YBaCuO layer. We will consider here the peculiarities in the interaction of light with the sublattice of apical oxygen O(4) in a quasi-two-dimensional layer of a $YBa_2Cu_3O_{7-x}$ crystal. The special attention paid to light scattering from vibrations of apical oxygen O(4) is due to the fact that superconducting pairing in 1 : 2 : 3 compounds is associated precisely with the fluctuation of the position of apical oxygen [13].

It is well known that the Raman-active A_g mode of *C*-vibrations of apical oxygen O(4) ions corresponds to a frequency of 500 cm^{-1} [3]. This mode is softened to a considerable extent with increasing oxygen deficit x [4] and is transformed into a broad peak near 480 cm^{-1} upon a transition to $YBa_2Cu_3O_6$ [3]. Above the line at 500 cm^{-1}, broadened lines are observed with maxima at frequencies of 580–600 and 630–650 cm^{-1} [3, 5, 6]; the symmetry of the Raman mode at 580 cm^{-1} is also A_g. The shape and intensity of these bands vary depending on the defectiveness of the oxygen sublattice. However, the above-mentioned bands appear in IR spectra as well; Cardona *et al.* [3] attribute the origin of the band 580–600 cm^{-1} to the mixing of IR-active modes with Raman-active modes (the specific mechanism of this mixing is not indicated). Ponosov *et al.* [7] associate the emergence of an additional band with the violation of selection rules in a defective structure and with the formation of superstructures as a result of ordering of oxygen vacancies. In [6], this band is attributed to valence vibrations of defective (with one oxygen ion missing) groups O(4)–Cu(1)–O(4). Gasparov *et al.* [8] explain the dependence of the intensity of the 500-cm^{-1} peak and the band in the vicinity of 590 cm^{-1} (“defective” mode) on the oxygen deficit parameter x proceeding from the model of a two-well potential of the O(4) ion for various temperature regimes.

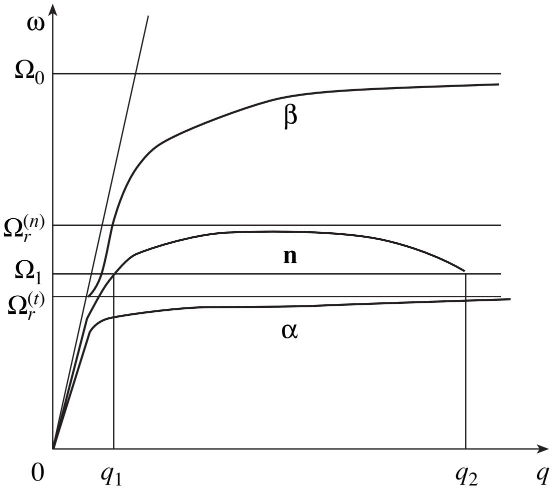
1063-7842/04/4910-1325$26.00 © 2004 MAIK “Nauka/Interperiodica”

1. MODEL

We consider light scattering from a quasi-two-dimensional crystalline $YBa_2Cu_3O_{7-x}$ structure in greater detail. Since the masses of Ba, Y, Cu, and O are in the ratio 8.6 : 5.6 : 4 : 1, we confine our analysis to a simple model, according to which the heavy ions of Ba, Y, and Cu form a rigid layered skeleton (each cell of the crystal contains six layers perpendicular to the *C* axis (Y, BaO, CuO_2, CuO_x, CuO_2, and BaO), while light oxygen ions vibrate.

In perfect crystals with an inversion center (including 1 : 2 : 3 compounds), Raman-active phonons correspond to symmetric (*g*-type) displacements of ions, while IR-active phonons correspond to antisymmetric (*u*-type) displacements. We analyze the role of bridge oxygen O(4) vibrating along the *C* axis in light scattering. Symmetric *g*-type displacements Q_g change the length of the bond in the O(4)–Cu(1)–O(4) group and, hence, determine the modulation of polarizability *X*, ensuring Raman scattering of light. Antisymmetric *u*-type displacements Q_u of O(4) ions are associated with corresponding oscillations of the dipole moment of ions, which interacts with light. Phonon polaritons are present in the crystal owing to the interaction of *u*-phonon and electromagnetic modes.

A different situation takes place in a real crystal with defects (vacancies, surfaces, etc.) in which the parity selection rules are not observed. A violation of the order in the oxygen sublattice leads to the emergence of singularities in the phonon states in the Raman spectra of a YBaCuO crystal and is responsible for the Raman activity of virtually all phonon modes. In particular, in the presence of oxygen vacancies on the O(4)–Cu(1)–V bridges or in the arrangement of the O(4)–Cu(1)–O(4) groups near the surface, normal vibrations of O(4) ions induce a dipole moment (thus, these ions are active in IR absorption); at the same time, these vibrations are associated with deformation of the chemical bond Cu(1)–O(4). The latter fact leads to Raman scattering of light at dipole-active vibrations of O(4). In all probability, such IR modes of the quasi-Raman type, which have become active as a result of symmetry breaking in the crystal structure, are manifested, according to [3], in the Raman spectra of YBaCuO.



Dispersion relations of electromagnetic $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and **n** modes localized in a quasi-two-dimensional crystal layer.

As a model problem, we consider the scattering of a luminous flux by a quasi-two-dimensional YBaCuO layer using the phenomenological approach developed in [14]. The frequency ω_I of light incident on the crystal surface is on the order of the characteristic electron frequency 10^{15} s^{-1}, but is much higher than the frequency Ω_u and Ω_g of vibrations of O(4) ions and of phonon polaritons Ω. The latter circumstance makes it possible to describe the modulation of electron polarizability determining Raman scattering from polaritons using the Placzek adiabatic approximation [15].

2. RESULTS AND DISCUSSION

(a) Dispersion of polaritons in bulk samples has been studied extensively [16], while this phenomenon in films whose thickness *d* is much smaller than the excitation radius still arouses interest. The interaction of an ultrathin crystalline layer with an electromagnetic wave of frequency Ω with wave vector $\mathbf{q}$ ($d \ll 2\pi/q$) in the plane of the film was studied in [14, 17, 18]. The field-induced polarization of the quasi-two-dimensional layer was described using a continual approximation. In this case, it is convenient to write the Maxwell equations with sources at the surface in the (ω, $\mathbf{z}$, z) representation. This allows us to eliminate the field components normal to the film, expressing them in terms of the planar components. This leads to two linearly independent systems of equations describing the *s* and *p* modes. The dispersion relations for one normal **n** mode and two planar modes, viz., the $\boldsymbol{\beta}$ mode ($\boldsymbol{\beta} = \mathbf{q}/q$ and the $\boldsymbol{\alpha}$ mode) $\boldsymbol{\alpha} = \mathbf{n} \times \boldsymbol{\beta}$), which are localized in such a layer, follow from the condition of solvability of these equations for a self-consistent description of the monolayer + field system, which are supplemented with constitutive equations with a model response function (see figure).

(b) Let normal **n** of the YBaCuO layer be oriented along the *C* axis. A TM-polarized light wave of frequency ω_i (in the IR spectral range, far away from magnetic dipole transitions) and wave vector $\mathbf{k}_i$, which is incident on the surface of the layer at angle Θ_i, is scattered at angle Θ_s by oscillations of the dipole moments of O(4) ions, generating normal polaritons $\Omega_n(\mathbf{q})$ localized in the layer (here and below, subscripts *i* and *s* indicate the incident and scattered electromagnetic wave). In terms of quasiparticles, an elementary Raman scattering process is reduced either to the reaction of decay of an exciting photon (ω_i, $\mathbf{k}_i$) into another photon (ω_s, $\mathbf{k}_s$) and a phonon polariton (Ω_n, $\mathbf{q}$) in the case of a Stokes process or to the generation of a photon (ω_s, $\mathbf{k}_s$)

as a result of the collision of a photon (ω_i, $\mathbf{k}_i$) and polariton (Ω_n, $\mathbf{q}$) in the case of an anti-Stokes process. The laws of conservation of energy and momentum corresponding to these processes have the form

$$\omega_s(\mathbf{k}_s) = \omega_i(\mathbf{k}_i) \pm \Omega_n(\mathbf{q}), \quad \mathbf{k}_s = \mathbf{k}_i \pm \mathbf{q}. \tag{1}$$

In the framework of approximation [14, 18], the dispersion relation for phonon **n**-polaritons in the IR spectral range under investigation is given by the equation

$$\sqrt{q^2 - \Omega^2/c^2} = 2\pi q^2 X_n(\Omega). \tag{2}$$

To specify the model, we choose the polarizability of the layer in the one-resonance approximation

$$X_n(\Omega) = X(\infty)(\Omega_0^2 - \Omega^2)/(\Omega_r^2 - \Omega^2). \tag{3}$$

Here, Ω_r is the resonance frequency and $\Omega_0 = \Omega_r\sqrt{X(0)/X(\infty)}$ is the transparency frequency of the layer. Analysis of Eqs. (2) and (3) shows that the dispersion curve is nonmonotonic and originates at point (0, 0) on the (q, Ω) plane. Subsequently, function $\Omega(q)$ increases and begins to decrease before it reaches the line $\Omega = \Omega_r$ (in the limit, to zero in the range of large values of q). The maximum value of frequency Ω_m of an **n**-polariton differs only slightly from Ω_r (by a value on the order of $(\Omega_r d/c)^2$. Consequently, we have

$$X(\Omega_m) = \sqrt{2}q_m/2\pi, \tag{4}$$

where the wave vector $q_m \approx \sqrt{2}\,\Omega_r/c$. The behavior of the dispersion curve described above shows that a certain frequency Ω_1 in the vicinity of the characteristic frequency $\Omega_r^{(n)}$ of the film polarizability corresponds to two waves with different wave vectors $q_1 \neq q_2$. The difference between the obtained additional wave from Pekar's "additional light waves" [19] is that the latter waves are due to nonlocality of the coupling between the polarization and the field, while the former wave is associated with the specific shape of the macrofield in the quasi-two-dimensional problem [20]. Thus, the Raman frequency shift $\Omega(\Theta) = \omega_s - \omega_i$ sharply decreases with decreasing scattering angle Θ_s from the value $\Theta_s = \Theta_i$ as well as upon its increase from the same value.

Let us analyze the maximal Raman shift for values of Ω near Ω_r (accordingly, for $q \longrightarrow q_m$), bearing in mind that the differential characteristic of light scattering attains its maximal value for $\Omega \longrightarrow \Omega_r$ [21] and taking into account the law of conservation of the planar component of the wave vector,

$$k_I \sin\Theta_i - k_s \sin\Theta_s = \pm q. \tag{5}$$

The latter equality can easily be transformed into the following relation: $\omega_i \sin\Theta_i - \omega_s \sin\Theta_s = \sqrt{2}\,\Omega_r$. We denote the difference between the angles of light scattering and reflection by $\Delta\Theta(\Theta_s = \Theta_i \pm \Delta\Theta)$. Taking into account the fact that $\Omega \ll \omega_i$, ω_s (as well as $\Delta\Theta \ll \Theta_i$, Θ_s), we arrive at the equality

$$(\Delta\Theta)^2 + 2\Delta\Theta\cot\Theta_i \approx 2\sqrt{2}\frac{\Omega_r}{\omega_i \sin\Theta_i}. \tag{6}$$

The solution to Eq. (6) has the form

$$\Delta\Theta = \pm\frac{\sqrt{2}\Omega_r}{\omega_i \cos\Theta_i}.$$

Consequently, for frequency $\omega_i = 1.1 \times 10^5$ cm^{-1} of a light beam incident on the surface of the crystal layer being studied at an angle of $\pi/3$ and scattered by lattice vibrations of frequency $\Omega_r = 500$ cm^{-1}, we obtain the following value of the difference between the angle of scattering and the angle of reflection: $\Delta\Theta \approx \pm 0.75°$.

(c) It was noted above that, for distorted YBaCuO crystal lattices (containing vacancies, surfaces, etc.), symmetric and antisymmetric combinations of the coordinates of displacements of O(4) ions become meaningless. Both normal vibrations stretch the Cu(1)–O(4) bond and also produce a dipole moment. The fraction of g- and u-components in the new normal modes $Q_{\tilde g}$ and $Q_{\tilde u}$ strongly depends on the relation $\alpha = (\Delta\Omega)^2/(\Omega_g^2 - \Omega_u^2)$,

$$Q_{\tilde u} = Q_u - \alpha Q_g, \quad Q_{\tilde g} = Q_g + \alpha Q_u, \tag{7}$$

and the corresponding frequencies have the form

$$\Omega_{\tilde u}^2 = \Omega_u^2 - (\Delta\Omega)^2/2, \quad \Omega_{\tilde g}^2 = \Omega_g^2 - (\Delta\Omega)^2/2. \tag{8}$$

Here, $\Delta\Omega$ is the correction to the vibrational frequency of the O(4) ions associated with lattice symmetry distortion. In accordance with the experimental results obtained in [5], this quantity amounts to about 1% of Ω_u, Ω_g.

In the framework of the anharmonic oscillator model, the dynamic equations describing vibrations of the O(4) ions along the C axis have the form

$$\begin{gathered} \ddot{Q}_{\tilde u} + \Gamma_1 \dot{Q}_{\tilde u} + Q_{\tilde u}^2 Q_{\tilde u} = (2e/m)E - (2\mu/m)Q_{\tilde g}Q_{\tilde u}, \\ \ddot{Q}_{\tilde g} + \Gamma_2 Q_{\tilde g} + Q_{\tilde g}^2 Q_{\tilde g} = -(\mu/m)Q_{\tilde u}^2, \end{gathered} \tag{9}$$

where e and m are the charge and mass of the O(4) ions and μ is the parameter of coupling between the g and u modes. Let us suppose that $\Gamma_1 = \Gamma_2 = \Gamma$ and that the electric field accompanying a phonon polariton is a biharmonic wave $E = E_1\exp(-i\Omega_1 t) + E_2\exp(-i\Omega_2 t)$ + c.c. (c.c. denotes the complex-conjugate term). In this case, we obtain the following steady-state solution to system of equations (9) accurate to terms quadratic in the field:

$$Q_{\tilde u(\tilde g)} = Q_{\tilde u(\tilde g)}^{(1)} + Q_{\tilde u(\tilde g)}^{(2)}; \tag{10}$$

here,

$$Q_{\tilde u}^{(1)} = (2e/m)E_l\exp(-i\Omega_l t)/(\Omega_{\tilde u}^2 - \Omega_l^2 - i\Omega_l\Gamma) + \text{c.c.} \quad (l = 1, 2), \tag{11a}$$

$$Q_{\tilde u}^{(2)} = 0, \quad Q_{\tilde g}^{(1)} = 0, \tag{11b}$$

$$\begin{aligned} Q_{\tilde g}^{(2)}(\Omega_1 \pm \Omega_2) &= -(2\mu/m) \\ &\times (2e/m)^2 E_1E_2/[(\Omega_{\tilde u}^2 - \Omega_1^2 - i\Omega_1\Gamma)(\Omega_{\tilde u}^2 - \Omega_2^2 \pm i\Omega_2\Gamma)] \\ &\times \exp[-i(\Omega_1 \pm \Omega_2)t]/[\Omega_{\tilde g}^2 - (\Omega_1 \pm \Omega_2)^2 - i(\Omega_1 \pm \Omega_2)\Gamma] \\ &+ \text{c.c.} \end{aligned} \tag{11c}$$

The displacement of the O(4) ions forming the surface density of the dipole moment and, hence, generating the scattering field at the Stokes and anti-Stokes frequencies is given by

$$Q_u = Q_{\tilde u}^{(1)} + \alpha Q_{\tilde g}^{(2)}. \tag{12}$$

Thus, since the intensity of the luminous flux is proportional to $|Q_u|^2$ (this follows from [17]), formulas (11) enable us to estimate the intensity of manifestation of singularities in the Raman spectra of ultrathin YBaCuO crystals. The observation of scattering of coherent electromagnetic waves from Stokes and anti-Stokes frequencies becomes possible in the vicinity of the resonance frequency.

CONCLUSIONS

The peculiarities of light scattering from vibrations of apical oxygen O(4) considered here follow from the dispersion relation (permitting the existence of an additional wave) for a phonon **n**-polariton localized in a thin layer of a 1 : 2 : 3 compound. The chosen model disregards the mixing in completely symmetric modes of vibrations of the oxygen ions belonging to the CuO_2 plane and to the O(4)–Cu(1)–O(4) bridge, as was done in [22], or the mixing of vibrations of the complex CuO_x plane—O(4)–Cu(1)–O(4) bridge [23]. Nevertheless, the example proposed here (scattering of a *p*-polarized luminous flux generating **n**-polaritons) makes it possible to analyze the consequences of mixing of *u* and *g* phonon modes.

It should be noted that the technique proposed here is also applicable for studying the scattering of an *s*-polarized electromagnetic wave generating the α component of surface polarization [14] and, hence, is of interest for studying Raman scattering from the vibrations of O(1)–Cu(1)–O(1) chains (arranged along the *B* axis of a YBaCuO crystal; in this case, α || *B*).

REFERENCES

1. D. Schweitzer, T. Bollmeier, B. Stritzker, *et al.*, Thin Solid Films **280**, 147 (1996).
2. J. Nam, R. A. Hughes, A. Dabkowski, *et al.*, Appl. Phys. Lett. **82**, 3728 (2003).
3. M. Cardona, L. Genzel, R. Liu, *et al.*, Solid State Commun. **64**, 727 (1987).
4. C. Tomsen, R. Liu, M. Bauer, *et al.*, Solid State Commun. **65**, 55 (1988).
5. V. A. Evseev and R. F. Konopleva, Preprint No. 1585, LIYaF (Institute of Nuclear Physics, Academy of Sciences of USSR, Leningrad, 1990).
6. M. V. Belousov, V. Yu. Davydov, S. F. Karmanenko, and O. V. Kosogov, Pis’ma Zh. Éksp. Teor. Fiz. **48**, 285 (1988) [JETP Lett. **48**, 316 (1988)].
7. Yu. S. Ponosov, G. A. Bolotin, O. V. Gurin, *et al.*, Pis’ma Zh. Éksp. Teor. Fiz. **48**, 380 (1988) [JETP Lett. **48**, 420 (1988)].
8. L. V. Gasparov, V. D. Kulakovskiĭ, V. B. Timofeev, and E. Ya. Sherman, J. Supercond. **8**, 27 (1995).
9. I. V. Shapiro, V. F. Shabanov, and A. N. Vtyurin, Preprint No. 534-F, IF im. L. V. Kirienko SO Akad. Nauk SSSR (Kirenskii Institute of Physics, Siberian Division, Academy of Sciences of USSR, Krasnoyarsk, 1988).
10. V. D. Kulakovskiĭ, O. V. Misochko, V. V. Timofeev, *et al.*, Pis’ma Zh. Éksp. Teor. Fiz. **46**, 460 (1987) [JETP Lett. **46**, 580 (1987)].
11. E. I. Kuznetsova, T. P. Krinitsina, S. V. Sudareva, *et al.*, Fiz. Met. Metalloved. **81**, 113 (1996).
12. *Physical Properties of High Temperature Superconductors*, Ed. by D. M. Ginsberg (World Sci., Singapore, 1989; Mir, Moscow, 1990), pp. 8–38.
13. J. Ranninger, Z. Phys. B **84**, 168 (1991).
14. V. V. Rumyantsev and V. T. Shunyakov, Kristallografiya **36**, 1342 (1991) [Sov. Phys. Crystallogr. **36**, 760 (1991)].
15. J. A. Giordmaire and W. Kaiser, Phys. Rev. **144**, 676 (1966).
16. A. Yariv and P. Yeh, *Optical Waves in Crystals: Propagation and Control of Laser Radiation* (Wiley, New York, 1984; Mir, Moscow, 1987).
17. V. V. Rumyantsev, Ukr. Fiz. Zh. **40**, 745 (1995).
18. V. V. Rumyantsev, E. Ya. Shtaerman, and K. V. Gumenik, in *Proceedings of the 7th International Conference on Nanometer-Scale Science and Technology (NANO-7/ECOSS-21), Malmo, 2002.*
19. S. I. Pekar, Otkrytiya, Izobret. **32**, 3 (1987).
20. V. V. Rumyantsev and V. T. Shunyakov, Physica B **176**, 156 (1992).
21. V. V. Rumyantsev and S. V. Yampolskiĭ, *Spectroscopy of Superconducting Materials*, Ed. by E. Faulques (ACS, Washington, 1999), pp. 131–138.
22. E. I. Rashba and E. Ya. Sherman, Sverkhprovodimost: Fiz. Khim. Tekh. **2**, 60 (1989).
23. J. Mustre de Leon, S. D. Conradson, I. Batistic, *et al.*, Phys. Rev. Lett. **65**, 1675 (1990).

Translated by N. Wadhwa

Technical Physics, Vol. 49, No. 10, 2004, pp. 1329–1334. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 84–89.
Original Russian Text Copyright © 2004 by Antonov.

OPTICS, QUANTUM ELECTRONICS

Acoustooptic Nonpolar Light Controlling Devices and Polarization Modulators Based on Paratellurite Crystals

S. N. Antonov

Institute of Radio Engineering and Electronics, Russian Academy of Sciences (Fryazino Branch),
Fryazino, Moscow oblast, 141190 Russia
e-mail: olga-ant@yandex.ru

Received February 5, 2004

Abstract—A complex of investigations and projects aimed at solving problems of highly effective acoustooptic control of nonpolarized laser radiation is generalized on the basis of anisotropic Bragg's diffraction in a TeO_2 crystal from a slow acoustic wave, which is characterized by extraordinarily high acoustooptic quality. Modulators of the zeroth diffraction order, modulators of the first Bragg's order, and a polarization-insensitive deflector are considered. The developed systems ensure the control of the polarization state of optical radiation. These systems are employed in setups and devices based on high-power solid-state and fiber lasers with a wavelength of 1.06 μm. *© 2004 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The highest efficiency of high-power commercial lasers (including solid-state and fiber lasers operating at a wavelength of 1.06 μm) is attained, as a rule, when radiation is either nonpolarized or possesses random polarization. This necessitates the development of devices that can effectively control such radiation. Paratelluride (TeO_2) single crystals in the case of anisotropic diffraction at a slow acoustic wave are record-holders with respect to practical application among the variety of available acoustooptic (AO) materials. This is due to phenomenally high AO quality M_2 in the case of good optical and acoustic properties of a crystal and the developed technology of its production. However, the high efficiency of this type of AO diffraction (up to 100%) is realized only when optical radiation incident on an AO cell is an optical eigenmode of the crystal, i.e., possesses a strictly definite polarization (linear, elliptical, or circular) depending on the geometry of the AO interaction. Consequently, the efficiency of diffraction for nonpolarized optical radiation cannot exceed 50% unless special measures are taken.

The methods developed in this study are based on the fact that nonpolarized optical radiation can be represented as the sum of two beams in the same direction and having the same intensity with orthogonally oriented polarization vectors. It is important to note that, with such a representation, these beams are incoherent and, hence, independent (the situation basically differs from the decomposition, say, of circularly polarized light into two linearly polarized orthogonal vectors). Thus, we sought AO solutions that permit the realization of anisotropic Bragg's diffraction for these two polarization components of a nonpolarized beam in the same device.

Specific calculations were made using the standard method of analysis of vector diagrams for a uniaxial optically active crystal (such as a TeO_2 crystal). The indicatrices of its refractive indices were approximated by ellipsoids of revolution [1],

$$\frac{n_1^2(\theta)\cos^2\theta}{n_o^2(1+\delta)^2} + \frac{n_1^2(\theta)\sin^2\theta}{n_e^2} = 1, \qquad (1)$$

$$\frac{n_2^2(\theta)\cos^2\theta}{n_o^2(1-\delta)^2} + \frac{n_2^2(\theta)\sin^2\theta}{n_e^2} = 1, \qquad (2)$$

where $n_1(\theta)$ and $n_2(\theta)$ are the refractive indices of the extraordinary and ordinary rays, respectively; θ is the angle to the optical axis; $\delta = \lambda_0\rho/2\pi n_o$ is the splitting coefficient; ρ is the specific rotation of polarization of light (rad/mm); and λ_0 is the wavelength of light in vacuum.

The following values of parameters were used: λ_0 = 1.06 μm, ρ = 254.8, n_o = 2.208, and n_e = 2.352. The velocity of sound propagating in the (110) plane for small deviations from the normal to the optical axis were approximated, following [2], by the function $\nu = \nu_0(1 + b\varphi^2)$, where b is the anisotropy parameter for elastic waves (b = 4.895), φ is the angle of deviation from the direction of the phase velocity of sound from the normal to the optical axis, and ν_0 is the velocity of sound propagating perpendicularly to the optical axis of the crystal ($\nu_0 = 0.617 \times 10^6$ mm/s).

1. MODULATOR OF ZERO-ORDER DIFFRACTION

In a number of technical projects using AO modulation of light intensity, it is necessary to minimize the optical losses introduced by a modulator. This concerns, for example, the modulation of the *Q* factor of a laser resonator or the problems associated with treatment of materials by a laser beam, when the power of light “delivered” to the surface of the material plays a decisive role. For this purpose, use is made of a modulator of the zeroth order of diffraction, whose working channel is transmitted light (zeroth order of diffraction). In the absence of sound, this diffraction order does not introduce losses of light and its important characteristic is the residual intensity of light in the transmitted beam (diffraction efficiency).

Figure 1 illustrates the operation of a high-efficiency zero-order modulator based on a TeO_2 crystal [3]. Nonpolarized optical radiation is incident on the crystal at a certain angle α. In the crystal, radiation splits into two orthogonally polarized eigenmodes with vectors $\mathbf{K}_o$ and $\mathbf{K}_e$, respectively, which correspond to the crystal indicatrices n_o and n_e. The problem is to find the parameters of a single acoustic wave with wave vector $\mathbf{q}$ directed along the [110] axis; the diffraction at this wave ensures the fulfillment of the equalities

$$\mathbf{K}_o \pm \mathbf{q} = \mathbf{K}_{ed}, \quad (3)$$

$$\mathbf{K}_e \pm \mathbf{q} = \mathbf{K}_{od}, \quad (4)$$

where $\mathbf{K}_{ed}$ and $\mathbf{K}_{od}$ are the wave vectors of orthogonally polarized “+” and “–” first Bragg’s orders of diffraction.

Figure 2 qualitatively illustrates the existence of joint solution to Eqs. (3) and (4). It shows the dependences of angles of incidence α_1 and α_2 for vectors $\mathbf{K}_o$ and $\mathbf{K}_e$, respectively, on the acoustic wave frequency *f*. It can be seen that, indeed, only one acoustic wave frequency f_0 exists, for which simultaneous diffraction of two orthogonally polarized components of nonpolarized light to the “+” and “–” first Bragg’s orders (angle of incidence α_0) is ensured. Such an AO modulator is essentially a “controllable polarization prism” since the polarization vectors of these orders are strictly orthogonal. The results of calculation of the values of f_0 for several laser wavelengths of practical importance are given in the table. The minimal lengths *L* of the converter, which ensure the Bragg diffraction regime (fulfillment of the inequality $Q \geq 4\pi$ for the Klein–Cook parameter) are also given in the table.

However, such diffraction alone cannot ensure the low residual intensity of light in the zeroth order in the presence of sound. This is due to the fact that real optical and acoustic beams are characterized by commensurate values of diffraction divergence [4] as well as the fact that the acoustic beam is slightly nonuniform both as regards the angular distribution of intensity and the coordinate distribution. In actual practice, a 90% diffraction efficiency in a single event of the AO interaction is a good result.

To substantially increase the diffraction efficiency, an AO modulator using a peculiarity of the TeO_2 single crystal of practical importance was developed. This peculiarity stems from the fact that the AO properties are identical for sound propagating along the orthogonal crystallographic axes [110] and $[1\bar{1}0]$. This makes it possible to obtain simultaneous diffraction in two mutually perpendicular directions for two independent acoustic beams in the same crystal (Fig. 3). It is important to note that the efficiency of zero-order diffraction in this case is equal to the product of the efficiencies on each of the two acoustic beams (by virtue of the independence of these AO interactions). Strictly speaking, the inclination of the crystal at Bragg’s angle in two orthogonal planes, which is necessary in this case, leads to a certain shift in the value of the rated frequency f_0; however, this shift is insignificant.

On the basis of the type of AO diffraction described above, a commercial modulator was manufactured for controlling high-power solid-state lasers with λ_0 =

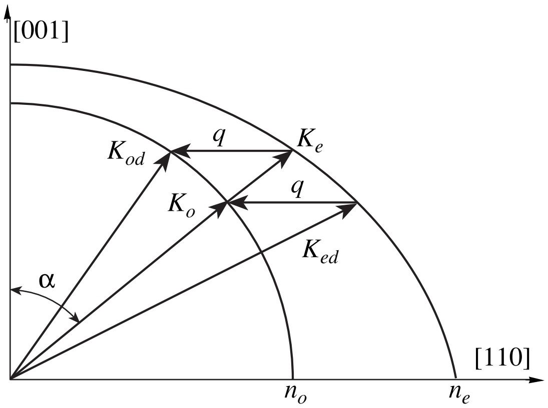


Fig. 1. Vector diagram illustrating the interaction of two optical eigenmodes at the same acoustic frequency.

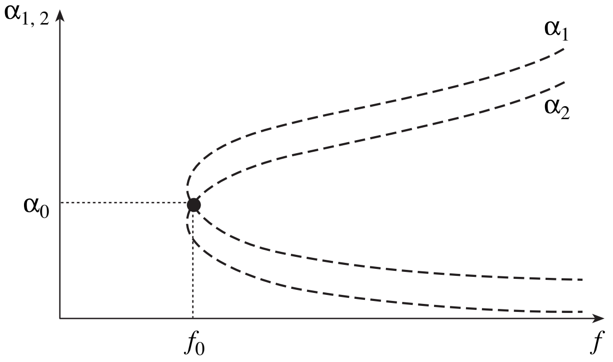


Fig. 2. Qualitative dependence of the angles of incidence of orthogonally polarized optical eigenmodes of the crystal on the frequency of the acoustic wave.

1.06 μm. The actual relative intensity of the zeroth diffraction order did not exceed 1–3%, depending on the laser beam parameter.

2. TWO-CRYSTAL MODULATOR OF THE FIRST BRAGG'S ORDER

One of the methods developed for deflecting nonpolarized light to the first Bragg's order is to use two identical AO cells connected in series in the optical scheme. The vector diagram shown in Fig. 4 demonstrates that such a method is possible in principle. Let us suppose that a single acoustic wave with vector **q** propagates in a TeO_2 crystal and two input optical beams with orthogonal polarizations $\mathbf{K}_e$ and $\mathbf{K}_o$ are incident on the left and right from the polarization axis. It follows from symmetry considerations the orthogonally polarized diffracted beams $\mathbf{K}_{ed}$ and $\mathbf{K}_{od}$ are deflected from the incident beams through strictly identical angles θ. It is important that the beams are deflected in the same direction. The diagram in Fig. 5 demonstrates the condition for diffraction at two identical successive AO cells (with the same frequencies of acoustic waves). Diffraction at the first crystal corresponds to the left-hand side of Fig. 4, and diffraction at the second crystal to the right-hand side. Thus, two diffracted orthogonally polarized beams of the same direction are formed at the outlet of the crystals; the combination of these beams gives a nonpolarized beam. The latter statement, however, is valid to within the spatial parallax emerging between the two beams (Fig. 6). In this figure, AO cells *1* and *2* are separated by distance L. It can be seen that the centers of the beams are displaced by Δd. Specifically, the spatial shift between the centers of the input beams is $\Delta d = 0.2$ mm for the laser modulator prepared in this way for the wavelength $\lambda_0 = 1.06$ μm at a sound frequency of 12 MHz and the spacing $L = 10$ mm between the crystals. Obviously, parallax Δd is significant when its value is commensurate with input beam aperture d.

The modulator designed here is a comparatively cheap technological appliance that has found wide applications. However, the parallax emerging in many cases restricts its application when, for example, a quick response of the modulator (small aperture d of the optical beam) is required. The removal of this limitation will be discussed in the next section.

3. ONE-CRYSTAL PARALLAX-FREE MODULATOR

This section is devoted to a polarization-insensitive modulator designed with a TeO_2 crystal and free of parallax for orthogonally polarized output beams [5].

The idea of the method is illustrated by the vector diagram shown in Fig. 7. Let nonpolarized optical radiation be incident on the crystal at a certain angle α to the optical axis. In the crystal, this radiation splits into two orthogonally polarized eigenmodes with vectors $\mathbf{K}_o$ and $\mathbf{K}_e$ corresponding to indicatrices n_o and n_e of the crystal. Let us suppose that the diffracted beams with vectors $\mathbf{K}_{od}$ and $\mathbf{K}_{ed}$ propagate in the same direction defined by angle β. The coupling between the incident and diffracted optical modes is executed by two independent acoustic beams with vectors $\mathbf{q}_1$ and $\mathbf{q}_2$, respectively. The wave vectors of the acoustic waves form angles γ_1 and γ_2 with the crystallographic direction [110]. In the general form, the problem is formulated as a search for the conditions under which the law of conservation of momentum is satisfied simultaneously for

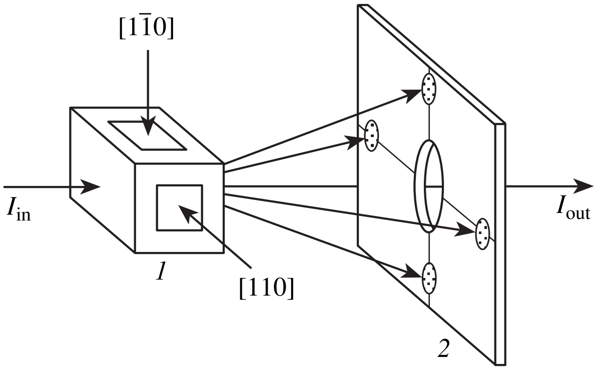


Fig. 3. Monocrystalline modulator for the zeroth diffraction order with two orthogonal acoustic waves: I_{in}—incident beam, I_{out}—output beam, *1*—TeO_2 crystal, *2*—diaphragm.

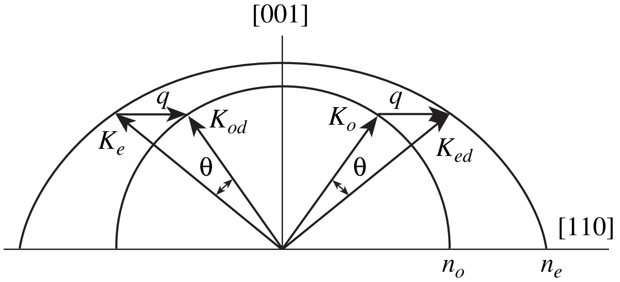


Fig. 4. Diffraction of two optical beams with orthogonal polarizations at the same acoustic wave in two quarters of the same TeO_2 crystal.

Rated values of f_0 and L as functions of the wavelength of light λ. The type of laser medium is indicated for each wavelength

λ, nm	Type of laser medium	f_0, MHz	L, mm
440	He–Cd	30.1	1.9
488	Ar	23.4	2.8
514	Ar	21.1	3.3
633	He–Ne	13.4	6.7
1060	Nd : YAG	5.5	24

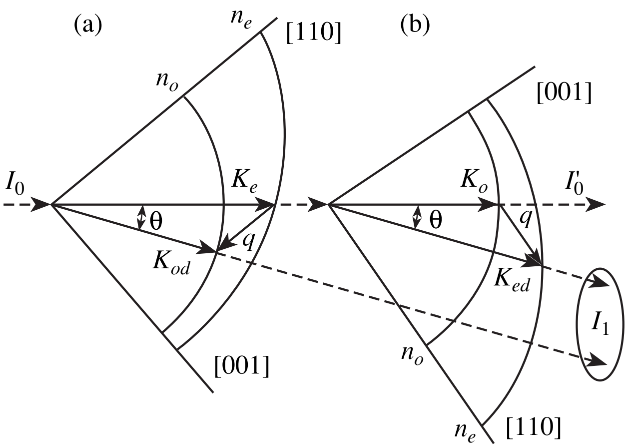


Fig. 5. Consecutive diffraction of a nonpolarized beam at two TeO_2 crystals with the formation of a single output beam: I_0—incident optical beam, I'_0—zeroth diffraction order; I'_1—resultant diffracted beam; (a) first crystal and (b) second crystal.

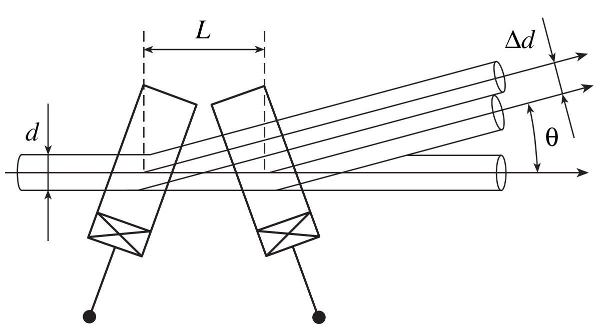


Fig. 6. Scheme of formation of a parallax in the output diffracted beam.

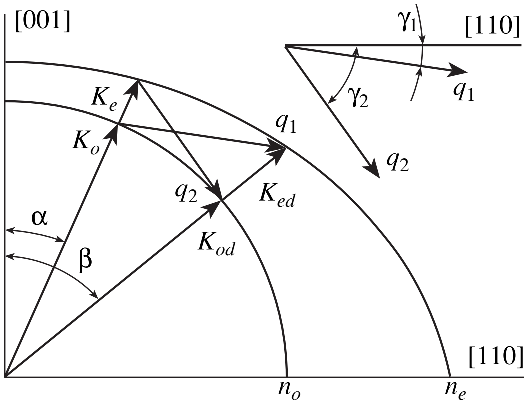


Fig. 7. Vector diagram of interaction of two optical eigenmodes at two acoustic waves and the angles of inclination of these waves relative to the [110] axis.

two independent diffraction processes:

$$\mathbf{K}_o + \mathbf{q}_1 = \mathbf{K}_{ed}, \tag{5}$$

$$\mathbf{K}_e + \mathbf{q}_2 = \mathbf{K}_{od}. \tag{6}$$

It was shown in [6] that this system of equations may have a solution for various combinations of angles γ_1 and γ_2. We will consider here the practically realized situation when $\gamma_1 = 4°$ and $\gamma_2 = 7°$ and the wavelength of light is $\lambda_0 = 1.06$ μm. The rated frequencies of acoustic beams with vectors $\mathbf{q}_1$ and $\mathbf{q}_2$ are equal to 34.655 and 35.981 MHz, respectively.

The absence of parallax in this AO modulator is ensured by the fact that the AO interaction of light and the two acoustic beams takes place in the same region of the crystal. This is ensured by the strong acoustic anisotropy of the TeO_2 crystal, i.e., considerable deflection of the acoustic energy flux from the [110] axis (by $\Delta\varphi$) for a small inclination of the acoustic wave front (the piezoelectric transducer plane) from this axis (through an angle of $\Delta\gamma$). For the diffraction plane in which the transducers are deflected, the ratio $\Delta\varphi/\Delta\gamma$ is approximately equal to 10 (for $\Delta\gamma < 7°$).

Figure 8 presents the scheme of such a modulator. Here, the "acoustic" end faces of the TeO_2 crystal *1* were inclined during the optical treatment so that the planes of piezoelectric transducers *2* and *3* form angles of 7° and 4° with the [110] axis. Acoustic anisotropy leads to deflection of the energy fluxes of acoustic waves through angles of approximately 60° and 40°, respectively, from the [110] direction. It can be seen that a region in which the acoustic beams intersect is formed at a small distance from the transducers and the AO interaction takes place precisely in this region.

The modulator samples prepared in this way completely confirmed the serviceability of the device. It should be noted, however, that this construction is technologically more complicated than a two-crystal modulator.

4. POLARIZATION-INSENSITIVE DEFLECTOR

Apart from modulation of high-power nonpolarized laser beams, practical problems associated with their scanning also exist. This concerns, for example, laser devices for material processing, in which the beam sweep (at least, in on coordinate) is performed by an AO deflector. The most optimal deflectors operate on TeO_2 crystals and employ anisotropic diffraction with a nonaxial geometry. With such diffraction, incident optical radiation must be characterized by linear polarization with a strict orientation relative to the crystal. To ensure polarization insensitivity, a technical scheme was developed, which is illustrated in Fig. 9. Here, input nonpolarized optical radiation with components $\mathbf{K}_o$ and $\mathbf{K}_e$ passes through two identical AO deflectors *1* and *2* with the same orientation relative to the beam. Plate *3* mounted between the deflectors rotates the polarization plane of light passing through it through 90°. The plate may be either a half-wave phase plate or a wafer made of an optically active crystal with a thickness corresponding to the rotation of polarization through 90° over the working wavelength of light. The deflectors operate strictly synchronously, since identical control frequencies (acoustic vectors **q**) are supplied to them. Thus, nonpolarized input radiation transmitted through deflector *1* is transformed into two beams with different directions: a zero-order beam and a deflected beam with orthogonal polarizations. After the passage through the polarization plate, the polarizations of both beams change by 90°; as a result, the nondeflected beam acquires the polarization corresponding to diffraction at crystal *2*, while the deflected beam does not interact with this crystal. At the outlet from the system, both beams deflected in each crystal form a nonpolarized diffraction order consisting of $\mathbf{K}_{od}$ and $\mathbf{K}_{ed}$ beams. Obviously, as in the case described in Section 2, a parallax appears in this case between the output beams. However, the light aperture for the deflectors in the diffraction plane is, as a rule, on the order of a few millimeters and the emerging parallax is insignificant.

5. AO MODULATION OF LIGHT POLARIZATION

Analysis of the developed optical schemes has made it possible to construct a new class of AO devices, viz., modulators of the polarization state of light. Let us consider this possibility for the two-crystal modulator presented in Figs. 5 and 6. Let linearly polarized radiation, whose polarization plane is inclined at an angle of 45° to the diffraction plane, be incident at the input of such a modulator. Obviously, such radiation can be decomposed into two coherent orthogonal linearly polarized components of the same intensity. Each component experiences Bragg diffraction at the corresponding crystal, and the output radiation will be polarized as a superposition of two diffracted beams with linear polarization. It is important to note here that, in the case of AO diffraction, the frequency of diffracted light is shifted by the frequency of the acoustic wave, the sign of the shift being determined by whether the acoustic front “surges toward” the light front or vice versa. In our case, the signs of the frequency shifts in the first and second AO cells are opposite and the total frequency difference for the output diffracted beams is approximately equal to the doubled frequency of the acoustic wave (approximately 70 MHz in the present case). Consequently, the polarization state of the output beam experiences modulation at this frequency and will pass via the following stages: linear-circular-linear (turned through 90° relative to the initial orientation).

Let us now consider the possibility of controlling the polarization of light using the appliance shown in Fig. 9. Let, as in the previous situation, linearly polarized radiation with a polarization plane inclined at 45° to the diffraction plane be incident on the input of the device. We assume that both deflectors operate at the same constant frequency of sound. Since the deflectors are identically oriented relative to the input optical beam, the frequency shift of the output beams will have the same magnitude and sign. Furthermore, we note that the diffracted beam acquires not only the frequency of the acoustic wave, but also its phase. Thus, by introducing an additional phase delay between the signals supplied to the deflectors, we can control the phase difference between the output diffracted beams. Consequently, it becomes possible to control the polarization state of output radiation (to rotate the linear polarization through 90° and to modify it into circular or elliptical polarization).

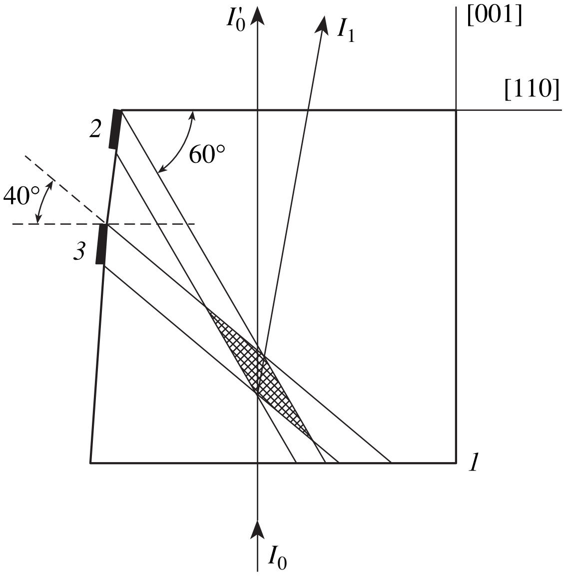


Fig. 8. Scheme of a one-crystal Bragg modulator: I_0—incident optical beam, I_0'—zeroth diffraction order, I_1—diffracted beam.

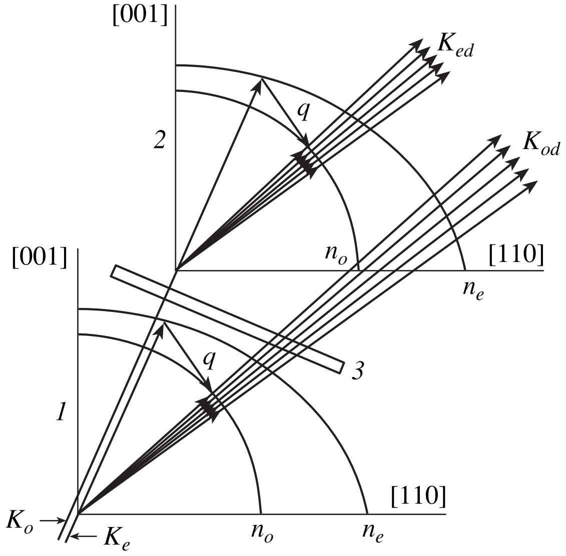


Fig. 9. Vector diagram for a two-crystal polarization-insensitive deflector.

The principles described here have formed the basis for practically implemented AO devices, which were used for developing new laser systems for image recording (Research and Production Center Alpha, Moscow; web site: http//www.alphalaser.ru). These systems employ high-power solid-state and fiber lasers operating at a wavelength of 1.06 μm and are intended for preparing flexographic matrices, photomasks, and laser markers.

This paper is based on the doctoral thesis defended by the author in 2003 [6].

REFERENCES

1. A. W. Warner, D. L. White, and W. A. Bonner, J. Appl. Phys. **43**, 4489 (1972).
2. V. P. Semenkov, Zh. Tekh. Fiz. **51**, 2090 (1981) [Sov. Phys. Tech. Phys. **26**, 1219 (1981)].
3. S. N. Antonov, V. M. Kotov, and V. N. Sotnikov, Zh. Tekh. Fiz. **61** (1), 168 (1991) [Sov. Phys. Tech. Phys. **36**, 101 (1991)].
4. L. N. Magdich and V. Ya. Molchanov, *Acoustooptic Devices and Their Applications* (Sov. Radio, Moscow, 1978) [in Russian].
5. S. N. Antonov and V. I. Mirgorodskiĭ, Zh. Tekh. Fiz. **74** (1), 84 (2004) [Tech. Phys. **49**, 83 (2004)].
6. S. N. Antonov, Doctoral Dissertation (Fryazino, 2003).

Translated by N. Wadhwa

Technical Physics, Vol. 49, No. 10, 2004, pp. 1335–1338. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 90–93.
Original Russian Text Copyright © 2004 by Riĭves, Svetlichnyĭ, Zhmenyak, Kel'man, Shpenik.

OPTICS,
QUANTUM ELECTRONICS

A Source of UV Radiation Based on Pulsed Discharge in a Xe–NaCl Mixture

R. B. Riĭves, E. A. Svetlichnyĭ, Yu. V. Zhmenyak, V. A. Kel'man, and Yu. O. Shpenik
Institute of Electronic Physics, National Academy of Sciences of Ukraine, Uzhgorod, 88017 Ukraine
e-mail: vkel@mail.uzhgorod.ua

Received February 24, 2004

Abstract—The design and operation parameters of an excimer lamp on a Xe–NaCl mixture are described. The emission spectrum of the lamp and its temporal characteristics are described. The XeCl excimer band at 308 nm dominates in the spectrum. The spectral distribution of energy within the band and its dependence on the experimental conditions are studied. The mechanism of excitation of the upper state of the excimer molecule is discussed. *© 2004 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Sources of spontaneous ultraviolet (UV) radiation are widely used in science and engineering. In particular, excimer incoherent lamps can compete in many cases with laser sources (primarily, in such fields as the initiation of photochemical and photophysical processes and destruction of harmful organic materials). In our opinion, such lamps have most promising applications in medicine and ecology, i.e., in the most sensitive fields that directly determine the quality of life.

At the present stage, a large number of samples of excilamps have been developed predominantly on binary gas mixtures emitting during excimer transitions in halogenides of inert gases. Nevertheless, the search for the chemical and partial composition of working mixtures and the methods of their pumping for obtaining maximal efficiencies and radiation power of such excimer lamps remains of utmost importance.

In our opinion, the proposal made in [1], where numerical simulation was carried out for a XeCl excimer lamp operating on a binary Xe–NaCl mixture pumped by a hard ionizer of the beam type, is especially attractive. It was shown that, under optimal conditions of excitation (a pressure of ~10 Torr, a specific energy contribution of ~60–125 W/cm^3, and a composition of the Xe–NaCl mixture of 20/1), the efficiency of the lamp attains a value of 40% and the emitted energy is 0.025 J/l.

It should be noted that the formulation of the problem of simulation of an excilamp with the above mixture composition was stimulated to a certain extent by publications [2, 3], in which this type of mixture was used in the experiments for the first time with NaCl injection into a supersonic plasma flow of xenon. Effective luminescence was attained on the 308-nm transition of the XeCl molecule. According to Alekhin *et al.* [2, 3], the excited state of the excimer molecule was mainly populated due to binary reactions of substitution of a Xe atom or ion for Na in the NaCl molecule.

In our study, we used the mixture proposed in [1–3] for constructing a low-pressure excilamp with excitation by a longitudinal high-voltage pulse-periodic discharge. This method obviously possesses considerable advantages over the beam or plasma method (as in [2, 3]) for practical applications. In addition, we emphasize that the chemical composition of the working Xe–NaCl mixture is itself a considerable advantage, since the rock salt used in this mixture as a halogen donor is a nontoxic substance under normal conditions.

EXPERIMENTAL TECHNIQUE

The excimer source is obtained on the basis of a vacuum-sealed gas-discharge tube (GDT) made of molten quartz and having an inner diameter of 12 mm and a length of the discharge gap between the electrode of 0.4 m. A high-purity powder of common salt NaCl was placed directly along the GDT. Coaxial water-cooled electrodes made of copper were glued at the opposite ends of the GDT. Radiation was extracted from the excilamp via quartz windows. The required temperature regime was created by using simultaneously self-heating (i.e., the heat liberated by the discharge) and heating from an external furnace. It should be noted that the temperature of the inner wall of the GDT required for obtaining a saturated vapor pressure of NaCl of 1 Torr must be not lower than 1138 K [4].

To excite a longitudinal pulse-periodic discharge in the excilamp, we used a thyratron oscillator with a TGI1-2000/35 commutator with resonant recharging of a storage capacitor of 1650 pF.

The emission parameters of the excilamp were recorded using a monochromator MDR-6, photodetec-

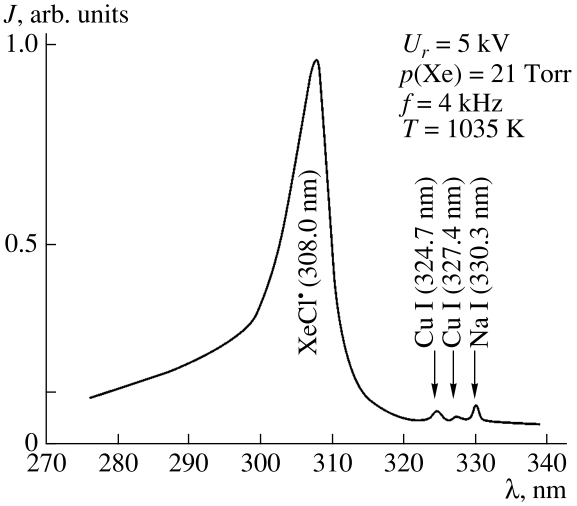


Fig. 1. Time-integrated emission spectrum $J(\lambda)$ of an excilamp operating on a Xe–NaCl mixture.

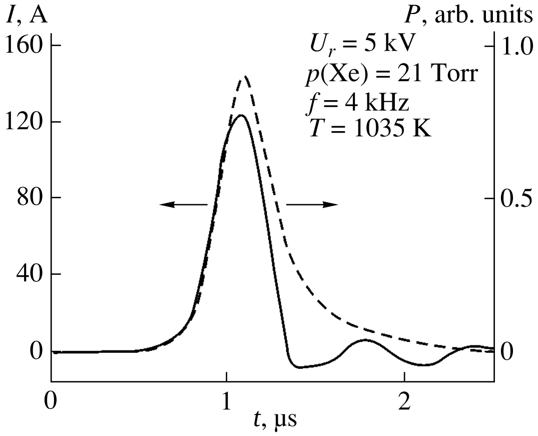


Fig. 2. Temporal shapes of current $I(t)$ and excimer emission $P(t)$ pulses.

tors FÉU-106 and 14ÉLU-FS, a power meter IMO-2N, and an oscilloscope S1-99.

SPECTRAL AND TEMPORAL FEATURES OF RADIATION FROM THE SOURCE

The most interesting feature of our studies is that not only beam pumping (proposed in [1]), but also the pumping of the Xe–NaCl lamp by a pulse-periodic discharge, leads to effective emission of UV radiation. Figure 1 shows the time-integrated emission spectrum $J(\lambda)$ of the excilamp. Here and below, the experimental conditions are indicated directly in the figure. The spectrum is confined to the segment 290–340 nm. The power emitted outside this range is insignificant. In this spectrum, the emission of the 308-nm XeCl excimer band is predominant. In addition to this band, the lines 330.2 + 330.3 nm of atomic sodium ($4p \longrightarrow 3s$ transition), as well as resonance lines 324.7 + 327.4 nm of the copper atom ($4p \longrightarrow 4s$ transition), are present in this spectral range. The emission of sodium is determined by the chemical composition of the working mixture used, while the emergence of copper in the discharge is stimulated by sputtering of copper electrodes as a result of ion bombardment and gas-transport reactions in the presence of the halogen. Such a "purity" of the spectrum is typical of discharges in inert gases with an impurity of metals, e.g., in lasers on self-terminating transitions. During heating, the spectral lines of inert gases gradually disappear in the emission spectrum almost completely in both cases. This effect is due to a decrease in the electron temperature in the discharge upon an increase in the concentration of easily ionized particles (in particular, atoms of metals).

The installation of a flat resonator with dielectric-coated mirrors with reflectances of 98% (totally reflecting mirror) and 30% (output mirror) at a wavelength of 308 nm did not change the shape of the spectral distribution of radiation within the excimer band. In particular, the band width at half-amplitude remained unchanged (the value of this quantity was ~5 nm, as in the cycle of publications [2, 3], with a plasma source) and was insensitive to change in the experimental conditions. This means that stimulated radiation was absent as expected, at least due to the low pressure of the mixture.

Figure 2 shows the temporal shapes of current $I(t)$ and power $P(t)$ pulses for excimer emission. The durations of the current pulse and the emission pulse at the base were ~400 ns and ~1 μs, respectively. The emission of the excimer band begins simultaneously with the evolution of current in the GDT, and the major part of radiation energy corresponds to the current pulse of the discharge.

The spectral structure of the excimer band unambiguously indicates that XeCl* molecules are generated in discharge in the lower vibrational states of the B term. However, in such a case, the excimer molecule can be formed only as a result of a binary substitution reaction of the type

$$\mathrm{Xe}^+ + \mathrm{NaCl} \longrightarrow \mathrm{XeCl}^* + \mathrm{Na}^+, \quad (1)$$

$$\mathrm{Xe}^* + \mathrm{NaCl} \longrightarrow \mathrm{XeCl}^* + \mathrm{Na}, \quad (2)$$

and not as a result of a three-particle ion–ion recombination reaction. The rate constants of these reactions were estimated in [1–3] at a level of $k = 10^{-10}$–10^{-9} cm^3/s. Although this estimate is insufficiently exact, we will use the average value ~5×10^{-10} cm^3/s. Provided that Xe$^+$ and Xe* leave the volume of the GDT exclusively as a result of pumping reactions (1) and (2), we find that the intensity decay time constant of excimer emission in afterglow amounts to $\tau = [kN(\mathrm{NaCl})]^{-1} = 5 \times 10^{-7}$ s. In this relation, N(NaCl) is the concentration of rock salt molecules, which is determined by the temperature in the GDT. In spite of its approximate nature, this estimate is in good agreement with the experimental data (Fig. 2) and, hence,

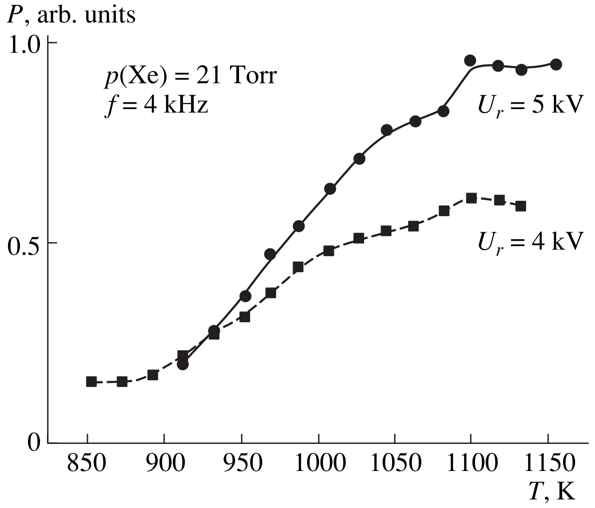


Fig. 3. Dependence of peak pulsed radiation power *P* of the excilamp on the temperature in the GDT.

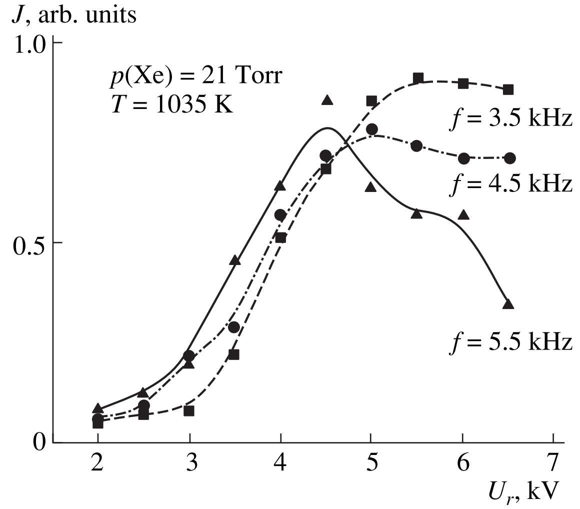


Fig. 4. Dependence of peak pulsed radiation power *P* of the excilamp on the rectifier voltage.

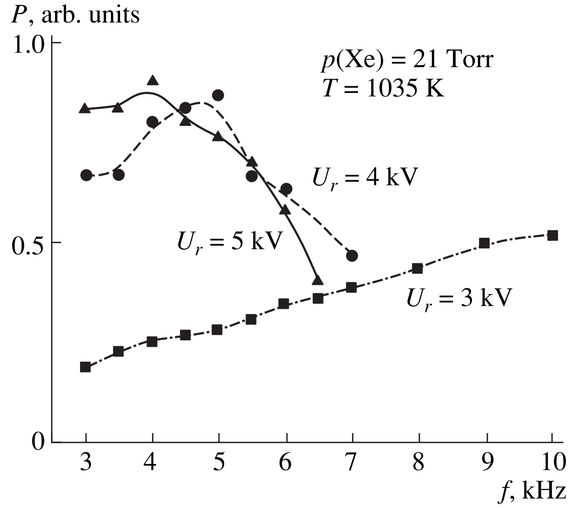


Fig. 5. Frequency dependence of peak pulsed radiation power *P* of the excilamp.

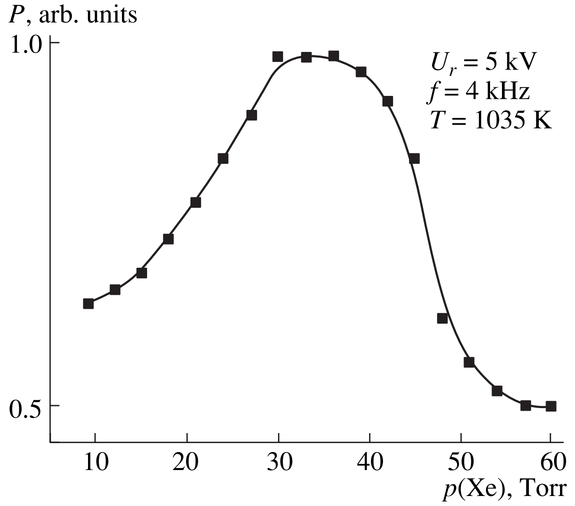


Fig. 6. Dependence of peak pulsed radiation power *p* of the excilamp on the xenon pressure.

may serve as additional evidence of the fact that pumping in our conditions is executed by substitution reactions.

OPTIMIZATION OF CONDITIONS FOR EXCITATION OF EXCIMER EMISSION

The external factors determining the UV radiation yield include xenon pressure *p*(Xe), GDT wall temperature *T*, voltage U_r of the rectifier, and excitation pulse repetition rate *f*. The combined action of all these factors ultimately determines the efficiency of the emitter. The results of measurements of peak pulse radiation power *P* of the excilamp as a function of the above-mentioned parameters are represented in Figs. 3–6. Most of the obtained dependences have extrema in the measuring range. This allows us to determine the factor point with the following parameters: pulse repetition rate *f* = 5 kHz, rectifier voltage U_r = 5 kV, xenon pressure *p*(Xe) = 30 Torr, and NaCl vapor pressure of 0.1 Torr (*T* = 1045 K); this point ensures the maximal emissive power of UV radiation for the emitter used in our experiments.

The mean power of the excimer emission of the lamp was estimated on the basis of the corresponding measurements from one of the GDT windows and taking into account the geometrical factor for the above optimal conditions. This mean power of spontaneous emission of the lamp amounted to several watts for an efficiency of ~1%.

We believe that the energy parameters of the emitter can be improved to a certain extent by choosing the appropriate mixture composition, increasing the working volume, etc. We also plan to employ a continuous glow discharge for creating an effective cw source of UV radiation on a Xe–NaCl mixture.

CONCLUSIONS

We developed an excilamp operating on the 308-nm transition of the XeCl molecule excited by a high-voltage pulse-periodic discharge on the basis of a mixture with nontoxic halogen donor Xe–NaCl. The mean radiation power of the 308-nm excimer band of XeCl was several watts. The spectral energy distribution in the band and the intensity decay rate of excimer emission in afterglow speak in favor of the fact that binary reactions of substitution of Xe for Na in the NaCl molecule play a dominating role in the population of the upper state of the XeCl* molecule.

REFERENCES

1. A. M. Boĭchenko, Kvantovaya Élektron. (Moscow) **29**, 163 (1999).
2. A. A. Alekhin, V. A. Barinov, Yu. V. Geras’ko, *et al.*, Zh. Tekh. Fiz. **63** (2), 43 (1993) [Tech. Phys. **38**, 80 (1993)].
3. A. A. Alekhin, V. A. Barinov, Yu. V. Geras’ko, *et al.*, Zh. Tekh. Fiz. **65** (5), 9 (1995) [Tech. Phys. **40**, 409 (1995)].
4. *Tables of Physical Data: Reference Book*, Ed. by I. K. Kikoin (Atomizdat, Moscow, 1976) [in Russian].

Translated by N. Wadhwa

Technical Physics, Vol. 49, No. 10, 2004, pp. 1339–1348. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 94–102.
Original Russian Text Copyright © 2004 by Dmitriev.

ACOUSTICS, ACOUSTOELECTRONICS

The Theory of Coupled Waves as a Universal Method for Designing Devices Operating on Surface Acoustic Waves

V. F. Dmitriev
Joint Stock Company Research Institute of Systems Engineering, St. Petersburg, 192102 Russia
e-mail: niis@mail.wplus.net

Received June 17, 2003

Abstract—A theory of devices operating on surface acoustic waves is proposed on the basis of modified equations for coupled modes. As a possible application of modified equations for coupled modes, a theory of dispersion acoustoelectronic delay lines with a curvilinear central line of electrode apertures is developed for interdigital transducers and reflector structures. Experimental studies are carried out and the experimental data are compared with the results of calculations. The experiments on the formation and compression of a linear frequency-modulated signal reveal a level of 38 dB for the side lobes in a compressed signal. *© 2004 MAIK “Nauka/Interperiodica”.*

INTRODUCTION

The coupled mode method is one of the effective approaches for designing devices operating on surface acoustic waves (SAW). The conventional method of coupled modes (see, for example, [1]), which is based on the derivation of a system of nonhomogeneous differential equations, unjustly complicates the solution of the problem of designing SAW-based devices with arbitrarily varying parameters. In the framework of this theory, it is difficult to take into account such factors as the varying period of the structure, arbitrary polarity of the electrodes connected to contact buses, apodization, and nonuniform broadening of the surface charge distribution over the electrodes of the structure. All these factors can be easily taken into account using a method based on modified equations for coupled modes operating with a structural unit. In addition, the proposed method is more promising as regards further complication of the initial model of the structure.

The theory developed here can be used for designing various SAW-based devices such as filters, dispersion delay lines, resonators, and appliances based on such devices. As piezoelectric elements, use can be made of strong piezoelectrics of the niobate and lithium tantalate type as well as weak piezoelectrics such as quartz or langasite. It should be noted that SAW-based filters of any type can be designed, including transversal filters, resonant filters operating on longitudinal modes, resonant filters on transverse modes, and staircase filters based on SAW resonators.

In [2], modified equations for coupled waves were used for constructing a theory for SAW-based filters using dispersion transducers with a staircase arrangement of topological elements. In [3], a theory was developed for filters operating on longitudinal resonance modes, while the theory for staircase filters using SAW resonators was described in [4]. In [5], the modified equations for coupled waves were used for constructing the theory of filters operating on weakly coupled longitudinal resonance modes.

Here, the modified equations for coupled waves are used to develop a theory for dispersion delay lines with a curvilinear central line of apertures of topological elements of interdigital transducers (IDTs) and reflector structures (RS’s).

The theory of dispersion acoustoelectronic delay lines (DADLs) with reflector structures in the form of stripes or grooves was proposed in [6] and in a simplified form in [7]. The theory proposed in [7] was based on the results of theoretical analysis of IMCON devices, which was carried out in [8]. The theory proposed in [6, 7] is based on the model of summation of waves reflected by each groove of one or two reflector structures. At the same time, the theory described in [6, 7] disregarded second-order effects. The most significant of such effects are multiple reflections in the reflector structure, phase velocity dispersion in the reflector structure, transformation of SAWs into bulk acoustic waves on reflecting elements, and diffraction of SAWs. In addition, the theory described in [6, 7] was based on a simplified model of calculation of IDT.

The empirical model proposed in [9, 10] makes it possible to take into account the effect of the mechanisms of transformation of SAWs into bulk acoustic waves in the theory of DADLs presented in [6, 7]. However, this model is simplified by nature and is valid only for DADLs with a linear topology.

The so-called slanted topology of a DADL was proposed in [11]. The use of slanted topology suspends to a certain extent the degradation of the frequency characteristics of wide-band DADLs with a large dispersion delay. The theory of this type of structures is based on the theory of DADLs with a linear topology and is distinguished by the fact that the structure splits into channels. Within each channel, calculations are made as for a DADL with a linear topology. Then the transfer coefficients for each channel are summed, which yields the total transmission coefficient for the DADL [11–13].

In this paper, the theory of a DADL with a curvilinear central line of electrodes in the DT and RS is constructed on the basis of the method of splitting of the initial structure into channels with subsequent summation of partial conductivities of each channel, which are calculated using the modified equations for coupled SAWs.

MODIFIED EQUATIONS FOR COUPLED SURFACE ACOUSTIC WAVES

Let the structure be given in the form of N metallic electrodes with arbitrarily alternating polarity, arbitrarily varying period, and overlapping of neighboring electrodes on the surface of a semi-infinite uniformly isotropic piezoelectric (Fig. 1). We also assume that a source of the signal with frequency ω and amplitude U_0 is connected on the left.

Let us consider the kth electrode of the IDT (Fig. 2). Let $R(z, \omega)$ and $S(z, \omega)$ be two coupled plane waves with wave number κ, which propagate in the electrode structure of the IDT so that $R(z, \omega)$ propagates in the direction of the z axis (direct wave) and $S(z, \omega)$ propagates in the direction opposite to the z axis (backward wave). Uniform plane waves can be represented in the form

$$R(z, \omega) = R(\omega)\exp(-j\kappa z), \tag{1}$$

$$S(z, \omega) = S(\omega)\exp(+j\kappa z), \tag{2}$$

where $R(\omega)$ and $S(\omega)$ are the complex amplitudes of the corresponding waves.

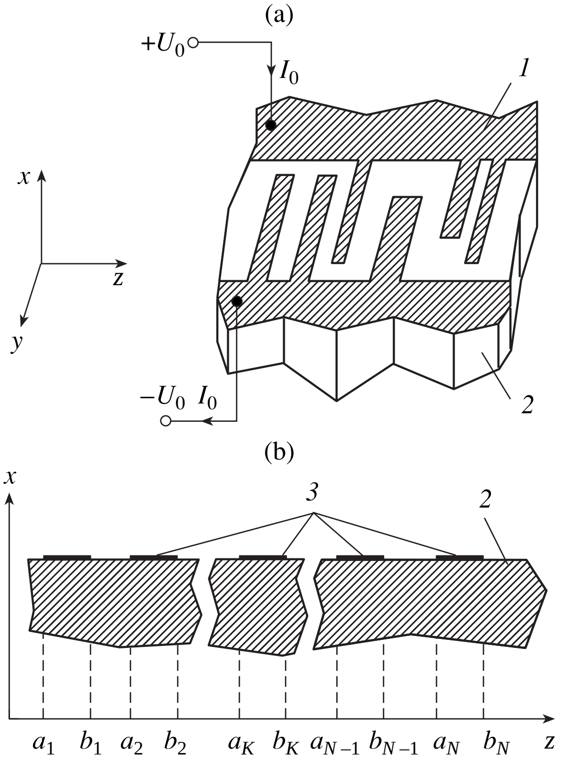


Fig. 1. (a) A fragment of a transducer at the surface of a semi-infinite piezoelectric and (b) a transducer in the xz plane. (*1*) Transducer fragment, (*2*) semi-infinite piezoelectric; (*3*) electrodes of the transducer; I_0 is the total current in the transducer.

Let waves $R_K(z, \omega)$ and $S_{K+1}(z, \omega)$ be incident on the kth electrode from left and right, respectively. Transmitted waves with complex amplitudes $S_K(\omega)$ and $R_{K+1}(z, \omega)$ are superpositions of partial waves formed due to reflection of wave $R_K(\omega)$ incident on the electrode, a part of wave $S_{K+1}(\omega)$ transmitted through the region of the kth electrode, and the wave formed due to the transformation of the energy of the ac electric field of frequency ω and amplitude U_0 produced by the source into a SAW.

Taking into account the corresponding phase factors as well as the mechanisms of reflection, transmission, and transformation of SAWs, we obtain the following expressions for the complex amplitudes of transmitted waves $S_K(\omega)$ and $R_{K+1}(\omega)$:

$$\begin{aligned} S_K(\omega) = {} & r_K\eta_{1K}(1-|\zeta_K^-|^2)^{1/2}\exp[-j(\kappa_E-\kappa_0)p_K]R_K(\omega) \\ & + \eta_{1K}(1-|r_K|^2)^{1/2}(1-|\zeta_K^+|^2)^{1/2} \\ & \times \exp[-j(\kappa_E-\kappa_0)p_K]S_{K+1}(\omega) \\ & + \xi_K^-(\kappa)\eta_{2K}\exp[-j(\kappa_E-\kappa_0)p_K/2]U_0, \end{aligned} \tag{3}$$

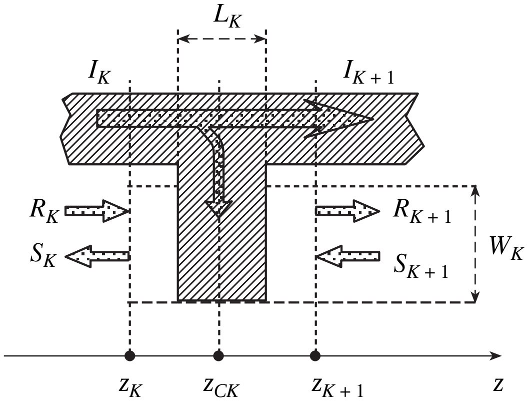


Fig. 2. The kth electrode of the transducer.

$$\begin{aligned}R_{K+1}(\omega) &= \eta_{1K}(1-|r_K|^2)^{1/2}(1-|\zeta_K^-|^2)^{1/2} \\ &\times \exp[-j(\kappa_E-\kappa_0)p_K]R_K(\omega) \\ &+ r_K\eta_{1K}(1-|\zeta_K^+|^2)^{1/2}\exp[-j(\kappa_E-\kappa_0)p_K]S_{K+1}(\omega) \\ &+ \xi_K^+(\kappa)\eta_{2K}\exp[-j(\kappa_E-\kappa_0)p_K/2]U_0.\end{aligned} \tag{4}$$

Here, r_K is the complex coefficient of reflection from the kth electrode; κ_E is the effective wave number of SAWs; $k_0 = 2\pi/p_K$, $p_K = z_{K+1} - z_K$; quantities $\xi_K^+(\kappa)$ and $\xi_K^-(\kappa)$ define the efficiency of direct transformation of SAWs at the kth electrode in directions $+z$ and $-z$, respectively; coefficients ξ_K^+ and ξ_K^- define the decrease in the amplitude of a wave upon its transition under the electrode due to inverse transformation; and $\eta_{1K} = W_{1K}/W_0$, $\eta_{2K} = W_{2K}/W_0$, where W_0 is the maximal aperture, W_{2K} is the overlap of the kth and $(k+1)$th electrodes; $W_{1K} = W_0$ if dummy electrodes are used; otherwise, $W_{1K} = W_{2K}$.

The phase factors of the terms associated with reflection (transformation) of waves determine the phase incursion from the center of wave reflection (transformation) z_{CK} to the corresponding boundary (z_K for $S_K(\omega)$ and z_{K+1} for $R_K(\omega)$, where z_K and z_{K+1} are the coordinates of the midpoint between the corresponding electrodes. The center of reflection (transformation) z_{CK} of SAWs is assumed to be at the center of the electrode.

We define the effective wave number on the segment between coordinates z_K and z_{K+1} as the mean value of the wavelength $\kappa_E = 2\pi p_K/[\lambda_0(p_K - L_K) + \lambda_M L_K] = \omega p_K/[V_0(p_K - L_K) + L_K V_M(h_M)] - j\alpha_K$, where V_0 is the SAW velocity on the free surface; λ_M and $V_m(h_M)$ are the SAW wavelength and velocity under the metallized surface, respectively; and α_K is the attenuation coefficient due to all loss factors operating during the SAW propagation in the electrode structure from coordinate z_K to coordinate z_{K+1} and reduced to unit length.

In contrast to analogous equations from [2–5], Eqs. (3) and (4) take into account the fact that the efficiencies of direct transformation of SAWs at the kth electrode in directions $+z$ and $-z$ may generally be different; i.e., $\xi_K^+(\kappa) \neq \xi_K^-(\kappa)$. This condition is determined both by the form of the Green's function of the SAW potential [14] and the redistribution of the surface current on the electrodes under the effect of the field of excited waves. In addition, in contrast to the equations considered in [2–5], Eqs. (3) and (4) take into account the decrease in the amplitude of waves as a result of their passage under the electrodes. Allowing for this effect is especially important in the case of strong piezoelectrics.

Current variations in the IDT bus are due to the transformation of direct and backward waves and the voltage drop across the electrode capacitance. The change in the current $\Delta I_K(\omega)$ in the contact IDT bus at the kth electrode is given by

$$\begin{aligned}\Delta I_K(\omega) &= I_K(\omega) - I_{K+1}(\omega) = [\xi_K^-(\kappa)]^*\eta_{2K} \\ &\times \exp[-j(\kappa_E-\kappa_0)p_K/2]R_K(\omega) \\ &+ [\xi_K^+(\kappa)]^*\eta_{2K}\exp[-j(\kappa_E-\kappa_0)p_K/2]S_{K+1}(\omega) \\ &+ \xi_K^-(\kappa)\eta_{2K}\exp[-j(\kappa_E-\kappa_0)p_K/2]S_K(\omega) \\ &+ \xi_K^+(\kappa)\eta_{2K}\exp[-j(\kappa_E-\kappa_0)p_K/2]R_{K+1}(\omega) \\ &+ j\omega\eta_{2K}(C_2/2)U_0,\end{aligned} \tag{5}$$

where the asterisk denotes a complex-conjugate quantity.

In accordance with the reciprocity principle, we assume in Eq. (5) that the functional dependences determining the direct and inverse transformation of SAWs are complex-conjugate quantities.

The first and second terms in Eq. (5) determine the current variation due to partial transformation of energy of a direct wave with amplitude $R_K(\omega)$ and backward wave with amplitude $S_{K+1}(\omega)$, which are incident on the electrode, into the energy of an ac electric field. The third and fourth terms in Eq. (5) determine the current variation due to the transformation of the energy of the ac electric field produced by the current into a direct wave with amplitude $S_K(\omega)$ and a backward wave with amplitude $R_{K+1}(\omega)$. The last term in Eq. (5) describes the change in the current in the IDT bus due to voltage drop across the static capacitance of the electrode, which is equal to $\eta_{2K}(C_2/2)$. Substituting Eqs. (3) and (4) into (5), we obtain

$$\begin{aligned}\Delta I_K(\omega) &= I_K(\omega) - I_{K+1}(\omega) = \eta_{2K}\exp\left[-j(\kappa_E-\kappa_0)\frac{p_K}{2}\right] \\ &\times \{[\xi_K^-(\kappa)]^* - \eta_{1K}(1-|\zeta_K|^2)^{1/2} \\ &\times [r_K\xi_K^+(\kappa) + (1-|r_K|^2)^{1/2}\xi_K^-(\kappa)] \\ &\times \exp[-j(\kappa_E-\kappa_0)p_K]\}R_K(\omega) \\ &+ \eta_{2K}\exp\left[-j(\kappa_E-\kappa_0)\frac{p_K}{2}\right]\{[\xi_K^+(\kappa)]^* \\ &- \eta_{1K}(1-|\zeta_K|^2)^{1/2}[r_K\xi_K^-(\kappa) + (1-|r_K|^2)^{1/2}\xi_K^+(\kappa)] \\ &\times \exp[-j(\kappa_E-\kappa_0)p_K]\}S_{K+1}(\omega) + \{j\omega\eta_{2K}(C_2/2) \\ &- \xi_K^+(\kappa)\xi_K^-(\kappa)\eta_{2K}^2\exp[-j(\kappa_E-\kappa_0)p_K]\}U_0.\end{aligned} \tag{6}$$

Let us consider the terms associated with the transformation of SAWs upon their passage through a IDT electrode. Frequency-dependent coefficients $\xi_K^+(\kappa)$ and $\xi_K^-(\kappa)$ determine the efficiency of transformation of the power of the voltage source, which is supplied to the *k*th electrode of the transducer, into the power of SAWs propagating in the +*z* and –*z* directions, respectively.

Using the Poynting complex theorem for the known distribution of the surface current on the IDT electrodes in the form $J(z, \kappa)$, we can obtain

$$[\xi_K^+(\kappa)]^2 = \frac{W_0}{U_0^2}\int_{a_K}^{b_K}\varphi_R^+(z,\kappa)J^*(z,\kappa)dz, \tag{7}$$

$$[\xi_K^-(\kappa)]^2 = \frac{W_0}{U_0^2}\int_{a_K}^{b_K}\varphi_S^-(z,\kappa)J^*(z,\kappa)dz. \tag{8}$$

Using the method of Green's functions, we can express potentials $\varphi_R^+(z, \kappa)$ and $\varphi_S^-(z, \kappa)$ in the form

$$\varphi_R^+(z,\kappa) = \frac{1}{-i\omega}\int_{-\infty}^{\infty}G^+(z-z_0,\kappa)J(z_0,\kappa)dz_0, \tag{9}$$

$$\varphi_S^-(z,\kappa) = \frac{1}{-i\omega}\int_{-\infty}^{\infty}G^-(z-z_0,\kappa)J(z_0,\kappa)dz_0, \tag{10}$$

where $G^+(z - z_0, \kappa)$ and $G^-(z - z_0, \kappa)$ are the Green's functions of a SAW for $\kappa > 0$ and $\kappa < 0$, respectively [14].

The self-consistent calculation of the surface current distribution $J(z, \kappa)$ on the electrodes, i.e., taking into account the edge effects, finite length of the IDT, and the back response for a piezoelectric, is described in [2].

Relations (3), (4), and (6) can be written in matrix form,

$$\begin{vmatrix} S_k(\omega) \\ R_{K+1}(\omega) \\ \Delta I_K(\omega) \end{vmatrix} = \begin{vmatrix} P^{(k)}(1,1) & P^{(k)}(2,1) & P^{(k)}(3,1) \\ P^{(k)}(2,1) & P^{(k)}(2,2) & P^{(k)}(3,1) \\ P^{(k)}(3,1) & P^{(k)}(2,3) & P^{(k)}(3,3) \end{vmatrix}\begin{vmatrix} R_K(\omega) \\ S_{K+1}(\omega) \\ U_0 \end{vmatrix},$$

where

$$\begin{aligned}
P^{(k)}(1,1) &= r_K\eta_{1K}(1-|\zeta_K^+|^2)^{1/2}\exp[-[j(\kappa_E-\kappa_0)]p_K];\\
P^{(k)}(1,2) &= +\eta_{1K}(1-|r_K|^2)^{1/2}(1-|\zeta_K^+|^2)^{1/2}\\
&\quad\times\exp[-j(\kappa_E-\kappa_0)p_K];\\
P^{(k)}(1,3) &= +\xi_K^-(\omega)\eta_{2K}\exp[-j(\kappa_E-\kappa_0)p_K/2];\\
P^{(k)}(2,1) &= \eta_{1K}(1-|r_K|^2)^{1/2}(1-|\zeta_K^+|^2)^{1/2}\\
&\quad\times\exp[-j(\kappa_E-\kappa_0)p_K];\\
P^{(k)}(2,2) &= +r_K\eta_{1K}(1-|\zeta_K^+|^2)^{1/2}\exp[-[j(\kappa_E-\kappa_0)]p_K];\\
P^{(k)}(2,3) &= +\xi_K^+(\omega)\eta_{2K}\exp[-j(\kappa_E-\kappa_0)p_K/2];\\
P^{(k)}(3,1) &= \eta_{2K}\exp\left[-j(\kappa_E-\kappa_0)\frac{p_K}{2}\right]\\
&\quad\times\{[\xi_K^-(\omega)]^*-\eta_{1K}(1-|\zeta_K|^2)^{1/2}[r_K\xi_K^+(\omega)\\
&\quad+(1-|r_K|^2)^{1/2}\xi_K^-(\omega)]\exp[-j(\kappa_E-\kappa_0)p_K]\};\\
P^{(k)}(3,2) &= +\eta_{2K}\exp\left[-j(\kappa_E-\kappa_0)\frac{p_K}{2}\right]\\
&\quad\times\{[\xi_K^+(\omega)]^*-\eta_{1K}(1-|\zeta_K|^2)^{1/2}[r_K\xi_K^-(\omega)\\
&\quad+(1-|r_K|^2)^{1/2}\xi_K^+(\omega)]\exp[-j(\kappa_E-\kappa_0)p_K]\};\\
P^{(k)}(3,3) &= j\omega\eta_{2K}(C_2/2)\\
&\quad-\xi_K^+(\omega)\xi_K^-(\omega)\eta_{2K}^2\exp[-j(\kappa_E-\kappa_0)p_K].
\end{aligned} \tag{11}$$

In this case, the *P* matrix of the IDT is determined on the whole by successive multiplication of the *P* matrices describing each electrode [3].

The above relations make it possible to calculate the initial conductance of the IDT as a part of a filter or a resonator with an arbitrarily varying period, the aperture of the electrodes along the IDT structure, an arbitrary direction of current in the electrodes, and the actual distribution of the surface current over IDT electrodes. It should be noted that the input conductance of the IDT is determined by element *P*(3, 3) of the overall *P* matrix of the SAW structure.

THE THEORY OF A DADL WITH A CURVILINEAR CENTRAL LINE OF IDT ELECTRODES AND REFLECTOR STRUCTURES

Let us consider a DADL with a curvilinear central line of the apertures of the IDT and RS electrodes (Fig. 3). We assume that the source of a sinusoidal signal of frequency ω, amplitude U_0, and internal resistance R_g is connected to transducer IDT-1. The signal detector with internal resistance R_n is connected to transducer IDT-2. Transducers IDT-1 and IDT-2 consist of electrodes with a smoothly varying period; each pair of electrodes excites a SAW with a definite frequency, which smoothly varies along the structures, and with a certain delay relative to the origin.

It should be noted that the law of variation of the period in the arrangement of IDT-2 electrodes must correspond to the period of arrangement of IDT-1 electrodes. The apertures of the electrodes in both IDT-1 and IDT-2 may be apodized independently in accordance with the required characteristics in the frequency range. The spatial arrangement, the period, and the aperture of reflector elements RA-1 and RA-2 must be

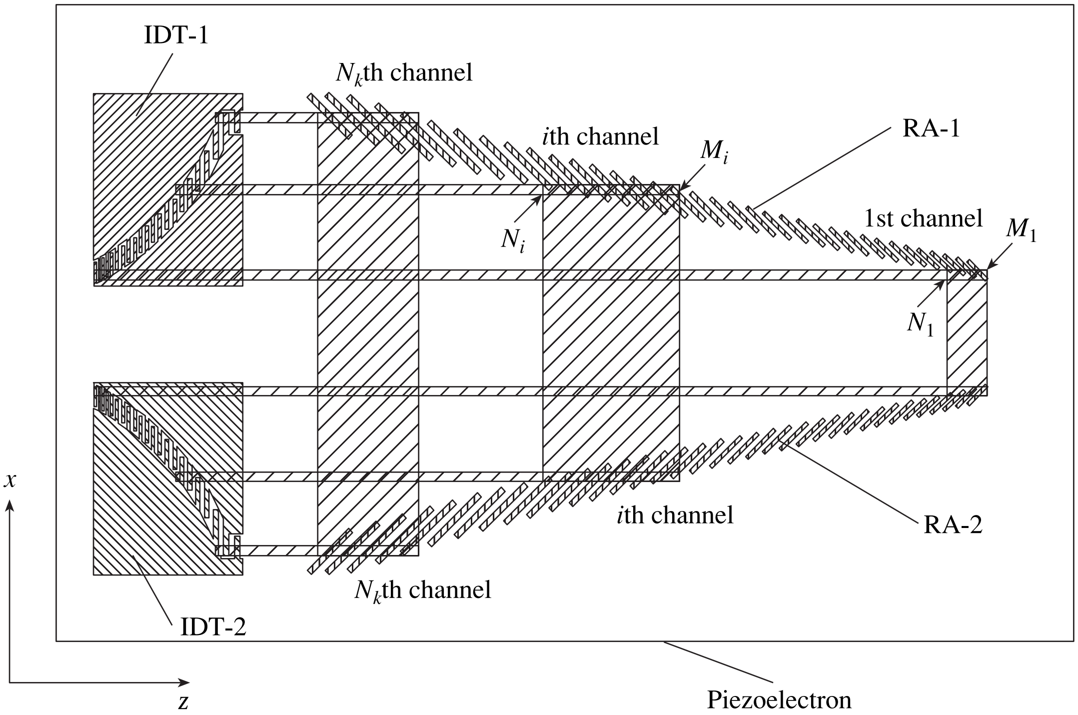


Fig. 3. DADL with a curvilinear central line of the apertures of the IDT and RS electrodes.

matched with the spatial arrangement and period of IDT-1 and IDT-2 electrodes, respectively. However, we do not impose any limitations on the spatial coordinates of the elements of DADL topology, since the theory presented here does not require such limitations. The above requirements must be met only for obtaining good characteristics of devices in the frequency and time ranges.

The theory of DADLs with a curvilinear mean line of the apertures of the IDT and RS electrodes involves the calculation of the conductivity matrix elements $Y(l_Y, m_Y)$ for the transducers and the transmission coefficient $K_{\mathrm{RS}}(\omega)$ of the reflector structure.

Let the mean line of IDT and RS electrodes of the dispersion delay line have a smoothly varying slope relative to the direction of propagation of SAWs along the structure (Fig. 3). The calculation of $Y(l_Y, m_Y)$ of a DADL with a tilted topology is based on a model in which the IDTs and RS's split into "channels" (Fig. 3). The splitting is carried out into N_K equal parts (only for convenience of calculations) in the overlap region IDT electrodes.

In accordance with the chosen model of partition into channels, the equivalent circuit of the DADL can be presented in the form shown in Fig. 4. The interaction between individual channels is accounted for by calculating $K^i_{\mathrm{RS}}(\omega)$ using the total aperture of the reflector element instead of the channel aperture. This assumption is substantiated by the fact that the transmission coefficient for the nearest channels slowly varies with channel number i and summation over all interacting channels is equivalent to an increase in the channel aperture.

At the first stage, we calculate the components of the conductivity matrix $\hat{Y}$ of the transducers in the DADL using a modified theory of coupled waves. We define the matrix components of the input conductivity of IDTs as the sum over all channels,

$$Y(l_Y, m_Y) = \sum_{i=1}^{N_K} Y_i(l_Y, m_Y), \tag{12}$$

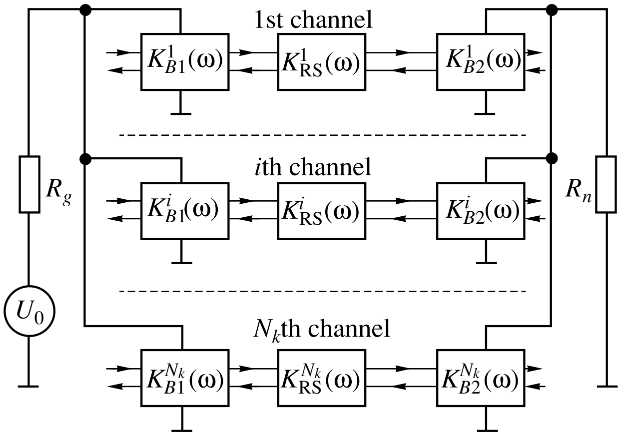


Fig. 4. Approximate equivalent electroacoustic circuit of a DADL with a curvilinear central line of the apertures of the IDT and RS electrodes.

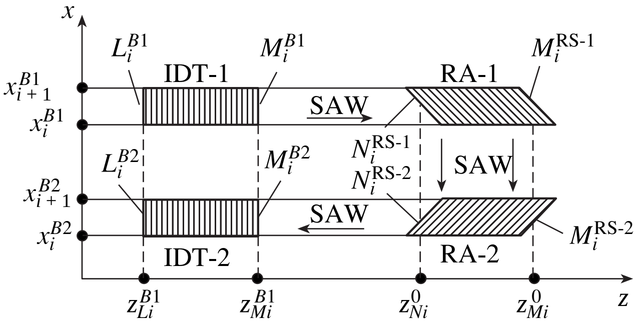


Fig. 5. The *i*th channel of a DADL with a curvilinear central line of the apertures of the IDT and RS electrodes.

where N_K is the number of channels; $Y(l_Y, m_Y)$ are the components of the conductivity matrix in the *i*th channel, $l_Y = 1, 2$ and $m_Y = 1, 2$.

We assume that the number of channels into which the structure splits is large enough for the result of calculation of frequency characteristics to remain unchanged upon an increase of N_K. It should be noted that the minimal channel width is bounded by a value equal to half the period of the structure.

The contribution $Y_i(l_Y, m_Y)$ from the *i*th channel to the total conductance of the SAW-based device can be determined by treating the *i*th channel as an independent SAW-based device (Fig. 5) and by using the components $P_i^{(B1)}(l_p, m_p)$ and $P_i^{(B2)}(l_p, m_p)$ of the P matrices of the input and output IDTs, as well as the transmission coefficients $K_{\text{RS}}^i(\omega)$ of the RS in the *i*th channel. This leads to the following expressions for the input conductance of the *i*th channel:

$$\begin{aligned} Y_i(1,1) &= P_i^{(B1)}(3,3) \\ &+ P_i^{(B2)}(1,1)P_i^{(B1)}(3,2)P_i^{(B1)}(2,3)/Y_{i0}, \end{aligned} \tag{13}$$

$$Y_i(1,2) = P_i^{(B1)}(3,2)P_i^{(B2)}(1,3)K_{\text{RS}}^i(\omega)/Y_{i0}, \tag{14}$$

$$Y_i(1,2) - P_i^{(B1)}(2,3)P_i^{(B2)}(3,1)K_{\text{RS}}^i(\omega)/Y_{i0}, \tag{15}$$

$$\begin{aligned} Y_i(2,2) &= P_i^{(B2)}(3,3) \\ &+ P_i^{(B1)}(2,2)P_i^{(B2)}(1,3)P_i^{(B2)}(3,1)/Y_{i0}, \end{aligned} \tag{16}$$

here, $Y_{i0} = [K_{\text{RS}}^i(\omega)]^2 - P_i^{(B1)}(2,2)P_i^{(B2)}(1,1)$.

Components $P_i^{(B1)}(l_p, m_p)$ and $P_i^{(B2)}(l_p, m_p)$ will be calculated using the coupled wave theory.

To synchronize the channels, we must take into account the initial phase for the first electrode of each IDT-1 (IDT-2) channel. For this purpose, we introduce the phase factors for the first electrodes of each channel into the components of the P matrix in relation (11):

$$P^{(L_i^{B1})}(1,1) = P^{(L_i^{B1})'}(1,1)[F_1^{(B1)}\{z_{i,1}^{(B1)}\}]^2, \tag{17}$$

$$P^{(L_i^{B1})}(1,2) = P^{(L_i^{B1})'}(1,2)F_1^{(B1)}\{z_{i,1}^{(B1)}\}, \tag{18}$$

$$P^{(L_i^{B1})}(1,3) = P^{(L_i^{B1})'}(1,3)F_1^{(B1)}\{z_{i,1}^{(B1)}\}, \tag{19}$$

$$P^{(L_i^{B1})}(2,1) = P^{(L_i^{B1})'}(1,2), \tag{20}$$

$$P^{(L_i^{B1})}(3,1) = P^{(L_i^{B1})'}(3,1)F_1^{(B1)}\{z_{i,1}^{(B1)}\}, \tag{21}$$

where the components of matrix $\hat{P}^{L_i^{(B1)'}}$ are defined by relations (11),

$$F_1^{(B1)}\{z_{i,1}^{(B1)}\} = \exp(j2\pi z_{i,1}^{(B1)}/\lambda_M - \alpha z_{i,1}^{(B1)}),$$

$$z_{i,1}^{(B1)} = z_{L_i}^{(B1)} - z_1^{(B1)}$$

is the distance between the first electrode in the *i*th channel of IDT-1 (with number L_i) and the first electrode in the entire IDT-1, and λ_M is the wavelength under the metallized surface; other components of the P matrix for the first electrode remain unchanged (as in relation (11)).

For the last electrodes in each channel of IDT-1, the components of the P matrix taking into account the phase factor have the form

$$P^{(M_i^{B1})}(1,2) = P^{(M_i^{B1})'}(1,2)F_2^{(B1)}\{z_{i,2}^{(B1)}\}, \tag{22}$$

$$P^{(M_i^{B1})}(1,2) = P^{(M_i^{B1})}(1,2), \tag{23}$$

$$P^{(M_i^{B1})}(2,2) = P^{(M_i^{B1})'}(2,2)[F_2^{(B1)}\{z_{i,2}^{(B1)}\}]^2, \tag{24}$$

$$P^{(M_i^{B1})}(2,3) = P^{(M_i^{B1})'}(2,3)F_2^{(B1)}\{z_{i,2}^{(B1)}\}, \tag{25}$$

$$P^{(M_i^{B1})}(3,2) = P^{(M_i^{B1})'}(3,2)F_2^{(B1)}\{z_{i,2}^{(B1)}\}, \tag{26}$$

where the components of matrix $\hat{P}^{(M_i^{B1})'}$ are defined by relation (11),

$$F_2^{(B1)}(z_{i,2}^{(B1)}) = \exp(j2\pi z_{i,2}^{(B1)}/\lambda_M - \alpha z_{i,2}^{(B1)}),$$

$$z_{i,2}^{(B1)} = z_N^{(B1)} - z_{M_i}^{(B1)}$$

is the distance between the last electrode of the entire IDT (with number N) and the last electrode in the *i*th channel (M_i), other components of the P matrix for the last electrode remaining unchanged.

Analogous expressions can also be derived for the first electrodes of IDT-2:

$$P^{(L_i^{B2})}(1,1) = P^{(L_i^{B2})'}(1,1)[F_1^{(B2)}\{z_{i,1}^{(B2)}\}]^2, \tag{27}$$

$$P^{(L_i^{B2})}(1,2) = P^{(L_i^{B2})'}(1,2)F_1^{(B2)}\{z_{i,1}^{(B2)}\}, \quad (28)$$

$$P^{(L_i^{B2})}(1,3) = P^{(L_i^{B2})'}(1,3)F_1^{(B2)}\{z_{i,1}^{(B2)}\}, \quad (29)$$

$$P^{(L_i^{B2})}(2,1) = P^{(L_i^{B2})'}(1,2), \quad (30)$$

$$P^{(L_i^{B2})}(3,1) = P^{(L_i^{B2})'}(3,1)F_1^{(B2)}\{z_{i,1}^{(B2)}\}. \quad (31)$$

Analogously, for the last electrode in each channel of IDT-2, we have

$$P^{(M_i^{B2})}(1,2) = P^{(M_i^{B2})'}(1,2)[F_2^{(B2)}\{z_{i,2}^{(B2)}\}], \quad (32)$$

$$P^{(M_i^{B2})}(1,2) = P^{(M_i^{B2})'}(1,2), \quad (33)$$

$$P^{(M_i^{B2})}(2,2) = P^{(M_i^{B2})'}(2,2)[F_2^{(B2)}\{z_{i,2}^{(B2)}\}]^2, \quad (34)$$

$$P^{(M_i^{B2})}(2,3) = P^{(M_i^{B2})'}(2,3)[F_2^{(B2)}\{z_{i,2}^{(B2)}\}], \quad (35)$$

$$P^{(M_i^{B2})}(3,2) = P^{(M_i^{B2})'}(3,2)[F_2^{(B2)}\{z_{i,2}^{(B2)}\}], \quad (36)$$

where the components of matrix $\hat{P}^{(L_i^{B2})'}$ are defined by relation (11).

We can now calculate the components $P_i^{(B1)}(l_p, m_p)$ and $P_i^{(B2)}(l_p, m_p)$ of the P matrix for the ith channel by multiplying the corresponding components for the electrodes in each channel,

$$P_i^{(B1)}(l_p, m_p) = \prod_{n = L_i^{B1}}^{M_i^{B1}} P_n^{(B1)}(l, m), \quad (37)$$

$$P_i^{(B2)}(l_p, m_p) = \prod_{n = L_i^{B2}}^{M_i^{B2}} P_n^{(B2)}(l, m), \quad (38)$$

where L_i^{B1}, L_i^{B2}, M_i^{B1}, and M_i^{B2} are the numbers of the first and last electrodes in each channel of IDT-1 and IDT-2, respectively (Fig. 5), and the symbol of the product implies the computation of successive products in accordance with Eqs. (8)–(16) from [3].

Let us now calculate the transmission coefficients $K_{RS}^i(\omega)$ of the reflector structure in the ith channel. Summation of partial waves reflected by inhomogeneities leads, in accordance with Fig. 4, to an expression for the transfer coefficient of the ith RS channel of the form

$$K_{RS}^i(\omega) \approx \sum_{n = N_i}^{M_i}\left\{\sum_{m = N_i}^{M_i} \gamma_{nm}\rho_m\rho_n\xi_n^i\xi_m^i \exp[-i\gamma_V(2z_{nm}^O - z_N^{B1} - z_N^{B2} + k_V x_n^O - k_V x_m^O) - i\varphi_{1nm}^i - i\varphi_{2nm}^i]\right\}, \quad (39)$$

where, in each ith channel, we sum the SAW passing from the first reflector element N_i of the channel to the last reflector element M_i of the channel; γ_{nm} is the overlap coefficient for the mth and nth grooves in the ith channel; ρ_n and ρ_m are the coefficients of reflection from the reflector element; ξ_n and ξ_m are the coefficient of SAW transmission through the reflector structure to the mth and nth grooves of the channel, respectively, which determine the fraction of power transmitted to the mth and nth groove; z_{nm}^O, x_n^O, z_m^O, and x_m^O are the coordinates of the overlap center for the nth and mth grooves, respectively; $\gamma_V = \beta - i\alpha(\omega)$, $\beta = \omega/V_{\parallel}$, $\alpha(\omega)$ is the attenuation coefficient for SAWs and $k_V = V_{\perp}/V_{\parallel}$, $V_{\perp}$ and $V_{\parallel}$ are the SAW velocities in the directions of z and x, respectively.

Let us consider the quantities appearing in relation (39) in greater detail. The transmission through the reflector element of the RS is accompanied by SAW reflection and scattering; since a fraction of SAW energy is transformed into bulk waves, the transmission coefficients are given by

$$\xi_m^i = \xi_{m0}^i \prod_{k = N_i}^{m}(1-\zeta_k)[1-\rho_k^2]^{1/2}, \quad (40)$$

$$\xi_n^i = \xi_{n0}^i \prod_{k = N_i}^{n}(1-\zeta_k)[1-\rho_k^2]^{1/2}, \quad (41)$$

where

$$\zeta_k = 1 - \exp[-\eta_k(\omega)(h_k/\lambda)]^2, \quad (42)$$

$$\xi_{m0}^i \approx \{(1-\zeta_m)[1-\rho_m^2]^{1/2}\}^{N_m}, \quad (43)$$

$$\xi_{n0}^i \approx \{(1-\zeta_n)[1-\rho_n^2]^{1/2}\}^{N_n}. \quad (44)$$

In relation (42), coefficient $\eta_k(\omega)$ defines the fraction of the power transferred into bulk waves on the nth reflector element of the RS, h_k is the depth (height) of the kth reflector element of the RS, and λ is the wavelength of SAWs at frequency $f = \omega/2\pi$. The method for calculating the frequency-dependent coefficient $\eta(\omega)$ is described in [15]. In formulas (43) and (44), N_n and N_m are the numbers of grooves through which a SAW passes during its propagation along the x axis in the upper and lower parts of the RS, respectively. Expressions (43) and (44) are written as approximate equalities, since the variation of ζ_m, ζ_n, ρ_m, and ρ_n is disregarded when ξ^i_{m0} and ξ^i_{n0} are calculated. Coefficients ξ_{m0} and ξ_{n0} account for attenuation of SAWs in the transverse direction and make a noticeable contribution only when an RS with a large aperture is used ($W_n^{RS} > 100\lambda$, where W_n^{RS} is the aperture of the nth groove of the RS, which is typical of DADLs with a linear topology).

Phase terms φ_{1nm} and φ_{2nm} in relation (39) take into account the change in the slope of the dispersion curve of the DADL upon a change in the velocity; in the case of sharp variations of the RS aperture (e.g., as a result of weight processing), they are responsible for deviations from a quadratic law on the phase–frequency characteristic. It should be noted that such distortions on the phase–frequency characteristic can be eliminated using "dummy" reflector elements.

It should be emphasized that the relations proposed above are also valid for designing DADLs with a linear topology as well as a curvilinear central line of electrode apertures and make it possible to account for the "large" aperture effect (SAW attenuation in the transverse direction of the RS due to scattering from reflector elements).

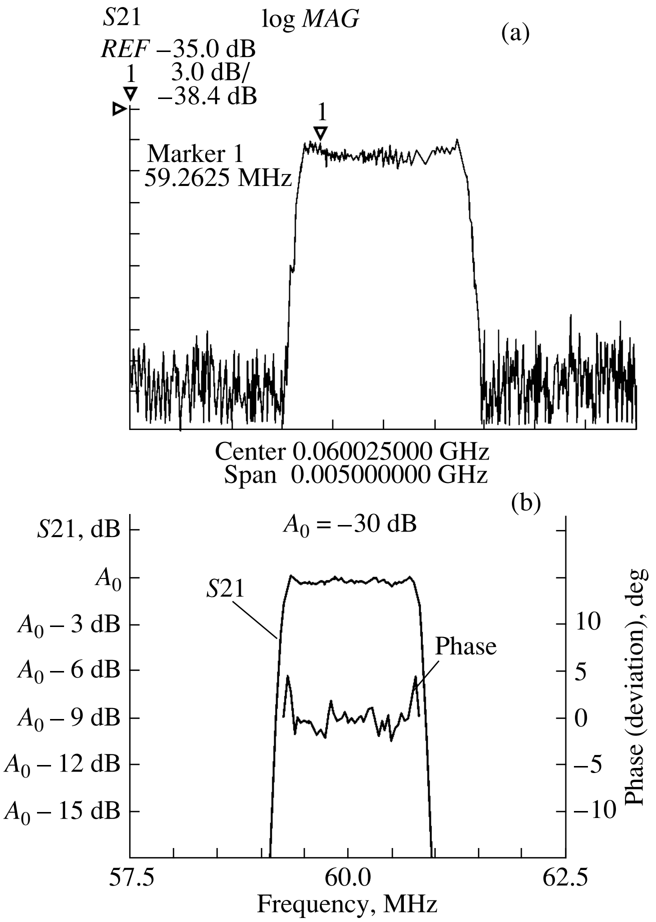


Fig. 6. The results of (a) measurements and (b) calculation of the frequency characteristic of a forming DADL. f_0 = 60 MHz, Δf = 1.5 MHz, T = 75 μs, and the slope of the dispersion curve is positive.

EXPERIMENTAL RESULTS

The theory described here was used for designing DADLs intended for the formation and compression of ultralong linear-frequency-modulated (LFM) signals with the following parameters: central frequency of 60 MHz, transmission band of 1.5 MHz, and LFM signal duration of 75 μs.

To construct a DADL with the above parameters, we choose a "linear" DADL topology with a reflector structure in the form of a sequence of grooves on the surface of a piezoelectric. To reduce distortions of the amplitude–frequency and phase–frequency characteristics in the reflector structures of the DADL, we used dummy grooves. For the piezoelectric material of the DADL, we chose lithium niobate of the *Y, Z* cut as the material ensuring minimal losses introduced into the DADL and the best quality of a compressed signal as compared to other suitable piezoelectric materials such as quartz or bismuth germanite. The transducer of the forming DADL consisted of 11 electrodes with an aperture of $120\lambda_0$. The reflector structure consisted of 5852 grooves with a depth of 0.14 μm, whose aperture corresponded to the aperture of the IDT electrodes. To suppress Fresnel pulsations, 15% of the RS lengths on the initial and final segments were apodized. To preserve the plane top of the amplitude–frequency characteristic and to keep the phase–frequency characteristic closest to the maximum possible extent to the required quadratic dependence in a frequency band of 1.5 MHz, the RS length (dispersion delay in the RS) was increased by 30%. To preserve the required slope of the dispersion characteristic, the RS frequency band at a level of –20 dB was also increased by 30%. Thus, the duration of the pulsed response of the DADL at a level of –3 dB was 86 μs for a frequency deviation in the spectrum of the LFM signal of ~1.65 MHz. The results of calculation of the transmission coefficient $S21(f)$ and the phase–frequency characteristic of the forming DADL are shown in Fig. 6b. The phase characteristic $\varphi(f)$ is given in the form of a deviation from the preset quadratic law,

$$\Delta\varphi(f) = \varphi(f) - \pi G_\Phi (f_{H\Phi} - f)^2 - 2\pi T_{H\Phi}(f_{H\Phi} - f),$$

where G_Φ = 0.0200 MHz/μs is the slope of the dispersion curve, $f_{H\Phi}$ = 59.250 MHz, and $T_{H\Phi}$ = 16.414 μs is the delay in the DADL–C at frequency $f_{H\Phi}$.

The frequency characteristics obtained on the experimental sample of forming DADL are shown in Fig. 6a. The amplitude–frequency characteristic was measured on the complex transmission coefficient meter NR-784. The working frequency band at a level of –3 dB was

1.65 MHz. The pulse response duration measured on the screen of the oscilloscope S1-108 was 86 μs at a level of –3 dB.

To obtain the minimal level of side lobes in a compressed signal from a compressing DADL (DADL–C), weight processing of the aperture of RS groove was used. Since the choice of weight processing is important for obtaining optimal parameters of a compressed signal, we discuss this in greater detail.

The decrease in the level of side lobes in a compressed signal due to weight processing is accompanied by a simultaneous slight expansion of the compressed signal. In the absence of weight processing, the minimal level of side lobes is –13.5 dB. The Maas weight function ensures the maximal suppression of side lobes in a compressed signal for its preset expansion due to weight processing. However, such a function cannot be physically realized since it has singularities at the edges. A physically realizable approximation of the Maas function is the Dolf–Chebyshev and Taylor functions, the Taylor function being more convenient for practical application in DADLs. The Taylor function is a practically realizable form of weight processing ensuring the minimal expansion of the compressed signal for a preset level of side lobes. The frequently used Hemming function is a rougher approximation of the Maas functions than the Taylor function. A disadvantage of the Taylor function is the absence of a decrease in the side lobe level α_s in the far-field zone (for $n \gg 1$). A decrease in the side lobe level in the far-field zone in proportion to $1/t$ is ensured by the Gauss function. The disadvantage of this function as compared to the Taylor function is a longer duration of a compressed signal at the same level of the nearest side lobes.

The contemporary technological level of manufacturing of DADLs makes it possible to obtain values of $\alpha_s \sim 40$ dB. Since the choice of the weight function with a larger theoretical level of α_s also leads to a stronger expansion of the compressed signal, the weight function is usually chosen with a theoretical value of α_s equal to 40–50 dB. In the DADL under investigation, the aperture of the RS grooves varies in accordance with the Gauss function with a pedestal of 0.008, which ensures the maximal attainable level of side lobe suppression in an ideal compressed LFM signal of –60 dB.

The results of calculation of the frequency characteristics of a compressing DADL are shown in Fig. 7a, while the frequency characteristics obtained on an experimental sample of DADL are shown in Fig. 7b. The values of $S21(f)$ were measured with the help of a complex transmission coefficient meter NR-784, while the phase–frequency characteristic was measured using a precision generator G4-176 and phase meter FK2-12. Phase characteristic $\varphi(f)$ of a compressing DADL is shown as a deviation from the preset quadratic law

$$\Delta\varphi(f) = \varphi(f) - \pi G_C(f_{HC} - f)^2 - 2\pi T_{HC}(f_{HC} - f),$$

where $G_C = -0.0200$ MHz/μs is the slope of the dispersion curve, $f_{HC} = 60.750$ MHz, and $T_{HC} = 7.87495$ μs is the delay of a compressing DADL at frequency f_H.

According to the results of measurement, the standard deviation of the phase characteristic from the preset law was 1.5°.

The experiments on the formation and compression of the LFM signal were performed with the DADLs whose frequency characteristics are shown in Figs. 6b and 7b. When a short rf pulse (of duration 0.1 μs) was fed to the input of the DADL–F with a positive slope of the dispersion curve, an LFM signal of duration 86 μs was obtained at the output of the forming DADL with a formed signal nonuniformity in the compressing DADL band of less than 1 dB. The formed LFM signal was amplified and fed to the input of a compressing DADL with a negative slope of the dispersion curve. As a result of the LFM signal compression, a level of side lobes of 38 dB was attained for a duration of the compressed signal of 1.1 μs at a level of –3 dB.

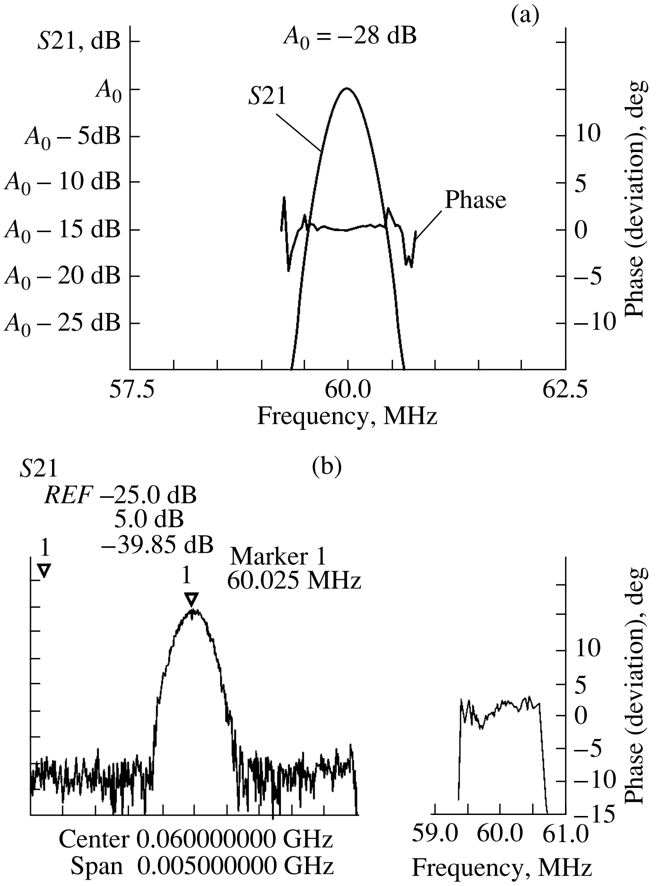


Fig. 7. The results of (a) measurements and (b) calculation of the transmission coefficient $S21$ and the phase–frequency characteristic of a compressing DADL. $f_0 = 60$ MHz, $\Delta f = 1.5$ MHz, $T = 75$ μs, the slope of the dispersion curve is negative, and the apodization function is Gaussian.

REFERENCES

1. S. V. Birykov, G. Martin, V. G. Polevoĭ, *et al.*, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 612 (1995).
2. V. F. Dmitriev, Zh. Tekh. Fiz. **72** (9), 93 (2002) [Tech. Phys. **47**, 1158 (2002)].
3. V. F. Dmitriev, Zh. Tekh. Fiz. **72** (11), 83 (2002) [Tech. Phys. **47**, 1427 (2002)].
4. V. F. Dmitriev, Zh. Tekh. Fiz. **72** (8), 95 (2002) [Tech. Phys. **47**, 1024 (2002)].
5. V. F. Dmitriev, Zh. Tekh. Fiz. **73** (2), 99 (2002) [Tech. Phys. **48**, 231 (2003)].
6. R. C. Williamson and H. I. Smith, IEEE Trans. Microwave Theory Tech. **21** (4), 195 (1973).
7. H. M. Gerard, O. W. Otto, and R. D. Weglein, *Proc. IEEE Ultrason. Symp*. (1974), pp. 197–202.
8. E. Sitting and C. Coquin, IEEE Trans. Sonics Ultrason. **15**, 111 (1968).
9. O. W. Otto and H. M. Gerard, Proc. IEEE Ultrason. Symp. (1977), pp. 596–401.
10. O. W. Otto and H. M. Gerard, J. Appl. Phys. **49**, 3337 (1978).
11. C. Waterkeyn and Gautier, Ultrason. Symp. (1983), pp. 195–199.
12. B. P. Potter and C. S. Hartmann, IEEE Trans. Sonics Ultrason. **26**, 411 (1979).
13. V. F. Dmitriev and I. S. Mitrofanov, in *Proceedings of the World Congress on Ultrasonics, Berlin, 1995*, pp. 435–438.
14. V. F. Dmitriev, Radiotekh. Élektron. (Moscow) **37**, 730 (1992).
15. V. F. Dmitriev, Zh. Tekh. Fiz. **65** (8), 111 (1995) [Tech. Phys. **40**, 809 (1995)].

Translated by N. Wadhwa

Technical Physics, Vol. 49, No. 10, 2004, pp. 1349–1353. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 103–107.
Original Russian Text Copyright © 2004 by Bulgakov, Meriutz, Ol'khovskiĭ.

Surface Electromagnetic Waves at the Interface between Two Dielectric Superlattices

A. A. Bulgakov*, A. V. Meriutz, and E. A. Ol'khovskiĭ****

** Usikov Institute of Radiophysics and Electronics, National Academy of Sciences of Ukraine,*
Kharkov, 61085 Ukraine

e-mail: bulgakov@ire.kharkov.ua

*** Kharkov Polytechnical Institute National Technical University,*
Kharkov, 61002 Ukraine

Received August 20, 2003

Abstract—The electrodynamic properties of the interface between two periodic dielectric superlattices are studied. It is shown that the interface may serve as a guide for an electromagnetic wave whose field decays exponentially on both sides of the plane of the interface. The field and power flux distributions, as well as the frequency dependence of the field penetration depth, are studied. *© 2004 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Layered media have been attracting the attention of researchers for a long time, because they are common in nature and their properties differ from those of homogeneous materials. Artificial layered media are new-generation materials with technologically controllable properties. Of special interest are periodic, quasi-periodic, and random layered structures. The physical properties of the periodic structures have been the subject of investigation in recent decades. Typically, a symmetry break in such materials generates new types of waves, such as surface waves, whose energy concentrates within the symmetry break region.

Infinite periodic structures are studied using the Floquet [1] and Bloch [2] theorems. However, these are inapplicable to asymmetric structures. Finite media were concerned in [3]. Later, the Abeles theorem was proved [4]. Semi-infinite periodic media were considered in [5, 6]. It was shown [6] that the relationship between the field components in semi-infinite and infinite superlattices is the same.

Surface waves at the interface between homogeneous and periodic media were theoretically studied in a number of papers. In one of the pioneering works, Tamm [5] predicted the existence of surface states at the surface of a crystal. Rayleigh acoustic surface waves were described in [6, 7]. Optical surface waves were predicted in [8] and observed in [9, 10]. Electromagnetic waves in semiconductor superlattices were studied in [4, 11].

In this paper, we theoretically demonstrate that the propagation of electromagnetic waves along the interface between two various dielectric superlattices is feasible and study the electrodynamic properties of these waves. The properties of these waves are of interest, since they offer scope for data transfer along interfaces in layered information systems, for contactless quality control during manufacturing, etc.

DISPERSION RELATION

Consider a contact between two semi-infinite layered periodic structures. Let either consist of alternat-

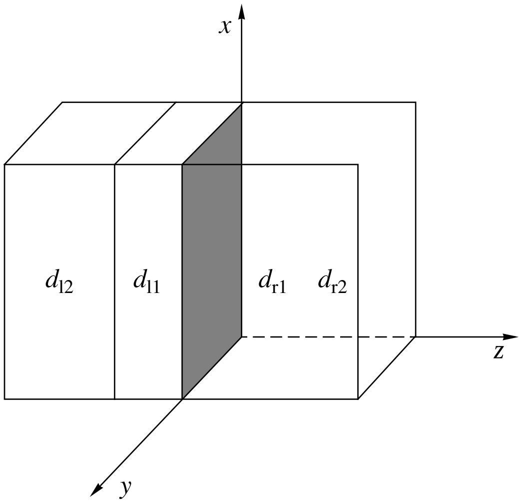


Fig. 1. Geometry of the problem (the interface is colored gray).

ing insulating layers of two sorts with different permittivities. The respective permittivities of one superlattice are $\varepsilon^{\mathrm{l1}}$ and $\varepsilon^{\mathrm{l2}}$; those of the other superlattice are $\varepsilon^{\mathrm{r1}}$ and $\varepsilon^{\mathrm{r2}}$ (here, superscripts “l” and “r” stand for the right- and left-hand superlattices, respectively). The geometry and coordinate system are shown in Fig. 1. We are interested in natural electromagnetic waves that propagate in this structure. These waves can be found by solving the Maxwell equations

$$\mathrm{curl}\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}, \quad \mathrm{curl}\mathbf{H} = \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} \tag{1}$$

for each of the layers.

At the interfaces between the layers and at the interface between the two superlattices, the boundary conditions involving the continuity of the tangential components of the electric and magnetic fields must be satisfied. Since the layers are assumed to be homogeneous along the y axis, we may put $\partial/\partial y = 0$. Then, the Maxwell equations are separated into two systems of equations for two independent polarizations with field components H_x, H_z, E_y and E_x, E_z, H_y. Below, we study the second type of polarization.

Consider the propagation of plane waves under the assumption that all the field components are proportional to $\exp(ik_x x + ik_z z - i\omega t)$. Substituting this dependence into the Maxwell equations yields the transverse wavenumbers for each of the layers:

$$k_z^{\alpha} = \sqrt{\frac{\omega^2}{c^2}\varepsilon_a - k_x^2} \quad (\alpha = l1, l2, r1, r2), \tag{2}$$

where c is the velocity of light.

The fields inside the layers are written as

$$H_y^{\alpha} = A_{\alpha}\exp(ik_z^{\alpha}z) + B_{\alpha}\exp(-ik_z^{\alpha}z). \tag{3}$$

From the Maxwell equations, components E_x^{α} and E_z^{α} can be expressed through H_y^{α}. It is convenient to describe the periodic structure by representing the fields through their values at $z = 0$. Below, we will use the transformation matrix m [4, 12]; for example, for the left-hand superlattice,

$$\begin{pmatrix} H_y^{\mathrm{l1}}(0) \\ E_x^{l1}(0) \end{pmatrix} = m^{(\mathrm{l})}\begin{pmatrix} H_y^{\mathrm{l2}}(d_{\mathrm{L}}) \\ E_x^{l2}(d_{\mathrm{L}}) \end{pmatrix}, \tag{4}$$

where d_{L} is the period of the corresponding (left-hand) superlattice.

Matrix m for the left-hand superlattice is given by

$$m^{(\mathrm{l})} = \begin{cases} m_{11}^{(\mathrm{L})} = \cos(k_z^{\mathrm{l1}}d_{l1})\cos(k_z^{\mathrm{l2}}d_{l2}) \\ -\dfrac{k_z^{\mathrm{l2}}\varepsilon^{\mathrm{l1}}}{\varepsilon^{\mathrm{l2}}k_z^{\mathrm{l1}}}\sin(k_z^{\mathrm{l1}}d_{l1})\sin(k_z^{\mathrm{l2}}d_{l2}), \\ m_{12}^{(\mathrm{L})} = -\dfrac{i\omega\varepsilon^{\mathrm{l2}}}{k_z^{l2}c}\sin(k_z^{\mathrm{l2}}d_{l2})\cos(k_z^{\mathrm{l1}}d_{l1}) \\ -\dfrac{i\omega\varepsilon^{\mathrm{l1}}}{k_z^{l1}c}\sin(k_z^{\mathrm{l1}}d_{l1})\cos(k_z^{\mathrm{l2}}d_{l2}), \\ m_{21}^{(\mathrm{L})} = \dfrac{k_z^{\mathrm{l1}}c}{i\omega\varepsilon^{l1}}\sin(k_z^{\mathrm{l1}}d_{l1})\cos(k_z^{\mathrm{l2}}d_{l2}) \\ +\dfrac{k_z^{\mathrm{l2}}c}{i\omega\varepsilon^{l2}}\cos(k_z^{\mathrm{l1}}d_{l1})\sin(k_z^{\mathrm{l2}}d_{l2}), \\ m_{22}^{(\mathrm{L})} = -\dfrac{k_z^{\mathrm{l1}}\varepsilon^{\mathrm{l2}}}{\varepsilon^{\mathrm{l1}}k_z^{\mathrm{l2}}}\sin(k_z^{\mathrm{l2}}d_{l2})\sin(k_z^{\mathrm{l1}}d_{l1}) \\ +\cos(k_z^{\mathrm{l1}}d_{l1})\cos(k_z^{\mathrm{l2}}d_{l2}). \end{cases} \tag{5}$$

For the right-hand superlattice, the transformation matrix is written in the same way.

To derive a dispersion relation, it is necessary to equate the tangential field components at the interface between the two superlattices at $z = 0$. In so doing, we will obtain two equations with four unknown coefficients. Two more equations can be obtained by establishing a relationship between components E_x^{α} and H_y^{α} inside either of the superlattices using the Floquet theorem:

$$\begin{pmatrix} H_y^{\mathrm{L}}(0) \\ E_x^{\mathrm{L}}(0) \end{pmatrix} = \begin{pmatrix} H_y^{\mathrm{L}}(0)\exp(i\bar{k}_{\mathrm{L}}d_{\mathrm{L}}) \\ E_x^{\mathrm{L}}(0)\exp(i\bar{k}_{\mathrm{L}}d_{\mathrm{L}}) \end{pmatrix}. \tag{6}$$

A similar formula can be derived for the right-hand superlattice. Here, $\bar{k}_{\mathrm{L,\,R}}$ are the Bloch wavenumbers for both superlattices, which can be found from the characteristic equation for either of the structures:

$$\begin{gathered} \exp(2i\bar{k}_{\mathrm{L,\,R}}d_{\mathrm{L,\,R}}) - (m_{11}^{(\mathrm{L,\,R})} + m_{22}^{(\mathrm{L,\,R})}) \\ \times\exp(i\bar{k}_{\mathrm{L,\,R}}d_{\mathrm{L,\,R}}) + 1 = 0. \end{gathered} \tag{7}$$

Our superlattices are semi-infinite; therefore, the Floquet theorem applies in this case, as was noted in the Introduction (the related arguments are adduced in [6]). Eventually, we arrive at a homogeneous system of equations with four unknowns. Equating the determinant of this system to zero, we get the dispersion relation

$$\frac{m_{12}^{(R)}\exp(i\bar{k}_R d_R)}{(1-m_{11}^{(R)}\exp(i\bar{k}_R d_R))}+\frac{m_{12}^{(L)}\exp(i\bar{k}_L d_L)}{(1-m_{11}^{(L)}\exp(i\bar{k}_L d_L))}=0. \tag{8}$$

It should be noted that surface waves may propagate along the interface if the fields fall ear from the interface between the superlattices. This condition is met when the opacity gaps of the superlattices overlap at least partially. The roots of Eq. (8) must be selected such that $\bar{k}_L$ and $\bar{k}_R$ lie in the opacity gaps and $\exp(i\bar{k}_L z)$ and $\exp(i\bar{k}_R z)$ decay with distance from the interface between the structures.

Figure 2 illustrates the zone pattern in both superlattices that is obtained by numerically solving Eq. (8) at $d_{l1} = 2.7 \times 10^{-4}$ m, $d_{l2} = 1.1 \times 10^{-4}$ m, $d_{r1} = 3.1 \times 10^{-4}$ m, $d_{r2} = 1.7 \times 10^{-4}$ m, $\varepsilon^{l1} = 10.5$, $\varepsilon^{l2} = 4.35$, $\varepsilon^{r1} = 8.5$, and $\varepsilon^{r2} = 2.5$. In Fig. 2, the frequency axes for the superlattices are coincident and the abscissa axes are directed oppositely. The zones in which the electromagnetic waves may propagate are hatched. Such a representation demonstrates the partially overlapping opacity gaps of the two lattices. It should be noted that there are two types of opacity gaps: the 0 gap and π gap. The former lies between the lines in which $\mathrm{Re}\bar{k}\,d = 2\pi n$; the latter, between the lines in which $\mathrm{Re}\bar{k}\,d = 2\pi n + \pi$. Figure 2 also plots the velocity-of-light lines $k_z^{r2} = 0$ (curves *1*) and $k_z^{l1} = 0$ (curves *2*).

Dispersion relation (7) was solved numerically. The results are shown in Fig. 2, where the thicker lines are the dispersion curves for the surface waves, which belong to the opacity gaps of both lattices. It is known that dispersion curves cannot break [13]. In this figure, the curves terminate when entering into the transmission zones in one of the superlattices (in the left or right plot). This means that the wave ceases to be surface in this superlattice.

SURFACE WAVE FIELDS AND FLUXES

In this section, we consider the field distribution in the layers for the surface waves associated with different zones. To find the fields, it is necessary to write the boundary conditions at the boundaries of the first periods of either lattice and take into account the relationship between field components $E_x^{L,R}$ and $H_y^{L,R}$ (6). Eventually, one obtains a system of eight equations, which are homogeneous algebraic equations. Let us express the field components through one of the coefficients, for example, A_{l1}, which is taken to equal unity. Then, one of the equations may be eliminated. By solving the system of seven equations, we find the unknown coefficients (see the Appendix).

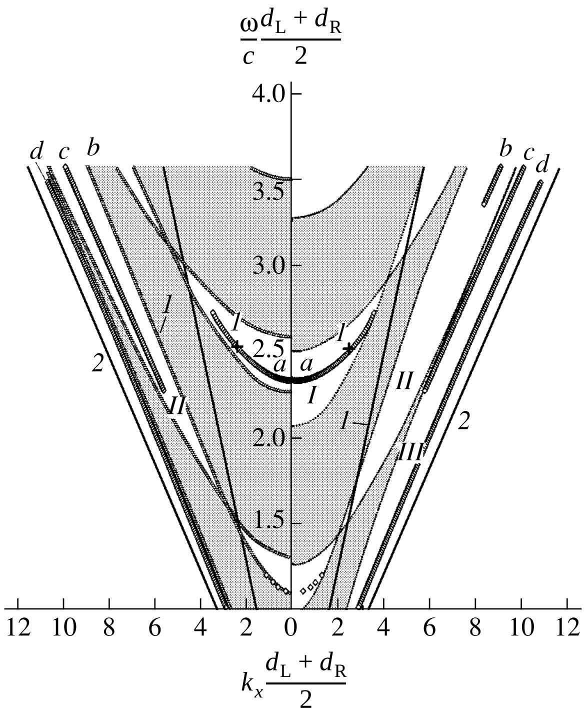


Fig. 2. Zone pattern in the lattices and the dispersion curves for the surface waves.

The surface wave fields associated with different opacity zones were studied numerically for the parameters given above. Wavenumbers k_z^{l1}, k_z^{l2}, k_z^{r1}, and k_z^{r2} may be both real and imaginary. This circumstance significantly affects the field distribution in the layers. If any of the transverse wave numbers is imaginary, the field distribution in the corresponding layer is of the surface type; i.e., the field amplitude reaches a maximum at the boundaries of this layer.

In the opacity zones, Bloch wavenumbers $\bar{k}_{L,R}$ are imaginary. By way of example, the surface wave field in the first zone is illustrated in Fig. 3. In the left-hand superlattice, the field decays with distance from the interface as $\exp(\mathrm{Im}\bar{k}_L z)$; in the right-hand superlattice, as $\exp(-\mathrm{Im}\bar{k}_R z)$. In this case, the period of the fields coincides with the period of the superlattices, because $\mathrm{Re}\bar{k}_{L,R} = 2\pi N$. Such a surface mode can be called the 0 mode.

If Bloch wavenumbers in both opacity zones can be represented as $\bar{k}_{L,R} + (2N+1)\pi/d_{L,R}$, we are dealing with the π mode, whose fields have a period twice as large as the period of the corresponding superlattice.

In Fig. 3, the superlattices are in contact when the permittivities of the layers at the interface are maximal: $\varepsilon^{l1} = 10.5$ and $\varepsilon^{r1} = 8.5$. In this case, at $z = 0$, the magnetic field is maximal, while the electric field turns to zero. The case where the permittivities of the layers at

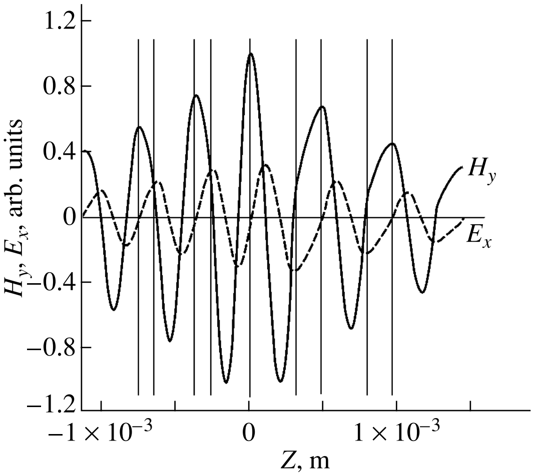


Fig. 3. Electromagnetic field distributions for the case where the contacting layers of the superlattices have maximal values of ε (point *1* in Fig. 2).

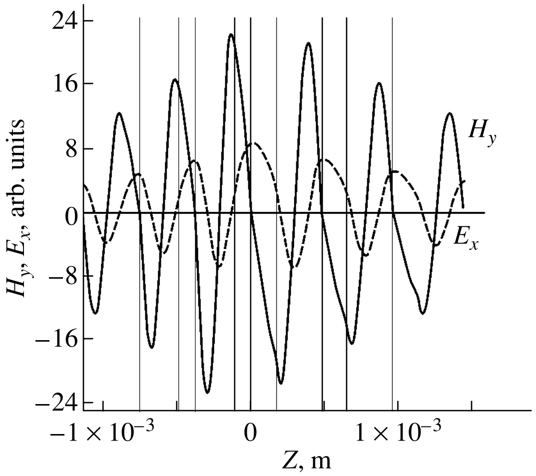


Fig. 4. Electromagnetic field distributions for the case where the contacting layers of the superlattices have minimal values of ε.

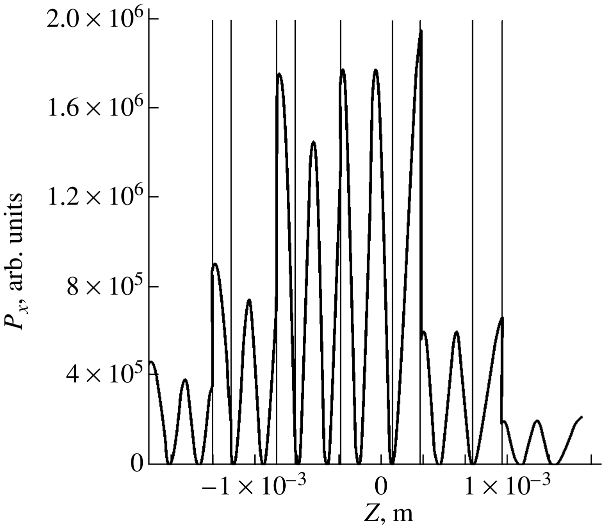


Fig. 5. Energy flux for point *1* in Fig. 2.

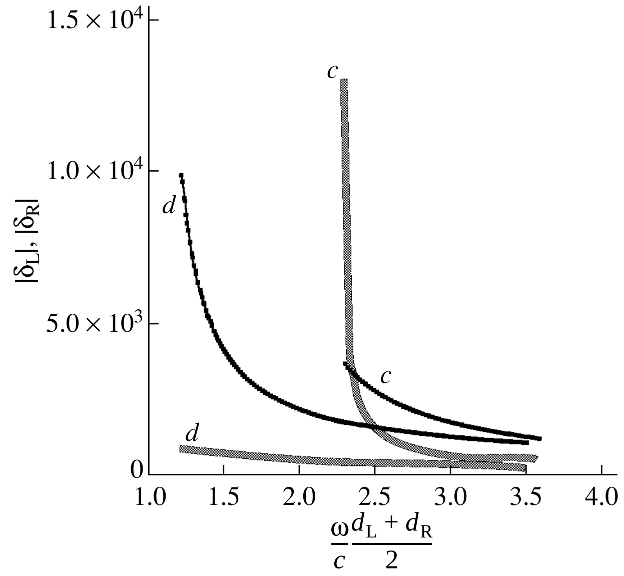


Fig. 6. Penetration depth for curves *c* and *d* in Fig. 1(thicker lines refer to the right-hand superlattice; thinner lines, to the left-hand superlattice).

the interface are minimal ($\varepsilon^{l1} = 4.35$ and $\varepsilon^{r1} = 2.5$) is shown in Fig. 4. Here, the situation is reverse: the electric field reaches a maximum at $z = 0$, while the magnetic field vanishes.

Dispersion curve *c* for the left-hand superlattice passes through the second opacity zone, where $\text{Re}\bar{k}_L = \pi$. For the right-hand structure, it falls into the third opacity zone, where $\text{Re}\bar{k}_R = 2\pi$. Accordingly (Fig. 2), the field pattern in the left-hand superlattice repeats each two periods; in the right-hand lattice, each alternate period. Furthermore, the fields in the given case differ in that their distribution in layers *l*1 and *r*1 is of the waveguide character, while in layers *l*2 and *r*2, of the surface type. It may be said that the fields leak (tunnel) through layers *l*2 and *r*2. This is the case of a mixed surface mode.

Figure 5 plots the energy flux along the *x* axis for point *1* in Fig. 2. The fluxes were calculated by the formula

$$P_x = -\frac{c}{4\pi}\text{Re}((H_y^{\alpha})^* E_z^{\alpha}). \tag{9}$$

The energy flux is maximal in the layers of the first period, i.e., in those adjacent to the interface, and exponentially decays with distance from the interface. The flux along the *z* axis turns out to be imaginary; hence, energy transfer in the *z* direction is absent.

It is of interest to consider the depth of electromagnetic wave penetration into the superlattices ($\delta_{\mathrm{L,R}} = 1/\bar{k}_{\mathrm{L,R}}$). The penetration depth increases as the dispersion curve approaches the boundaries of the opacity zones and becomes infinite when the surface wave changes to the bulk wave (Fig. 6). Figure 6 plots curves *c* and *d*, which lie in zones *II* and *III*.

CONCLUSIONS

We theoretically demonstrate that surface electromagnetic waves may propagate along the interface between two different superlattices. The wave field amplitudes exponentially decay with distance from the interface on either side. Three types of the surface waves are shown to exist: the 0 modes, the π modes, and the mixed modes. In the 0 mode, the field distribution repeats each alternate period of the superlattices; in the π mode, each two periods; and in the mixed mode, each alternate period in one of the superlattices and each two periods in the other. The energy flux distribution over the layers for the surface wave is analyzed, and the depth of field penetration into the superlattices is calculated. The penetration depth decreases as the dispersion curve associated with the surface wave moves away from the boundary of the transmission zone.

APPENDIX

The coefficients in the formulas for the electromagnetic field in the first layer of the left-hand superlattice are given by

$$A_2 = -iA_1\frac{\cos(k_z^{l1}d_{l1})\cos(k_z^{l2}d_{l1})\exp(i\bar{k}_{\mathrm{L}}d_{\mathrm{L}}) + \dfrac{k_z^{l1}\varepsilon^{l2}}{\varepsilon^{l1}k_z^{l2}}\sin(k_z^{l1}d_{l1})\sin(k_z^{l2}d_{l1})\exp(i\bar{k}_{\mathrm{L}}d_{\mathrm{L}}) - \cos(k_z^{l2}d_{\mathrm{L}})}{\sin(k_z^{l1}d_{l1})\cos(k_z^{l2}d_{l1})\exp(i\bar{k}_{\mathrm{L}}d_{\mathrm{L}}) - \dfrac{k_z^{l1}\varepsilon^{l2}}{\varepsilon^{l1}k_z^{l2}}\cos(k_z^{l1}d_{l1})\sin(k_z^{l2}d_{l1})\exp(i\bar{k}_{\mathrm{L}}d_{\mathrm{L}}) + \dfrac{k_z^{l1}\varepsilon^{l2}}{\varepsilon^{l1}k_z^{l2}}\sin(k_z^{l2}d_{\mathrm{L}})}, \quad \text{(A1)}$$

for the second layer of the left-hand superlattice,

$$B_1 = A_1\frac{\cos(k_z^{l2}d_{\mathrm{L}})}{\exp(i\bar{k}_{\mathrm{L}}d_{\mathrm{L}})} + iA_2\frac{k_z^{l1}\varepsilon^{l2}}{\varepsilon^{l1}k_z^{l2}}\frac{\sin(l_z^{l2}d_{\mathrm{L}})}{\exp(i\bar{k}_{\mathrm{L}}d_{\mathrm{L}})}, \quad \text{(A2)}$$

$$B_2 = iA_1\frac{\sin(k_z^{l2}d_{\mathrm{L}})}{\exp(i\bar{k}_{\mathrm{L}}d_{\mathrm{L}})} + A_2\frac{k_z^{l1}\varepsilon^{l2}}{\varepsilon^{l1}k_z^{l2}}\frac{\cos(l_z^{l2}d_{\mathrm{L}})}{\exp(i\bar{k}_{\mathrm{L}}d_{\mathrm{L}})}. \quad \text{(A3)}$$

Formulas for the right-hand superlattice are obtained by changing the respective subscripts.

REFERENCES

1. G. Floquet, Ann. de I'Ecole Normale, Ser. 2 **12**, 47 (1883).
2. F. Bloch, J. Physik **52**, 555 (1928).
3. L. Brillouin, *Wave Propagation in Periodic Structures* (Dover, New York, 1953; Inostrannaya Literatura, Moscow, 1959).
4. F. G. Bass, A. A. Bulgakov, and A. P. Tetervov, *High-Frequency Properties of Semiconductors with Superlattices* (Nauka, Moscow, 1989) [in Russian].
5. I. E. Tamm, Zh. Éksp. Teor. Fiz. **3**, 34 (1933).
6. I. M. Livshits and L. N. Rozentsveig, Zh. Éksp. Teor. Fiz. **18**, 1012 (1948).
7. A. A. Maradudin, Festkoerperprobleme **20**, 25 (1981).
8. D. Kossel, J. Opt. Soc. Am. **56**, 1434 (1966).
9. W. Ng, P. Yeh, P. Chen, and A. Yariv, Appl. Phys. Lett. **32**, 370 (1978).
10. A. A. Bulgakov and V. R. Kovtun, Opt. Spektrosk. **56**, 769 (1984) [Opt. Spectrosc. **56**, 471 (1984)].
11. Yu. A. Romanov, Zh. Tekh. Fiz. **42**, 1804 (1972) [Sov. Phys. Tech. Phys. **17**, 1447 (1972)].
12. M. Born and E. Wolf, *Principles of Optics*, 4th ed. (Pergamon, Oxford, 1969; Nauka, Moscow, 1973).
13. *Physical Acoustics: Principles and Methods*, Ed. by W. P. Mason (Academic, New York, 1968; Mir, Moscow, 1973), Vol. 1.

Translated by A. Khzmalyan

Technical Physics, Vol. 49, No. 10, 2004, pp. 1354–1359. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 108–112.
Original Russian Text Copyright © 2004 by Shagaev.

RADIOPHYSICS

A Method for Calculating the Characteristics of Surface Magnetostatic Waves in Anisotropic Ferromagnetic Films

V. V. Shagaev

Research Institute of Electronics Materials, Kaluga, 248650 Russia

e-mail: met@postklg.ru

Received December 30, 2003

Abstract—A method is proposed for obtaining an approximate dispersion equation for a surface magnetostatic wave in weakly anisotropic ferromagnetic films. The derived equation provides an explicit analytic description of the wave frequency on the tensor components of effective demagnetizing anisotropy factors. The approximate dispersion equation is used for studying the angular dependences of frequency in tangentially magnetized films of cubically anisotropic ferrites. The results can be used for developing spin-wave devices. © *2004 MAIK “Nauka/Interperiodica”.*

INTRODUCTION

Surface magnetostatic waves (SMSWs) propagating in thin ferrite films are studied in detail and used widely in spin-wave electronic devices [1–4]. The dispersion relations for waves being excited play an important role in the development and analysis of physical models for corresponding devices [5]. A surface wave in an isotropic tangentially magnetized film is the only magnetostatic mode propagating at right angles to the magnetizing field. It is possible to derive an analytic expression for the SMSW frequency. Allowance for the magnetic anisotropy of the film material changes dispersion relation qualitatively [6]. In particular, the spectrum of magnetostatic waves in the given geometry contains not only the surface mode, but also bulk modes, and the dispersion equations for frequency cannot be transformed in this case into an explicit analytic expression. This circumstance considerably complicates an analysis of wave characteristics.

This study is aimed at developing convenient methods for calculating the dispersion relations for SMSWs in anisotropic ferromagnetic films. We will derive an approximate dispersion equation and compare it with exact equations.

BASIC RELATIONS

The geometry of the problem is illustrated in Fig. 1. A ferromagnetic film of thickness d and of infinitely large size in the plane is magnetized to saturation by an external constant magnetic field. The magnetization vector $\mathbf{M}_0$ forms an angle θ with the plane of the film. A magnetization wave with wave vector $\mathbf{k}$ propagates in the plane of the film at an angle φ to the projection of $\mathbf{M}_0$. We use two systems of coordinates. In the system xyz, the y axis is parallel to the plane of the film and $z \parallel \mathbf{M}_0$. In the system $\xi\eta\zeta$, the ξ axis is directed along the normal $\mathbf{n}$ to the film and $\eta \parallel \mathbf{k}$. The dispersion equations (DEs) of magnetostatic waves are derived by integrating simultaneously the Maxwell equations in the magnetostatic approximation and the corresponding electrodynamic boundary conditions as well as the linearized equation of motion for magnetization, disregarding exchange and losses. As a result, we obtain the DEs

$$\tan\left[\frac{kd\sqrt{(\mu_{\xi\eta}+\mu_{\eta\xi})^2-4\mu_{\xi\xi}\mu_{\eta\eta}}}{2\mu_{\xi\xi}}\right] = \frac{\sqrt{(\mu_{\xi\eta}+\mu_{\eta\xi})^2-4\mu_{\xi\xi}\mu_{\eta\eta}}}{\mu_{\xi\eta}\mu_{\eta\xi}-\mu_{\xi\xi}\mu_{\eta\eta}-1}, \quad (1)$$

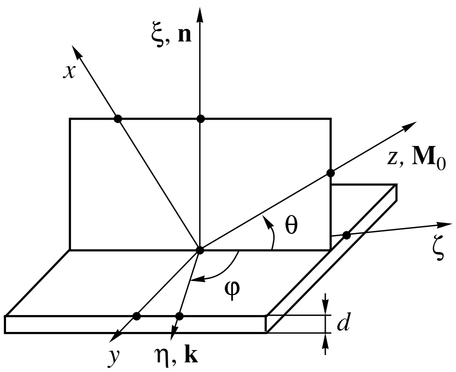


Fig. 1. Coordinate systems used in deriving the dispersion equations for magnetostatic waves.

1063-7842/04/4910-1354$26.00 © 2004 MAIK “Nauka/Interperiodica”

for $(\mu_{\xi\eta} + \mu_{\eta\xi})^2 \geq 4\mu_{\xi\xi}\mu_{\eta\eta}$ and

$$\exp\left[\frac{kd\sqrt{4\mu_{\xi\xi}\mu_{\eta\eta} - (\mu_{\xi\eta} + \mu_{\eta\xi})^2}}{\mu_{\xi\xi}}\right]$$
$$= \frac{\left[\sqrt{4\mu_{\xi\xi}\mu_{\eta\eta} - (\mu_{\xi\eta} + \mu_{\eta\xi})^2} - 2\right]^2 + (\mu_{\xi\eta} - \mu_{\eta\xi})^2}{\left[\sqrt{4\mu_{\xi\xi}\mu_{\eta\eta} - (\mu_{\xi\eta} + \mu_{\eta\xi})^2} + 2\right]^2 + (\mu_{\xi\eta} - \mu_{\eta\xi})^2} \tag{2}$$

for the opposite inequality.

In these equations, we use the components of the magnetic susceptibility tensor $\hat{\mu}$ of the film material in system $\xi\eta\zeta$. The expressions for the components of tensor $\hat{\mu}$ can be derived in the simplest way in the coordinate system xyz [7]:

$$\hat{\mu} = \begin{pmatrix} \mu_x & \mu_s + i\mu_a & 0 \\ \mu_s - i\mu_a & \mu_y & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\mu_x = 1 + \frac{4\pi M_0\omega_x}{\Omega}, \quad \mu_y = 1 + \frac{4\pi M_0\omega_y}{\Omega},$$

$$\mu_s = -\frac{4\pi M_0\omega_s}{\Omega}, \quad \mu_a = \frac{4\pi M_0}{\Omega}\left(\frac{f}{g}\right),$$

$$\omega_x = H_{\text{in}} + M_0(N^a_{yy} - N^a_{zz}),$$

$$\omega_y = H_{\text{in}} + M_0(N^a_{xx} - N^a_{zz}), \quad \omega_s = M_0 N^a_{xy},$$

$$H_{\text{in}} = \frac{(\mathbf{H}_e\mathbf{M}_0) - 4\pi(\mathbf{M}_0\mathbf{n})^2}{M_0}, \quad \Omega = \omega_x\omega_y - \omega_s^2 - \left(\frac{f}{g}\right)^2.$$

Here, g is the gyromagnetic ratio (it is sometimes used in the form of $2\pi g$; $g = 2.8$ MHz/Oe for the electron spin), $\mathbf{H}_e$ is the external magnetic field strength, N^a_{ij} are the tensor components of effective demagnetizing anisotropy factors, and f is the wave frequency. In passing to the coordinate system $\xi\eta\zeta$, the components of tensor $\hat{\mu}$ are transformed as follows:

$$\left.\begin{aligned} \mu_{\xi\xi} &= \mu_x\cos^2\theta + \sin^2\theta \\ \mu_{\eta\xi} &= [(1-\mu_x)\sin\theta\cos\varphi + (\mu_s + i\mu_a)\sin\varphi]\cos\theta \\ \mu_{\xi\eta} &= [(1-\mu_x)\sin\theta\cos\varphi + (\mu_s - i\mu_a)\sin\varphi]\cos\theta \\ \mu_{\eta\eta} &= (\mu_x\sin^2\theta + \cos^2\theta)\cos^2\varphi \\ &\quad - \mu_s\sin 2\varphi\sin\theta + \mu_y\sin^2\varphi. \end{aligned}\right\} \tag{3}$$

We will consider below the DEs for $\mathbf{M}_0 \perp \mathbf{n}$ and $\mathbf{k} \perp \mathbf{M}_0$; i.e., $\theta = 0$ and $\varphi = \pi/2$. Concerning the magnetization geometry, the following remark should be made. In an experimental study of properties of magnetostatic waves, the orientation of the external magnetizing field $\mathbf{H}_e$ and not the orientation of magnetization $\mathbf{M}_0$ is usually controlled. In an anisotropic uniformly magnetized film, for $\mathbf{H}_e \perp \mathbf{n}$ and $\mathbf{k} \perp \mathbf{H}_e$, angles θ and φ specifying the orientation of $\mathbf{M}_0$ may differ from values $\theta = 0$ and $\varphi = \pi/2$. It is clear from physical considerations that this difference is smaller the larger the strength of the magnetizing field as compared to the effective magnetic anisotropy fields is. A detailed analysis of this question forms a separate problem and is beyond the scope of our analysis of dispersion relations. It should only be noted that the formulas given above can be used for calculating the angular derivatives of frequency, which turn to zero for $\theta = 0$ and $\varphi = \pi/2$ ($df/d\theta = 0$ and $df/d\varphi = 0$). Consequently, the corrections to the dispersion relations, which are associated with allowance for small deviations of angles from the values considered here, will have the second order of smallness.

Assuming that conditions $\theta = 0$ and $\varphi = \pi/2$ are satisfied and substituting these values into expressions (3) and then into relations (1) and (2), we obtain, after simple transformations, the DEs required for our analysis.

(1) A surface wave (or, according to [3], a mixed wave for $N^a_{xy} \neq 0$) has the dispersion relation

$$kd = \frac{f^2 - f_0^2}{2\sqrt{(f^2 - f_1^2)(f^2 - f_2^2)}} \times \ln\left[1 + \frac{4\sqrt{(f^2 - f_1^2)(f^2 - f_2^2)}}{(4\pi M_0 g)^2 - \left(\sqrt{f^2 - f_1^2} + \sqrt{f^2 - f_2^2}\right)^2}\right], \tag{4}$$

where

$$\left(\frac{f_0}{g}\right)^2 = [H_{ez} + M_0(N^a_{yy} - N^a_{zz})][H_{ez} + 4\pi M_0 + M_0(N^a_{xx} - N^a_{zz})] - (M_0 N^a_{xy})^2, \tag{5}$$

$$\left(\frac{f_{1,2}}{g}\right)^2 = \left(\frac{f_0}{g}\right)^2 + 4\pi M_0^2\left[\frac{N^a_{xx} - N^a_{yy}}{2} \pm \sqrt{\frac{(N^a_{xx} - N^a_{yy})^2}{4} + (N^a_{xy})^2}\right]. \tag{6}$$

The spectrum of this type of wave occupies the frequency interval

$$f_1 < f < f_\infty = gH_{ez} + \frac{gM_0}{2}(4\pi + N^a_{xx} + N^a_{yy} - 2N^a_{zz}), \tag{7}$$

and the wave vectors are bounded by the values

$$kd > k_{svd} = \frac{2(f_1^2 - f_0^2)}{(4\pi M_0 g)^2 - (f_1^2 - f_2^2)}.$$

(2) The frequency interval $f_2 < f < f_0$ contains the spectrum of backward bulk waves, while the interval $f_0 < f < f_1$ contains the spectrum of direct bulk waves.

For $f = f_1$ and $k = k_{sv}$, the dispersion relation for the principal mode of a bulk wave is smoothly transformed into the dispersion relation for a surface wave. The DE for the principal mode has the form

$$kd = \frac{f^2 - f_0^2}{\sqrt{(f^2 - f_2^2)(f_1^2 - f^2)}}$$

$$\times \arctan\left[\frac{2\sqrt{(f^2 - f_2^2)(f_1^2 - f^2)}}{(4\pi M_0 g)^2 + f_1^2 + f_2^2 - 2f^2}\right], \qquad (8)$$

$$0 \le k \le k_{sv}, \quad f_0 \le f \le f_1.$$

We will compare the frequency intervals occupied by bulk waves and the wave formed by the principal mode of a direct bulk wave for $f_0 \le f \le f_1$ and a surface wave for $f_1 < f < f_\infty$ (this compound wave will be henceforth referred to as a SMSW):

$$\frac{f_1^2 - f_2^2}{f_\infty^2 - f_0^2} = \frac{16\pi\sqrt{(N_{xx}^a - N_{yy}^a)^2 + (2N_{xy}^a)^2}}{(4\pi + N_{xx}^a - N_{xy}^a)^2 + (2N_{xy}^a)^2}.$$

This expression implies that, in films of weakly anisotropic ferrites, for which $|M_0 N_{ij}^a| \ll 4\pi M_0$, the major part of the spectrum of magnetostatic waves with wave vectors perpendicular to the magnetization is occupied precisely by the SMSW.

APPROXIMATE DISPERSION EQUATION

Let us derive an explicit approximate expression for f from Eqs. (4) and (8). As the initial approximation, we choose the dispersion relation

$$f^2 = f_0^2 \exp(-2kd) + f_\infty^2[1 - \exp(-2kd)], \qquad (9)$$

where f_0 and f_∞ are defined by formulas (5) and (7).

Equation (9) successfully approximates the dispersion relation for a surface wave. Indeed, for the limiting values of the wave vector $k \longrightarrow 0$ and $k \longrightarrow \infty$, Eq. (9) gives correct expressions for the boundary frequencies of the spectrum of the SMSW. In addition, in some cases, Eq. (9) is exact for values $0 < k < \infty$ also. Such a coincidence is observed for isotropic ferromagnetic films when $N_{ij}^a = 0$ as well as for anisotropic films, when the magnetization vector is directed along high-symmetry axes characterized by the relation $N_{xy}^a = 0$ and $N_{xx}^a = N_{yy}^a$. For example, in monocrystalline films, such directions will coincide with the tetrad and higher order crystallographic symmetry axes. Thus, a further construction of an approximate DE will be associated with taking into account component N_{xy}^a and the difference $(N_{xx}^a - N_{yy}^a)$ in this equation. We transform Eq. (4) to the form (analogous transformations can also be carried out for Eq. (8))

$$kd = kd(f^2, \sigma, \varepsilon) \equiv \frac{f^2 - f_0^2}{2\Phi(f^2, \sigma, \varepsilon)} \ln\left\{1\right.$$

$$\left. + \frac{\Phi(f^2, \sigma, \varepsilon)[(4\pi M_0 g)^2 - 2(f^2 - f_0^2) + \sigma + 2\Phi(f^2, \sigma, \varepsilon)]}{(4\pi M_0 g)^2 (f_\infty^2 - f^2)}\right\},$$

where the following notation is introduced to simplify the form of the equation:

$$\Phi(f^2, \sigma, \varepsilon)$$

$$= \sqrt{(f^2 - f_0^2)^2 - (f^2 - f_0^2)\sigma - (4\pi M_0 g)^2 \varepsilon},$$

$$\sigma = 4\pi(M_0 g)^2 (N_{xx}^a - N_{yy}^a), \quad \varepsilon = (M_0 g N_{xy}^a)^2.$$

It can easily be verified that the transformed DE is converted into the initial approximation (9) for $\sigma = \varepsilon = 0$. The approximation linear in parameters σ and ε can be calculated by the formula

$$f^2 = f_0^2 \exp(-2kd) + f_\infty^2[1 - \exp(-2kd)]$$

$$-\left[\frac{\partial(kd)}{\partial f^2}\right]^{-1}\left[\frac{\partial(kd)}{\partial\sigma}\sigma + \frac{\partial(kd)}{\partial\varepsilon}\varepsilon\right].$$

Calculations and substitution of the expressions for the derivatives taken at $\sigma = \varepsilon = 0$ leads to the final relation,

$$f^2 = f_0^2 \exp(-2kd) + f_\infty^2[1 - \exp(-2kd)] \qquad (10)$$

$$- P(kd)\sigma - R(kd)\varepsilon,$$

where

$$P(kd) = \frac{\exp(-2kd)}{2}\left[\frac{2kd}{1 - \exp(-2kd)} - 1\right],$$

$$R(kd) = \frac{1}{[\exp(2kd) - 1]}$$

$$\times \left[\frac{4kd}{1 - \exp(-2kd)} + \exp(-2kd) - 3\right].$$

It should be noted that $R(kd) \longrightarrow 0$ both for $kd \longrightarrow 0$ and for $kd \longrightarrow \infty$ and that Eq. (10), analogously to Eq. (9), leads to exact expressions for the boundary frequencies of the SMSW spectrum. Dispersion relation (10) has a physically clear structure—it consists of the DE, whose form is typical of isotropic films or high-symmetry directions of the magnetization vector in an anisotropic film, and additional terms allowing for the "low-symmetry" parameters N_{xy}^a and $(N_{xx}^a - N_{yy}^a)$. Pay attention to the fact that the dependence on the magnetizing field strength is contained in Eq. (10) only in the "high-symmetry" part. This circumstance is in accordance with the physical concepts concerning the decrease in the effect of magnetic anisotropy as we pass to the limit $H_e \longrightarrow \infty$.

ANISOTROPY OF THE SPECTRUM OF SURFACE WAVES IN FILMS WITH CUBIC MAGNETIC ANISOTROPY

After the substitution of expressions for f_0, f_∞, σ, and ε into Eq. (10), the latter will explicitly define the dependence of f on kd as well as on the anisotropy tensor components N_{ij}^a. Thus, the application of an approximate DE instead of the exact DE allows us to substantially simplify the analysis of the effect of magnetic anisotropy of films on the spectrum of SMSWs.

Let us consider by way of an example films with a cubic magnetic anisotropy, which are oriented along the planes passing through an axis of the ⟨110⟩ type. It should be noted in connection with the problem being considered that the basic material used in spin-wave electronics is yttrium iron garnet (YIG, $Y_2Fe_5O_{12}$) belonging to weakly anisotropic ferrites with a cubic symmetry of the crystal lattice. Films with crystallographic orientations of the {111} type have found the widest application. The results of analysis of anisotropy in the SMSW spectrum in YIG films with the {111} orientation are given in [8]. At the same time, some characteristics of spin-wave devices can be improved by using films with other orientations. In particular, it was noted in [9, 10] that the thermal stability of the SMSW spectra for films with the {110} and {100} orientations is higher than for {111} films. All these orientations ({111}, {110}, and {100}) are particular cases of the model under investigation.

Figure 2 illustrates the model for a film with the tangential {110} axis. The crystallographic orientation of the magnetization vector is specified by two angles: angle ψ measured from the ⟨110⟩ axis in the plane of the film and angle δ defining the inclination of the {100} plane passing through the ⟨110⟩ axis under consideration to the plane of the film. Cubic anisotropy will be taken into account using only one constant. Films with the {111} orientation will be considered separately, since two constants must be taken into account for such films. The components of tensor N_{ij}^a appearing in the DE are described by the expressions (these expressions are derived, analogously to [8], on the basis of the formulas connecting components N_{ij}^a with the cosines of the angles between the coordinate axes in system xyz and the axes [100], [010], and [001] of the cubic crystal)

$$\left.\begin{aligned}
M_0 N_{xx}^{c(1)} &= -2\frac{K_1^c}{M_0}\left\{1 - \frac{1+3\cos 2\delta}{16}[(1+3\cos 2\delta) + 3(1-\cos 2\delta)\cos 2\psi]\right\},\\
M_0 N_{yy}^{c(1)} &= -\frac{K_1^c}{M_0}\left\{1 + \frac{1+3\cos 2\delta}{32}[(1+3\cos 2\delta) + 3(5-\cos 2\delta)\cos 4\psi]\right\},\\
M_0 N_{zz}^{c(1)} &= -\frac{K_1^c}{M_0}\left\{1 + \frac{1+3\cos 2\delta}{32}[(1+3\cos 2\delta) + 4(1-\cos 2\delta)\cos 2\psi - (5-\cos 2\delta)\cos 4\psi]\right\},\\
M_0 N_{xy}^{c(1)} &= \frac{3}{2}\frac{K_1^c}{M_0}\sin 2\delta\left[\cos 3\psi + \frac{1+3\cos 2\delta}{8}(\cos\psi - \cos 3\psi)\right],
\end{aligned}\right\} \quad (11)$$

where K_1^c is the first constant of cubic magnetic anisotropy.

Figure 3 shows examples of the angular dependences of the SMSW frequencies calculated using Eq. (10) with substitution of expressions (11). In our calculations, we used the values of magnetic parameters for YIG crystals. Additional analysis of calculated data revealed that the largest difference between exact and approximate DE is manifested in the vicinity of $kd = 1$. In Fig. 3, this difference does not exceed 1% relative to angular variations of frequency associated with the effect of cubic anisotropy (i.e., relative to the difference $\max\{f(\psi)\} - \min\{f(\psi)\}$). The following two peculiarities are worth noting. The first corresponds to

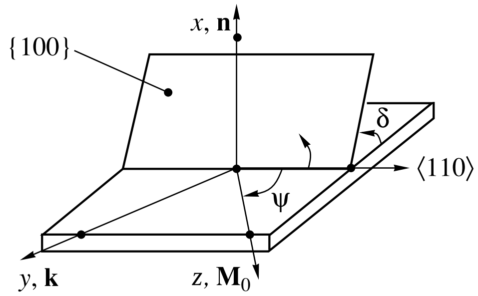


Fig. 2. Model of a ferromagnetic film.

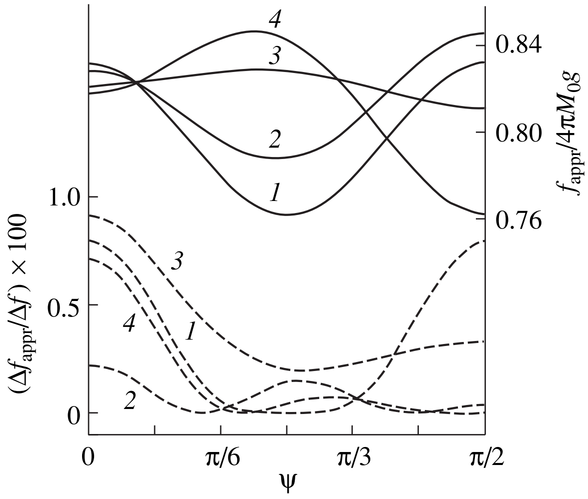


Fig. 3. Angular dependences of the SMSW frequency for $kd = 1$ in a cubic anisotropic film. Dependences $f_{appr}(\psi)$ are calculated using approximate DE (10) (solid curves). In calculating errors (dashed curves), we used the dependences $f(\psi)$ derived from exact equation (4) as well as the differences $\Delta f_{appr} = f_{appr} - f$ and $\Delta f = \max\{f(\psi)\} - \min\{f(\psi)\}$. Magnetic parameters: $4\pi M_0 = 1750$ G, $K_1^c/M_0 = -43$ Oe, and $H_e/4\pi M_0 = 1/3$. Figures on the curve correspond to the following values of angle δ: (*1*) 0, (*2*) π/6, (*3*) π/3, and (*4*) π/2.

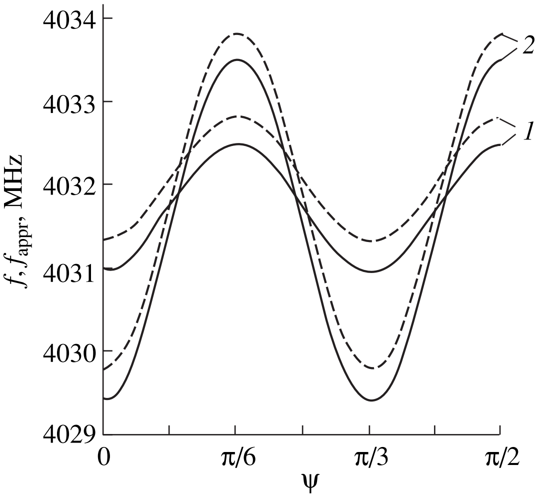


Fig. 4. Angular dependences of the SMSW frequency for $kd = 1$ in a film of a cubic crystal with the {111} orientation. Solid and dashed curves are calculated using approximate and exact dispersion equations (10) and (4), respectively. The values of magnetic parameters are the same as in Fig. 3. Curves *1* take into account only the first anisotropy constant, while curves *2* account for the first and the second anisotropy constants (the latter was taken as $K_2^c/M_0 = -2$ Oe).

films with orientations {100} and {110} (δ = 0 and δ = π/2). For these orientations, the angular dependence of frequencies is the most significant. Nevertheless, the approximate DE describes this dependence to a high degree of accuracy. The second peculiarity concerns films with the {111} orientation. In this case, $(1 + 3\cos 2\delta) = 0$ and, in accordance with relations (11), the dependence on angle ψ is observed only for component $N_{xy}^{c(1)}$. In contrast to diagonal components, N_{xy}^{a} appears in the DE in quadratic form; for this reason, the angular dependence associated with this component is the weakest. In this case, as shown in [8], the analysis of anisotropy of the SMSW spectrum requires the inclusion of not only the first, but also the second, constant of cubic magnetic anisotropy. The tensor components of effective demagnetizing factors corresponding to the inclusion of the second constant (K_2^c) have the following form (the expressions were derived in the same way as expressions (11)):

$$M_0 N_{xx}^{c(2)} = \frac{K_2^c}{6M_0},$$

$$M_0 N_{yy}^{c(2)} = \frac{K_2^c}{18M_0}(1 + 5\cos 6\psi),$$

$$M_0 N_{zz}^{c(2)} = \frac{K_2^c}{18M_0}(1 - \cos 6\psi),$$

$$M_0 N_{xy}^{c(2)} = \frac{\sqrt{2}}{6}\frac{K_2^c}{M_0}\cos 3\psi.$$

The results shown in Fig. 4 visually demonstrate the effect of the second constant on the angular dependence of the SMSW frequencies in {111} films. It should be noted that the approximate DE successfully describes the "amplitude" and "phase" of angular variations of frequencies.

Let us introduce into the film model the uniaxial anisotropy in addition to cubic anisotropy. For example, both types of anisotropy can be present in YIG films [11]. We will assume that the principal axis of uniaxial anisotropy is directed along the normal to the film. Then, in the geometry of tangential magnetization, the tensor components of the effective demagnetizing factors of uniaxial anisotropy have the form

$$M_0 N_{xx}^{u} = -\frac{2K_1^u}{M_0}, \quad N_{yy}^{u} = N_{zz}^{u} = N_{xy}^{u} = 0, \qquad (12)$$

where K_1^u is the first uniaxial anisotropy constant.

Examples of the dependences calculated on the basis of expressions (11) and (12) are given in Fig. 5.

The approximate DE derived here takes into account magnetic anisotropy in the most general form and, hence, can be applied for monocrystalline ferromagnetic films with any type of the crystal lattice and with an arbitrary crystallographic orientation. The simple

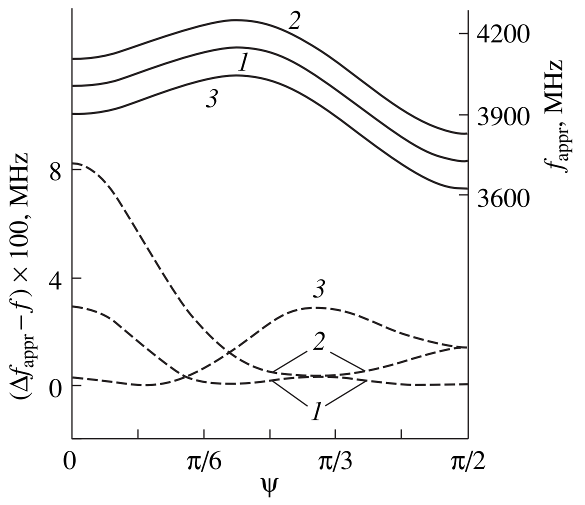


Fig. 5. Angular dependences of the SMSW frequency for $kd = 1$ in a film with cubic and uniaxial anisotropy. The film orientation is {110} ($\delta = \pi/2$). Dependences $f_{appr}(\psi)$ are calculated using approximate DE (10) (solid curves), while the $f(\psi)$ dependences are based on exact DE (4). The dashed curves describe the difference $f_{appr}(\psi) - f(\psi)$. The magnetic parameters are the same as in Fig. 3. The figures on the curves correspond to the following values of the uniaxial anisotropy constant: (*1*) $K_1^u = 0$, (*2*) $K_1^u = K_1^c$, and (*3*) $K_1^u = -K_1^c$.

analytic form of the approximate DE makes it possible to substantially simplify analysis of the process occurring in anisotropic films with the participation of SMSWs. Thus, our results can be used for improving the characteristics of ferromagnetic films and in development of spin-wave devices.

REFERENCES

1. R. W. Damon and J. R. Eshbach, J. Phys. Chem. Solids **19**, 308 (1961).
2. B. Schneider, Phys. Status Solidi B **51**, 325 (1972).
3. A. S. Beregov, Izv. Vyssh. Uchebn. Zaved. Ser. Radioélektronika **27** (10), 9 (1984).
4. S. Ishak, Proc. IEEE **76**, 171 (1988).
5. A. N. Slavin and Yu. K. Fetisov, Zh. Tekh. Fiz. **58**, 2210 (1988) [Sov. Phys. Tech. Phys. **33**, 1343 (1988)].
6. I. V. Zavislyak, V. M. Talalaevskiĭ, and L. V. Chevnyuk, Fiz. Tverd. Tela (Leningrad) **31**, 319 (1989) [Sov. Phys. Solid State **31**, 906 (1989)].
7. A. G. Gurevich, *Magnetic Resonance in Ferrites and Antiferromagnets* (Nauka, Moscow, 1973) [in Russian].
8. O. A. Chivileva, A. G. Gurevich, and L. M. Émiryan, Fiz. Tverd. Tela (Leningrad) **29**, 110 (1987) [Sov. Phys. Solid State **29**, 61 (1987)].
9. V. V. Shagaev, Zh. Tekh. Fiz. **68** (10), 99 (1998) [Tech. Phys. **43**, 1226 (1998)].
10. G. G. Bondarenko and V. V. Shagaev, Perspekt. Mater., No. 5, 33 (2000).
11. A. S. Beregov and E. V. Kudinov, Élektron. Tekh., Ser. Élektron. SVCh, No. 6, 41 (1986).

Translated by N. Wadhwa

Technical Physics, Vol. 49, No. 10, 2004, pp. 1360–1363. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 113–116.
Original Russian Text Copyright © 2004 by Bondarenko, Gabdullin, Gnuchev, Davydov, Korablev, Kravchik, Sokolov.

SURFACES, ELECTRON AND ION EMISSION

Emissivity of Powders Prepared from Nanoporous Carbon

V. B. Bondarenko*, P. G. Gabdullin*, N. M. Gnuchev*, S. N. Davydov*, V. V. Korablev*, A. E. Kravchik, and V. V. Sokolov****

** St. Petersburg State Polytechnical University, ul. Politekhnicheskaya 29, St. Petersburg, 195251 Russia*

e-mail: disurf@phtf.stu.neva.ru

*** Federal State Unitary Organization Russian Research Center Prikladnaya Khimiya,*
St. Petersburg, 197198 Russia

Received February 26, 2004

Abstract—Powders prepared from nanoporous carbon are promising for creating cold emitters, which are essential to the development of reliable next-generation monitors. The results of an experimental study of the temperature and time dependences of the emission current from nanoporous carbon coatings are reported. It is shown that the stable emission may last at least 20 h under continuous operation if the emission current density does not exceed 0.6 mA/cm^2 at room temperature and an accelerating field strength of 800–1200 V/mm. The highest values of the unstable-in-time current density vary from 2.5 to 3.2 mA/cm^2. © *2004 MAIK “Nauka/Interperiodica”.*

INTRODUCTION

Efficient field-emission cathodes are made of different carboniferous materials, such as nanodimensional fibers, diamond-like films, and porous structures [1–11]. However, the materials listed suffer from substantial disadvantages: complex production technology [12, 13], high cost, the need for high electric fields to provide acceptable emission currents, and emission current instability.

In this work, we report the results of studying nanoporous carbon (NPC)—a material that is to a great extent free of these disadvantages.

PREPARATION OF NANOPOROUS CARBON AND ITS BASIC PROPERTIES

The emitters used in this work were made using simple, cheap, and reliable technology. NPC specimens were prepared by chlorination of silicon, titanium, and boron carbides and molybdenum semicarbide:

$$\mathrm{MeC} + n/2\mathrm{Cl}_2 \longrightarrow \mathrm{MeCl}_n + \mathrm{C}.$$

Chlorination was accomplished in a direct-flow graphite reactor at 870 K for Mo_2C, 1070 K for TiC, and 1270 K for B_4C. The degree of removal of the element was determined by weighing (from the loss of the solid component weight). Table 1 lists a number of properties of the initial carbides and the NPC yield after chlorination. The yield is seen to correlate well with the carbon content in the related carbide.

As follows from the X-ray diffraction data, the NPC obtained from the carbides contains only the carbon phase with a paracrystalline structure. This structural type features a random arrangement of hexagonal carbon-atom monolayers linked through atoms of amorphous carbon.

To remove residual chlorine, the NPC powders were heat-treated in the hydrogen flow at 770 K. A high degree of purification of the powders was confirmed by the photoelectron spectroscopy data, which show that the carbon content in the powder is as high as 98.5–99.0%.

Based on the data of adsorption structural analysis, the NPC prepared from titanium and silicon carbides can be referred to microporous systems, as indicated by the nitrogen adsorption isotherm of type I (according to the IUPAC classification). The material prepared from boron carbide and molybdenum semicarbide is referred to mesoporous systems, as indicated by the nitrogen adsorption isotherm of type IV (Table 2).

Table 1. Properties of initial carbides and NPC yields after clorination

Carbide	Mean particle size, μm	Density, g/cm^3	Carbon content, mas %	NPC yield after clorination, mas %
SiC	0.8	3.21	30.0	29.8
TiC	2.0	4.91	20.1	19.5
B_4C	2.0	2.52	21.2	20.8
Mo_2C	40.0	8.9	5.9	5.5

Table 2. Properties of powders from nanoporous carbon

Initial carbide	Apparent density, g/cm^3	Pycnometric density, g/cm^3	Volume of pores, cm^3/cm^3	Sorption capacity with respect to benzene, cm^3/cm^3	Specific surface, m^2/g	Diameter of pores, nm	Type of nitrogen adsorption isotherm
SiC	0.91	2.15	0.58	0.61	1000	0.7	I
TiC	0.99	2.20	0.55	0.60	1100	0.8	I
B_4C	0.55	2.20	0.75	0.76	1310	4.0	IV
Mo_2C	0.53	2.15	0.77	0.75	1740	4.0	IV

EXPERIMENTAL

The specimens to be tested for emissivity were prepared as follows. A sheet of molybdenum was covered by a thin layer of orthophosphoric acid H_3PO_4 (binder), which, in turn, was covered by the powder. Those particles of the powder not adhered to the sheet were blown away. During heating in a vacuum, the binder burned up and the particles adhered to the substrate constituted an emitter. The tests were carried out in a stainless steel chamber at a pressure varying from 2×10^{-7} to 1×10^{-9} Pa. The specimens were placed on holders made of chemically pure tantalum sheets.

A tungsten-wire helical heater was provided under each of the specimens to raise the emitter temperature to 720–770 K. Heating could be combined with measurements; that is, the current emitted by the specimen surface toward the cylindrical molybdenum anode could be measured. The round-ended anode of diameter d = 6 mm had a polished working (end) face. The distance between the specimen and end face of the anode was 1.0 ± 0.3 mm.

It was assumed in the calculations that the emitting region is just below the anode. The emission surface area was equal to $S \approx \pi(d/2)^2 = 0.283\ \text{cm}^2 \approx 0.3\ \text{cm}^2$. The fact that it is the visual appearance of this region that changed after the emission experiments (it became darker than the rest of the specimen surface) supports the validity of this assumption.

The temperature was measured with a thermocouple whose junction was tightly pressed against the holder in the immediate vicinity of the specimen edge.

EMITTER ACTIVATION

The as-prepared specimens did not emit at room temperature and sweeping field strengths E_A < 2000 V/mm. When activating the emitter, one should take into account a number of factors, three of which are ranked below in order of decreasing significance: heating of the specimen, delay time, and application of a positive (sweeping) voltage to the anode.

The most efficient way of activating the emitter is heating combined with the application of a sweeping potential to the anode. Figure 1 shows two typical time dependences of the emission current, where the current during activation either appears suddenly (almost stepwise) (Fig. 1a) or smoothly grows (Fig. 1b).

The structural reconstruction taking place in the material activated by heating lasts from 20 min to several hours. Basically, the specimens can be activated without applying a voltage to the anode. In this case, however, activation takes one and a half or even twice as much time as when the voltage is applied and the

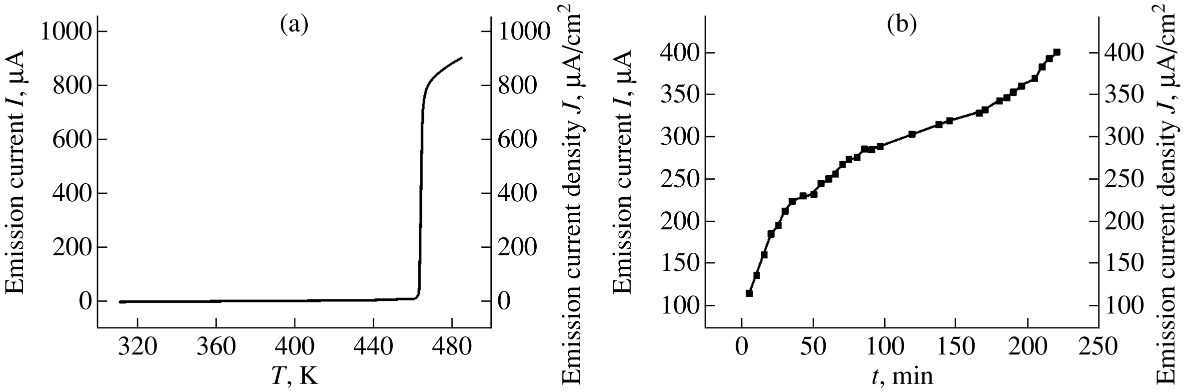


Fig. 1. Temperature and time dependences of the emission current during activation. (a) $E_A \approx 500$ V/mm, the total tine of heating is 60 min and (b) T = const = 530 K, $E_A \approx 1000$ V/mm.

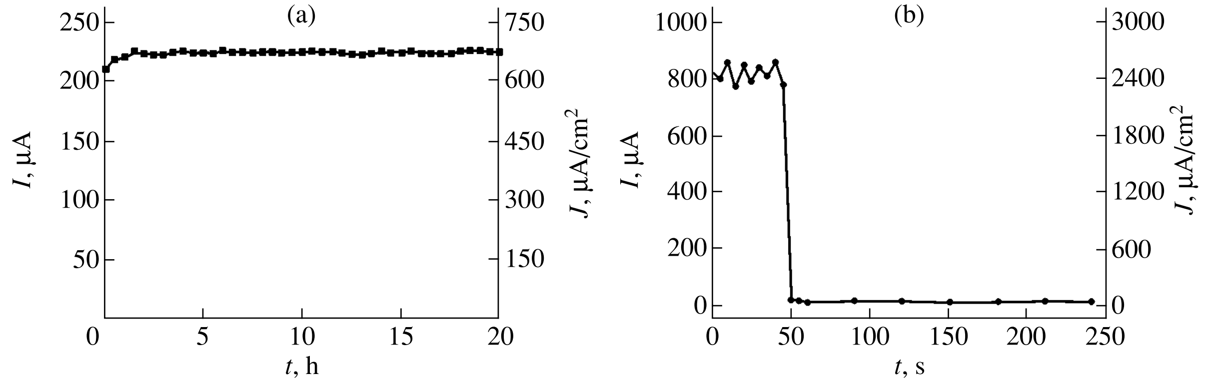


Fig. 2. Time dependences of the emission current at T = 290 K and $E_A \approx$ 1000 V/mm. The ordinates on the right show the corresponding current density values.

emission current is as low as several tens of microamperes (versus 400–800 μA, or 1.2–2.4 mA/cm^2, as shown in Fig. 1). The activation of the specimens can be enhanced by application of a voltage to and moderate heating of the anode (370–470 K); however, heating is optional.

EFFECT OF VARIOUS FACTORS ON THE EMISSION CURRENT

(i) Effect of the specimen temperature on the emission current. Heating of the specimen generates an emission current. As the temperature is varied between 290 and 670 K, the I–V characteristics of the activated emitter remain concave (which is typical of field and thermionic emissions) but the absolute values of the current grow. For E_A = 1000 V/mm, the current density increases from 30–45 μA/cm^2 at T = 320 K to 300–400 μA/cm^2 at T = 490 K.

(ii) Effect of the field strength. For low field strengths (E_A < 400 V/mm), the emission may appear, if at all, only if the specimen is slowly and thoroughly (for several hours) heated. If the field E_A = 500–1000 V/mm is applied to the heat-preactivated cathode, the emission current grows with time and slowly reaches saturation. Thus, the application of a moderate voltage to the anode raises the emissivity of the cathode. If, however, the field strength is high (E_A = 600–2000 V/mm or higher), the emission becomes unstable and eventually comes down to zero. If the emission remains weak (no activation) and the anode is under a smoothly increasing voltage, a short (0.5–1.0 s) current spike (5–40 μA/cm^2) is observed at E_A = 1600–2400 V/mm. Subsequently, the emission is impossible to cause without heating.

In the case of the preactivated emitter, the current becomes more and more unstable when approaching a maximal value. The amplitude of its “random” oscillations (these oscillations do look like random) builds up, reaching the value of the very current. At some time the current drops below 10 μA again and does not rise any longer, remaining at this level.

(iii) Temporal dependences. The onset and buildup of the emission even from the preactivated specimen is an inertial process if the cathode was not heated and remained inoperative for a time. For E_A = 500–800 V/mm, the time it takes for the initial current I_A = 10–20 μA to reach its maximal value (500–900 μA) is several tens of seconds.

When the current density was low (J_A = 300–600 μA/cm^2), the emission current might persist for several hours, varying within ±2% (Fig. 2a). However, for current densities above 1000 μA/cm^2, the current was extremely unstable (Fig. 2b).

CONCLUSIONS

The emission from the surface of NPC powders has the following features. (i) It was established that the stable field-emission current of density up to 600 μA/cm^2 may be achieved; (ii) preactivation of the surface is requisite for high current densities; (iii) the onset of emission is an inertial process; (iv) at low current densities (<600 μA/cm^2), the emission is indefinitely stable; (v) at high current densities (>1000 μA/cm^2), it is unstable, the relative amplitude of the random oscillations building up with increasing E_A; and (vi) sweeping field strengths above 600–800 V/mm jeopardize the emission (for a reason other than electrical breakdown).

NPC-based emitters may find wide application, specifically because they are cheap and simple to fabricate.

REFERENCES

1. C. Lea, J. Phys. D **6**, 1105 (1973).
2. J.-M. Bonard, J.-P. Salvetat, T. Stockli, *et al.*, Appl. Phys. A **69**, 245 (1999).

3. O. P. Monteiro, V. P. Mammana, M. C. Salvadori, *et al.*, Appl. Phys. A **71**, 121 (2000).
4. A. A. Talin, K. A. Dean, and J. E. Jaskie, Solid-State Electron. **45**, 963 (2001).
5. A. N. Obraztsov, I. Yu. Pavlovskiĭ, and A. P. Volkov, Zh. Tekh. Fiz. **71** (11), 89 (2001) [Tech. Phys. **46**, 1437 (2001)].
6. W. I. Milne, K. B. K. Teo, M. Chhowalla, *et al.*, Curr. Appl. Phys. **1**, 317 (2001).
7. N. S. Xu, Y. Chen, S. Z. Deng, *et al.*, J. Phys. D **34**, 1597 (2001).
8. J.-M. Bonard, M. Croci, C. Klinke, *et al.*, Carbon **40**, 1715 (2002).
9. G. Pirio, P. Legagneux, D. Pribat, *et al.*, Nanotechnology **13**, 1 (2002).
10. A. N. Obraztsov, A. P. Volkov, K. S. Nagovitsyn, *et al.*, J. Phys. D **35**, 357 (2002).
11. Yu. V. Gulyaev, Vestn. Akad. Nauk **73**, 389 (2003).
12. D. Laplaze, L. Alvarez, T. Guillard, *et al.*, Carbon **40**, 1621 (2002).
13. W. K. Maser, A. M. Benito, and M. T. Martinez, Carbon **40**, 1685 (2002).

Translated by V. Isaakyan

Technical Physics, Vol. 49, No. 10, 2004, pp. 1364–1367. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 117–121.
Original Russian Text Copyright © 2004 by Sorokin, Volkov.

SURFACES, ELECTRON AND ION EMISSION

Nanoisland Nucleation in Thermal Spike Regions on the Material Surface Irradiated by Swift Heavy Ions

M. V. Sorokin and A. E. Volkov

Russian Research Center Kurchatov Institute, pl. Akademika Kurchatova 1, Moscow, 123182 Russia
e-mail: sorokin@dni.polyn.kiae.ru

Received March 23, 2004

Abstract—The energy that is lost by swift heavy ions (SHIs) in a material highly excites its electronic subsystem, while the ion subsystem initially remains almost unperturbed. Subsequent energy transfer from the excited electrons to target atoms may cause a short-term local temperature rise (thermal spike), which, in turn, may induce phase transitions in the nanodimensional region near the ion trajectory. The possibility of nanoisland nucleation within such spikes on the material surface exposed to SHIs is studied. Presumably, the nanoislands appear when the characteristic nucleation time is shorter than the time of spike region cooling. It is shown that the maximal nucleation rate may be observed at a distance of the center of the spike. This may result in the annular distribution of the islands around the SHI trajectory. *© 2004 MAIK “Nauka/Interperiodica”.*

INTRODUCTION

Recently, it has been discovered that, when passing through a material, swift heavy ions with energies exceeding 1 MeV/nucleon and masses $N > 50m_p$, where m_p is the proton mass, affect the decomposition kinetics of supersaturated solid solutions and cause the formation of nanodimensional precipitates in metals [1, 2], insulators [3], and semiconductors [4]. This effect correlates with the electron losses of the incident particles (electron losses prevail at such energies) and demonstrates new possibilities for controlling the formation kinetics of anisotropic nanostructures.

With such energies and masses of projectiles, over 90% ($S_e > 10$ KeV/Å) of the lost energy goes into the electronic subsystem excitation, while the ionic subsystem initially remains almost unperturbed. Subsequent energy transfer from the electrons excited to the target atoms may raise the temperature for a short time [5–13] (thermal spike) in the nanodimensional region near the ion trajectory. The effect of a thermal spike due to the relaxation of high-energy electron excitations in SHI tracks on nanoprecipitation in supersaturated solid solutions is considered in [5]. The temperature dependence of the precipitation rate typically has a sharp peak at elevated temperatures [14]. Because of this, it was assumed [5] that nanoprecipitates appear in SHI tracks when the material in the track heats up to the point where the characteristic time of precipitation (clustering) becomes shorter than the track cooling time.

Swift heavy ions may also produce thermal spikes on the irradiated surface. Similarly to the spike in the bulk of the material, the surface spike may also influence the decomposition of the supersaturated surface solutions of impurity atoms and/or adatoms, thus favoring the formation of 2D or 3D nanoislands.

In this work, we studied the characteristic time of 2D nanoisland nucleation as a function of temperature. Based on the results obtained, the system’s parameters and irradiation conditions that are the most favorable for the nucleation of nanoislands in SHI-induced surface thermal spikes were determined.

It is shown that, at certain parameters of the system, the maximal nucleation rate may be observed at the periphery of a nanodimensional thermal spot. Such inhomogeneity may give rise to the annular distribution of the nanoislands near the ion track.

MODEL

The state of a surface solution is characterized by the concentration of impurity atoms and/or adatoms. Let us assume that this concentration exceeds the thermodynamically equilibrium level for the irradiation temperature (the solution is supersaturated).

We also assume that the relaxation of the electronic subsystem excited by SHIs appreciably heats up a nanodimensional surface region near the ion trajectory. In this paper, we omit the issues concerning melting of the material and start our consideration from the instant in time when the temperature in the spike region drops below the melting point. One more assumption is that the nucleating nanoislands are much smaller than the region heated, so that the influence of temperature gradients on the nucleation kinetics is disregarded.

As the temperature rises, the diffusion mobility of impurity atoms and/or adatoms grows. On the other hand, the supersaturation of the solution decreases, since the equilibrium concentration is an increasing function of temperature. The competition between these processes causes a maximum of the nucleation rate, which is observed in a certain temperature range depending on the material constants [14]. Thus, for a certain set of the parameters of the impurity atom (and/or adatom) solution, the local heating of the solution may markedly cut the characteristic time of nanoisland nucleation around the SHI track under specific irradiation conditions. It is evident that island formation in this region occurs if the nucleation time is shorter than the time of spike region cooling. Therefore, it is reasonable to compare the minimum time of surface spot cooling with the maximal time of nanoisland nucleation in the spot.

The characteristic time of cooling is given by the expression

$$t_T = \frac{R_T^2}{\chi_i}, \tag{1}$$

where R_T is the initial radius of the region heated and χ_i is the thermal conductivity of the material.

According to the experimental data and analytical estimates, the radius of the thermal spike ranges from several to several tens of nanometers [11, 15–17]. In order to estimate the minimal time of cooling, we use the thermal conductivity values that are typical of metals ($\chi_i = 10^{-3}$–10^{-2} cm^2/s):

$$t_T \approx 10^{-9}\text{–}10^{-10}\ \text{s}. \tag{2}$$

In the concentrated solutions, even two bonded impurity atoms may form a stable cluster. Structural irregularities present on the surface also assist in nanoisland nucleation. However, we consider here the slowest uniform nucleation at a weak supersaturation to estimate the maximal possible nucleation time. In this case, the nucleation can be described as the overcoming of an energy barrier the height of which depends on the supersaturation of the impurity solution and on temperature. As the characteristic nucleation time (t_c), we take the mean time taken for the nuclei to reach the critical size corresponding to this barrier. This time is given by [18]

$$t_c(n_{\min} \longrightarrow n_c) \simeq \left(\frac{2}{\pi}\left|\frac{d^2\varphi}{dy^2}\right|_{y = n_c}\right)^{-1/2} \frac{\exp[\varphi(n_c)]}{2\pi\nu\exp(-U/T)}. \tag{3}$$

Here, ν is the oscillation frequency of the adatoms and U is the barrier for impurity migration over the surface. Due to the rapid (on the order of nanoseconds) cooling of the spike region, the nuclei insignificantly, if at all, disturb the impurity spatial distribution. In this case, nucleation barrier $\varphi(n_c)$ and Zel'dovich factor $|d^2\varphi/dy^2|_{y = n_c}$ are defined by the following expressions [14, 19–22]:

$$\varphi(n_c) = \frac{\pi\sigma^2 s}{\ln(c_i/C)T(T_0 - T)}, \tag{4}$$

$$\left|\frac{d^2\varphi}{dy^2}\right|_{y = n_c} = \frac{\ln^3(c_i/C)}{2\pi\sigma^2 s}\frac{(T_0 - T)^3}{T}. \tag{5}$$

Here, $T_0 = \Phi/\ln(c_i/C)$ and the nucleation occurs only at $T < T_0$; T is the temperature measured in energy units; c_i and C are the concentrations of impurity atoms in the nanoislands and on the surface, respectively; Φ is the Gibbs free energy of impurity dissolution at the boundary of a semi-infinite layer; s is the nanoisland surface area per impurity atom; and σ is the specific excess Gibbs energy of the nanoisland boundary. It is assumed that the islands have the form of single-layer disks.

SYSTEM PARAMETERS FAVORING NANOISLAND NUCLEATION

The temperature dependence of nucleation time $t_c(T)$ (Fig. 1) has a deep minimum whose position depends on the energy parameters characterizing the dissolution and diffusion of impurity atoms near the island boundary, i.e., on Φ, σ, and U. From curves *1–7* in Fig. 2, we can determine their values that are the most favorable for the nucleation of the nanoislands in the region of subnanosecond thermal spikes.

For example, the increase in Φ increases the driving force of the phase transition and reduces the characteristic nucleation time (in Fig. 2, curves *1*, *4*, and *5* correspond to Φ = 1.00, 1.25, and 0.75 eV, respectively). These plots show that, when Φ is high, the nucleation time does not exceed a nanosecond over a wide domain in the plane of parameters (T, $\log(c_i/C)$).

The dependence of nucleation time t_c on surface tension coefficient σ is illustrated in Fig. 2 by curves *1*, *6*,

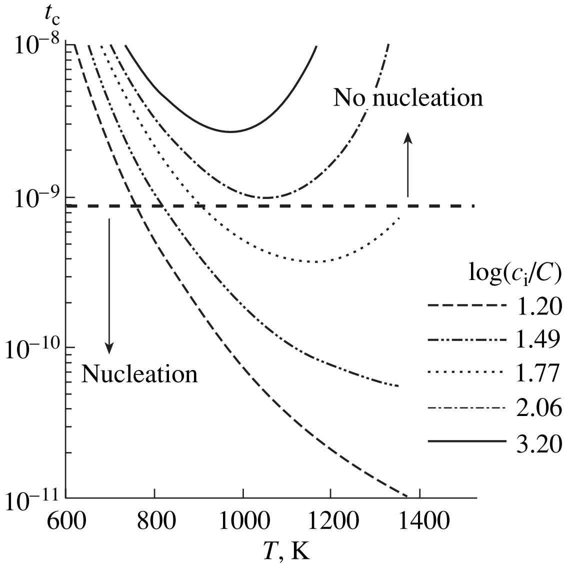


Fig. 1. Nucleation time t_c vs. temperature T and supersaturation ($\log(c_i/C)$) at U = 0.7 eV, $\sigma = 1 \times 10^{-5}$ erg/cm, and Φ = 1 eV.

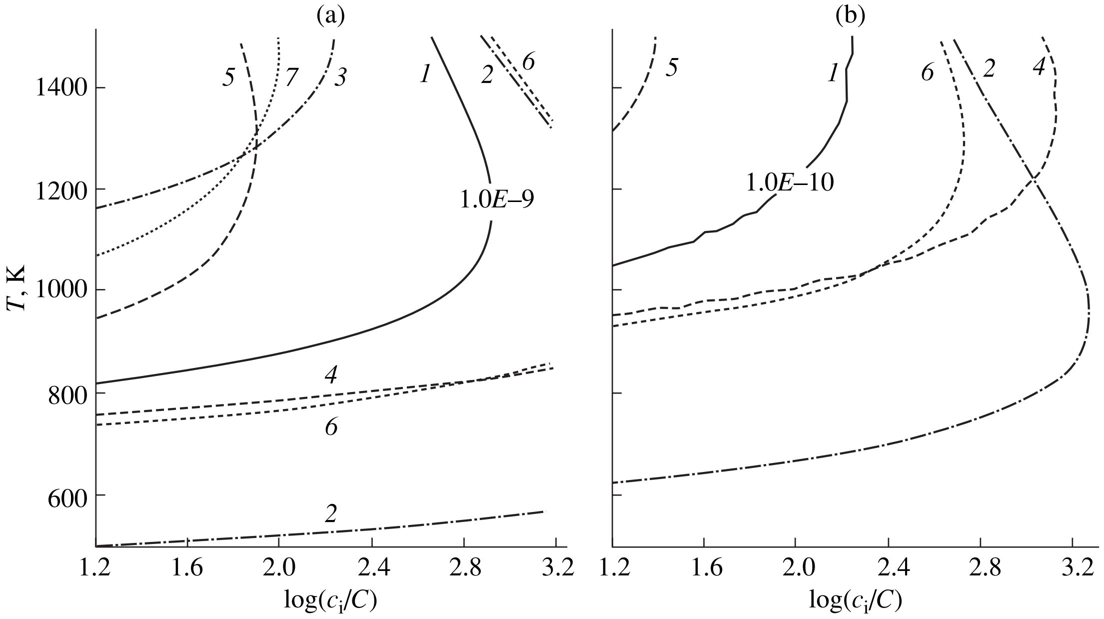


Fig. 2. Nucleation time (t_c) isolines bounding the domain (on the right) where the nucleation time is less than (a) 10^{-9} s and (b) 10^{-10} s. Lines *1*–*3* are drawn for $\sigma = 1 \times 10^{-5}$ erg/cm, $\Phi = 1$ eV, and migration barrier U = (*1*) 0.7, (*2*) 0.4, and (*3*) 1 eV. Lines *1*, *4*, and *5* are drawn at $\sigma = 1 \times 10^{-5}$ erg/cm, $U = 0.7$ eV, and impurity solubility Φ = (*1*) 1, (*4*) 1.25, and (*5*) 0.75 eV. Lines *1*, *6*, and *7* are drawn at $U = 0.7$ eV, $\Phi = 1$ eV, and surface tension coefficient σ = (*1*) 1×10^{-5}, (*6*) 0.75×10^{-5}, and (*7*) 1.25×10^{-5} erg/cm.

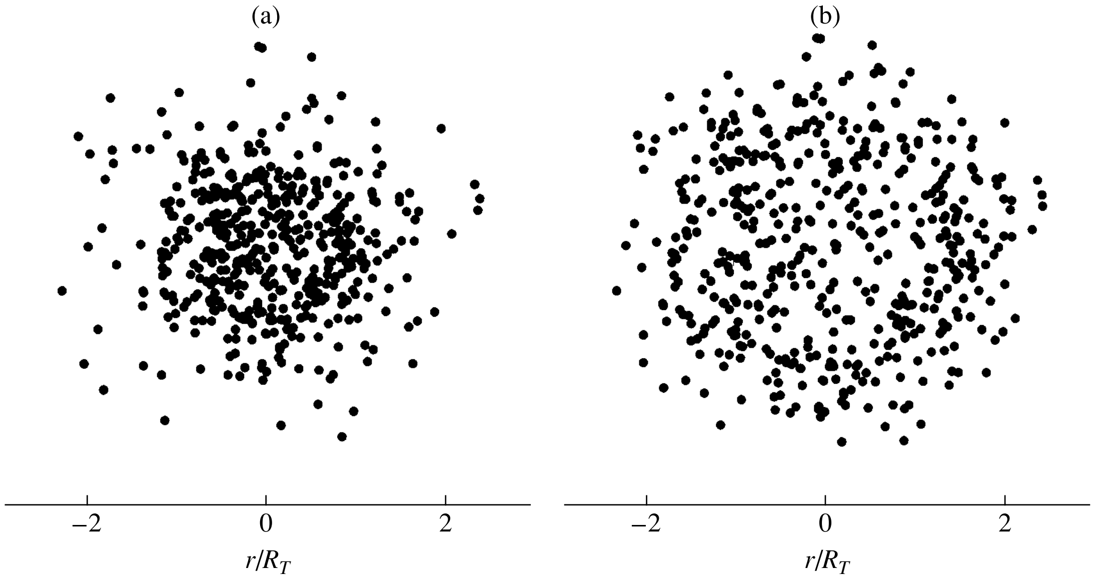


Fig. 3. Nucleus distribution near the SHI track.

and *7* ($\sigma = 1 \times 10^{-5}$, 0.75×10^{-5}, and 1.25×10^{-5} erg/cm, respectively). It should be noted that, when the impurity concentration is about 1%, the nucleation time appears to be less than a nanosecond even if the surface tension coefficient is high ($\sigma > 1.25 \times 10^{-5}$ erg/cm).

The diffusion barrier height considerably affects the nucleation kinetics. From Fig. 2 (curves *1*–*3*), it follows that a low barrier to impurity diffusion on the surface in the spike region may substantially activate nucleation of the nanoislands.

FORMATION OF SURFACE ANNULAR STRUCTURES NEAR SWIFT HEAVY ION TRACKS

Initially, the thermal spike heats only the central part of the SHI track, while its periphery remains cool for a

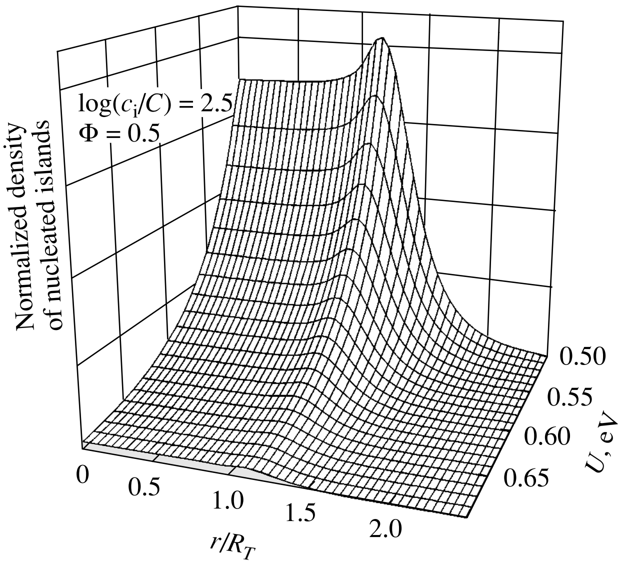


Fig. 4. Amount of nanoislands as a function of the system parameters. $T_0 = \Phi/\ln(c_i/C)$ and $T_{\max}$ is the initial temperature at the center of the region heated.

time. Subsequent heat transfer cools down the spike's center and heats up the periphery. It may so happen that the temperature providing the maximum nucleation rate persists at the periphery for a longer time than at the center. Figures 3a and 3b illustrate the case when the normalized number of nanoislands at the periphery of the cooling spike ($r/R_T \approx 1$) is larger than that at the center (annular nucleation).

As follows from Fig. 4, pronounced off-center peaks appear in the nucleus distribution at $T_0 \simeq 0.3T_{\max}$ ($T_{\max}$ is the initial temperature at the center, nucleation is assumed to occur at $T < T_0$). In this case, a ring of nuclei may form at a low solution energy ($\Psi < 0.6$ eV) and high diffusion mobility ($U < 0.5$ eV) of the impurity atoms and/or adatoms. In particular, when the maximal temperature in the SHI track is close to the melting point ($T_{\max} \approx T_m$, where $T_m = 1500$–2000 K is the melting point), the narrow dip in the temperature dependence of nucleation time is bound to lie between 600 and 800 K. As the system evolves (for example, after annealing), such an initial annular (nonuniform) spatial distribution of the nanoislands may result in the appearance of nanodimensional annular heterostructures on the surface irradiated.

ACKNOWLEDGMENTS

This work was financed by a program in support of junior scientists from the Russian Research Center Kurchatov Institute, grant no. 14, 2002.

REFERENCES

1. A. Barbu, P. Pareige, and V. Jacquet, Nucl. Instrum. Methods Phys. Res. B **146**, 278 (1998).
2. A. Iwase, T. Hasegawa, T. Tobita, *et al.*, Nucl. Instrum. Methods Phys. Res. B **195**, 309 (2002).
3. E. Valentin, H. Bernas, C. Ricolleau, and F. Creuzet, Phys. Rev. Lett. **86**, 99 (2001).
4. P. I. Gaiduk, A. Nylandsten Larsen, C. Trautmann, and M. Toulemonde, Phys. Rev. B **66**, 045316 (2002).
5. A. E. Volkov, Nucl. Instrum. Methods Phys. Res. B **193**, 376 (2002).
6. V. L. Ginzburg and V. R. Shabanovskiĭ, Dokl. Akad. Nauk SSSR **100**, 445 (1955).
7. F. Seitz and J. S. Koehler, Solid State Phys. **2**, 305 (1956).
8. M. I. Kaganov, I. M. Lifshits, and L. V. Tanatarov, Zh. Éksp. Teor. Fiz. **30**, 232 (1956) [Sov. Phys. JETP **3**, 216 (1956)]; J. Nucl. Energy, Part A: Reactor Sci. **12**, 69 (1960).
9. I. A. Baranov, Yu. V. Martynenko, S. O. Shchepelevich, and Yu. N. Yavlinskiĭ, Usp. Fiz. Nauk **156**, 178 (1988) [Sov. Phys. Usp. **31**, 1015 (1988)].
10. M. Toulemonde, C. Dufour, and E. Paumier, Phys. Rev. B **46**, 14362 (1992).
11. A. E. Volkov and V. A. Borodin, Nucl. Instrum. Methods Phys. Res. B **146**, 137 (1998).
12. G. Schiwietz, G. Xiao, P. L. Grande, *et al.*, Nucl. Instrum. Methods Phys. Res. B **146**, 131 (1998).
13. A. E. Volkov and V. A. Borodin, Nucl. Instrum. Methods Phys. Res. B **193**, 381 (2002).
14. J. Frenkel, *Kinetic Theory of Liquids* (Clarendon, Oxford, 1946; Nauka, Leningrad, 1945).
15. *Laser Spectroscopy and Photochemistry of Metal Surfaces*, Ed. by H. L. Dai and W. Ho (World Sci., Singapore, 1995).
16. A. I. Ryazanov, A. E. Volkov, and S. Klaumünzer, Phys. Rev. B **51**, 12 107 (1995).
17. A. I. Ryazanov, H. Trinkaus, and A. E. Volkov, Phys. Rev. Lett. **84**, 919 (2000).
18. C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences* (Springer-Verlag, Berlin, 1985; Mir, Moscow, 1986).
19. Ya. B. Zel'dovich, Zh. Éksp. Teor. Fiz. **12**, 525 (1942); Acta Physicochim. URSS **18**, 1 (1943).
20. E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics* (Pergamon, Oxford, 1981; Nauka, Moscow, 1979).
21. *Solids Far from Equilibrium. Collection Alea-Saclay: Monographs and Texts in Statistical Physics,* Ed. by C. Godreche (Cambridge Univ. Press, Cambridge, 1992).
22. A. A. Chernov, E. I. Givargizov, Kh. S. Bagdasarov, V. A. Kuznetsov, L. N. Dem'yanets, and A. N. Lobachev, *Modern Crystallography III: Crystal Growth* (Nauka, Moscow, 1980; Springer, Berlin, 1984).

Translated by A. Sidorova

Technical Physics, Vol. 49, No. 10, 2004, pp. 1368–1374. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 122–127.
Original Russian Text Copyright © 2004 by Kulikov.

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

The Conduction Current in a Metal–Insulator–Metal Structure

V. D. Kulikov
Tomsk Polytechnical University, Tomsk, 634050 Russia
e-mail: kulikov@list2.epd.tpu.edu.ru

Received November 25, 2003

Abstract—The flow of the conduction current in a metal–insulator–metal structure under X-ray and optical excitation is considered. Accumulation of positive charge at the negative electrode in the KBr and CsI crystals is studied. The method of discharge currents after the X-ray and optical excitation is suggested for estimation of the near-surface charge. It is shown that the values of charge transported by the conduction current and those determined from the measured discharge-current signal are in satisfactory agreement. The lifetimes of near-surface charges of holes and anionic vacancies and the values of the electric-field strength at the metal–insulator interface are estimated. It is assumed that electroneutrality of the sample is established owing to the motion of electrons from the surface into the bulk over dislocations. © *2004 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Metal–insulator–metal (MIM) structures are widely used in optoelectronic devices and experimental studies of radiation-induced conductivity in high-resistivity materials. Ohmic contacts can be formed in MIM structures; the current flowing through the structure is a through current due to injection of electrons from a metal electrode. If the band gap of insulators is wide (6–14 eV), a nonohmic contact forms at the metal–insulator interface in MIM structures; this contact retards the transport of electrons from the metal into the insulator. Free charge carriers are generated in the crystal itself, and the ohmic current is related to the motion of nonequilibrium charge carriers in an electric field. The second insulator–metal contact makes it possible for electrons to leave the sample. When a conduction current flows through the structure, a positively charged layer of holes is formed at the metal–insulator interface owing to the transport of electrons into the sample bulk. If there is no injection from the electrode, the positive charge of the holes is equal to the charge of the escaped electrons. The near-surface charge of the holes screens an external electric field, so that the problem of reduction and termination of the conduction current in the sample arises. In the conditions of a high electric-field strength (~10^6 V/cm) near the junction, the pattern of the current flow can be more complicated, owing to effects such as tunneling of electrons from the metal electrode, electric breakdown, and so on. At present, available quantitative data on the amplitude and spatial–temporal distributions of electric field and charge in MIM structures during generation of the conduction current are lacking. In particular, this situation can be attributed to the lack of relevant experimental methods for estimating the charge and field in insulators and to the complexity of the theoretical calculations.

A photoelectric method based on a variation in the potential of the semiconductor surface irradiated with optical photons is a promising method for studying the surface charges. At extremely high levels of illumination (10^{20}–10^{22} photon cm^{-2} s^{-1}), all external fields are completely screened with nonequilibrium charge carriers; simultaneously, the energy bands become completely flat at the surface [1].

The objective of this study was to gain insight into the processes of positive-charge accumulation at the interface between the insulator and the alkali-halide crystal in the presence of the conduction current. We intended to estimate the magnitude of the near-surface charge, the electric-field strength at the metal–insulator junction, and the characteristic time needed for screening the field in the sample. We were also going to clarify the mechanism of the establishment of electroneutrality in the sample. The photoelectric method is used to measure the surface charge in insulators.

EXPERIMENTAL

The method for measuring the current–voltage (I–V) and photoelectric characteristics is illustrated in Fig. 1a. We used the high-purity KBr and CsI single crystals grown from the melt of the corresponding salt of extra-pure grade. The samples were thin platelets with an area of 15 × 15 mm^2 and a thickness of ~300–350 μm. Aluminum electrodes 8 mm in diameter and a guard ring (in order to eliminate the surface current) were formed on the sample surface using either sputtering-assisted deposition or Al foil cemented to the surface with a hermetic (a silicon-organic compound).

Similarly, Al electrodes were cemented to the metal substrate; the contact capacitance was found to be equal to $C_1 \approx 2.5 \times 10^{-11}$ F. The thickness of the hermetic-compound layer was $d \approx 60$ μm. We also used contacts in the shape of an electrode grid pressed to the sample.

The method for measuring the *I–V* characteristics under X-ray and optical excitation was described previously [2, 3]. The X-ray radiation was generated by irradiating the ~300 μm-thick Al target with an electron beam; the highest electron energy was ~0.3 MeV, the pulse width (at the half-height) was $\Delta t \sim 20$ ns, and the electron-current density was ~300 A/cm^2. The smallest pulse separation was ~40 s. The volume density of generation of charge carriers *G* in alkali-halide crystals was ~10^{23} cm^{-3} s^{-1}. A pulsed voltage *V* with duration of ~20 μs and with negative polarity was applied to electrode *4* of the sample (Fig. 1a). An electron pulse was generated in ~8 μs after the voltage was applied.

When measuring the photoconductivity, we irradiated the sample with an X-ray pulse in order to generate the color centers. The initial concentration of the *F* color centers in KBr was ~3×10^{14} cm^{-3} [2]. Then (in 10 s), a voltage pulse and a photopulse were applied. A single-mode Nd:YAG laser was used in the experiment. The interelectrode space of the sample was illuminated either through the side face or through the contact grid. The energy density of laser radiation at the fundamental wavelength of 1.06 μm was ~0.04 J/cm^2 at the sample with a laser-pulse width of ~30 ns (at the half-height). The F^- color centers were excited at the fundamental wavelength. In order to excite the *F* centers, we used a lithium niobate crystal to convert the laser radiation to the second harmonic with a wavelength of 0.53 μm and a photon-flux density of ~10^{24} cm^{-2} s^{-1}.

In a MIM structure (Fig. 1a), the applied voltage is divided into following components when a current pulse is generated:

$$\varphi_1 + \varphi_s + \varphi_V + \varphi_R = V. \tag{1}$$

Here, φ_1, φ_s, φ_V, and φ_R are the potential drop across the hermetic-compound layer, the value of potential at the sample surface, the potential drop across the sample bulk, and the potential drop across the resistor *R*, respectively.

In order to describe the transient processes in a first approximation, we can introduce an effective capacitor C_e formed by positive charge in the sample and negative charge of electrons in the metal electrode *4*. For a structure with a cemented electrode, we disregard the quantity φ_s compared to φ_1, so that the potential drop across the effective capacitor $\varphi_e = \varphi_1$ ($\varphi_e \approx \varphi_s$ for the cemented electrode). Expression (1) can then be rewritten as

$$\frac{Q}{C_e} + i(R_V + R) = V, \tag{2}$$

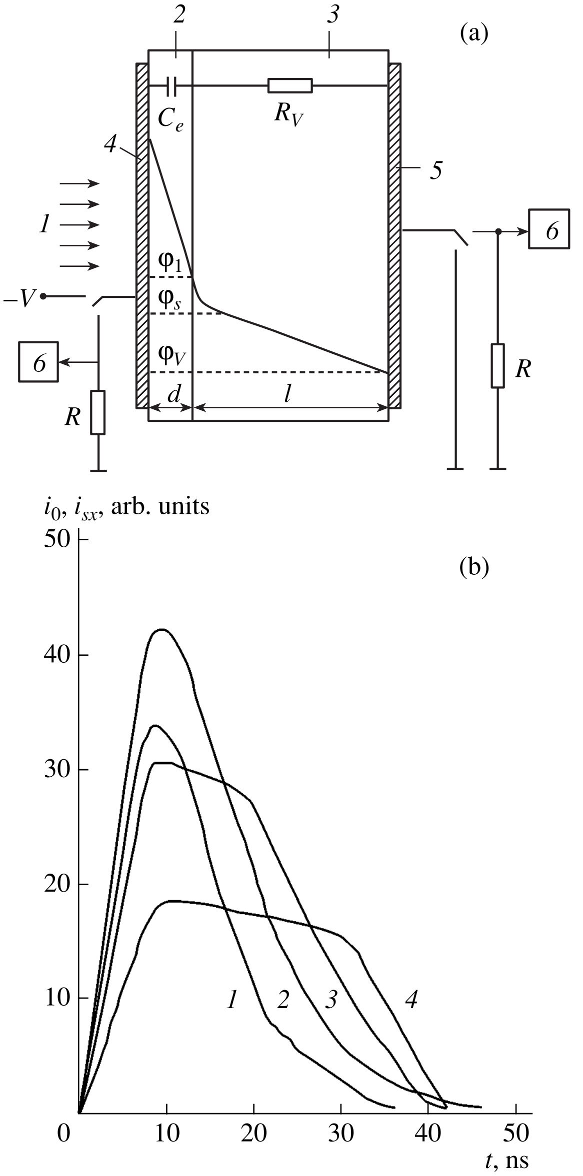


Fig. 1. (a) Schematic representation of the experimental setup: (*1*) the X-ray radiation; (*2*), the hermetic; (*3*), the sample of an alkali-halide crystal; (*4*) and (5), electrodes; (*6*), the oscilloscope; and $R = 50$ Ω, the load resistance. (b) The pulses of the conduction-current density i_0 at the electric-field strengths of (*1*) 2×10^4 and (*2*) 5×10^4 V/cm and of the discharge current i_{sx} for the KBr samples that have (*3*) cemented and (*4*) deposited electrodes and are subjected to the X-ray excitation.

where *Q* is the charge density at the capacitor plate, $i = \partial Q/\partial t$ is the current density in the structure, and R_V is the sample-bulk resistance. A solution to (2) at the charge-accumulation stage is given by $Q = C_eV[1 - \exp(-t/\tau_1)]$, where $\tau_1 = (R_V + R)C_e$ is the time constant of the circuit. This constant controls the field-screening time in a MIM structure. The complete field screening is attained in times ~$3\tau_1$. The current density $i \approx$

$V/R_V(1 - t/\tau_1)$ is nearly constant at the radiation-pulse width $\Delta t \ll \tau_1$. After completion of both the radiation and external-voltage pulse, a negative charge is located at the electrode *4*; this charge compensates the positive charge in the sample bulk. The potential drop φ_e can be measured, for example, by discharging the capacitor C_e through the series-connected resistors R_V and R. The expression that describes the kinetics of the discharge current density for capacitor C_e can be written as $i_s = Q_s\tau_1^{-1}\exp(-t/\tau_1)$, where $Q_s = C_e\varphi_e$ is the charge density at electrode *4*. The initial amplitude $i_s \approx \varphi_e/(R_V + R)$, and the charge-relaxation time equals $\sim\tau_1$. In insulators, $R_V \gg R$ even under irradiation, so that only a small fraction of φ_e drops across the resistor R. In this case, we can use the dependence $\varphi_e = Q_s/C_e$ in order to estimate φ_e. The charge density Q_s is determined by integrating (with respect to time) either the density i_s of the discharge current that drains off electrode *4* or the current density i. At the metal–insulator interface, we have

$$Q_s = \varepsilon_1\varepsilon_0 E_1 = \varepsilon\varepsilon_0 E_s, \tag{3}$$

where $\varepsilon_1\varepsilon_0$ and $\varepsilon\varepsilon_0$ are the absolute dielectric constants of the hermetic and the crystal, respectively, and E_1 and E_s are the electric-field strengths in the hermetic's layer and at the crystal surface, respectively.

The method for measuring the discharge current in the KBr and CsI samples consisted in the following: after termination of the conduction-current pulse, electrode *4* was disconnected from the voltage source and was connected to the oscilloscope, while electrode *5* was connected to ground (Fig 1a). The structure was then irradiated with an X-ray pulse; we measured the density of discharge current i_{sx} that was originally generated by the X-ray pulse and flowed through the resistor R. The measurement time was ~40 s.

The KBr samples were used in photoelectric measurements. Positively charged vacancies are generated and accumulated when the *F* color centers are ionized by light. A layer of *F* centers is formed as a result of excitation of F^- centers. After termination of the photocurrent pulse, electrode *4* was connected to the oscilloscope and electrode *5* was connected to ground. The structure was then irradiated with a laser pulse, and the photodischarge-current density i_{sp} was measured. The measurement time was ~10 s.

The polarity of the pulses of discharge currents induced by the X-ray and optical photons corresponded to the drain of negative charge off electrode *4*.

In a KBr crystal, a near-surface charge due, for example, to holes can relax as a result of irradiation with an X-ray or optical photons. The values of the peak amplitude of the discharge current after irradiation with X-ray and optical photons were in satisfactory agreement.

A study of the starting surface of the samples by the discharge-current method showed that there was no initial band bending within the experimental accuracy. The band bending was also not observed after separate effects of radiation and an external electric field with a strength as high as $\sim 5 \times 10^4$ V/cm.

RESULTS AND DISCUSSION

According to previous reports [4–6], holes in alkali-halide crystals self-localize rapidly (in 10^{-12} s) after exposure to ionizing radiation and form a configuration of the X_2^- type with neighboring atoms (here, *X* stands for a halide atom); this configuration is also referred to as the V_k center. The V_k centers capture free electrons with the resulting formation of self-bound excitons in an excited state. Later on, the self-bound excitons either relax to a lower energy state with subsequent radiative recombination or dissociate into two radiation defects from the excited state; these defects are referred to as the *F–H* color centers. The aforementioned dissociation occurs in $\sim 10^{-11}$ s. An *F* center is a positively charged anionic vacancy that localized an electron. If an *F* center captures an electron, an F^- color center is formed. An *F* center is neutral, whereas an F^- center is charged negatively [5]. The hole mobility is close to zero. Only free electrons are involved in conduction until these electrons either recombine with the V_k centers or are captured by the *F* centers or initial defects in the crystal. Bivacancies (cationic and anionic vacancies that are close to each other) can be considered as initial defects in alkali-halide crystals [6].

The lifetime of the *F* and F^- centers in the KBr and CsI crystals subjected to the X-ray excitation and also the lifetime of near-surface charges of holes and anionic vacancies after the completion of currents induced by the X-ray and optical photons were determined from the decay of the peak amplitude of pulses of the photoconductivity current and the discharge current measured at different instants of time after termination of the excitation pulse. A decrease in the amplitude of signals with time is described satisfactorily by the exponential dependence $\sim\exp(-t/\tau_2)$, where τ_2 is the lifetime of particles. The lifetime of the *F* centers was equal to ~11 min. The behavior of photoconductivity under excitation of the F^- centers was an exception. An increase in the photocurrent was observed for ~11 min with subsequent slow decrease in the current. This phenomenon can be attributed to the recharging of the *F* and F^- color centers. The charge of holes decreased most rapidly. We have $\tau_2 \sim 4.8$ min for both the deposited and cemented electrodes in the KBr crystal at room temperature. The lifetime of the near-surface hole charge in the CsI samples is shorter than ~40 s; however, τ_2 increased to ~30 min at the liquid-nitrogen temperature. The lifetime of the near-surface charge of anionic vacancies is longest (~46 min).

In our experiments, we did not observe a signal of photodischarge current after photoionization of the F^-

centers, which was consistent with the absence of a charge at the near-surface layer of neutral *F* color centers.

Dependences of the peak amplitude of the conduction-current density i_0 and the discharge-current densities i_{sx} and i_{sp} on the electric-field strength *E* for the KBr and CsI samples are shown in Fig. 2. One of the contacts on the KBr platelet was deposited, and the other contact was cemented. In the case of the CsI platelet, both contacts were cemented. Differences between the currents induced by X-ray radiation in the sample irradiated on the sides of the deposited and cemented contacts were no larger than 10%. Experiments with the CsI sample were performed at the liquid-nitrogen temperature. The *I*–*V* characteristics of the KBr and CsI crystals (curves *1*, *4* in Figs. 2a, 2b) are nonlinear in the range from 2×10^4 to 5×10^4 V/cm. Starting with a certain characteristic electric-field strength ($E \sim 5 \times 10^4$ V/cm), the curves $i(E)$ become nearly ohmic. The current-pulse width at $E \sim 2 \times 10^4$ V/cm (Fig. 1b) is limited by the duration of irradiation with X-ray photons. As the electric-field strength increases, a persistent current component is observed after termination of the X-ray pulse. The nonlinear shape of the *I*–*V* characteristics in the KBr and CsI crystals in the fields with strengths of 10^4–10^5 V/cm is most likely caused by an increase in the charge-carrier concentration due to the transfer of electrons from the trapping centers to the insulator's conduction band via the impact-ionization mechanism. According to estimations [7], the activation energy of aforementioned centers in CsI is ~0.05 eV and their concentration is $\sim 10^{14}$ cm^{-3}. It is conceivable that the trapping centers are initial crystalline defects, i.e., closely spaced anionic and cationic vacancies. Heat treatment of the CsI crystal for 1 h at 600°C led to an increase in the absorption in the α band by a factor of ~1.2 (the α band is the region of absorption by halide ions located near anionic vacancies [6]); simultaneously, the conduction-current density decreased by a factor of 3. This fact can be attributed to separation of bivacancies and capture of electrons by anionic vacancies with formation of the *F* color centers. A heat treatment of the KBr sample for 3 h at 600°C did not result in variations in both the absorption in the α band and the conductivity, which is caused by a higher activation energy for motion of anionic vacancies in KBr (~0.92 eV) compared to that in CsI (~0.39 eV) [6].

Taking into account that the charge-carrier lifetime $\sim 10^{-11}$–10^{-9} s [2–4] is shorter than the excitation-pulse width, we determined the steady-state electron concentration n_0 from the expression $i_0 = en_0\mu E$, where *e* is the elementary charge and μ is the electron mobility (for μ, we used data reported previously [5, 8]). In the case of X-ray excitation of KBr, the charge-carrier concentration increased by a factor of 2.6 (from $\sim 1.6 \times 10^{12}$ to $\sim 4.2 \times 10^{12}$ cm^{-3}) in the nonlinear portion of the *I*–*V* characteristic. In the CsI crystal at 77 K, the concentration n_0 varied from 2.4×10^{11} to $\sim 4.2 \times 10^{11}$ cm^{-3}, respectively. Photoexcitation of the *F* color centers in KBr at $E \sim 5 \times 10^4$ V/cm gave rise to $n_0 \sim 3.3 \times 10^{12}$ cm^{-3}. A large value of n_0 under photoexcitation is caused by a comparatively long lifetime of electrons in the conduction band of the insulator ($\sim 10^{-9}$ s [2]), which is a consequence of linear recombination of charge carriers in the case of capture by the color centers.

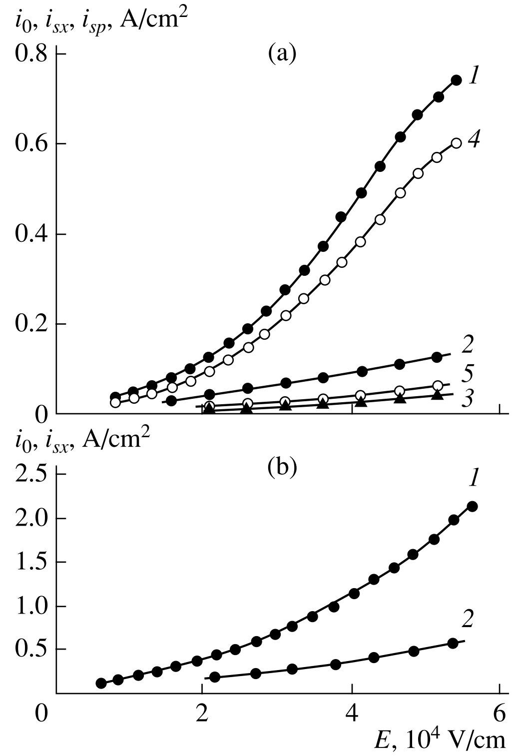


Fig. 2. Dependences of the (*1*, *4*) conduction-current density i_0 and the discharge currents (*2*, *3*) i_{sx} and (*5*) i_{sp} on the electric-field strength in the (a) KBr and (b) CsI crystals under the (*1*–*3*) X-ray and (*4*, *5*) laser excitation within the absorption band corresponding to the excitation of the *F* color centers. The amplitude *3* is increased by a factor of two.

The available data on the radiation-induced conductivity make it possible to estimate the values of the time constant $\tau_1 \sim C_1R_V$. At $E \sim 5 \times 10^4$ V/cm, $R_V \approx 4.4 \times 10^3$ Ω at the conductivity peak in KBr, so that $\tau_1 \approx 10^{-7}$ s. In CsI, $R_V \approx 1.8 \times 10^3$ Ω and $\tau_1 \approx 4.2 \times 10^{-8}$ s. The values of τ_1 exceed the current-pulse rise time ~10 ns in KBr and CsI by factor of 10 and 4, respectively. We may assume that the major portion of the external voltage drops across the insulator bulk.

The amplitude and temporal properties of the X-ray-excited i_{sx} and photoexcited i_{sp} discharge currents include the following. The current densities i_{sx} and i_{sp} (Figs. 2a, 2b) for the KBr and CsI crystals with the cemented contact (curves *2*, *5*) and deposited contact (curve *3*) increase somewhat superlinearly as the electric-field strength increases. The known lifetime of the near-surface charge τ_2 makes it possible to estimate the current densities i_{sx} and i_{sp} at the time instant of the charge-formation completion. In the case of the charge relaxation of holes in KBr and CsI, the i_{sx} amplitude should exceed the measured value by a factor of 1.4 and 1.1, respectively; the amplitude of i_{sp} is nearly constant if we are dealing with relaxation of anionic vacancies. In the context of the *RC*-circuit model, we can estimate the ratio between the amplitude values of the discharge-current density and the current density in the structure $i_{sx}/i \approx \Delta t/\tau_1$ after the conduction-current completion; this ratio is found to be equal to ~0.1 for KBr and ~0.25 for CsI (the used value of R_V corresponds to $E \sim 2 \times 10^4$ V/cm). The experimental ratios of the peak amplitudes of i_{sx} and i_{sp} to i_0 in the region of the high electric-field strength $E \sim 5 \times 10^4$ V/cm in KBr with the cemented contact are equal to ~0.2 and ~0.1, respectively; in CsI with the cemented contact, we have ~0.25. The value of this ratio is equal to ~0.07 for KBr with deposited contact. In the region of a low electric-field strength ($E \sim 2 \times 10^4$ V/cm), experimental ratio of the current densities i_s/i_0 exceed the estimates by a factor of ~2–2.5.

Three-to-five probing X-ray or optical pulses were required in order to ensure the complete drain off of the near-surface charge. The kinetic curve of the surface-charge relaxation was simulated by the plot of decay of the peak-pulse amplitudes i_{sx} and i_{sp} plotted along the time axis with steps of 30 ns. The relaxation dependences can be satisfactorily linearized if represented as $\ln i_s(t)$. In the KBr sample with the cemented contact, the relaxation time of holes τ_1 was equal to ~115 ns at a low electric-field strength ($E \sim 10^4$ V/cm) and to ~280 ns at a high electric-field strength ($E \sim 5 \times 10^4$ V/cm). In CsI, the relaxation times are shorter: ~60 and ~100 ns, respectively. The relaxation time for the charge of anionic vacancies in KBr is equal to ~140 ns at a low electric-field strength and to ~350 ns at a high electric-field strength. The near-surface charge of holes in the KBr sample with deposited electrode relaxes most slowly; the charge-relaxation time equals to ~1.1 μs. The aforementioned values of the charge-relaxation time are consistent with the estimate $\tau_1 \approx C_1R_V$ (the value of R_V corresponds to $E \sim 2 \times 10^4$ V/cm); this estimate is equal to ~290 ns for KBr and 71 ns for CsI. Undoubtedly, the measurement accuracy is lowered if short probing pulses are used. It is noteworthy that the relaxation time of the near-surface charge formed at low electric-field strengths E is smaller by a factor of 2–2.5 than the value of τ_1 for the charge formed at a larger value of E. Conceivably, the extent of the near-surface charge increases owing to the negative charge localized at the traps. The effective capacitance decreases, which leads to an increase in the amplitude of the discharge current and to a decrease in the charge-relaxation time.

In this study, the values of the density of the charge transported by the conduction current Q_0 were compared with the density of the surface charge Q_s at the time instant of the current-generation termination. Taking into account the summation of charge at $V = 500$ V ($E = 1.56 \times 10^4$ V/cm) in KBr with cemented electrode, we obtain the satisfactory agreement between the values of $Q_s = 1.2 \times 10^{-9}$ C/cm^2 and $Q_0 = 1.3 \times 10^{-9}$ C/cm^2. At a high voltage $V = 1500$ V, we obtain $Q_s = 7.7 \times 10^{-9}$ C/cm^2, which corresponds to ~70% of the charge $Q_0 = 1.13 \times 10^{-8}$ C/cm^2. In CsI at $V = 500$ V ($E \sim 1.6 \times 10^4$ V/cm), the value of charge $Q_s = 4.1 \times 10^{-9}$ C/cm^2 agrees satisfactorily with $Q_0 = 4.83 \times 10^{-9}$ C/cm^2. At 1500 V, $Q_s = 2.15 \times 10^{-8}$ C/cm^2 and $Q_0 = 2.86 \times 10^{-8}$ C/cm^2. It can be seen that the values of the charges Q_s and Q_0 differ by no more than ~30%. We may assume that $Q_s \approx Q_0$ within the experimental accuracy. This result indicates that there is no significant injection of electrons from the metal electrode at the stage of the conduction-current generation and localization of the positive charge at the negative electrode. The accuracy of estimating the value of Q_s for the samples with a deposited electrode is low.

The value of the potential φ_1 after completion of the conduction current was determined from the data on the charge densities Q_s and Q_0 and the capacitance C_1. The value of Q_s can be used to obtain $\varphi_1 \approx 26$ V in KBr at $V = 500$ V. At a higher voltage $V = 1500$ V, one can use the charge Q_0 to find that $\varphi_1 = 226$ V, which amounts to 15% of V. According to formula (3), the surface field strength $E_s = 2.5 \times 10^4$ V/cm and $E_1 = \varphi_1/d \approx 3.7 \times 10^4$ V/cm. As the conductivity increases, the potential drop across the hermetic's layer increases. For CsI at 1500 V, the charge Q_0 can be used to obtain $\varphi_1 \approx 571$ V (38% of the applied voltage V), so that $E_1 \approx 10^5$ V/cm. It is worth noting that, during the flow of the current, the values of E_1, E_s, and φ_1 are smaller than the above estimates, since the field strengths and the potential drop are controlled by the surface-charge density $\rho = \int_0^l (p - n)dx$ (here, n and p are the concentration profiles of electrons and holes in the sample).

The values of φ_s can be determined using the calculated estimates. It is assumed that, for the MIM structure (Fig. 1a) with the deposited electrode, all the quantities depend on the single coordinate x, which is reckoned from the left-hand surface of the sample. The spatial distributions of the concentrations n and p in the sample with quadratic recombination of charge carriers can be found from the continuity equations (4) and (5), the Poisson equation (6), and the boundary condition for the problem (7):

$$\frac{\partial n}{\partial t} = G - npv\sigma - \frac{\partial}{\partial x}\left(D\frac{\partial n}{\partial x} + \mu nE\right), \tag{4}$$

$$\frac{\partial p}{\partial t} = G - npv\sigma, \tag{5}$$

$$\frac{\partial E}{\partial x} = \frac{e(p-n)}{\varepsilon\varepsilon_0}, \qquad (6)$$

$$n = n_s \quad \text{at} \quad x = 0. \qquad (7)$$

Here, D is the diffusion coefficient for electrons, v is the thermal velocity of electrons, σ is the cross section for the electron–hole recombination, and n_s is the surface electron concentration.

Let us express the electron concentration as $n = n_0\exp(e\varphi/kT)$, where φ is the potential, k is the Boltzmann constant, and T is temperature. In the steady-state case, the hole concentration $p = n_0^2/n$ and, taking into account Einstein's relation $\mu = eD/kT$, the system of Eqs. (4)–(6) can be reduced to the following equation for n:

$$\frac{\partial^2 n}{\partial x^2} + \left(\frac{\partial n}{\partial x}\right)^2\frac{1}{n} + \frac{n^2}{L^2 n_0} - \frac{n_0}{L^2} = 0. \qquad (8)$$

Here, $L = (kT\varepsilon\varepsilon_0/e^2n_0)^{1/2}$ is the Debye screening length.

Equation (8) was solved using the Bulirsch–Stoer method [9]. Figure 3 shows the concentration profiles for holes in a KBr crystal with $n_0 = p_0 = 4.2 \times 10^{12}$ cm^{-3} and $L = 1.27$ μm for the following values of the electric-field strength at the insulator surface: $E_s = 10^3$, 10^4, and 10^5 V/cm. As E_s increases, the hole concentration p increases predominantly in the near-surface layer with a thickness of $(0.02–0.04)L$; this concentration at $x = 0$ is equal to about 7×10^{13}, 2.6×10^{15}, and 6×10^{16} cm^{-3} for the above-mentioned field strengths. In this case, the total hole density $P_s = \int_0^l p\,dx$ and the charge density ρ take almost the same values of 4.6×10^9, 2.83×10^{10}, and 2.8×10^{11} cm^{-2} (for P_s) and 2.8×10^9, 2.8×10^{10}, and 2.8×10^{11} cm^{-2} (for ρ). The values of the surface potential $\varphi_s = \int_0^l E\,dx$ tend to a certain limiting value as E_s increases: $\varphi_s \approx 0.07$, 0.16, and 0.21 V. At a depth $x \sim 3L$, the hole concentration $p \approx p_0$, so that the value of E is close to zero. Assuming that the spatial distributions of the surface charge in the dynamic and static conditions differ insignificantly in the KBr sample at $E \sim 5 \times 10^4$ V/cm, one can use the experimental estimate of the hole density $P_s = Q_0/e \approx 7.1 \times 10^{10}$ cm^{-2} to find that $\varphi_s \approx 0.18$ V. If an effective capacitance $C_e = Q_0/\varphi_s \approx 10^{-8}$ F is introduced for the junction between the deposited contact and insulator, the time needed for screening the electric field equals $\tau_1 \approx 10^{-4}$ s.

Electroneutrality in the sample can be established only owing to injection of external electrons from the surface. It is well known that tunneling sets in at the electric-field strength ~10^6 V/cm; evidently, the tunneling effect can be disregarded in the case under consideration. Most likely, the mechanism of the charge-carrier transport is related to the activation character of motion of electrons over energy levels introduced by dislocations into the band gap of the crystal. This inference is confirmed by an increase in the lifetime of the charge in CsI when the sample is cooled to the liquid-nitrogen temperature.

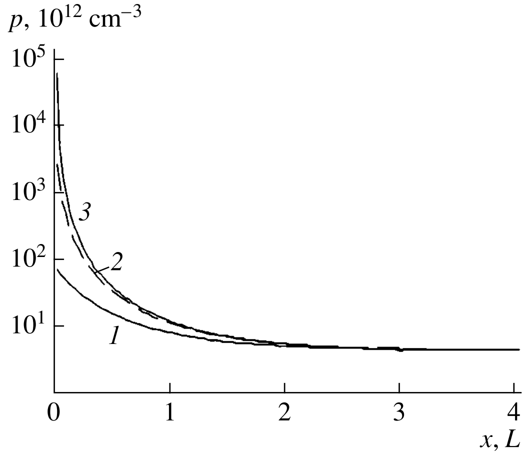


Fig. 3. Concentration profiles of holes p in the KBr sample for the surface electric-field strength equal to E_s = (*1*) 10^3, (*2*) 10^4, and (*3*) 10^5 V/cm.

CONCLUSION

Accumulation of positive surface charge at the negative electrode is a specific feature of the flow of current in a metal–insulator–metal structure. I suggested a method for measuring the current induced by the X-ray or optical excitation in order to determine the surface charge in wide-gap materials; this method is based on the initiation of the discharge current for an effective capacitor at the metal–insulator junction and on the estimation of the drained-off charge. The values of the charge and the conductivity of the sample were used to calculate the electric-field strength at the surface and the surface potential. According to the data obtained in this study, in a KBr sample that had a deposited electrode and was subjected to X-ray radiation for ~30 ns at an electric-field strength of ~5×10^4 V/cm, the conduction-current density is as high as ~0.6 A/cm^2; the electric-field strength and the potential at the metal–insulator junction are equal to ~2.5×10^4 V/cm and ~0.18 V, respectively; the effective capacitance of the junction equals ~10^{-8} F; and the time needed for screening of the field is equal to ~10^{-4} s. The small values of the surface electric-field strength and the values of the charge transported by the conduction current that almost coincide with those determined from the discharge-current signal indicate that there is no tunneling injection of electrons from the negative electrode during the conduction-current flow. Most likely, the electroneutrality of the sample is established owing to the motion of electrons from the surface over dislocations.

ACKNOWLEDGMENTS

I thank Yu.V. Lisyuk for his useful participation in discussions of the results of this study.

REFERENCES

1. V. A. Zuev and V. G. Popov, *Photoelectrical MIS Devices* (Radio i Svyaz', Moscow, 1983) [in Russian].
2. V. D. Kulikov and Yu. V. Lisyuk, Zh. Tekh. Fiz. **63** (7), 74 (1993) [Tech. Phys. **38**, 568 (1993)].
3. V. D. Kulikov, Fiz. Tverd. Tela (St. Petersburg) **43**, 1580 (2001) [Phys. Solid State **43**, 1644 (2001)].
4. R. T. Williams, J. N. Bredford, and W. L. Faust, Phys. Rev. B **18**, 7038 (1978).
5. E. D. Aluker, D. Yu. Lusis, and S. A. Chernov, *Electron Excitations and Radioluminescence in Alkali-Halide Crystals* (Zinatne, Riga, 1979) [in Russian].
6. Ch. B. Lushchik and A. Ch. Lushchik, *Relaxation of Electron Excitation with Defect Formation in Solids* (Nauka, Moscow, 1989) [in Russian].
7. V. D. Kulikov, Izv. Vyssh. Uchebn. Zaved. Fiz., No. 1, 74 (2001).
8. C. H. Seager and D. Emin, Phys. Rev. B **2**, 3421 (1970).
9. E. M. Kudryavtsev, *Mathcad 2000 Pro* (DMK, Moscow, 2001) [in Russian].

Translated by A. Spitsyn

Technical Physics, Vol. 49, No. 10, 2004, pp. 1375–1380. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 128–133.
Original Russian Text Copyright © 2004 by Kryachko, Levin, Tatarintsev, Bormontov.

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

A Noncontact Method to Study the Charge State of the Semiconductor–Insulator Interface

V. V. Kryachko†, M. N. Levin, A. V. Tatarintsev, and E. N. Bormontov
Voronezh State University, Voronezh, 394693 Russia
e-mail: levin@lev.vsu.ru

Received November 26, 2003

Abstract—We present a method for investigating the charge state of the semiconductor–insulator interface using the measurements of the contact potential difference between the surface of the insulator film and a vibrating probe. In this method, the surface electrostatic potential is changed through charge variation at the outer surface of the insulator. The charge value is determined by the time of the structure exposure to a corona discharge. The method is applied for investigating the effect of exposure of a silicon–silicon dioxide system to radiation and pulse magnetic fields. *© 2004 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The charge state of the semiconductor–insulator (SI) interface, which above all is understood as the fixed charge stored in the insulator and the energy distribution of the surface states (SS's) density over the semiconductor band gap, determines directly the parameters of the principal elements of the integrated circuits. Elaboration and improvement of the methods for controlling the charge state of the SI interface continues to be a necessary condition for progress in microelectronics.

The problems of studying the SI interface in the metal-insulator-semiconductor (MIS) structure were solved successfully by using the capacitance spectroscopy techniques, among which the principal ones are the Terman method of measurement of the high-frequency capacitance-voltage (*C*–*V*) characteristics [1], the Nicollian–Goetzberger method that involves the measurement of the total conductivity [2], and the Berglund method involving the low-frequency or quasi-static *C*–*V* characteristics recording [3]. The study of the SI interface without deposition of a metal electrode is no less important for semiconductor device technology and microelectronics. The principal difficulties in solving this problem are due to the fact that the charge properties of such structures are extremely sensitive to probing actions.

The aim of this study is the development of a method for investigating the charge state of the semiconductor–insulator interface without the formation of MIS structure.

METHODOLOGY

It is known that the distribution of the surface state density over the semiconductor band gap N_{ss} (SS spectrum) is determined as

$$N_{ss}(\psi_s) \equiv -q^{-1}\frac{\partial Q_{ss}(\psi_s)}{\partial \psi_s}, \qquad (1)$$

where ψ_s is the surface potential, Q_{ss} is the charge stored in SS's, and q is the electron charge.

The SS spectrum can be obtained expressing the SS charge $Q_{ss}(\psi_s)$ from the equation that describes the steady-state potential distribution in a MIS structure,

$$V_g = \varphi_{ms} + \psi_s - C_{0x}^{-1}(Q_{0x} + Q_{sc}(\psi_s, \varphi_b) + Q_{ss}(\psi_s)), \qquad (2)$$

and differentiating it with respect to ψ_s,

$$N_{ss}(\psi_s) = \frac{C_{0x}}{q}\left[\left(\frac{\partial \psi_s}{\partial V_g}\right) - 1\right] - \frac{C_{sc}(\psi_s, \varphi_b)}{q}. \qquad (3)$$

Here, V_g is the gate voltage of the MIS structure, φ_{ms} is the contact potential difference (CPD) between the metal gate and semiconductor facing of the MIS structure being studied, C_{0x} is the specific capacity of the insulator layer, Q_{0x} is the total charge stored in the insulator per unit area, Q_{sc} is the charge stored in the depletion region of the semiconductor, φ_b is the bulk potential of the semiconductor, and C_{sc} is the capacitance of the depletion region of the semiconductor.

The problem of obtaining the SS spectrum is reduced to the experimental determination of the dependence of semiconductor surface potential ψ_s on gate voltage V_g of the MIS structure.

In the recognized capacitance methods for studying the charge state of MIS structures, the $\psi_s(V_g)$ dependence is determined from the experimental capacitance–voltage (*C*–*V*) characteristics using theoretical dependencies of the capacitance of the semiconductor depletion region on the surface potential.

---

† Deceased.

1063-7842/04/4910-1375$26.00 © 2004 MAIK "Nauka/Interperiodica"

Unlike the variation of the surface potential or band bending ψ_s in the semiconductor of the MIS structure, which is accomplished by the simple setting of gate voltage V_g, the variation of the band bending in the SI structure represents the independent problem.

In the proposed method for investigating the charge state of the SI interface, the surface potential of the semiconductor is changed through the charge variation at the outer surface of the insulator film. The value of this charge is set by the time of SI structure exposure to a corona discharge. It is essential that no data on the absolute value of the charge placed at the insulator surface be required to obtain SS spectrum at the SI interface. It is sufficient to vary this charge and to measure, not the charge itself, but the corresponding values of the potential of the surface V_C and the surface potential ψ_s in the semiconductor. The potential of the surface V_C is determined employing CPD measurements by the Kelvin probe method [4] with the calibrated probe, which scans the surface of the semiconductor wafer, and the band bending in the semiconductor is determined from the difference of the V_C values measured in the dark and under illumination with infrared or visible light, V_C^L, which ensures photovoltage saturation corresponding to the flat-band condition

$$\psi_s = V_C - V_C^L. \tag{4}$$

The derivation of the parametric dependence $\psi_s(V_g)$ for a SI structure in such a manner in order to obtain SS spectrum and the effective charge in the insulator is the kernel of the proposed method.

Let us consider the technique for determining the charge characteristics of the semiconductor–insulator interface employing the contact potential difference method. In the presence of a difference in the work functions for the metal and the semiconductor, the potential of the insulator surface in the MIS structure will be determined as follows:

$$V_C - \varphi_{ms} = V_{0x} + \psi_s. \tag{5}$$

The following equality should be fulfilled to ensure electroneutrality in the structure:

$$Q_0 + Q_{sc} + Q_{ss} + Q_{0x} = 0. \tag{6}$$

Here, V_{0x} is the voltage drop across the insulator layer of thickness d_{0x} and Q_0 is the charge at the outer surface of the insulator. The analytical dependence of the semiconductor space charge $Q_{sc}(\psi_s, \lambda)$ at a specified doping level is known [5].

The charge Q_{0x} stored in the insulator is determined by the integral

$$Q_{0x} = \int_0^{d_{0x}} \rho(x)dx, \tag{7}$$

where $\rho(x)$ is the depth distribution of charge in the insulator (coordinate $x = 0$ corresponds to the surface of the insulator).

The voltage drop V_{0x} across the insulator layer is determined by the integral

$$V_{0x} = \int_0^{d_{0x}} E(x)dx, \tag{8}$$

whereas the electric field $E(x)$ in the insulator can be obtained by solving the Poisson's equation

$$\frac{dE}{dx} = \frac{\rho(x)}{\varepsilon_{0x}}. \tag{9}$$

Integrating Eq. (9), we obtain

$$E(x) = E(0) + \frac{1}{\varepsilon_{0x}}\int_0^x \rho(x)dx, \tag{10}$$

where ε_{0x} is the absolute permittivity of the insulator.

Substitution of Eq. (10) into equality (8) gives

$$V_{0x} = E(0)d_{0x} + \frac{1}{\varepsilon_{0x}}\int_0^{d_{0x}} dx \int_0^x \rho(x)dx. \tag{11}$$

Taking the integral in expression (11) by parts, we obtain

$$V_{0x} = E(0)d_{0x} + \frac{Q_{0x}}{C_{0x}} - \frac{1}{C_{0x}}\int_0^{d_{0x}} (x/d_{0x})\rho(x)dx. \tag{12}$$

Then, using the equality

$$E(0)d_{0x} = \frac{Q_0}{C_{0x}} = -\frac{Q_{sc} + Q_{ss} + Q_{0x}}{C_{0x}}, \tag{13}$$

we obtain from Eqs. (5) and (13)

$$\begin{aligned} V_C - \varphi_{ms} = {} & -\frac{Q_{sc}(\psi_s, \lambda) + Q_{ss}(\psi_s)}{C_{0x}} \\ & -\frac{1}{C_{0x}}\int_0^{d_{0x}} (x/d_{0x})\rho(x)dx + \psi_s. \end{aligned} \tag{14}$$

The integral appearing on the right-hand side of equality (14) represents the effective charge in the insulator reduced to the SI interface,

$$Q_{0x\,\mathrm{cff}} = \int_0^{d_{0x}} (x/d_{0x})\rho(x)dx. \tag{15}$$

Taking into account Eq. (15), relation (14) can be written as follows:

$$Q_{ss}(\psi_s) = -C_{0x}(V_C - \varphi_{ms} - \psi_s) - Q_{sc}(\psi_s) - Q_{0x\,\mathrm{eff}}. \tag{16}$$

It was found that the surface states at the Si–SiO$_2$ interface are caused by the so-called P_b centers, i.e., trivalent silicon atoms (–Si≡Si$_3$), which have one unsaturated bond. These P_b centers are amphoteric, exhibiting donor properties in the bottom half of the band gap and acceptor properties in the upper half [6, 7]. As the surface Fermi level crosses the middle of the band gap, i.e., when surface potential ψ_s is equal to bulk potential φ_b (middle gap mode), there is no charge at the P_b centers. In this mode, the SS charge is $Q_{ss\,mg} \equiv Q_{ss}(\psi_s = \varphi_b) = 0$. Then, in the context of approximation on the amphoteric origin of SS's at the Si–SiO$_2$ interface, to determine the effective charge in the oxide layer, the middle gap mode should be ensured by changing the surface potential ψ_s through the variation of the charge Q_0 at the outer surface of the insulator layer, e.g. by treatment of the structure with a corona discharge.

Equality (16) makes possible the obtaining of the expression for calculating the effective charge in the insulator layer $Q_{0x\,\mathrm{eff}}(\psi_s = \varphi_b)$ for the middle gap mode,

$$\begin{aligned} &Q_{0x\,\mathrm{eff}}(\psi_s = \varphi_b) \\ &= -C_{0x}(V_{Cmg} - \varphi_{ms} - \varphi_b) - Q_{sc}(\psi_s = \varphi_b, \lambda). \end{aligned} \quad (17)$$

Relation (4) enables us to rewrite Eq. (17) as follows:

$$\begin{aligned} &Q_{0x\,\mathrm{eff}}(\psi_s = \varphi_b) \\ &= -C_{0x}(V^L_{Cmg} - \varphi_{ms}) - Q_{sc}(\psi_s = \varphi_b, \lambda). \end{aligned} \quad (18)$$

Differentiating equality (16) with respect to surface potential ψ_s, we obtain formula (1) for calculating the energy density of surface states.

Probe scanning over the surface of the semiconductor wafer makes possible the obtaining of the planar distribution of the potential and the estimation of the character of its fluctuations, as well as fluctuations of the surface charge.

Since the fluctuations of the surface charge and potential stemming from the random character of the local actions of technological processes are statistically independent at various points at the semiconductor surface, their distributions can be described by the Gaussian distribution [8]

$$P_\psi \psi_s = \frac{1}{\sigma_\psi \sqrt{2\pi}} \exp\left(-\frac{(\psi_s - \overline{\psi_s})^2}{2\sigma_\psi^2}\right), \quad (19)$$

$$P_Q(Q_s) = \frac{1}{\sigma_Q \sqrt{2\pi}} \exp\left(-\frac{(Q_s - \overline{Q_s})^2}{2\sigma_Q^2}\right), \quad (20)$$

where $\overline{\psi_s}$ and $\overline{Q_s}$ are the average values of the surface potential and surface charge, respectively, and σ_ψ and σ_Q are the mean square deviations of the potential and charge, respectively.

The main contribution to fluctuations of the surface charge Q_s in the Si–SiO$_2$ structures comes from the fluctuations of the fixed charge of the oxide Q_{0x}. Then, using the Nicollian–Goetzberger model [8], we can calculate the parameters of the Gaussian distribution for the charge stored in the insulator Q_{0x} from the results of the direct measurements of the parameters of surface potential distribution ψ_s as

$$\overline{Q_{0x}} = \frac{W^2 \sigma_Q^2}{q}, \quad (21)$$

$$\sigma_Q = \sigma_\psi (C_{0x} + C_{sc} + C_{ss}), \quad (22)$$

where $W = (2\varepsilon_s \overline{\psi_s}/qN_B)$ is the average thickness of the depletion region, ε_s is the permittivity of the semiconductor, and N_B is the doping impurity concentration in the semiconductor.

IMPLEMENTATION OF THE METHOD

To realize the proposed method, a computer-aided measuring device is constructed, which comprises a noncontact scanning millivoltmeter with a Kelvin probe and a setup for automatic maintenance of the separation between vibrating probe and surface being analyzed.

The functional scheme of the device is shown in Fig. 1. The device consists of the following modules and blocks: desktop *1.3* with probing head *1.1*, low-frequency generator *5*, high-frequency generator *6*, preamplifier *2*, tracking module *3*, power amplifier *4*, and compensation module *7*.

The desktop is intended for fixing the semiconductor wafer being analyzed, which can be 100 or 150 mm in diameter. The probing head is an electromagnetic vibrator consisting of a movable cylinder fixed with bracing wires in a frame with a magnetic system. Vibra-

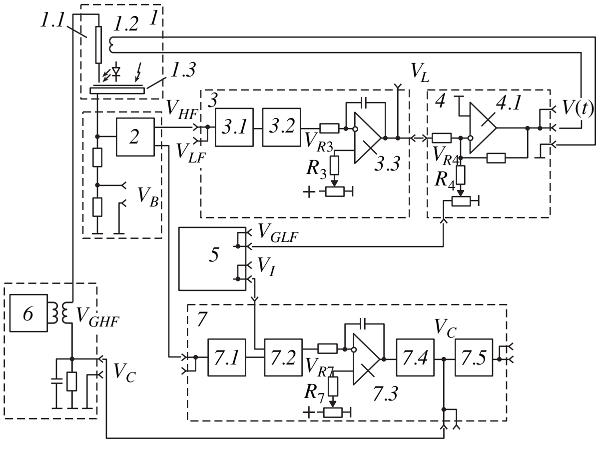


Fig. 1. Functional scheme of the measuring device.

tor coil *1.2* is situated on the cylinder. A platinum probe with a flat face of diameter d = 100–300 μm is attached to the end of the cylinder. The face of the probe is polished, cleaned, and thermally treated. This ensures the permanency of the work function of the probe material. The probe calibration is performed using a material with the known work function, in particular, mercury or graphite. In the course of measurements, the probe performs vibrations above the sample surface with an amplitude of about 5 μm. The frequency of these vibrations is determined by the design of the probing head, and for this particular construction of the device it is equal to 800 Hz [9]. The distance between the vibrating probe and the surface being analyzed is $d_0 \approx 0.1d$. This distance is kept constant with an accuracy of ±1 μm during the probe movement above the surface being studied as its relief changes within ±25 μm. The velocity of the probe movement above the sample surface is 4 mm s^{-1}, and the range of CPD measurement is ±5 V. A corona discharge unit, which furnishes a means for surface charging by positive or negative charge, and an optical unit for illuminating the surface up to the attainment of the flat-band conditions are embedded into the probing head.

Low-frequency generator *5* forms the sinusoidal electric signal V_{GLF} of frequency 800 Hz (probe swinging signal), which is fed to vibrator coil *1.2* through power amplifier *4* and ensures vibrations of the probe with a preset frequency and amplitude. Besides, the generator produces a rectangular signal with the adjustable phase shift V_I to ensure the operation of the synchronous detector of the compensation module.

High-frequency generator *6* forms the signal V_{GHF} of frequency 500 kHz required for the operation of the tracking module that ensures the automatic maintenance of the distance d_0 between the probe and the surface being studied.

Preamplifier *2* has two channels and serves for selective amplification of the low-frequency component of the measured signal and the high-frequency component of the signal that is used for operation of the tracking system. The V_B socket positioned on the preamplifier board serves for applying bias voltage to the desktop circuit intended for calibration of the device or variation of the operating range of the compensation voltage V_C measurement if the CPD value falls beyond the operating range of the device.

The tracking module is intended for the formation of the control signal V_L that ensures a preset value of the gap between the probe and the sample.

The power amplifier forms the signal $V(t)$ supplied to the coil of the magnetoelectric system of the probe vibrator to control the amplitude of its vibrations and the gap value.

The compensation module serves to form the constant compensation voltage V_C, which is equal in magnitude to the contact potential difference between the local part of the surface being studied and the measuring surface of the probe.

The measuring device operates as follows. The offset current, which has two components, flows in the circuit in response to the probe vibration in the presence of an electric field in the gap. The value of the low-frequency component of frequency 800 Hz (frequency of probe vibration) is determined by the potential of the surface being studied. The value of the high-frequency component (of 500 kHz frequency) is determined by the deviation from the preset value of the gap between the probe and the sample. Each of these components is enhanced by the preamplifier.

The high-frequency signal component V_{HF} comes from the output of the preamplifier to the input of the tracking module, where it is enhanced by resonant *LC*-amplifier *3.1* and detected by amplitude detector *3.2*. The voltage, which is practically linearly related to the V_{HF} value and inversely proportional to the probe gap value, is generated at the output of the amplitude detector. This voltage is applied to the inverting input of tracking module integrator *3.3*; the reference voltage V_{R3} presetting the value of the operating probe gap with the help of regulator $R3$ is applied to noninverting input of the integrator. The difference of these voltages is enhanced by operational amplifier *3.3*, at whose output the tracking signal V_L is formed.

The tracking signal V_L at the input of the power amplifier is summed with the swinging signal of the probe vibrator V_{R4}, whose value is set by the probe amplitude vibration regulator $R4$. At the output of the power amplifier loaded on the vibrator coil, the voltage $V(t)$ is formed, which constitutes the sum of the constant component of tracking voltage and the harmonic component of the probe swinging, $V(t) = V_- + V_\sim$, where V_- and $V_\sim$ are the constant and harmonic components of the signal $V(t)$, respectively. The value V_- is intended for probe gap stabilization, and the value $V_\sim$ sets the amplitude of the probe vibrations.

As the probe gap changes, the tracking system forms an output signal in such a way that the value of constant component V_- flowing in the circuit of the vibrator coil is changed compensating thereby the gap variation.

The low-frequency component of the signal V_{LF} comes from the preamplifier output to the input of active band-pass filter *7.1*. The amplified signal is detected by synchronous detector *7.2* controlled by the rectangular pulses from the low-frequency generator (V_I). At the output of the synchronous detector, a unidirectional signal is formed whose polarity is determined by the relation between the work functions of the sample being studied and the platinum probe. This signal is integrated by integrator *7.3* and passes to the output of the compensation module via transistor voltage follower *7.4*. Regulator $R7$ of the integrator is intended for voltage compensation at the output of the synchronous

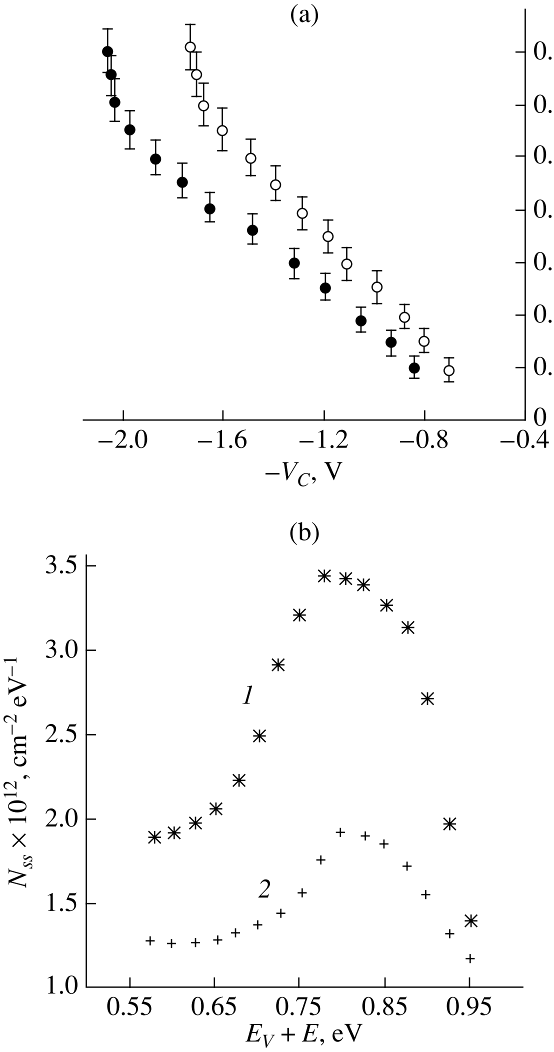


Fig. 2. Experimental dependencies of the surface potential ψ_s on the compensation voltage V_C (a) and calculated SS spectra (b) of the irradiated samples. The measurements were performed in 10^3 (*1*) and 10^5 s (*2*) of irradiation.

detector as $V_{LF} = 0$ at the input of the compensation module.

Thus, at the output of the compensation module, the compensation voltage V_C is formed, which is equal but opposite to the contact potential difference between the probe and the surface being analyzed. This voltage is applied to the probe through the feedback circuit, with the result that the electric field in the gap is fully compensated and the offset current in the measuring circuit is terminated. The voltage V_C comes to a computer input through analog-digital converter *7.5*.

APPLICATION OF THE METHOD

The elaborated method was used to determine the radiation-induced modification of the SS spectrum and to investigate the statistical parameters of the charge distribution over the surface of a semiconductor wafer.

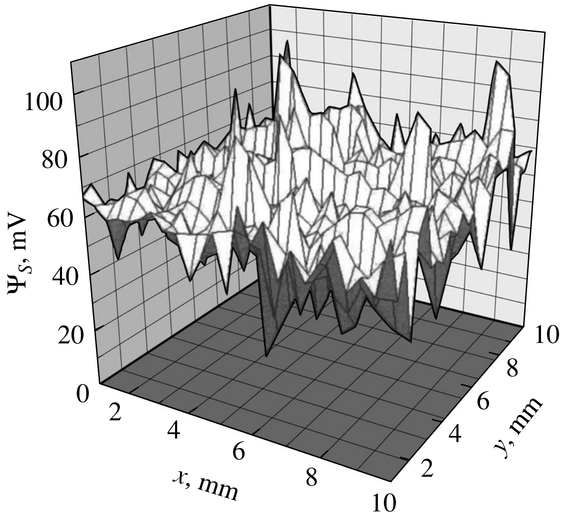


Fig. 3. Potential relief $\psi_s(x, y)$ of a silicon–silicon dioxide interface.

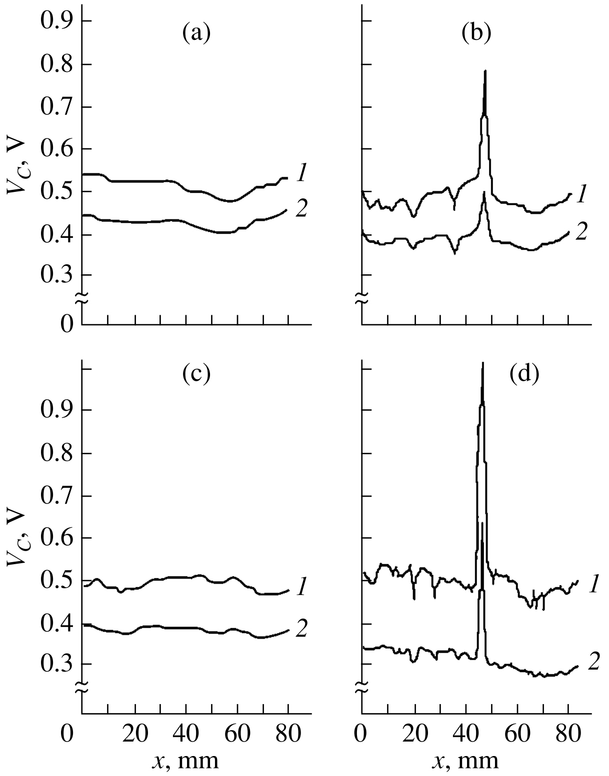


Fig. 4. Distribution of the compensation voltage V_C along the diameter of the wafer: (a) initial; (b, c) in 60 s and 50 h after exposure to corona discharge; (d) after the exposure to a PMF. (*1*, *2*) Measurement in a dark and with illumination of the semiconductor surface, respectively.

Polished ⟨100⟩-oriented silicon wafers of KÉF-7.5 and KDB-12 types 100 mm in diameter were used in the measurements. The wafers were oxidized according to the standard procedure for the gate insulator formation. The oxide thickness was from 40 to 98 nm. The SS energy spectrum was changed by exposure to soft X-ray radiation with the energy $E_x \sim 20$ keV of quanta and a dose up to 2×10^2 C kg^{-1} using an X-ray setup IRIS-M3. The measurements were performed both at individual points and during continuous scanning over the surface being studied.

Experimental dependencies of surface potential ψ_s on compensation voltage V_C measured at different instants after irradiation, as well as SS spectra at the SI interface calculated from these dependencies, are presented in Fig. 2. The energy distribution of the density of radiation-induced SS's consists of a broad peak with the maximum near $E = E_V + 0.75$ eV. Such an SS spectrum is typical of the known P_b centers generated by radiation at the Si–SiO$_2$ interface [10].

The typical potential relief of the oxide–semiconductor interface over an area of 1 cm^2 of one of the wafers being studied is shown in Fig. 3. Note that the character of potential distribution ψ_s was reproducible both within the area of one wafer and in different wafers.

The distribution function of the surface potential fluctuations obtained as a result of the treatment of experimental data presented in Fig. 3 is a Gaussian function with the parameters $\overline{\psi_s} = 65$ mV and $\sigma_\psi = 15$ mV.

The parameters of the Gaussian distribution for the stored charge calculated using formulas (21), (22) have the values $\overline{Q_{0x}} = 7.2 \times 10^{-8}$ C cm^{-2} and $\sigma_Q = 2.5 \times 10^{-9}$ C cm^{-2}.

The high sensitivity of the method to the charge state of the semiconductor surface makes possible, in addition to the quantitative measurements, the express-control of the quality of semiconductor wafers to be performed in order to reveal the latent defects, which are manifested as local inhomogeneities of the surface potential. The distributions of compensation voltage V_C along the diameter of a silicon KDB-12 wafer before and after the exposure of its bounded area to a corona discharge are shown in Fig. 4. It is obvious from this figure that the exposure to the discharge results in the formation of the local inhomogeneity at the semiconductor surface, whose position is fixed and can be determined from the variation of the V_C value (Fig. 4b). After two days, the peculiarities introduced by the exposure to the corona discharge were no longer manifested in a given point (Fig. 4c). However, subsequent treatment with a pulsed magnetic field (PMF) makes it possible to detect the area of the exposure to the corona discharge by an abrupt local variation of V_C. The increase of the values of compensation voltage and surface band bending can be explained by the increase of the local density of the positively charged surface states at the interface.

CONCLUSIONS

(1) A new noncontact method for studying the charge state of the semiconductor–insulator interface was developed using the Kelvin probe measurement of the CPD and surface electrostatic potential modulation by the corona discharge.

(2) In order to realize this method, a computer-aided measuring device was designed, comprising a noncontact scanning millivoltmeter with a Kelvin probe and automatic maintenance of the distance between the oscillating probe and the surface being analyzed.

(3) The potentialities of the method were demonstrated for studying the effect of exposure of the silicon–silicon dioxide system to radiation and pulsed magnetic fields.

REFERENCES

1. I. M. Terman, Solid-State Electron. **5**, 285 (1962).
2. E. N. Nicollian and A. Goetzberger, IEEE Trans. Electron Devices **12**, 108 (1965).
3. C. N. Berglund, IEEE Trans. Electron Devices **13**, 701 (1966).
4. G. P. Peka, *Physical Phenomena on Semiconductor Surface* (Vishcha Shkola, Kiev, 1984) [in Russian].
5. S. Sze, *Physics of Semiconductor Devices* (Wiley, New York, 1981; Mir, Moscow, 1984).
6. P. M. Lenahan and P. V. Dressendorfer, J. Appl. Phys. **55**, 3495 (1984).
7. M. A. Jupina and P. M. Lenahan, IEEE Trans. Nucl. Sci. **37**, 1650 (1990).
8. E. N. Nicollian and A. Goetzberger, Bell Syst. Tech. J. **46**, 1055 (1967).
9. V. N. Budko and V. V. Kryachko, Kondens. Sredy Mezhfaznye Granitsy **5**, 155 (2003).
10. *Ionizing Radiation Effects in MOS Devices and Circuits*, Ed. by T. P. Ma and P. V. Dressendorfer (Wiley Interscience, New York, 1989).

Translated by M. Lebedev

Technical Physics, Vol. 49, No. 10, 2004, pp. 1381–1383. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 134–136.
Original Russian Text Copyright © 2004 by Petrakov, Kryazhev.

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

X-ray Phase-Contrast Study of Dissolution of NaCl Crystals

A. P. Petrakov and A. A. Kryazhev
Syktyvkar State University, Syktyvkar, 167001 Russia
e-mail: petrakov@syktsu.ru

Received February 11, 2004

Abstract—The X-ray phase contrast method is used to study changes in a solution during the dissolution of NaCl crystals. The effect of the solution concentration on the refractive index for X rays is studied experimentally. The possibility of determining the refraction decrement in a local region of the solution is demonstrated. *© 2004 MAIK "Nauka/Interperiodica".*

INTRODUCTION

X rays are electromagnetic waves with a very short length (~1 Å) and, hence with a high penetrability. The phase front of an electromagnetic wave passing through an object experiences distortions leading to refraction of X rays. The angle of refraction depends on the radiation energy and on the electron density of the object [1].

The phase contrast method is so sensitive that it makes it possible to detect the change in the density of solutions upon dissolution of crystals. The dissolution of the crystals is accompanied by a deviation of the density and composition in individual microscopic regions from their mean values over the entire volume. Fluctuations may be so strong that the observation of light scattering in the visible spectral range becomes possible [2].

The mass transfer occurring during the dissolution of crystals leads to a change in the concentration in the space of the medium. By monitoring this parameter in the course of dissolution, it is possible to determine the process dynamics and to study the effects of interaction between the crystal and the solution.

This study is aimed at analysis of local changes in the physical properties of a solution during dissolution of NaCl crystals with the help of the phase contrast method.

EXPERIMENTAL TECHNIQUE

Experiments were made on a three-crystal X-ray diffractometer. The monochromator was a silicon single crystal with symmetric triple reflection (111). The analyzer was also a silicon single crystal, but with single analogous reflection. To increases the sensitivity, the analyzer was deflected from the exact Bragg position towards smaller angles by a quantity equal to the half-width of the diffraction reflection curve (DRC) of the analyzer [3]. Radiation of CuK_{α_1} was used.

After the monochromator, an X ray passed through a $4 \times 10 \times 30$-mm plane-parallel polyethylene cell filled with distilled water. A nylon rod of radius 0.7 mm was mounted at the center of the cell. NaCl single crystals weighing 0.051, 0.105, and 0.161 g (±0.001 g) were placed one by one at a distance of 1 mm from the rod. Each crystal was immersed in fresh water. In all experiments, the dependence of X-ray radiation intensity on the time of crystal dissolution was recorded. An X ray passed through the rod boundary at the same level with the crystals being dissolved. The schematic of the experiment is shown in Fig. 1. The X-ray radiation intensity was recorded by a scintillation detector with a time step of 1 s during 1000 s after the beginning of dissolution of the crystals.

EXPERIMENTAL RESULTS AND DISCUSSION

Figure 2 shows the dependence of the X-ray radiation intensity on the time of dissolution of single crystals. The intensity recorded by the detector before

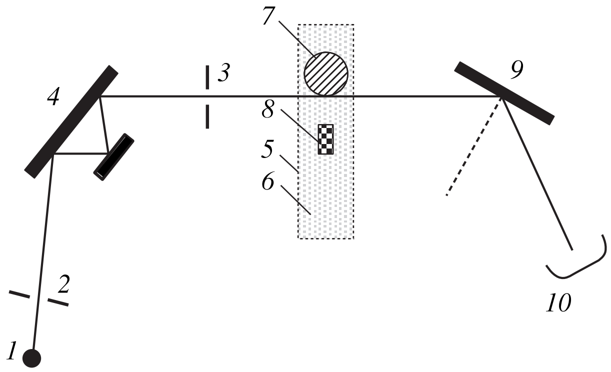


Fig. 1. Schematic diagram of the experiment: (*1*) radiation source; (*2*, *3*) slits, (*4*) crystal–monochromator; (*5*) polyethylene plane-parallel cell; (*6*) solution; (*7*) nylon rod; (*8*) crystal to be dissolved; (*9*) crystal–analyzer; (*10*) detector.

1063-7842/04/4910-1381$26.00 © 2004 MAIK "Nauka/Interperiodica"

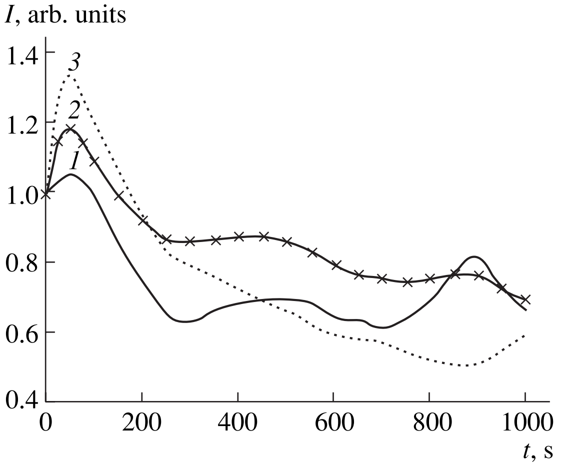


Fig. 2. Dependence of the X-ray radiation intensity recorded by the detector on the time of dissolution of the crystals with a weight of 0.051 (*1*), 0.105 (*2*), and 0.161 g (*3*).

immersion of a crystal in the solution was taken as unity. The statistical error of measurements was 0.3%. All curves display a peak near 50 s. The peak height increases with the curve number. Subsequently, the X-ray radiation intensity on all the curves decreases with time almost uniformly up to 250 s, after which the decrease slows down. Curves *1* and *2* exhibit even an increase in intensity in the region of 500 and 900 s. The increase in intensity on curve *3* is observed only after 900 s.

The table contains instants of time at which the extrema of the refraction decrement δ are observed, as well as the corresponding values of n/n_0 (n is the electron density of the solvent at the instant of measurement and n_0 is the electron density of water). It can be seen from the table that the first extremum was observed for all curves in the region of 50 s. The second extremum on curves *1* and *2* was formed after 300 s of dissolution, while all subsequent extrema appeared periodically over 200 s.

Values of refractive decrement δ and relative electron density n/n_0 at extremal points in Fig. 1

t, s	Curve number	Extremum	$\delta \times 10^{-6}$	n/n_0
50	*1*	max	3.254	1.023
	2	max	3.259	1.025
	3	max	3.263	1.026
300	*1*	min	3.185	1.001
	2	min	3.193	1.004
	3	–	–	
500	*1*	max	3.215	1.011
	2	max	3.222	1.013
	3	–	–	
700	*1*	min	3.189	1.003
	2	min	3.191	1.003
	3	–	–	
900	*1*	max	3.262	1.025
	2	max	3.253	1.023
	3	min	3.185	1.001

Water has a large dielectric constant ~78.3 at 25°C [2]. This explains the high solubility of polar compounds. Water is characterized by a short-range order; in the given approximation, we can speak of the structure of water. During dissolution, ions are arranged in the structure of water or in the voids, or substitute for H_2O molecules. Dissolution leads to violation of the structure of water due to mismatching between the sizes of the ions and water molecules as well as due to forced orientation of water molecules. Ions may cause not only damage to the structure, but also its ordering. This is due to the fact that the interaction between the ions and water molecules leads to a displacement of the electron density to the cation; as a result, the interaction of water molecules from the first hydrate layer with water molecules from the second layer becomes stronger. When NaCl is dissolved, Na^+ ions tend to order the structure, while Cl^- ions strive to break it. Both destruction and ordering affect the processes accompanying the dissolution of the crystals.

The phase contrast method is based on the variation of the coefficient of reflection from the crystal–analyzer as a result of refraction of X rays in the object being studied. The strongest refraction takes place at a boundary between the media of a rounded shape [4]; for this reason, the intensity of X-ray radiation was measured at the boundary of the nylon rod. In the experimental setup used here, during displacement of the cylindrical rod in air in a direction perpendicular to the X ray, a maximum was observed when the rod entered the X-ray region and a minimum was detected as the rod left the X-ray region [5]. The experiments were made in the vicinity of the minimum.

Dissolution of crystals is accompanied by an increase in the density of the solution and, hence, in the refraction decrement. As a result, the difference in the refraction decrements at the rod boundary decreases, which is accompanied by an increase in the radiation intensity. The intensity of the X ray reflected from the crystal–analyzer is defined [6, 7] as

$$I = I_a R(\Delta Q - \beta),$$

where I_a is the intensity associated with absorption, $R(\alpha)$ is the DRC from the crystal–analyzer, ΔQ is the angle of rotation of the analyzer, and β is the angle of refraction, which depends on the variation of the refraction decrement.

The increase in intensity observed for all curves during the first 50 s is due to the increase in the density of the solution at the rod boundary as a result of dissolution of crystals. The height of the peak increases with the crystal mass. This is due to the fact that the area of the crystal surface also increases with its mass. An increase in the surface of dissolution is accompanied by an increase in the concentration of dissolved ions.

The subsequent decrease is associated with an increase in the distance to the nearest face of the crystal due to its dissolution. The distance becomes so large that the ions that have arrived in the region under investigation diffuse to the remaining part of the cell at a rate higher than the rate of arrival of new ions from the crystals being dissolved.

Subsequent retardation in the decrease of the intensity and even its increase are caused by hydrodynamic jets. Our experiments were made in an open cell, which must inevitably be accompanied by evaporation of hydrodynamic jets that increase the concentration of dissolved particles in some regions of the solution [8].

Using the formula given above, we can derive the dependence of the refractive decrement of the solvent on the dissolution time (the refractive decrements of nylon and water are 3.52×10^{-6} and 3.18×10^{-6}, respectively). For this purpose, we must know the time dependence of the intensity associated with absorption alone. This quantity was measured in experiments carried out without a crystal–analyzer. The refractive decrement determined in this way varied with time in the same way as the intensity in Fig. 1. The maximal and minimal values are given in the table. The table also contains the values of n/n_0 showing the variation of the electron density of the solvent. These values were obtained using the fact that the refractive decrement is proportional to the electron density.

It can be seen from the table that, at the extrema corresponding to maxima, the electron density of the solution substantially differs from the analogous value for water. At the minima, the values are also higher than for water, but insignificantly so. Alteration of extrema with a period of 200 s after 300 s of dissolution on curves *1* and *2* indicates that evaporation of the solvent leads to the formation of regions in which the electron density increases and decreases periodically in time.

The late attainment of the minimal value on curve *3* indicates that the processes caused by the evaporation of the solvent depend on the weight of the crystals being dissolved. This can be explained as follows. Relatively stable aquacomplexes formed in the solution consist of ions surrounded by water molecules. In the course of dissolution, the ion concentration in a certain region attains a value at which ions compete for water molecules. This struggle for water molecules affects the process of generation of hydrodynamic jets. The increase in the intensity on curve *3* (Fig. 1) after 900 s of dissolution indicates that the competition for water molecules cannot rule out the generation of hydrodynamic jets, but can postpone the beginning of their formation. The increase in the intensity on curves *1* and *2* takes place much earlier than on curve *3*, since the ion concentration in the dissolution of the corresponding crystals is lower than the limit beyond which the competition for water molecules takes place.

The investigations performed here lead to the following conclusions. X-ray phase-contrast studies make it possible to measure the refraction decrement (which varies in the process of dissolution of NaCl crystals) in local regions of solutions. Evaporation of solvent from the open surface causes periodic variation of the electron density in individual regions of the solution.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, project nos. 03-03-16239 and 02-04-64688.

REFERENCES

1. T. J. Davis, D. Gao, T. E. Gureyev, *et al.*, Nature Lett. **373**, 595 (1995).
2. T. G. Petrov, E. B. Treĭvus, Yu. O. Punin, and A. P. Kasatkin, *Crystal Growth from Solutions* (Nedra, Leningrad, 1983) [in Russian].
3. V. A. Bushuev and A. P. Petrakov, Kristallografiya **46**, 209 (2001) [Crystallogr. Rep. **46**, 171 (2001)].
4. V. A. Bushuev, V. N. Ingal, and E. A. Belyaevskaya, Kristallografiya **41**, 808 (1996) [Crystallogr. Rep. **41**, 766 (1996)].
5. A. P. Petrakov, Zh. Tekh. Fiz. **73** (5), 84 (2003) [Tech. Phys. **48**, 607 (2003)].
6. V. A. Bushuev, V. N. Ingal, and E. A. Belyaevskaya, Kristallografiya **43**, 586 (1998) [Crystallogr. Rep. **43**, 538 (1998)].
7. T. E. Gureyev and S. W. Wilkins, Nuovo Cimento D **19**, 545 (1997).
8. V. I. Rakin, *Spatial Irregularities in Crystal-Forming Structure* (UrO RAN, Yekaterinburg, 2003) [in Russian].

Translated by N. Wadhwa

Technical Physics, Vol. 49, No. 10, 2004, pp. 1384–1386. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 137–139.
Original Russian Text Copyright © 2004 by Polyakov, Bikbaev, Zuev.

BRIEF COMMUNICATIONS

Visualization of Inhomogeneities in a Plastic Flow by the Fields of Decorrelation and Scintillation Rates of Video Speckles

S. N. Polyakov, S. A. Bikbaev, and L. B. Zuev

Institute of Physics of Strength and Material Science, Siberian Division,
Russian Academy of Science, Tomsk, 634021 Russia
e-mail: levzuev@mail.tomsknet.ru

Received February 18, 2004

Abstract—We consider the algorithms of *in situ* methods for visualizing the localization zones of plastic deformation, as well as experimental dependences of visualizing signals and parameter (which are inversely proportional to the quality of measurements) on the deformation being visualized. Typical examples of the result of visualization of a plastic flow are given in the case of propagation of Chernov–Lüders bands. *© 2004 MAIK “Nauka/Interperiodica”.*

Inhomogeneity and localization of plastic flows have been investigated in recent years mainly by using two-exposure speckle photography, which does not permit an analysis of the process in real time. General interest in the problem of inhomogeneity of plastic flows has stimulated the development of numerous methods of electronic speckle-interferometry, which does not employ the photographic process (see, for example, [2, 3]). However, all these methods are characterized by a high noise level in output data and are restricted to an analysis of a plastic flow in a narrow range of strain increments. These drawbacks are obviously insurmountable, since they stem from decorrelation of speckles caused by an inseparable combination of displacement of points on the surface (the informative part of the signal) and the change in the surface relief, which is inevitable in plastic deformation (noise).

The search for alternative methods of mapping pictures of nonuniform deformation has led to two versions of direct application of the decorrelation effect for observing deformation localization zones. The first approach was realized in our earlier works [4–6], where the nonuniform strain distribution over the sample was visualized by the field of the mutual decorrelation selection factor D of video signal S, which can be calculated *in situ* within $(m \times n)$-pixel neighborhoods of the points of speckle-video images of the deformed surface being mapped and which is defined as

$$\begin{aligned} D_{ijk}^{(m \times n,\, p)} = 1 - \big|\langle (S_{ijk} - \langle S_{ijk}\rangle)/\sigma S_{ijk} \\ \times (S_{ij(k-p)} - \langle S_{ij(k-p)}\rangle)/\sigma S_{ij(k-p)}\big|\rangle. \end{aligned} \tag{1}$$

Here, i and j are the coordinates in the pixel representation, k is the video frame number, $p \geq 1$ is the frame shift (the number of frames between the speckle images being compared), and σS_{ijk} is the standard deviation of the video signal within the $(m \times n)$-pixel neighborhood of points being mapped. Algorithm (1) ensured the visualization of the Chernov–Lüders bands (Fig. 1) and reconstruction of the positions of their fronts with an error below 1.5–2.0 ps for an interframe increment $\Delta\varepsilon \geq 3 \times 10^{-5}$ in the total tensile strain and with linear regressive smoothing of the coordinates of fronts. However, the range of applicability of the method also proved to be limited by the nonlinearity of the dependence of D on strain components ε_{xx} and ε_{yy} lying in the plane of observation and its saturation for $\varepsilon \geq 0.00075$ (Fig. 2, curve *2*).

These drawbacks in the direct application of the speckle decorrelation effect for visualizing a plastic flow could be overcome by improving the procedure of counting the number M of speckle scintillations, which was described in [7]. In this case, a plastic flow was successfully visualized *in situ* by the field of speckle scintillation rate, which is determined by ensemble-averaging of statistically independent simultaneous measurements in the $(m \times n)$-pixel neighborhood of the point being mapped:

$$\dot{\overline{M}}{}_{ij(k-p/2)}^{(m \times n,\, p)} = -\frac{1}{p}\overline{M}{}_{ij(k-p)}^{(m \times n)} + \frac{1}{p^2}\int\limits_{k-p}^{k} dq\, \overline{M}{}_{ijq}^{(m \times n)}. \tag{2}$$

The scintillation itself in the course of each independent measurement is defined as an event of an interframe (in the interval $[k-1, k]$) transition of video signal S_{ijk} via the corresponding counting threshold $(1 + \beta)\bar{S}_{ijk}^{(m' \times n')}$, defined by a certain relative level β of video signal S_{ijk} averaged over the $(m' \times n')$-pixel neighborhood of the point being mapped. For transitions of S_{ijk} via counting β-threshold (both in the upward and downward direction), current values of the number of scintillations are calculated according to the rule $M_{ijk}(\beta) =$

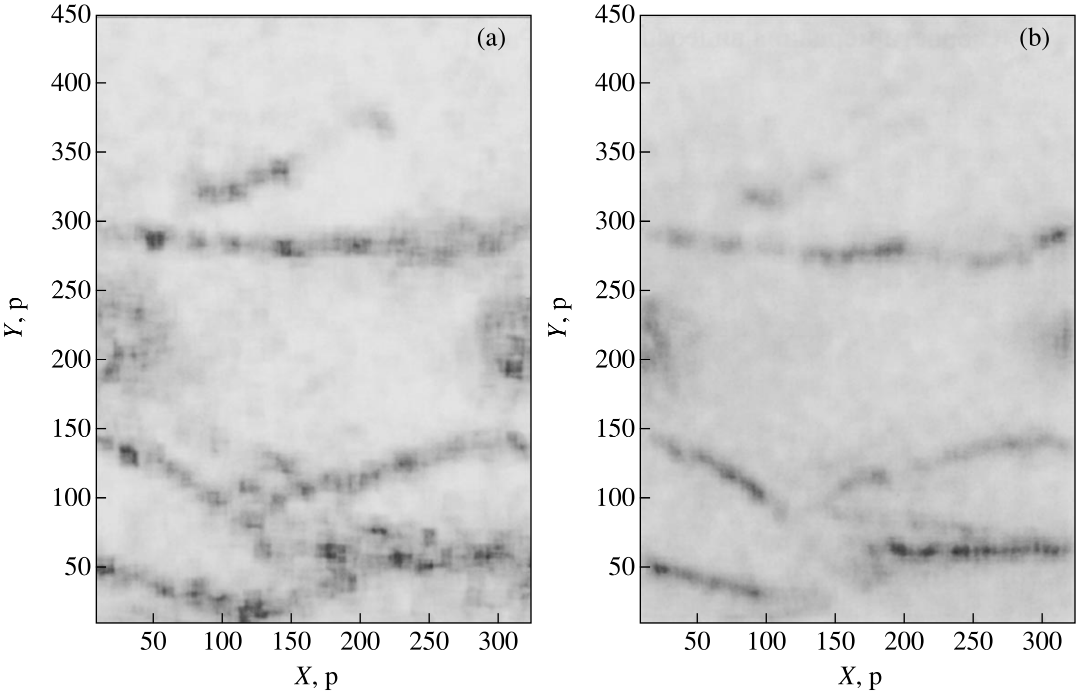


Fig. 1. Visualization of a plastic flow in the case of uniform extension of a plane sample made of low-carbon steel with symmetric transverse cuts, carried out with spatially confined averaging (over the $11 \times 11p$ neighborhood, 1 p = 8 μm) (a) by the computational decorrelation method and (b) by the scintillation rate field of video speckles (11 statistically independent simultaneous measurements; regressive smoothing parameter $p = 5$). In both cases, the fronts of the Lüders bands and other regions of localized deformation are reflected by darkening of grey images.

$M_{ij(k-1)}(\beta) + 1$ and the neighborhood is chosen from the requirement of its statistical representability for determining the mean level S of the speckle field, which approaches the general sampling average. It is found that, in choosing the neighborhood, we can confine our analysis to only two or three mean speckle sizes since a decrease in the number of speckles inordinately increases the dispersion of quantity $M(\beta)$, while an increase in the number of speckles reduces the quality of visualization because of nonuniform illumination of the object. Averaging of $M(\beta)$ over ensemble β is carried out by taking into account the weight coefficients defined on the elastic segment of tension in the course of obtaining gauges $M(\beta) = M(\beta, \varepsilon_{xx})$. A linear regressive smoothing of current values of $\overline{M}_{ijk}^{(m\times n)}$ (2) realized as a result of integration based on the Gauss method in the interval $[k - p, k]$ facilitates suppression of noise associated with the random distribution of the speckle field if the assumption that the strain increment, linear in time holds. An analogous result is attained by averaging the interframe increment of the number of scintillations in an $(m \times n)$-pixel neighborhood, provided that the ε_{xx} and ε_{yy} components are parallel to the observed plane surface of the object being studied.

We assume that the process of speckle decorrelation determined by the time evolution of plastic deformation is a random steady-state Gaussian process, is characterized by the same distribution of the time derivative, and describes the scintillation dynamics by the law [8] $M \sim t/\tau$ (τ is the correlation interval of the process). In this case, taking into account the dependence $\overline{M}_{ijk}^{m\times n}(\varepsilon_{xx})$ derived by us (see Fig. 2, curve *1*), interframe increments $\overline{M}_{ijk}^{(m\times n)} - \overline{M}_{ij(k-p)}^{(m\times n)}$ can be juxtaposed to the strain increment; this allows us to interpret the signal defined by algorithm (2) as a linear function of strain components $\dot{\varepsilon}_{xx}$ and $\dot{\varepsilon}_{yy}$.

Thus, algorithm (2) ensures visualization of deformation with only one temporal parameter, viz., time mark $k - p/2$. For a linearly increasing strain, the width p of the regressive smoothing interval fails to explicitly determine the value of the visualizing signal (the speckle scintillation rate) in contrast to algorithm (1) whose signal D is directly connected with the interframe shift p. Consequently, algorithm (2) is one-parametric by nature, which considerably simplifies the collection of data on the space–time structure of the deformation field, simplifies their interpretation, and permits the estimation of strain accumulation by integrating the signal with respect to time if the increment of scintillations is a linear function of ε_{xx} and ε_{yy}.

Owing to the linearity of the dependence $\overline{M}_{ijk}^{(m\times n)}(\varepsilon_{xx})$, algorithm (2) is obviously advantageous as compared to speckle-decorrelation algorithm (1) as regards the precision and informativity of the strain

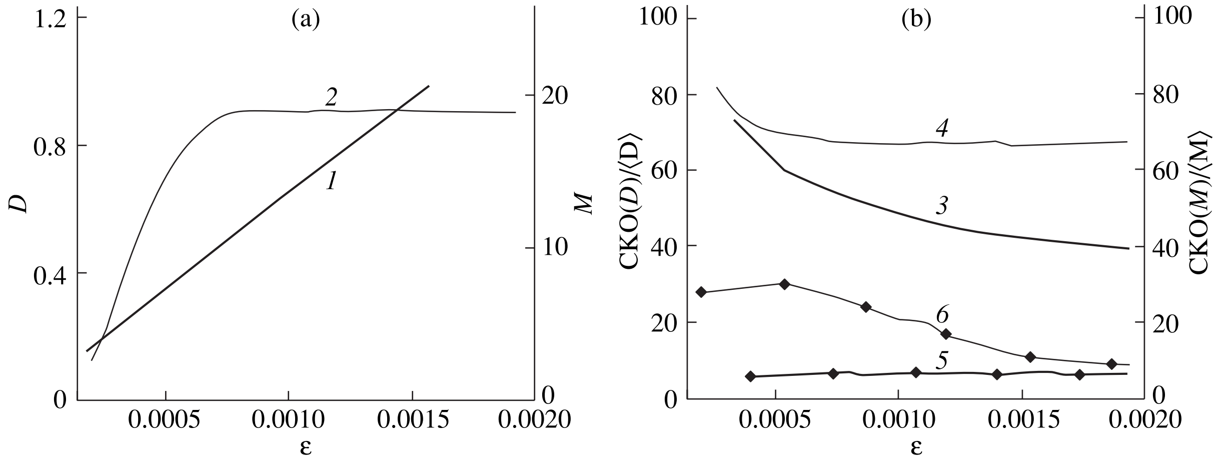


Fig. 2. Experimentally established dependences on the increment of total strain ε of (a) visualization signal (*1*—*M*, *2*—*D*) for algorithms (1) and (2) and (b) parameters of reciprocal quality of measurements; (*3*—CKO(*M*)/⟨*M*⟩, 1 × 1*p*; *4*—CKO(*D*)/⟨*D*⟩, 2 × 2*p*; *5*—CKO(*M*)/⟨*M*⟩, 21 × 21*p* (*p* = 5); *6*—CKO(*D*/⟨*D*⟩), 20 × 20*p* (*p* = 1).

field map. The obvious advantage of algorithm (2) in respect to the noise level (which follows from the lower value of the standard deviation and its independence of strain; see curves *3* and *5* in Fig. 2) is equally important. These circumstances illustrate the obvious advantages of visualization potentialities of this algorithm in the case of a plastic flow (see Fig. 1).

REFERENCES

1. L. B. Zuev, Ann. Phys. **10**, 965 (2001).
2. P. K. Rastogi, Opt. Lasers Eng. **26** (1), 93 (1997).
3. M. Sjudahl, *Digital Speckle Pattern Interferometry and Related Techniques* (Wiley, New York, 2001), pp. 289–336.
4. V. V. Gorbatenko, S. N. Polyakov, and L. B. Zuev, Zavod. Lab. **67** (7), 29 (2001).
5. L. B. Zuev, S. N. Polyakov, and V. V. Gorbatenko, Proc. SPIE **4900**, Part 2, 1197 (2002).
6. S. N. Polyakov, V. V. Gorbatenko, E. L. Lopaev, and L. B. Zuev, Avtometriya **39** (5), 102 (2003).
7. A. P. Vladimirov, Doctoral Dissertation (Yekaterinburg, 2002).
8. V. I. Tikhonov, *Spikes of Random Processes* (Nauka, Moscow, 1970) [in Russian].

Translated by N. Wadhwa

Technical Physics, Vol. 49, No. 10, 2004, pp. 1387–1389. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 74, No. 10, 2004, pp. 140–142.
Original Russian Text Copyright © 2004 by Lomtev.

BRIEF COMMUNICATIONS

Surface Electromagnetic Waves at a Superfluid Liquid—Normal Metal Interface

A. I. Lomtev
Donetsk Physicotechnical Institute, National Academy of Sciences of Ukraine, Donetsk, 83114 Ukraine
e-mail: lomtev@kinetic.ac.donetsk.ua

Received February 18, 2004

Abstract—Surface electromagnetic waves, which can propagate along the plane interface between a superfluid liquid and a normal metal, are investigated. The dispersion relations for surface waves with different polarizations of the optical anisotropy of the superfluid liquid are obtained and their possible frequency ranges are determined. It is shown that anisotropic optical contribution to the dielectric constant of the superfluid liquid can be determined from experiments on excitation of surface electromagnetic waves. *© 2004 MAIK "Nauka/Interperiodica".*

Investigations of surface electromagnetic waves propagating along plane interfaces of media with different dielectric properties have aroused great interest [1–3]. The term "surface wave" was introduced for the first time in connection with the theory of radio-wave propagation near the earth's surface. In 1907, it was theoretically shown [4] that a slow surface wave in the range of radio frequencies could propagate along a conducting earth (or sea) surface. This wave does not fundamentally differ from a surface wave propagating along a wire.

It is also known that a surface *H* wave (*P* polarization) can propagate along the interface of two media with positive and negative dielectric constants (ε_1 and $-|\varepsilon_2|$), respectively. Then, the wave decays deep into both media [5].

Two kinds of macroscopic movement with velocities of the normal v_n and superfluid components v_s are possible in the superfluid liquid. Hence, the permittivity tensor can depend on the relative velocity $w = v_n - v_s$. Thus, optical anisotropy can appear in a homogenous thermodynamically equilibrium liquid. The optical anisotropy is expressed by [6]

$$\delta\varepsilon_{ij} = \lambda w_i w_j, \tag{1}$$

where $\delta\varepsilon_{ij}$ is the nonscalar part of the permittivity and λ is the constant of optical anisotropy (to be more precise, a function of temperature and pressure).

In the phonon range of fairly low temperatures, the constant of optical anisotropy is determined by the expression [6]

$$\lambda_P = -(\varepsilon_1 - 1)^2 \frac{\rho_{Pn}}{\rho} \frac{1}{s^2}, \tag{2}$$

where ε_1 is the isotropic part of the permittivity, ρ is the density of the liquid, ρ_{Pn} is the phonon part of the normal component density, and *s* is the velocity of sound.

In the high-temperature roton range, the constant of optical anisotropy is expressed by [6]

$$\lambda_R = -(\varepsilon_1 - 1)^2 \frac{\rho_{Rn}}{\rho} \frac{p^0}{5T\Delta}, \tag{3}$$

where ρ_{Rn} is the roton normal density, p_0 and Δ are the momentum and energy of rotons, and *T* is the temperature.

The total permittivity tensor of the superfluid liquid, $\tilde{\varepsilon}_{1ij} = \varepsilon_1\delta_{ij} + \lambda w_i w_j$, can be reduced to the principal axes when it is expressed in the form

$$\tilde{\varepsilon}_{1ij} = (\varepsilon_1 + \lambda w_i^2)\delta_{ij}. \tag{4}$$

The principal values of this tensor are positive, since $\varepsilon_1 \geq 1$, while the anisotropy term $|\lambda| w_i^2 \ll 1$.

The permittivity of a normal metal can be expressed by [7, 8]

$$\tilde{\varepsilon}_{2ij} = \varepsilon_2\delta_{ij} = [1 - \omega_p^2/\omega(\omega + i/\tau)]\delta_{ij}, \tag{5}$$

where $\omega_p = (4\pi ne^2/m^*)^{1/2}$ is the electron plasma frequency, *n* is the electron density, *e* is the electron charge, m^* is the effective mass of charge carriers, and τ is the electron relaxation time (the time between consecutive collisions of electrons with defects or impurities).

Equation (5) for the permittivity describes the contribution of intraband transitions to the permittivity of almost free electron gas of a metal or *n*-type semiconductor. In very pure metals (especially at helium temperatures), $\tau \approx 10^{-9}$ s, while for typical metals $\omega_p \approx 10^{16}$ s^{-1} [8]. Hence, in the range of anomalous disper-

1063-7842/04/4910-1387$26.00 © 2004 MAIK "Nauka/Interperiodica"

sion of a metal for frequencies $\tau^{-1} \ll \omega - \omega_p$, where $\omega \approx 10^{15}$–10^{16} s^{-1}, the imaginary part of the permittivity in Eq. (5) can deliberately be neglected and damping can be regarded as weak. In this case, permittivity (5) is a real and negative function of the frequency,

$$\tilde{\varepsilon}_{2ij} = \varepsilon_2\delta_{ij} = -(\omega_p^2/\omega^2 - 1)\delta_{ij}, \tag{6}$$

and the metal is a surface-active medium [9].

A metal of the alkaline group, for example Na, can be used as the metal. Its spherical Fermi surface ensures isotropy of metal permittivity.

Let us consider the possibility of propagation of surface electromagnetic H waves along the superfluid liquid–normal metal interface. In the coordinate system, in which the tensors of the contacting media (4) and (6) are diagonal, we take an interface as the xy plane and the wave propagates along the x axis, while field H is parallel to the y axis. For waves of P polarization proportional to $\exp(-i\omega t)$ with nonzero components $\{E_x, H_y, E_z\}$, the Maxwell equations are written in the form

$$\begin{gathered}\partial H_y/\partial z = i(\omega/c)D_x, \quad \partial H_y/\partial x = -i(\omega/c)D_z,\\ \partial E_x/\partial z - \partial E_z/dx = i(\omega/c)H_y,\end{gathered} \tag{7}$$

where $D = \tilde{\varepsilon}E$.

Let the half-space $z > 0$ be filled with a superfluid liquid with a positive permittivity ($\tilde{\varepsilon}_1$), while the half-space $z < 0$ is filled with a metal with a negative permittivity ($\tilde{\varepsilon}_2$). We look for a field in the wave decaying at $z \longrightarrow \pm\infty$ in the form

$$H_1 = H_0\exp(ikx - \kappa_1 z) \quad \text{at} \quad z > 0, \tag{8}$$

$$H_2 = H_0\exp(ikx + \kappa_2 z) \quad \text{at} \quad z < 0, \tag{9}$$

where k, κ_1, and κ_2 are real-valued quantities. The boundary condition $H_1(z = 0) = H_2(z = 0)$ has already been met, while the condition for E_x continuity leads to the dispersion equation $k = k(\omega)$,

$$\frac{1}{\varepsilon_{1xx}}\frac{\partial H_1}{\partial z} = \frac{1}{\varepsilon_2}\frac{\partial H_2}{\partial z} \quad \text{at} \quad z = 0 \tag{10}$$

or

$$\frac{\kappa_1}{\varepsilon_{1xx}} = \frac{\kappa_2}{|\varepsilon_2|}. \tag{11}$$

As will be shown below, $\kappa_2 > \kappa_1$; surface waves exist if the inequality $|\varepsilon_2| > \varepsilon_{1xx}$, which imposes an upper bound on the range of possible frequencies $\omega < \omega_p/(1 + \varepsilon_{1xx})$, is satisfied. Hence, the surface wave frequency should satisfy the inequalities $\tau^{-1} \ll \omega - \omega_p/(1 + \varepsilon_{1xx})$.

According to the Maxwell equations (7) for field (9), the expression for κ_2 has the form

$$\kappa_2^2 = k^2 + \frac{\omega_p^2 - \omega^2}{c^2}. \tag{12}$$

Let us now assume that relative velocity $w = (w_x, 0, 0)$. For this direction of the relative velocity, the expression for κ_1 according to the Maxwell equations (7) for field (8) takes the form

$$\kappa_1^2 = \frac{\varepsilon_1 + \lambda w_x^2}{\varepsilon_1}\left(k^2 - \frac{\omega^2}{c^2}\varepsilon_1\right). \tag{13}$$

Taking into account Eqs. (12) and (13), dispersion equation Eq. (11) can be expressed by

$$k^2(\omega) = \frac{\omega^2(\omega_p^2 - \omega^2)\varepsilon_1[\omega^2(\varepsilon_1 + \lambda w_x^2) + \omega_p^2 - \omega^2]}{c^2[(\omega_p^2 - \omega^2)^2 - \omega^4\varepsilon_1(\varepsilon_1 + \lambda w_x^2)]}. \tag{14}$$

If the relative velocity is directed along the z axis, $w = (0, 0, w_z)$ and the relation for κ_1 can be expressed by

$$\kappa_1^2 = \frac{k^2\varepsilon_1}{\varepsilon_1 + \lambda w_z^2} - \frac{\omega^2}{c^2}\varepsilon_1, \tag{15}$$

while taking into account Eqs. (12) and (15), dispersion equation (11) can be expressed by

$$k^2(\omega) = \frac{\omega^2(\omega_p^2 - \omega^2)(\varepsilon_1 + \lambda w_z^2)[\omega^2(\varepsilon_1 - 1) + \omega_p^2]}{c^2[(\omega_p^2 - \omega^2)^2 - \omega^4\varepsilon_1(\varepsilon_1 + \lambda w_z^2)]}. \tag{16}$$

In the general case of polarization of the relative velocity, when $w = (w_x, 0, w_z)$, taking into account Eq. (12), the expression for κ_1,

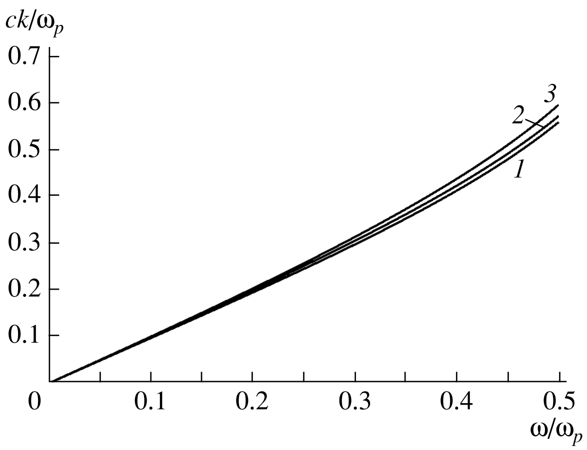
$$\kappa_1^2 = (\varepsilon_1 + \lambda w_x^2)\left(\frac{k^2}{\varepsilon_1 + \lambda w_z^2} - \frac{\omega^2}{c^2}\right), \tag{17}$$

leads to dispersion relation (11) in the form

$$k^2(\omega) = \frac{\omega^2(\omega_p^2 - \omega^2)(\varepsilon_1 + \lambda w_z^2)[\omega_p^2 + \omega^2(\varepsilon_1 + \lambda w_x^2 - 1)]}{c^2[(\omega_p^2 - \omega^2)^2 - \omega^4(\varepsilon_1 + \lambda w_x^2)(\varepsilon_1 + \lambda w_z^2)]}. \tag{18}$$

It is seen from comparison of dispersion equations (14), (16), and (18) that optical anisotropy of a superfluid liquid results in the dependence of the form of the surface wave spectrum on the polarization type of optical anisotropy.

Dispersion curves of the surface electromagnetic waves as numerical solutions to Eqs. (14), (16), and (18) are shown in Fig. 1. It is seen that the maximal discrepancy in the spectral branches caused by the polarization type of the liquid helium optical anisotropy can be registered at frequencies close to frequency $\omega \approx \omega_p/(1 + \varepsilon_{1xx})$. It follows from Eq. (12) for κ_2 and disper-



Dependence of the reduced wave vector ck/ω_p on the reduced frequency ω/ω_p for surface electromagnetic waves. (*1*) Solution to dispersion equation (18), (*2*) solution to dispersion equation (16), (*3*) solution to dispersion equation (14) for parameters $\varepsilon_{1xx} \approx 1$ and $|\lambda| w_x^2 \approx |\lambda| w_z^2 - 10^{-3}$ in the allowed frequency range $10^{-5} \le \omega/\omega_p < 0.5$.

sion relations (14), (16), and (18) that the penetration depth of the surface wave electromagnetic field in the metal is $\delta = \kappa_2^{-1} \approx c/\omega_p = c/\sqrt{3}\, v_F k_s \gg r_D$, where $v_F \approx 10^5$–10^6 cm/s is the Fermi velocity of electrons and $k_s^{-1} = r_D$ is the Debye screening radius. However, the electron mean free path in the metal is $l = \tau v_F = (\tau\omega_p/\sqrt{3})r_D \gg r_D$. The ratio $\delta/l = (c/v_F)/\omega_p\tau \approx 10^{-2}$–$10^{-3}$. Due to the inequality $\delta \ll l$, one can assume that a surface wave in the metal is subjected to the regime of anomalous skin effect.

The studied waves with dispersion relations (14), (16), and (18) are surface waves like the waves [4] propagating along the plane interface between an optically anisotropic superfluid liquid and an isotropic normal metal in the visible frequency range.

If the volume of the superfluid liquid is experimentally available, the contribution of optical anisotropy of liquid helium can easily be determined by measuring the optical birefringence for linearly polarized radiation [6] or by measuring the rotation angle of the polarization ellipse for elliptically polarized light. If only the surface of the liquid is experimentally available due to some reasons, anisotropic contributions λw_x^2 and λw_z^2 to the permittivity of the superfluid liquid can easily be determined from dispersion equations (14) and (16) using optical experiments on surface wave excitation.

As can easily be shown, the propagation of surface E waves (S polarization) with components $\{H_x, E_y, H_z\}$ along the interface between the superfluid liquid and the normal metal is impossible in principle.

ACKNOWLEDGMENTS

In conclusion, the author acknowledges S.V. Tarasenko for useful discussion and thanks Yu.V. Medvedev and I.B. Krasnyuk for their attention and support.

REFERENCES

1. V. L. Falko, S. I. Khankina, and V. V. Yakovenko, Phys. Lett. A **209**, 118 (1995).
2. M. F. Pascual, W. Zierau, T. F. Leskova, and A. A. Maradudin, Opt. Commun. **155**, 351 (1998).
3. N. L. Dmitruk, S. V. Mamykin, and O. V. Rengevych, Appl. Surf. Sci. **166**, 97 (2000).
4. L. A. Vaĭnshteĭn, *Electromagnetic Waves* (Sov. Radio, Moscow, 1957), Sect. 63 [in Russian].
5. L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Nauka, Moscow, 1970; Pergamon Press, Oxford, 1982), Sect. 88.
6. A. F. Andreev, Pis'ma Zh. Éksp. Teor. Fiz. **31**, 191 (1980) [JETP Lett. **31**, 175 (1980)].
7. D. Pines, *Elementary Excitations in Solids* (Benjamin, New York, 1963; Mir, Moscow, 1965).
8. P. M. Platzman and P. A. Wolff, in *Solid State Physics*, Suppl. 13 (Academic, New York, 1973; Mir, Moscow, 1975).
9. *Surface Polaritons*, Ed. by V. M. Agranovich and D. L. Mills (North-Holland, Amsterdam, 1982; Nauka, Moscow, 1985).

Translated by M. Astrov