Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1008–1012. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 974–978.
Original Russian Text Copyright © 2005 by Gorobeĭ, Korsukov, Luk'yanenko.

PROCEEDINGS OF THE CONFERENCE DEDICATED TO S. N. ZHURKOV (1905–1997)

Energy Criterion for Relief Formation on a Strained Surface

N. N. Gorobeĭ*, V. E. Korsukov, and A. S. Luk'yanenko****
** St. Petersburg State Polytechnical University, Politekhnicheskaya ul. 29, St. Petersburg, 195251 Russia*
*** Ioffe Physicotechnical Institute, Russian Academy of Sciences, Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia*
e-mail: Vjacheslav.Korsukov@pop.ioffe.rssi.ru

Abstract—The strain-induced roughening revealed by scanning tunneling microscopy on the surface of elastically strained Ge(111) crystals and $Fe_{70}Cr_{15}B_{15}$ amorphous alloy foils is discussed. It is demonstrated that the strained-induced roughening can be considered a channel for a decrease in the elastic energy of the strained surface. The diffusion mechanism of relief formation on a strained Ge(111) surface is proposed and justified.

1. INTRODUCTION

The Zhurkov kinetic concept of fracture is based on the postulate that, on the atomic level, fluctuations of the thermal energy play a decisive role in the fracture of a mechanically loaded solid [1]. In this respect, Zhurkov believed that an important step toward justifying the thermal fluctuation mechanism was to reveal the thermal fluctuations and their manifestations using the newest methods for analyzing the structure and atomic dynamics of the crystal lattice. Such potentialities are provided in particular by a rich arsenal of analytical methods devised for solving problems in surface physics. Owing to its physical nature, the surface itself is an interesting object of investigation from the standpoint of defect formation. The mere fact that the surface is a free boundary suggests its lower thermodynamic and mechanical stabilities. All the above was the reason why almost immediately after the advent of a new technique, namely, scanning tunneling microscopy (STM), Zhurkov initiated investigations into the dynamics of the relief of a mechanically strained surface on a nanometer scale with the use of the STM method and took an active part in these works.

Zhurkov *et al.* [2] were the first to undertake an STM investigation of the dynamics of a Ge(111) surface relief under biaxial lateral tension. It was found that the strain-induced relief is formed at relatively small loads of the order of 0.1 GPa. The strain-induced relief is comprised of roughnesses (mounds) whose lateral and vertical (perpendicular to the surface plane) sizes are of the order of several hundreds and several tens of nanometers, respectively. The evolution of mounds with time was investigated in close detail. It was revealed that, under a permanent applied load, the mounds grow to the aforementioned sizes in approximately one hour and then remain unchanged. The roughening is a reversible process. After the load is removed, the roughness relaxes almost to the initial level in approximately ten hours. More recently, similar processes of formation of a strain-induced relief under uniaxial tension were observed on the surface of foils prepared from the iron-based amorphous alloy $Fe_{70}Cr_{15}B_{15}$ [3, 4]. In this case, depending on the applied load, the formation of a strain-induced relief can proceed according to several scenarios to the extent that this process ceases to be stabilized even when the vertical sizes of roughnesses reach a micrometer scale (at loads of approximately 1 GPa).

In the present work, we studied germanium and amorphous alloys. These materials exhibit an elastic behavior over a wide range of loads. Hence, we can assume that, at loads of approximately 0.1 GPa, the contribution made by the plastic strain to the bulk properties of the specimens will be insignificant. Therefore, the observed evolution of the relief is a surface process and its explanation should be sought in the manifestation of the structural features and physical properties of the surface due to the presence of the free boundary.

Before proceeding to the formation of a strain-induced roughness, we make a number of important comments. Actually, the surface relief of a strained solid is subject to nanometer-scale fluctuations and these fluctuations of the surface relief, in any interpretation, are manifestations of the aforementioned fluctuations of the thermal energy. However, the fluctuations of the thermal energy are not destructive but, contrastingly, in some cases (at relatively small loads), lead to the formation of stable spatial aggregates in the structure of the surface relief. This is not surprising because the processes of fracture, recovery, and formation of a structure always proceed "hand in hand" and the final result depends on which of these processes dominates under given conditions.

The revealed phenomenon, i.e., the relief formation on an elastically strained Ge(111) surface, is in many respects similar to the phenomenon of island epitaxy, which has been extensively investigated both experimentally and theoretically [5, 6]. Great interest expressed by researchers in island epitaxy stems from the unique features of this phenomenon. On the one hand, island epitaxy hinders the formation of homogeneous regions in the heterostructure. On the other hand, under specific conditions, island epitaxy can be used for the controlled formation of a surface relief, for example, in the form of quantum dots. Thus, the problem formulated in terms of the mechanics of strained solids somewhat unexpectedly turned out to be closely related to the problems of microelectronics.

The similarity between the strain-induced roughening and island epitaxy suggests that these phenomena have a common physical nature. In this work, we extend the previously proposed interpretation of strain-induced roughening, which is based on the principle of minimum energy of a strained surface. In general terms, our approach is consistent with the universally accepted theoretical explanation of island epitaxy. The difference lies in the nature of the mechanical surface stresses, which are responsible for the mound nucleation. In the case of heterostructures, these are the stresses arising in an epitaxial layer due to the mismatch between the lattice parameters of the layer and the substrate. In our experiments, the strain in the surface layers arises from *in situ* mechanical loading of the specimen.

Our interpretation of the mound nucleation on a strained surface can be divided into two parts, which will be referred to as energetics and kinetics. In the former (energetic) part, we justify the direction of the process toward an increase in the roughness size. In this case, the process is treated as a channel for a decrease in the free energy of the strained surface. In the latter (kinetic) part, we consider specific physical mechanisms of mass transfer that, in principle, can explain the observed spatiotemporal characteristics of the strain-induced roughening.

2. ENERGETICS OF ROUGHNESSES

Let us evaluate the energy balance for the strain-induced roughening approximately. Within the mechanical approach, by analogy with the Griffith analysis of an elastic crack [7], we estimate only the mechanical energy and ignore the entropy contribution. As the first step in analyzing the energetics of the strain-induced roughening, we discuss a simple example in which the energy balance initiates a specific mechanism responsible for the structural transformation of a solid, namely, plastic deformation of a mechanically loaded specimen. We consider the experimental situation in which a cylindrical specimen is subjected to uniaxial tension after the weight P is suspended from the specimen. The plastic deformation results in an elongation of the specimen at a constant volume $V = Sl = \text{const}$, so the cross section of the specimen (on the average) decreases. The change in the energy of the system (specimen + weight) is determined by the decrease in the potential energy of the weight and the change in the elastic energy of the specimen due to the variation in its cross section; that is,

$$\Delta W = -P\Delta l + \Delta W_{\text{el}}. \qquad (1)$$

Here, we take into account the constant volume and the relationship $\sigma = P/S$ and obtain

$$\Delta W_{\text{el}} = \Delta\left(\frac{\sigma^2}{2E}\right)V \cong \frac{P^2}{ES}\Delta l, \qquad (2)$$

where E is the Young's modulus. The final energy balance is invariably negative,

$$\Delta W \cong P\Delta l\left(-1 + \frac{\sigma}{E}\right) < 0, \qquad (3)$$

because the inequality $\sigma < E$ is satisfied. Thus, we can make the obvious inference that, under mechanical loading, the elongation of the specimen leads to a decrease in the energy of the entire system. The next question arises as to which mechanism provides the high efficiency of the process. It is known that the main physical mechanism responsible for the above scenario for a decrease in the energy is plastic deformation of the specimen. However, for the materials under investigation (germanium, iron-based amorphous alloy), the channel for a decrease in the energy due to plastic deformation appears to be closed over a wide range of loads. In our experiments, such is indeed the case: the material in the bulk of the specimen is in an elastically strained state, and the crystal structure can undergo transformations only in surface layers.

When structural transformations in the bulk of the specimen are “frozen” and the specimen is in an elastically strained state, the bulk of the specimen can be treated as a “substrate” for surface atomic layers and the strain of the substrate can be considered to be specified and fixed. For now, the boundary between the substrate and the surface layer is drawn conventionally because the boundary location depends on the physical mechanism of transformation of the surface structure. Our concern here is only with the change in the surface relief due to the transformation of the surface structure. Let us compare the mechanical energy of the initially planar, strained surface layer with the energy of a surface layer that has the same volume but is “reconstructed” into an array of mounds. For simplicity, we examine a more or less regular structure with pyramidal mounds.

Next, we introduce a number of simplifying assumptions which can be formulated in an explicit form. First and foremost, we assume that, after the pyramidal relief is formed, the material contained in the pyramids appears to be in an unloaded unstrained state. To put it differently, the relief reconstruction is accompanied by the dissipation of elastic energy over the entire surface layer under consideration. Of course, the pyramid bases in contact with the laterally strained substrate are also strained and the strain is transferred to overlying atomic layers. However, as will be shown below, this interlayer transfer of the lateral strain decays rather rapidly and, hence, can be disregarded. This is the essence of our first approximation. As a result, the elastic-energy gain per pyramid can be represented in the form

$$\Delta W_{\text{el}} \cong -\frac{\sigma^2}{2E}V, \tag{4}$$

where V is the pyramid volume.

Now, we analyze the process under consideration from the standpoint of the energy consumption. The energy consumption is caused by the increase in the surface area after roughening. The physical nature of the increase in the surface area during roughening is somewhat different from that in the course of crack formation in terms of the Griffith theory. The mechanical rupture of chemical bonds in the bulk of the crystal is obviously a more energy-consuming process as compared to the detachment and migration of atoms over the surface. However, as before, we use a purely mechanical approach within which the surface energy can be determined from the mechanical estimate [8]

$$\gamma \sim \frac{Ea}{5}, \tag{5}$$

where a is the lattice parameter. In this case, we also disregard the possible dependence of the surface energy on the crystallographic orientation of the surface region under investigation. This is our second simplifying assumption. As a consequence, the net energy in the formation of the pyramid can be estimated as

$$\Delta W_{\text{surf}} = \gamma\left(4 \times \frac{1}{2}A\sqrt{h^2 + \frac{A^2}{4}} - A^2\right), \tag{6}$$

where A is the linear size of the pyramid base and h is the pyramid height. The final energy balance for the strain-induced roughening (per pyramid) can be written in the approximation of a small pyramid height ($h \ll A$) in the following form:

$$\Delta W \cong -\frac{\sigma^2}{6E}A^2h + 2\gamma h^2. \tag{7}$$

It can be seen from relationship (7) that the energy gain ($\Delta W < 0$) is reached in the case when the vertical size of the roughnesses (at a fixed lateral size of the roughnesses and at a specified load) does not exceed the critical value

$$h_m \cong 10\left(\frac{\sigma}{E}\right)^2\frac{A^2}{a}, \tag{8}$$

where estimate (6) for the surface energy is taken into account.

In the above analysis, the lateral size of roughnesses A was assumed to be fixed but remained uncertain. The theoretical grounds for estimating the quantity A will be considered in the next section. Here, the experimentally determined lateral size of roughnesses ($A_{\text{av}} \sim 300$ nm) on the Ge(111) surface is taken for estimation. According to expression (8), the maximum size of roughnesses, which is allowed by the energy balance on the Ge(111) surface at $\sigma = 0.1$ GPa ($a \sim 0.3$ nm), is estimated as $h_m \sim 3$ nm. Recall that, according to experimental data, the mean vertical size of strain-induced roughnesses on the Ge(111) surface is equal to several tens of nanometers. This size is one order of magnitude larger than the above theoretical estimate. Such a discrepancy should not be considered serious. This discrepancy only confirms our inference that the surface energy calculated in the framework of the mechanical approach [see formula (5)] is overestimated. The main result holds true: there is a range of geometric parameters of the strain-induced roughness in which the surface energy decreases. It should also be noted that, in the case when the actual size of roughnesses h exceeds the size permissible by the principle of energy gain, there should occur an inverse process, i.e., deroughening. It is this process that is experimentally observed on the Ge(111) surface at relatively small loads.

In closing the examination of the energy balance, we go beyond the approximation of small vertical sizes of roughnesses and write the exact expression for the critical vertical size of roughnesses:

$$h_m \cong \frac{A^2}{a}\frac{10\left(\frac{\sigma}{E}\right)^2}{1 - \left(\frac{A}{4a}\right)^2\left(\frac{\sigma}{E}\right)^4}. \tag{9}$$

For small loads and, correspondingly, small vertical sizes of roughnesses, this expression transforms into relationship (8). However, it is worth noting that, according to formula (9), the dependence of the roughness size on the load exhibits a critical behavior. Actually, when the load reaches the critical value,

$$\sigma_* \sim E\sqrt{\frac{a}{4A}}, \tag{10}$$

the vertical sizes of roughnesses are not limited from above. For the same parameters a, A, and E ($E_{\text{Ge}} \cong 1.3 \times 10^2$ GPa), the critical load is estimated as $\sigma_* \cong 1.5$ GPa. Note that, under loads of the order of 1 GPa, roughnesses on the surface of the $Fe_{70}Cr_{15}B_{15}$ amorphous alloy foils grow at a high rate and reach micrometer-scale sizes without a noticeable tendency toward stabilization. This process of strain-induced roughening results in the fracture of the specimen, which is initiated on the surface.

3. DIFFUSION KINETICS

The energy criterion for strain-induced roughening (considered in the previous section) indicates only a possible decrease in the surface energy due to the transformation of the surface relief. Solving the problem associated with the implementation of this possibility is outside the scope of the simple energy balance treatment and, as was noted above, calls for the kinetic approach. Now, we should concretize our consideration. In this respect, we will primarily discuss the data obtained for the Ge(111) face. It is assumed that the main mechanism responsible for the strain-induced roughening is the surface diffusion assisted by external loading. Therefore, it is necessary to answer two questions. First, can the diffusion mechanism of mass transfer on the surface be efficient enough to ensure the experimentally observed spatiotemporal characteristics of strain-induced roughening? Second, what is the role of mechanical loading in the initiation of the diffusion mechanism? It is known that, in the absence of loading, no transformations are observed on the surface at room temperature, so that the surface diffusion is “frozen.”

For loads $\sigma \sim 0.1$ GPa used in our experiments with Ge(111) crystals, the lattice strain (in the crystal bulk) is estimated as $\varepsilon \sim 10^{-3}$. This strain is substantially weaker than that required for breaking chemical bonds. Moreover, for this value of the volume strain, the role played by the loading in the initiation of the diffusion process remains unclear. Hence, we make the inference that the surface structure should involve sources of local strains. Strain concentrators can be edges of the features existing in the surface relief. In this case, the increase in the strain is caused by the mismatch between the lengths of chemical bonds in two layers located one above the other when the lower layer (conventionally, the substrate) is characterized by a uniform strain ε_b, whereas the overlying layer (conventionally, adlayer) is an island with laterally limited sizes. According to the calculations, the local strain ε_i in the vicinity of the adlayer edge can be two orders of magnitude larger than the strain of the substrate, provided the lateral size of islands is estimated as $A \sim 10^2$ nm. This estimate justifies the lateral size used for the pyramid base in the previous section. Note that the strain of overlying adlayers decays exponentially away from the substrate. Therefore, if the initial surface is not ideally planar (this is always the case), the strain is enhanced at the edges of the existing islands. Moreover, surface semiconductor layers (with a thickness of several micrometers or several tens of micrometers) at approximately identical loads and room temperature undergo considerable plastic (microplastic) deformation [9], which also leads to the formation of strain concentrators. It is expected that this phenomenon can also manifest itself in our experiments.

It remains now to answer the last question as to how strain concentrators initiate the diffusion mechanism of mass transfer. The surface diffusion of an atom is limited by its detachment with the formation of a vacancy–adatom pair (an adatom is an atom that is rather weakly bonded to the surface). The activation rate of vacancy–adatom pairs is estimated from the real parameters of surface diffusion on the Ge(111) face. According to [10], the activation energy of surface diffusion on the Ge(111) face at room temperature is estimated as $Q_0 \cong$ 1 eV. The subsequent diffusive motion of the adatom over the surface is characterized by a relatively low activation energy; as a result, an activation barrier of ~1 eV should be attributed to the stage of adatom formation. More specifically, this activation energy should be assigned to the atoms located at the edges of atomic terraces, because the diffusion parameters for a rough surface are measured from the deroughening rate.

Let us first estimate the expectation time for detachment of such an edge atom from the initial surface in the absence of external loading (hereafter, we will consider room temperature). This time can be estimated from the formula

$$\tau = \tau_0 e^{-\frac{Q_0}{kT}}, \tag{11}$$

where τ_0 is the period of atomic vibrations. The physical meaning of formula (11) is obvious: the preexponential factor determines the frequency of attempts to overcome the barrier per unit time, and the exponential function describes the Boltzmann probability of a favorable outcome of the attempt. The period of atomic vibrations is taken to be $\tau_0 \sim 10^{-12}$ s. Then, the expectation time for activation of the vacancy–adatom pair on the initial surface in the absence of external loading turns out to be of the order of 10^5 s, (i.e., one day). For this expectation time, the surface relief remains stable. However, as the activation energy Q_0 decreases by only 10%, the expectation time decreases to 10^3 s (one hour). When the activation energy decreases by 25%, the expectation time decreases to 10 s. According to the above estimate of the local strain ε_i, similar effects can be observed in the vicinity of the bases of the mounds in the existing roughness. Simple estimates also demonstrate that this is sufficient to provide the mass transfer revealed on the Ge(111) surface by scanning tunneling microscopy. Therefore, the surface diffusion in the given case is an efficient mechanism of strain-induced roughening.

The analysis of the results obtained for the surface of the uniaxially strained foils produced from the iron-based amorphous alloy [3, 4] shows that they have much in common with those for the Ge(111) surface. Obviously, the corresponding data should be interpreted in the same manner. In the case of foils, the evolution of the relief is also initiated by external loading and the onset of the evolution obeys the energy criterion. However, taking into account that the observed phenomenon exhibits a multiscale (from a nanometer scale to a micrometer scale) behavior, the monoatomic diffusion cannot be treated as a dominant physical mechanism of mass transfer. At the same time, this phenomenon cannot be completely explained, for example, within the theory of plastic deformation. In our opinion, the best explanation is provided by an intermediate variant which can be referred to as cluster diffusion. Cluster diffusion is considered to mean not only the surface diffusion of islands (clusters) but also the possible diffusive mass transfer inside the defective surface layer. This brings up a number of important questions about the role played by external loading in the initiation and promotion of the diffusion process. In particular, the question arises as to whether the decrease in the activation energy of the process due to strain concentrators is of particular significance. All these problems call for further investigation.

Turning back to the idea of the decisive role played by fluctuations of the thermal energy in the structural evolution of strained solids, we can argue that this role is not reduced to the initiation of fracture. The strain-induced roughening is a process that is also initiated by external loading but has a "creative" nature. Investigations have revealed that, in some cases, this process provides a means for forming a quasi-periodic structure of the surface relief [2]. The main physical mechanism responsible for the formation of quasi-periodic structures, at least on the Ge(111) surface, is associated with the surface diffusion. Surface diffusion is a stochastic process in which the elementary event (activation of adatom–vacancy pairs) is an activation process; i.e., this event is a manifestation of the aforementioned thermal-energy fluctuations. In our opinion, it is deeply symbolic that a large contribution to the recognition of the creative function of thermal fluctuations was made by Serafim Nikolaevich Zhurkov.

REFERENCES

1. V. R. Regel, A. I. Slutsker, and É. E. Tomashevskiĭ, *Kinetic Nature of the Strength of Solids* (Nauka, Moscow, 1974) [in Russian].
2. S. N. Zhurkov, V. E. Korsukov, A. S. Luk'yanenko, B. A. Obidov, V. N. Svetlov, and A. P. Smirnov, Pis'ma Zh. Éksp. Teor. Fiz. **51** (6), 324 (1990) [JETP Lett. **51** (6), 370 (1990)].
3. V. E. Korsukov, A. S. Luk'yanenko, B. A. Obidov, and V. N. Svetlov, Pis'ma Zh. Éksp. Teor. Fiz. **55** (10), 595 (1992) [JETP Lett. **55** (10), 621 (1992)].
4. V. E. Korsukov, A. S. Luk'yanenko, B. A. Obidov, V. N. Svetlov, and E. V. Stepin, Pis'ma Zh. Éksp. Teor. Fiz. **57** (6), 343 (1993) [JETP Lett. **57** (6), 358 (1993)].
5. C. W. Snyder, B. G. Orr, D. Kessler, and L. M. Sander, Phys. Rev. Lett. **66** (23), 3032 (1991).
6. J. H. van der Merwe, J. Appl. Phys. **34**, 117 (1963).
7. A. A. Griffith, Philos. Trans. R. Soc. London, Ser. A **221**, 163 (1921).
8. V. I. Vladimirov, *Physical Theory of Plasticity and Strength* (Leningr. Politekh. Inst., Leningrad, 1975), Chap. 2 [in Russian].
9. V. P. Alekhin, *Physics of Plasticity and Strength of Surface Layers of Materials* (Nauka, Moscow, 1983) [in Russian].
10. S. Z. Bokshteĭn, *Diffusion and Structure of Metals* (Metallurgiya, Moscow, 1973) [in Russian].

Translated by O. Borovik-Romanova

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1013–1019. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 979–985.
Original Russian Text Copyright © 2005 by Melker, Kornilov.

PROCEEDINGS OF THE CONFERENCE DEDICATED TO S. N. ZHURKOV (1905–1997)

Molecular Dynamics Study of the Destruction of Single-Walled Carbon Nanotubes under Tension

A. I. Melker and D. A. Kornilov

St. Petersburg State Polytechnical University, Politekhnicheskaya ul. 29, St. Petersburg, 195251 Russia
e-mail: NDTCS@spes.stu.neva.ru

Abstract—The deformation and destruction of single-walled open carbon nanotubes are investigated in the framework of the molecular dynamics model of charges localized at bonds, which takes into account the atomic (ionic) and electronic degrees of freedom. This approach allows one to study excited electronic states induced by electronic transitions and to investigate both the ionic and electronic subsystems simultaneously. The structural transformations of nanotubes with (*10*, *2*) chirality under tension are investigated, and the tensile stress–strain diagrams of these nanotubes in the temperature range 100–1000 K are calculated. It is established that, at low and moderate temperatures, the destruction of the nanotubes has a similar nature: the interatomic covalent bonds are broken in the normal cross section of the nanotube. At high temperatures, the nanotube undergoes a scission along the stretching axis. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The considerable interest expressed in the design of instruments and devices based on nanostructures has raised problems regarding the description and prediction of the mechanical properties of their components. As a rule, these problems have been solved using molecular dynamics simulation (see, for example, [1, 2] and references therein). Until recently, molecular dynamics simulations and analytical calculations were conveniently performed with the Tersoff–Brenner empirical potential and modifications of it. However, the key factor in molecular dynamics studies is appropriate choice of the potential of interatomic interactions [3]. In the majority of molecular dynamics simulations of atomic systems, the contribution of the electronic subsystem is taken into account by implicitly assuming that the electrons are always in the ground state. In order to avoid this disadvantage, we elaborated a molecular dynamics model of charges located at bonds which takes into account both atomic (ionic) and electronic degrees of freedom. The fundamental difference between our model and other currently available models is that, within classical mechanics, the model of charges at bonds provides a means for investigating both atomic (ionic) and electronic subsystems simultaneously. This approach accounts for the contribution from the electronic subsystem even in the case when the electrons are in the excited state or when the system undergoes electronic transitions. The proposed model makes it possible to observe the variation in the configuration on the electronic level and to extend information thus derived to the atomic level closely related to the experiment [4–7].

This model is especially useful in investigating nonequilibrium electronic–atomic processes with a strong correlation and localization in space and time, particularly in studying fullerenes and nanotubes. The advantage of the new approach is that the structure is not postulated *a priori* and, moreover, can undergo significant transformations with time [4–8]. However, the main advantage of the model is that it offers considerable possibilities for analyzing the influence of electric and magnetic fields on the fullerenes and nanotubes under dynamic conditions [9].

In the present study, we considered the destruction of single-walled chiral carbon nanotubes. All calculations were carried in the framework of the new molecular dynamics method, which was developed previously to investigate the self-organization of fullerenes and nanotubes [5–7].

2. COMPUTER SIMULATION

We simulated the transformation of the structure of the nanotubes simultaneously with the mechanical properties during the process of their deformation. In addition to the conventional tensile stress–strain diagrams, we calculated the mean force acting on the atoms of the nanotube in a direction perpendicular to the stretching axis. In order to obtain the tensile stress–strain diagrams, we developed a special grip (Fig. 1). The main features of this grip are as follows.

(1) The number of carbon atoms contained in the grip is no less than 20% of the total number of atoms in the nanotube.

(2) The grip as a whole is displaced by a distance Δr = 0.001–1.0r_0 only in the Nth step of integration, where r_0 is the equilibrium distance between the nearest neighbor atoms (the C–C bond length) and N varies

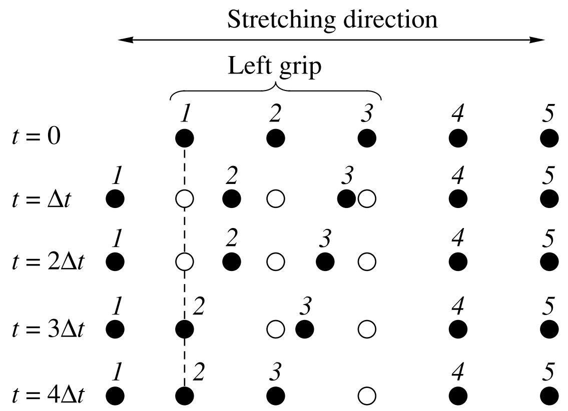


Fig. 1. Schematic diagram illustrating the flexible grip. Numerals near the circles are the numbers of atomic layers located perpendicular to the tensile force. All atoms within the same layer are displaced equally. The sequential positions of the atoms of the grip between the two main steps of integration are shown from the top down.

from 500 to 10 000. The parameters Δr and N determine the rate of deformation.

(3) In each step, the atoms of the grip are displaced by different distances. Atoms *1* are displaced by a distance Δr, atoms *2* by a distance $\Delta r/2$, atoms *3* by a distance $\Delta r/4$, etc.

(4) The nth atom of the grip in the N/nth step is additionally displaced by a distance $\Delta r_n = \Delta r/2^n$. This results in displacement of the grip as a whole in the Nth step.

(5) For all other steps of integration, excluding the motion of the grip itself, the atoms of the grip are assumed to be rigidly fixed.

This procedure was referred to as the construction of a flexible movable grip. Such a grip excludes breakage of the nanotube at the boundary of the grip.

We investigated the destruction of nanotubes of length $21r_0$ with (*10*, *2*) chirality. It should be noted that the chirality of single-walled nanotubes is designated by two indices and is defined as follows. A point on the atomic planar lattice with indices (l, m) is assumed to correspond to the origin of the coordinates (0, 0) by rolling a sheet cut from this plane. This results in the formation of a cylinder with an axis of rotation of the order of l and of a step m. In the case when $m = 0$, the tube is achiral. The simulation was carried out in the temperature range 100–1100 K, and the rate of deformation was 2.5×10^{-7} nm/ps. The initial configuration of the chiral nanotube is given in Fig. 2. In this figure, the small-sized spheres designate bonding electron pairs located at the midpoints of covalent bonds and the nonbonding (lone) electrons; the large-sized spheres correspond to carbon atoms. In other respects, the computational procedure is similar to that used before in the study of the destruction of achiral single-walled carbon nanotubes [8].

3. RESULTS OF THE COMPUTER SIMULATION

Figures 3 and 4 illustrate the evolution of the structure and the tensile stress–strain diagrams of a nanotube at a temperature of 100 K. It can be seen from these figures that the destruction occurs through the breakage of interatomic bonds independently in two nonadjacent regions located in the vicinity of the normal cross sections of the nanotube and is accompanied by local compression of the damaged regions. This process leads to considerable fluctuations in the tensile force. A similar tendency seems to exist also at the beginning of the process occurring at a temperature of 500 K, but then the destruction is localized in one of the regions shown in Fig. 5a, whereas in another region in the time interval 400–800 ps, there appear anomalously stretched bonds (Fig. 5b). The tensile stress–strain diagrams are practically identical to each other (Fig. 6). In contrast to the destruction at low and moderate temperatures, the destruction occurring at a temperature of 1100 K evolves in a different manner. The destruction proceeds in a considerable more extended region and looks more like a longitudinal shear than a normal detachment (Fig. 7). In this case, the tensile stress–strain diagrams also differ significantly (Fig. 8). The destruction is very similar to that observed upon destruction of a chain crystal simulating a highly oriented polymer (see [10, p. 199]). On the macroscopic level, this destruction corresponds to a scission of the polymer along the stretching axis.

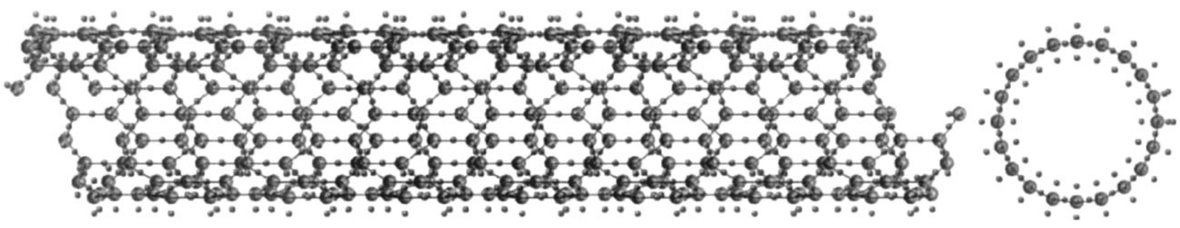
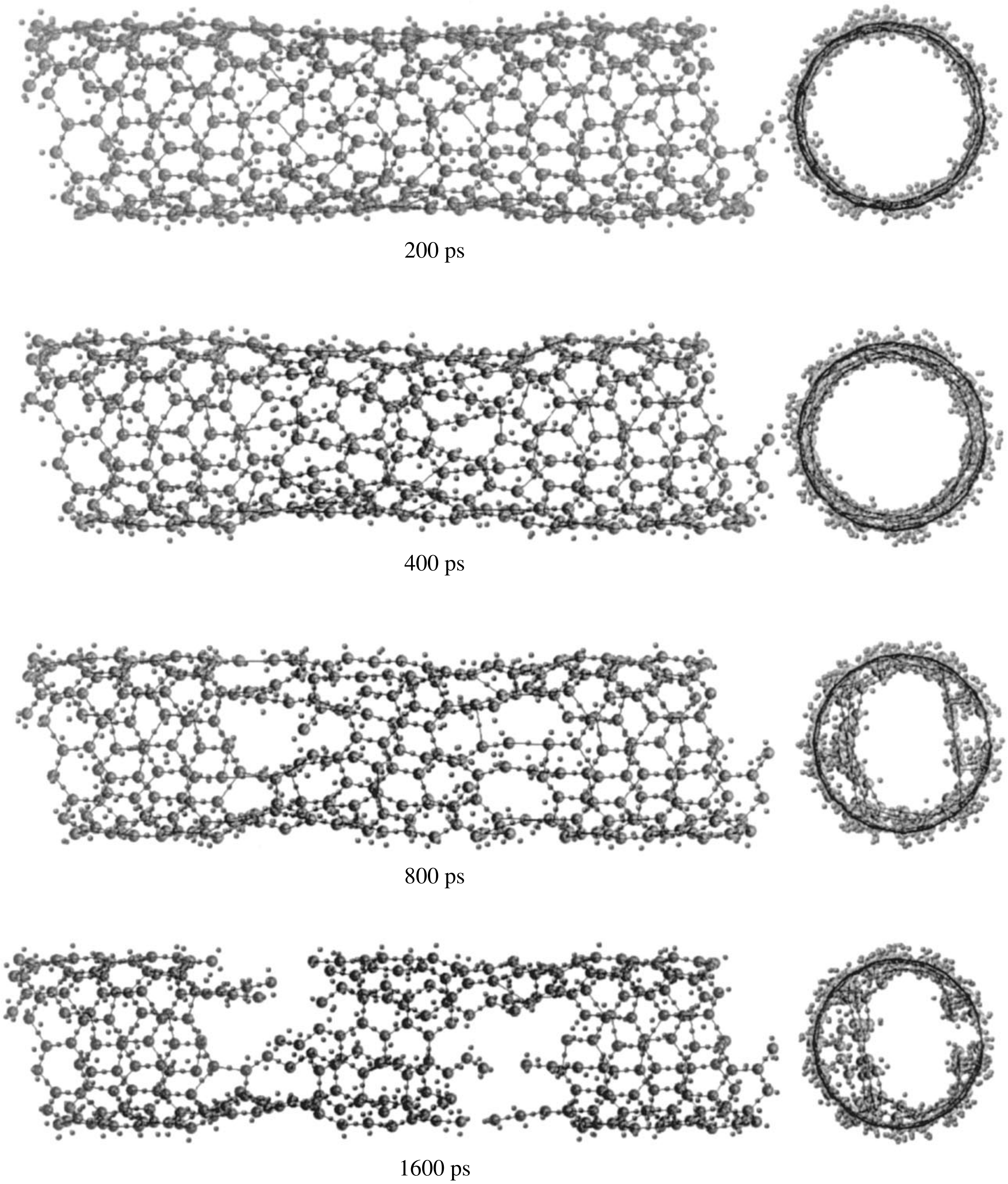


Fig. 2. Initial configuration of the nanotube with (*10*, *2*) chirality.



200 ps

400 ps

800 ps

1600 ps

Fig. 3. Structural transformation of the nanotube in the course of stretching at a temperature of 100 K.

4. DISCUSSION

First and foremost, we compare the character of the destruction of the chiral and achiral nanotubes. In the achiral nanotubes, at all temperatures, the destruction takes place in a narrow region located along the normal to the tensile force and has a brittle nature [8]. The thickness of this region, which is approximately equal to r_0 at low temperatures, somewhat increases with an increase in the temperature and becomes approximately equal to $(2–3)r_0$ at 1300 K. The tensile stress–strain diagrams at all temperatures are identical to each other and bear a resemblance to the tensile stress–strain diagrams of anharmonic chains of atoms (see [10, p. 81]). Strictly speaking, the tensile stress–strain diagrams calculated by the molecular dynamics method are not generally accepted, because they include forces localized on the scale of a hexagonal unit cell. These forces generate residual microstresses, which are called residual

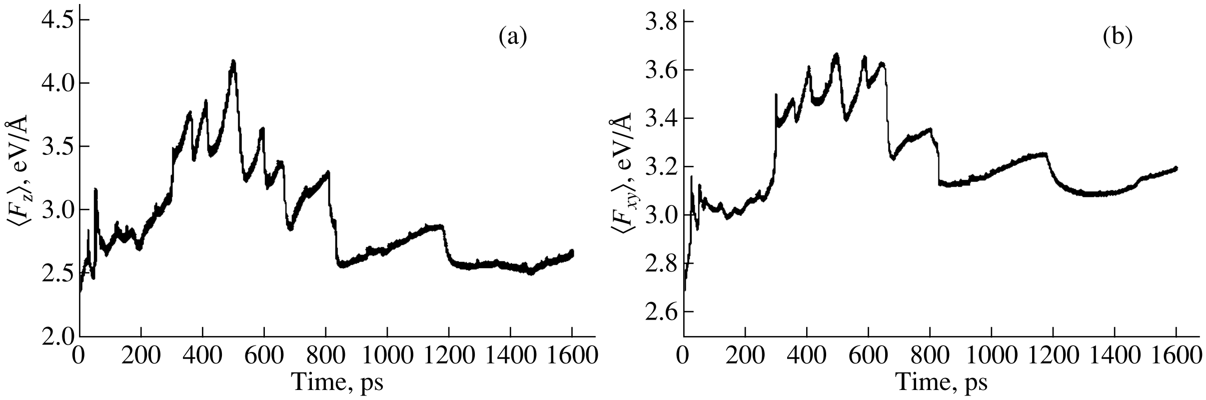


Fig. 4. Tensile stress–strain diagrams of the nanotube subjected to (a) a tensile force acting along the stretching axis *z* and (b) a compressive force normal to the stretching axis at a temperature of 100 K.

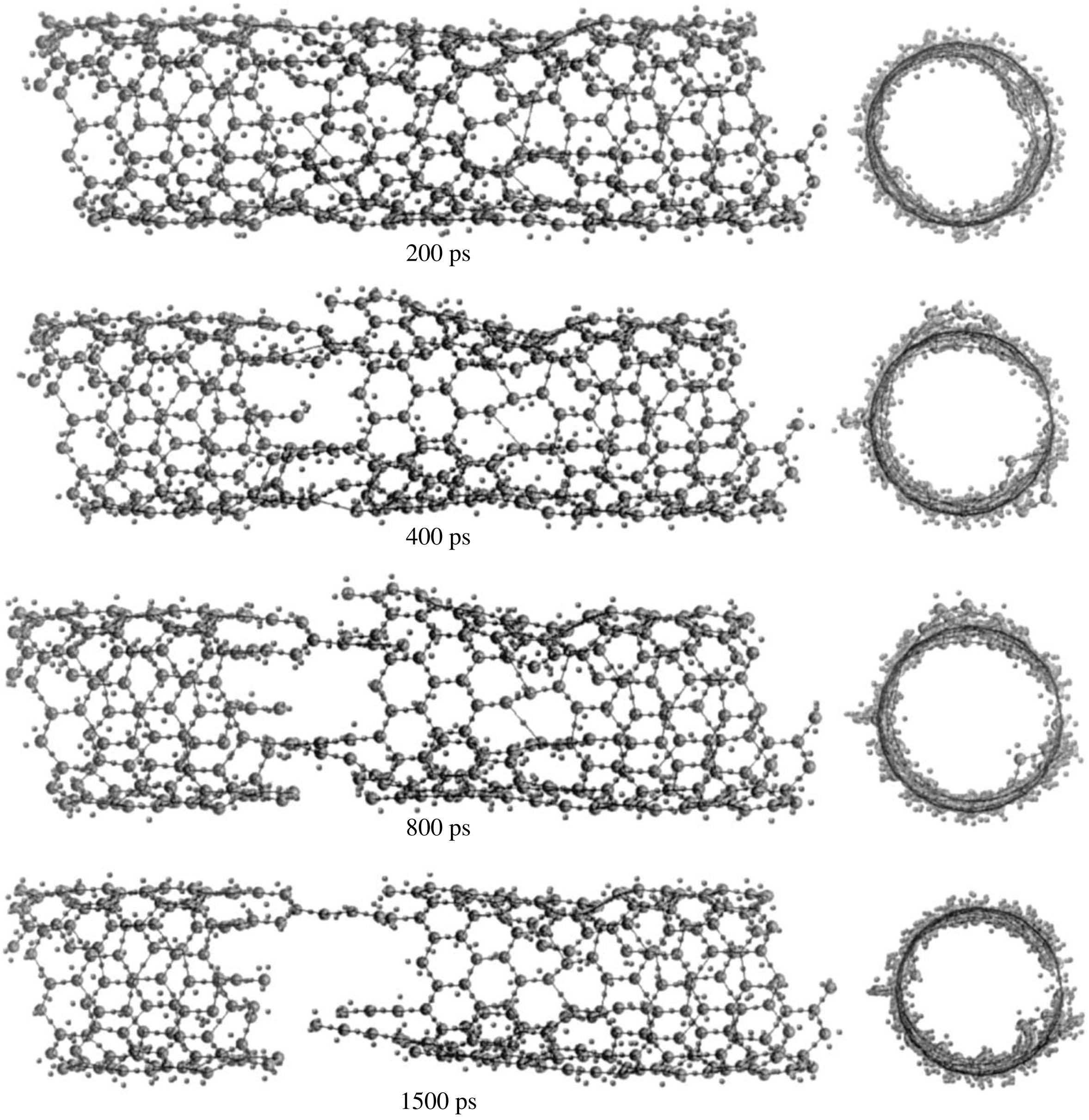


Fig. 5. Structural transformation of the nanotube in the course of stretching at a temperature of 500 K.

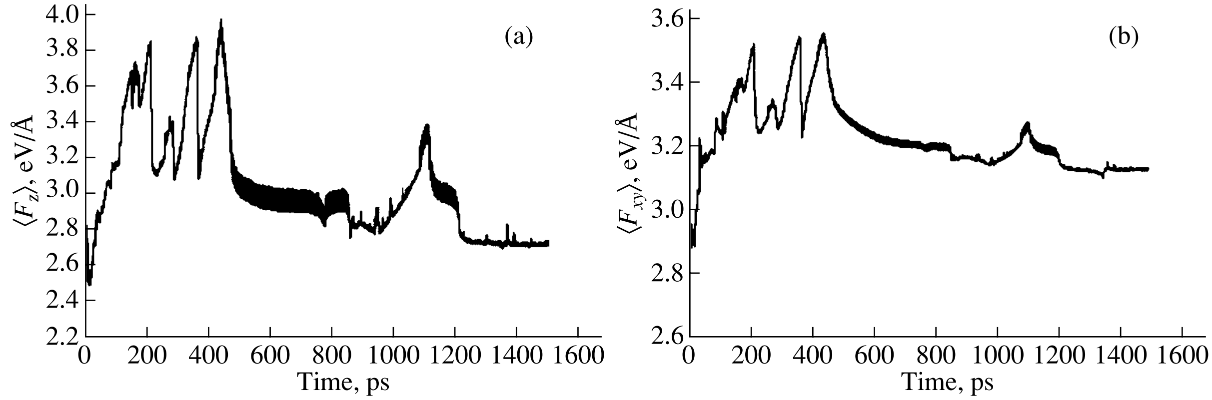


Fig. 6. Tensile stress–strain diagrams of the nanotube subjected to (a) a tensile force acting along the stretching axis *z* and (b) a compressive force normal to the stretching axis at a temperature of 500 K.

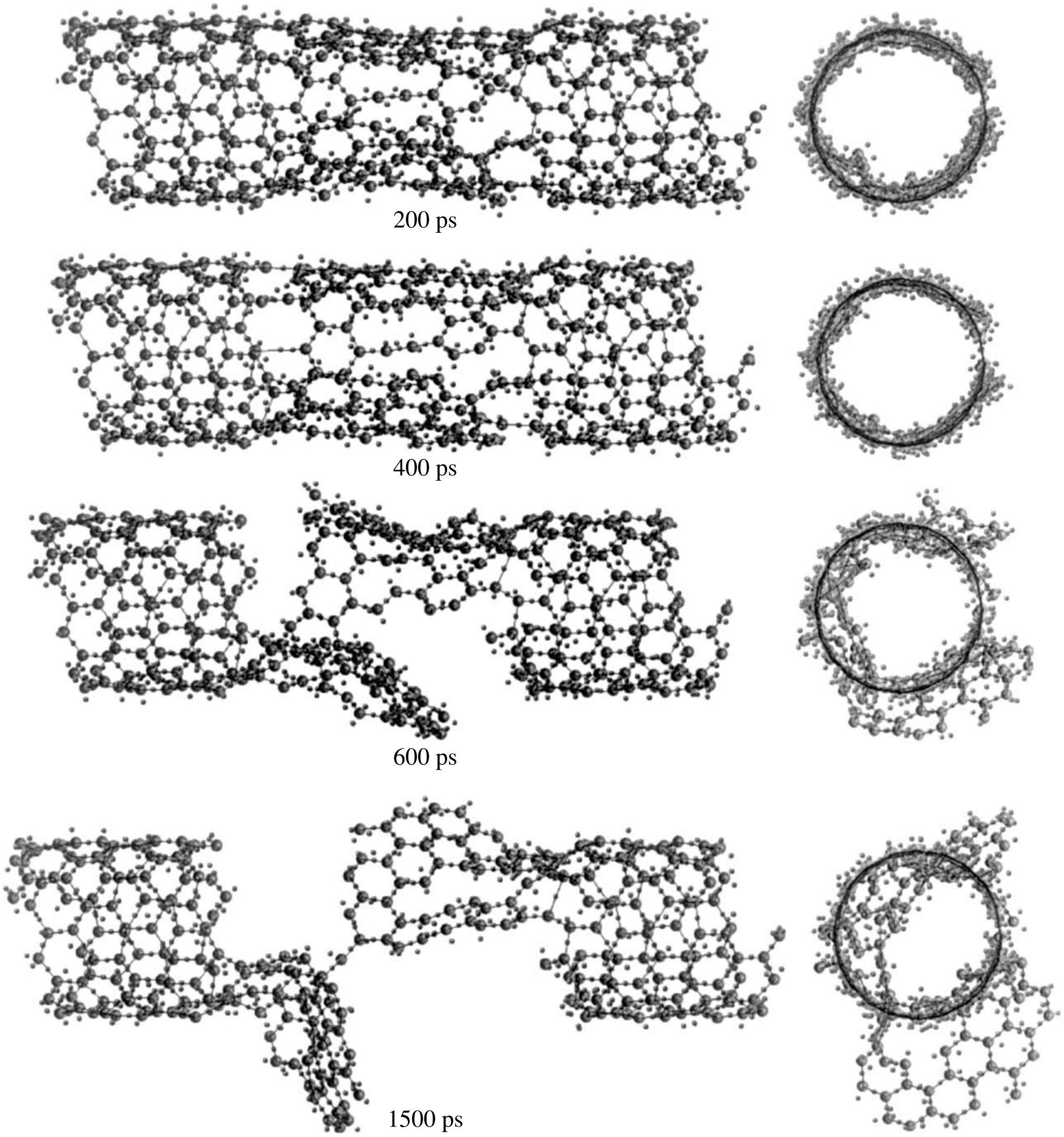


Fig. 7. Structural transformation of the nanotube in the course of stretching at a temperature of 1100 K.

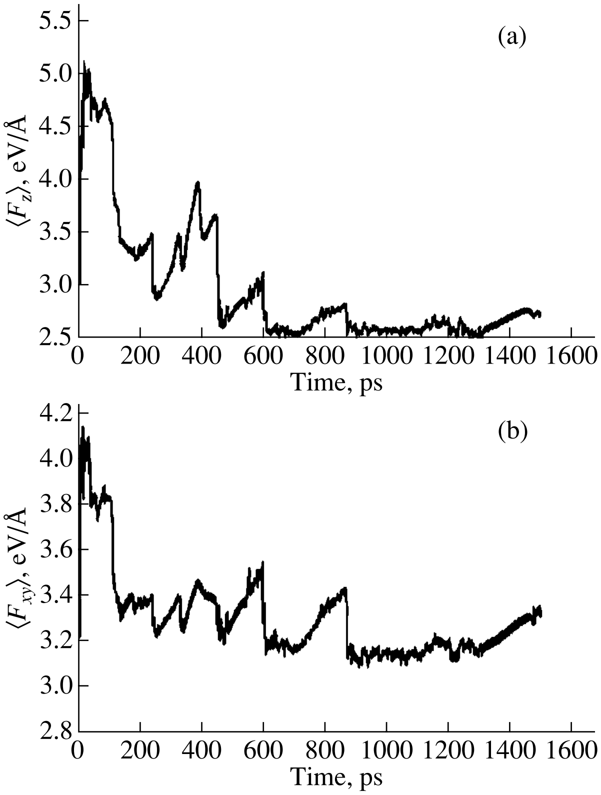


Fig. 8. Tensile stress–strain diagrams of the nanotube subjected to (a) a tensile force acting along the stretching axis z and (b) a compressive force normal to the stretching axis at a temperature of 1100 K.

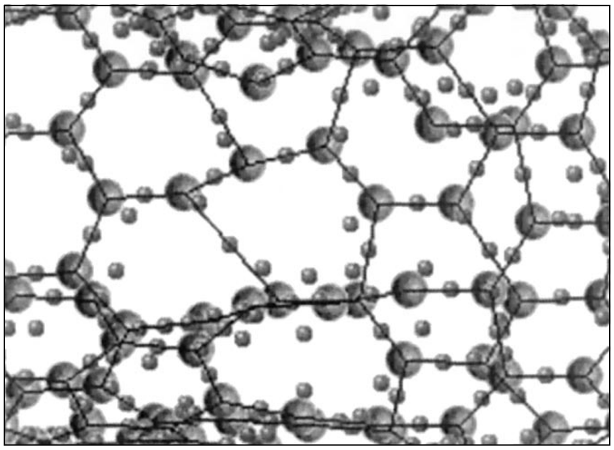


Fig. 9. Anomalously stretched bonds (at the center and in the lower part of the figure) upon tension at a temperature of 500 K for 600 ps.

stresses of the third kind. In order to change over to conventional tensile stress–strain diagrams, it is necessary to subtract the constant component associated with these residual stresses from the "molecular dynamics" diagrams. The analysis of the results presented in Figs. 4, 6, and 8 and in the corresponding figures given in our previous paper [8] demonstrates that the value of this component is approximately equal for the chiral and achiral nanotubes. When we subtracted this constant component from the molecular dynamics diagrams and then compared the tensile stress–strain diagrams, we found that the ultimate strength of the chiral nanotubes is less than that of the achiral nanotubes by a factor of approximately 5. Let us now elucidate what the reasons are for this difference.

Within the theory of elasticity, a chiral nanotube is a screw dislocation in a hollow right circular cylinder for which the Burgers vector b is proportional to the chirality index m and the cylinder itself is a hollow hard rod. The screw dislocation aligned parallel to the axis of the cylinder of radius R induces a torque moment M and twists this rod through an angle α. In the cylindrical system of coordinates, these quantities are given in the following form (see [10, p. 102]):

$$M = \mu b\frac{R^2}{2}, \quad \alpha = \frac{b}{\pi R^2}, \quad u_\theta(r, z) = \frac{b}{\pi R}z,$$

$$u_z(r, \theta) = \frac{b}{2\pi}\theta, \quad \sigma_{\theta z} = \frac{\mu b}{2\pi R}.$$

Here, μ is the shear modulus, u is the displacement, and σ is the stress. It follows from these formulas that the twisting increases with a decrease in the cylinder radius. For this reason, the long and thin filamentary crystals containing the screw dislocation greatly twist. This effect is referred to as Eshelby twisting. Moreover, the last formula suggests that, in a chiral nanotube, there occurs a shear stress localized on the scale of the whole nanotube. According to the stress classification, this residual microstress is called the stress of the second kind. For a nanotube with (l, m) chirality, this stress is of the order of $\mu m/l$, which, in our case, is ~$\mu/5$. According to the estimate obtained by Frenkel (see [10, p. 91]), this value is close to the theoretical shear strength. As can be seen from the initial structure of the nanotube (Fig. 2), the hexagonal unit cells are considerably distorted as a result of the Eshelby effect. The analysis of the structures presented in Figs. 2, 3, 5, and 7 (200 ps) demonstrates that, first, the deformation develops as a transverse shear. The sharp decrease in the strength of the chiral nanotubes as compared to that of the achiral nanotubes can be explained by the fact that the applied tensile stresses are summarized with the twisting microstresses of the second kind, which are close to the critical stresses.

One more circumstance should be mentioned. The tensile stress–strain diagrams at a temperature of 500 K exhibit strong fluctuations in the time intervals 400–800 and 1100–1200 ps (Fig. 8). It is in these time intervals that anomalously stretched bonds arise in the structure. They become particularly noticeable if we increase the scale of the image (Fig. 9). It can be assumed that both phenomena are related to each other.

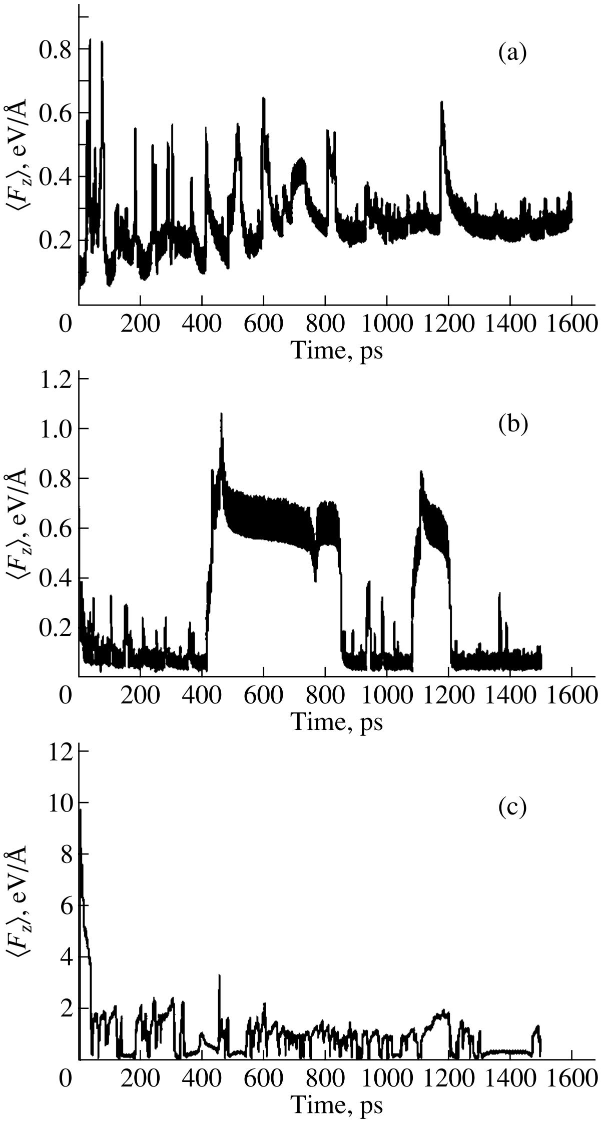


Fig. 10. Thermal fluctuations of the modulus of the interaction force between the atoms in the active part of the nanotube between the grips at temperatures of (a) 100, (b) 500, and (c) 1100 K.

In order to verify this assumption, we calculated the force of interaction between the atoms located along the stretching axis in the active part of the nanotube between the grips (Fig. 10).

It can be seen from Fig. 10 that, at 500 K in the time intervals 400–800 and 1100–1200 ps, the force of interatomic interaction in the active part of the nanotube is substantially greater than that in other time intervals. It follows from Figs. 6 and 10 that the ratio of this force to the tensile force is approximately equal to 0.6/3.0 = 0.2. Let us discuss this phenomenon. For the Morse potential

$$\varphi(r) = \varphi_0\{\exp[-2\alpha(r-r_0)] - 2\exp[-\alpha(r-r_0)]\},$$

the elastic moduli of the conventional and anomalously stretched bonds are given, respectively, by (see [10, p. 84])

$$k_1 = \frac{\alpha^2\varphi_0}{2\ln 2}, \quad k_2 = \frac{\alpha^2\varphi_0}{2(4-\ln 2)}.$$

Here, φ_0 is the energy of dissociation of the interatomic bond, $\alpha^2\varphi_0$ is the elasticity constant, and r_0 is the equilibrium distance between the atoms. In this case, we have

$$\frac{k_2}{k_1} = \frac{\ln 2}{(4-\ln 2)} \cong 0.2,$$

which coincides with the ratio between the above forces.

It is known (see [10, p. 81]) that anomalously stretched bonds are local concentrators of potential energy. The formation of anomalously stretched bonds takes place even when the tensile force is less than the ultimate strength. This slows down the destruction because part of the strain energy is absorbed by anomalously stretched bonds. In our case, this phenomenon manifests itself in the slowing down of the decrease in the strength after the ultimate strength is reached; consequently, the partially damaged tube can still sustain some, but a smaller, load.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education and Science of the Russian Federation within the program "Controlled Synthesis of Fullerenes and Other Atomic Clusters."

REFERENCES

1. P. Zhang, Y. Huang, H. Gao, and K. C. Huang, Trans. ASME **69**, 454 (2002).
2. T. Xiao and K. Liao, Nanotechnology **14**, 1197 (2003).
3. A. I. Melker, in *Polymerwerkstoffe 2000*, Ed. by W. Grellmann, H. Kausche, J. Kressler, and J. Vogel (Martin-Luther Universität, Halle, 2000), p. 394.
4. A. I. Melker, S. N. Romanov, and D. A. Kornilov, Mater. Fiz. Mekh. **2** (1), 42 (2000).
5. D. A. Kornilov, A. I. Melker, and S. N. Romanov, Proc. SPIE **4348**, 146 (2001).
6. A. I. Melker, D. A. Kornilov, S. N. Romanov, and N. A. Izotova, Proc. SPIE **5127**, 76 (2003).
7. A. I. Melker, Proc. SPIE **5400**, 54 (2004).
8. D. A. Kornilov and A. I. Melker, Proc. SPIE **5400**, 103 (2004).
9. K. I. Zhaldybin, D. M. Kalinin, and A. I. Melker, Proc. SPAS **8**, B23 (2004).
10. A. I. Melker, *Dynamics of Condensed Matter* (S.-Peterburg Akad. Nauk Probl. Prochn., St. Petersburg, 2004), Chap. 1 [in Russian].

Translated by O. Moskalev

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1020–1027. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 986–993.
Original Russian Text Copyright © 2005 by El'yashevich.

PROCEEDINGS OF THE CONFERENCE DEDICATED TO S. N. ZHURKOV (1905–1997)

Structure and Long-Term Mechanical Properties of Oriented Polyethylene

G. K. El'yashevich

Institute of Macromolecular Compounds, Russian Academy of Sciences, Bol'shoĭ pr. 31, St. Petersburg, 199004 Russia

Abstract—The formation of a supramolecular structure is studied in highly oriented polyethylene samples produced using two different methods based on melt extrusion. Relations between the degree of orientation, structural features, and mechanical properties of oriented samples are established. Zhurkov's concepts regarding the kinetic nature of the strength of solids are used to analyze the effect of the method employed to fabricate a supramolecular structure on the behavior of the samples upon loading and their long-term mechanical properties. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

It is now generally accepted that high mechanical properties of fibers and films made of flexible-chain polymers can be obtained when a system becomes oriented under the action of uniaxial tension. In the absence of molecular orientation, flexible-chain polymers are known to crystallize into folded crystallites connected by through chains (i.e., their chains fold). It is chain folding during crystallization that causes a low degree of linkage between the crystallites and the formation of an amorphous phase with through chains, which is a trouble spot of the system. A low fraction of through chains (according to Flory's estimations [1], at most 50% of the chains pass from every crystallite to a neighboring one) is a factor that limits the mechanical properties of the system. To increase these properties, it is necessary to increase the number of through chains and, especially, the number of load-bearing chains, whose number is less than 5% of the total number of chains in the cross section of a sample [2]. This problem does not arise when the self-folding of macromolecules is replaced by their unfolding. Since a molecule passes from one crystallite to another upon unfolding, virtually all chains in amorphous regions become through; this results in a sharp increase in the linkage between crystallites and, as a consequence, in an increase in the mechanical properties of the samples. As has been shown by many researchers [3–5], the transition from folding crystallization to chain-unfolding crystallization can be realized under conditions of molecular orientation. In this case, the supramolecular structure depends on both the degree of orientation and the initial structure of a system.

The problems of strengthening of flexible-chain polymers, in particular, commercial polymers, such as polyethylene and polypropylene, have been dealt with in many studies in Russia and abroad because of their scientific and practical importance. Two scientific schools have contributed substantially to the development of the scientific foundations and practical realization of improved mechanical properties of flexible-chain polymers: one is the school of the Ioffe Physicotechnical Institute (IPI), Russian Academy of Sciences, headed by Zhurkov, and the other is the school of the Institute of Macromolecular Compounds (IMC), Russian Academy of Sciences, headed by Frenkel. Researchers belonging to these schools formulated basic principles for the transition of flexible-chain polymers into an oriented state and determined key factors affecting this transition.

Studies of the supramolecular structure and structural transformations in crystallizing flexible-chain polymers have shown that these transformations are accompanied by the formation of certain structural elements and the failure of others. Therefore, two competing processes can simultaneously occur during uniaxial tension: strengthening as a result of orientation-induced straightening of molecules and their parallel arrangement along the tensile axis and failure as a result of the breaking of load-bearing through chains and the cleavage of crystallites. The competition between these two processes specifies the efficiency of orientation used to improve the mechanical properties. The main problem in this case is to attain a situation where the process of strengthening dominates over failure. The initial structure of a sample subjected to orientation plays a significant role. It has been found that failure of a polymer sample can be avoided during deformation if its initial state does not have a developed supramolecular structure in the form of a large number of coarse crystallites but rather contains fine crystallites (which can easily be rearranged) or is structureless. The domelike shape of the crystallization rate versus temperature curve [6] demonstrates that there are two regions where the crystallization rate decreases to low values and where such a structureless state can be achieved; more specifically, in the left branch, near the glass transition temperature T_g, and in the right branch, near the melting temperature

T_m. Accordingly, researchers have developed and realized strengthening processes in which the initial systems intended for subsequent orientation were formed in these two temperature ranges in the kinetic curve.

Researchers at the IPI [7, 8] designed the process of oriented drawing, where the initial structure was taken to be a structure that forms in the lower temperature branch of this curve, namely, an amorphous sample or a sample with an isotropic fine-crystallite structure prepared through crystallization at temperatures near T_g under substantial supercooling. Using this quenching, an amorphous, structureless state could be attained in polypropylene which has a sufficiently high glass transition temperature (about 0°C). Polyethylene has a very low T_g and a high crystallization rate [6]; therefore, even deep quenching cannot result in an amorphous state. However, a fine spherulite structure can be formed and then placed under tension to perform unfolding in crystallites and obtain a fibrillar structure based on unfolded-chain crystallites. This transformation of a quenched structure makes it possible to avoid failure if drawing is conducted at low tension rates and in a few stages, with the temperature being successively increased in each subsequent stage.

Another process of oriented crystallization (designed at the IMC) involves the higher temperature branch of the kinetic curve near T_m. In this process, a supercooled melt is subjected to molecular orientation in its highly elastic state [9–11]. This process occurs at a high tension rate, so orientation precedes crystallization and the crystallization takes place from a preliminarily oriented melt. As a result of tension, molecules in the melt acquire partly or completely unfolded configurations, which are fixed by crystallization to form crystallites with unfolded chains. The higher the tension of the melt before crystallization, the higher the degree of chain unfolding and the higher the proportion of the crystallites. The process of oriented crystallization ends in solidification at the melting temperature of fibrillar crystallites. The noncrystallized portion of the sample solidifies to form folded crystallites when the temperature decreases to their melting temperature [11].

Both processes (oriented crystallization initiated by melt tension at a near-T_m temperature and oriented drawing in which samples crystallized well below T_m is subjected to uniaxial tension) lead to the formation of highly organized supramolecular structures containing crystallites with unfolded chains. However, these processes differ substantially in terms of the shape, size, and linkage of the structural elements. These different structures undergo noticeably different long-term changes in their mechanical properties. To analyze these differences, the supramolecular structures of highly oriented polyethylene samples produced using these two methods were studied in [12, 13]. Samples of both types were prepared from commercial LPPE linear polyethylene with $M_w = 2.35 \times 10^5$ and $M_w/M_n = 6–8$.

2. SUPRAMOLECULAR STRUCTURE PRODUCED BY ORIENTED DRAWING

To study the effect of oriented drawing on the structure of oriented films, we prepared samples using melt extrusion at 270–290°C and jet drawing to 1.1–1.3. The extruded melt from the spinneret was immediately directed into a quenching bath having a temperature of 10°C. As shown in [8], these conditions are optimal for the formation of an isotropic structure in samples. The film produced was subjected to one-, two-, or three-stage drawing on a laboratory gradient heater. The first stage of drawing was carried out at 110°C. The temperature of this stage, in which necking (i.e., transition to an oriented state [14]) occurs, was chosen because segment mobility appears in polyethylene crystallites at this temperature [15]; therefore, conditions are created for chain unfolding and the transformation from folded into fibrillar crystallites. In the second and third stages, the drawing temperatures were 115 and 120°C, respectively.

The supramolecular structure of the samples was studied using scanning electron microscopy on a MINISEM microscope (AKASHI, Japan). The internal structure of the samples was examined by the “peeling” method [16]. These studies showed that the main element of the supramolecular structure of the samples produced by oriented drawing is a microfibrillar aggregation extended along the orientation axis. At low draw ratios ($\lambda_d = 6–9$), the sample structure is inhomogeneous and reflects the temperature gradient of drawing. In the surface layers, which were in direct contact with the heater, the structure is more perfect and consist of close-packed, well-oriented macroscopic fibrils 0.8- to 1.0-μm thick. Deeper layers have a looser, weakly organized structure and contain voids 0.03–0.14 μm in size. An increase in the draw ratio in the first stage to $\lambda_d = 14$ results in the formation of a highly oriented, microfibrillar structure (Fig. 1a). At these draw ratios, the transition from a fine-crystallite (fine-spherulite) initial structure to a fibrillar structure with unfolded-chain crystallites is complete [8, 12, 17]. The main building blocks of the latter structure are filamentary 0.1-μm-thick microfibrils separated by microvoids and extended along the tensile axis. The microfibrils are closely packed into macroscopic fibrils, and individual microfibrils are connected by transverse stressed through chains, which provide linkage between structural elements in the bulk of a sample.

In the second stage of drawing to a draw ratio $\lambda_d = 25$, the microfibrils are additionally oriented and their length increases. However, the transverse through chains disappear and the linkage between fibrils weakens; as a result, the sample cleaves easily to form individual microfibrils upon loading (Fig. 1b). With a further increase in the orientation in the third stage of drawing to $\lambda_d = 30$, the weakening of the linkage between fibrils leads to loosening of the fibrillar structure and its structural elements are deformed to differ-

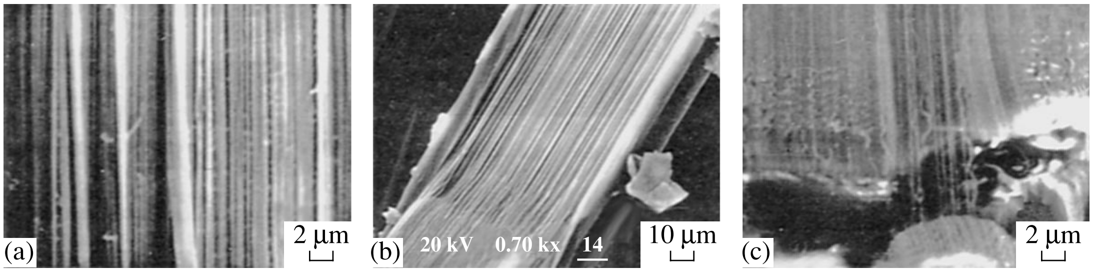


Fig. 1. Electron micrographs of the peeled surface of samples produced by oriented drawing at λ_d equal to (a) 14.2, (b) 21.3, and (c) 30.

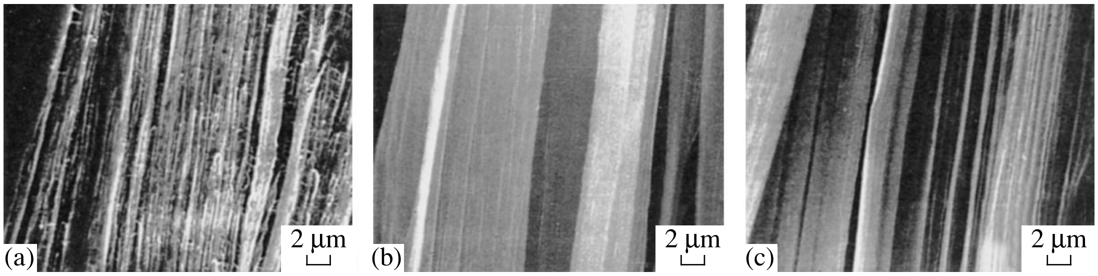


Fig. 2. Electron micrographs of the peeled surface of samples produced by oriented crystallization at λ_m equal to (a) 6, (b) 9, and (c) 16.5.

ent degrees. These factors result in failure and partial relaxation of the most stressed structural elements with time. Figure 1c shows numerous structural details in the form of curved aggregations after this relaxation.

As shown by x-ray diffraction, the increase in the oriented-draw ratio is accompanied by a significant increase in the fraction of a crystalline phase, which reaches 85% in the samples drawn to $\lambda_d = 30$. The mechanical properties also increase with λ_d: the elastic modulus increases continuously to $E = 35$ GPa at $\lambda_d = 30$, and the strength becomes maximum ($\sigma = 1.15$ GPa) at $\lambda_d = 20–22$ and then decreases to 1.0 GPa. We may assume that the rupture of overstressed building blocks causes a certain decrease in the strength at high draw ratios and that an increase in the degree of orientation provides a continuous increase in the elastic modulus, which becomes slower at $\lambda_d > 20$.

3. SUPRAMOLECULAR STRUCTURE PRODUCED BY ORIENTED CRYSTALLIZATION

Oriented crystallization is a continuous one-stage process with combined orientation and solidification zones. According to this method, a polyethylene melt is extruded at 180°C with a minimum possible jet-draw ratio, which is close to unity. Then, the melt was directed to rolls heated to 110°C, which ensured its effective deceleration before subsequent tension and crystallization induced by this orientation effect [10, 11]. The roll temperature was chosen on the basis of the dependence of the strength on the crystallization temperature determined earlier [12]. In our case, this temperature corresponds to a supercooled melt, whose high viscosity allows its tension to a high degree.

We studied samples produced at different degrees of melt tension before crystallization. Electron micrographs show that the samples produced at the lowest degrees of melt tension ($\lambda_m = 6.0$) consist of extended stripelike macroscopic fibrils 5- to 10-μm thick connected by stretched, thin, filamentary through molecules (strands) (Fig. 2a). The stripelike macroscopic fibrils are separated by extended wedge microvoids 1-μm wide and 10-μm long. The edges of these voids are connected by strands. The strands connecting void edges with macroscopic fibrils in the planes that are either normal or parallel to the surface lead to the formation of a dense viscous structure that is difficult to split.

At high degrees of tension ($\lambda_m = 9.0$ and 16.5, for Figs. 2b and 2c, respectively), the sample acquires a stripelike homogeneous supramolecular structure. However, strands between structural elements disappear. A comparison of Figs. 2a and 2c shows that, as the degree of orientation increases, voids between stripelike aggregations disappear and the stripelike building blocks become smoother and narrower (from 1 to 5 μm). The fibril length increases, the thickness of the fibrils making up stripes decreases, and their packing becomes denser and more perfect. However, even at $\lambda_m = 16.5$ (Fig. 2c), certain disorientation is retained in the form of bent macroscopic fibrils and macroscopic fibrils deviating from the main direction.

X-ray diffraction studies of orientationally crystallized samples show that, at $\lambda_m < 12$, the large spatial period is 250 Å and is independent of the degree of orientation. At higher degrees of orientation of the melt before crystallization, this period increases sharply to 400 Å. This increase in the crystallite size is accompanied by a decrease in the intensity of discrete small-angle scattering; at $\lambda_m > 14$, the small-angle reflection disappears, which indicates a denser packing of the amorphous phase in the intrafibrillar space and the formation of unfolded-chain crystallites [12, 18].

All these changes indicate the formation of a more perfect structure and significantly improve the mechanical properties. Indeed, the strength and elastic modulus of the samples produced at $\lambda_m = 6$ are 0.33 and 3.0 GPa, respectively, whereas those of the samples produced at $\lambda_m = 16.5$ are 1.05 and 13.3 GPa, respectively. As shown in [18], at these degrees of orientation in the first stage of crystallization, fibrillar crystals form during cooling of an oriented melt at temperatures below the melting temperature of the fibrillar crystals but above T_m of folded crystals. As the temperature decreases further, all the chains that have not been involved into the fibrillar crystals undergo folding crystallization. The higher the tension of the melt before crystallization, the greater the proportion of fibrillar crystals in a sample and the higher its mechanical properties.

4. COMPARISON OF THE SUPRAMOLECULAR STRUCTURES OF THE SAMPLES PRODUCED THROUGH ORIENTED CRYSTALLIZATION AND ORIENTED DRAWING

Thus, electron-microscopic studies demonstrate significant differences in the structures of samples produced by the two methods indicated above, namely, differences in the shape, size, and linkage of building blocks. In the orientationally crystallized samples, the building blocks are rather coarse monolithic stripelike aggregations, which do not peel when the samples are subjected to peeling. The building blocks of orientationally drawn samples are individual thin macroscopic fibrils, which are separated into thinner microfibrils due to peeling under load.

These differences are caused by the substantially different mechanisms of structure formation inherent to these two methods. The supramolecular structure of orientationally crystallized samples forms from a melt with homogeneously distributed crystallization centers, and this structural homogeneity remains unchanged at all degrees of melt tension. These samples are characterized by strong linkage between building blocks irrespective of the degree of orientation, since fibrils consisting of unfolded chains penetrate one another to form a continuous spatial skeleton embedded in the matrix of folded crystals and noncrystallized chain segments. This skeleton penetrates through a sample and takes the applied to the sample during deformation; therefore, the mechanical properties of the sample are specified by the properties of this rigid skeleton and its volume fraction. The strength of disordered polyethylene is known to be less than 0.1 GPa. Although the volume fraction of the skeleton in ordered polyethylene does not exceed 15–20% [10], this amount is sufficient to increase its strength by one order of magnitude.

The structure of orientationally crystallized samples forms from an oriented melt at a near-melting temperature, i.e., under near-equilibrium conditions. Therefore, the chains that are not involved in the crystalline skeleton can relax. Moreover, the high mobility of macromolecules under these conditions and the absence of ordered building blocks makes it possible to tension the viscous medium without the formation of voids and other defects and to form monolithic crystalline aggregations under conditions of rapid and free entrapment of molecules by these aggregations. The sizes of growing crystallizing regions are only controlled by the rate of material solidification, which controls crystallization. The crystalline skeleton formed under these conditions provides for high time stability of the mechanical properties of the samples.

The structure of samples produced by oriented drawing results from restructuring of the material crystallized at significant supercooling under conditions far from equilibrium. In this case, chain relaxation is hindered. Therefore, multiple nucleation during crystallization (when one chain penetrates several crystallites) leads to a large number of stressed through chains (strands). The fracture of these strands upon drawing can induce defects and free chain ends. Moreover, when chains are subject to orientation and unfolding in folded crystallites upon drawing, a tensile force is applied to each crystallite. Thereafter, unfolded chains are rearranged to form fibrillar aggregations under conditions of low chain mobility in the solidified material and short-term exposure to an elevated temperature. Difficult transverse motion of chains during the formation of these aggregations results in very fine building blocks (microfibrils), which manifest themselves during the separation of their coarser aggregations (macroscopic fibrils) upon loading in the process of peeling (Fig. 1c).

However, the possibility of reaching higher draw ratios (up to $\lambda_d = 30$) and, hence, higher degrees of orientation in multistage oriented drawing as compared to the draw ratios typical of single-stage oriented crystallization (up to $\lambda_m = 15–17$) provides improved mechanical properties of the orientationally drawn samples in spite of their inherent defects and voids.

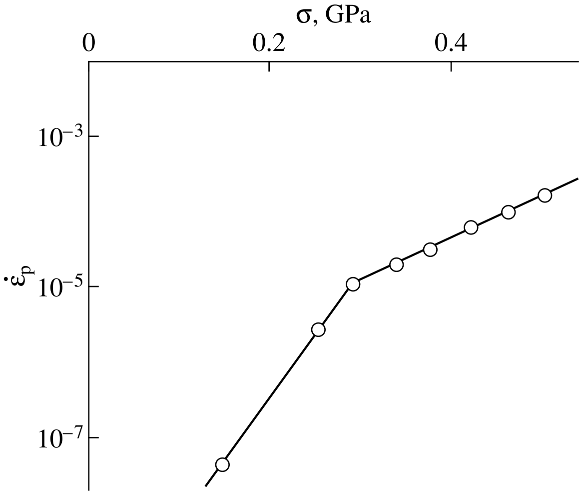


Fig. 3. Dependence of the plateau strain rate $\dot{\varepsilon}_p$ on the stress σ.

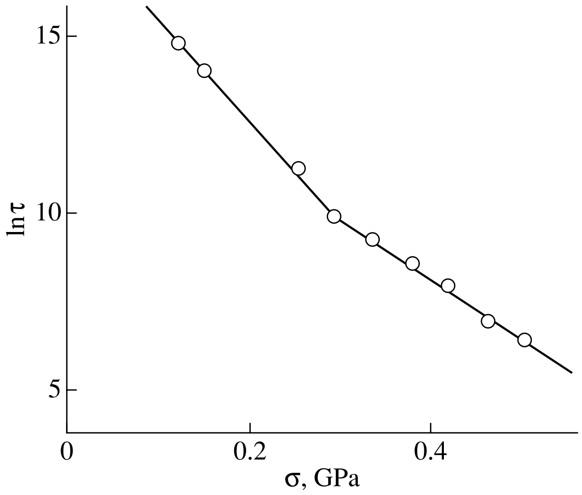


Fig. 4. Dependence of the logarithm of the time to failure τ on the stress σ.

5. LONG-TERM MECHANICAL PROPERTIES OF ORIENTED SAMPLES

Although the orientationally crystallized and orientationally drawn samples contain both fibrillar and folded crystallites [12, 18], the differences in their supramolecular structures substantially affect their response to an applied load and their behavior under a load. This behavior was analyzed under creep conditions [19, 20].

5.1. Deformation Behavior of Samples Produced by Oriented Drawing

In samples produced by oriented drawing, the linkage between the main building blocks containing unfolded-chain crystallites (macro- and microfibrils) is not high enough to prevent their displacement under a load. Therefore, the deformation behavior of these samples is characterized by creep at a load exceeding a certain critical value. We observed virtually no creep in the samples with a strength of 0.9 GPa at loads below ~0.13 GPa (which is about 15% of the breaking stress). At these loads, the sample elongation was less than 2% at times to failure varying from one to a few months, depending on the load.

The sample deformation was studied under a constant load at stresses above the critical value [19]. It was shown in [20] that, over a wide stress range, the dependences of the strain rate $\dot{\varepsilon}$ on the relative elongation have a plateau (corresponding to steady-state-creep conditions) and that the logarithm of $\dot{\varepsilon}$ depends linearly on the applied load σ in the plateau. This dependence is described by the Eiring equation

$$\ln\dot{\varepsilon}_p = \ln(\dot{\varepsilon}_0/2) - (\Delta U/kT) + (\sigma V/kT),$$

where $\dot{\varepsilon}_p$ is the value of $\dot{\varepsilon}$ in the plateau, $\dot{\varepsilon}_0$ is a constant, ΔU is the activation energy, and V is the activation volume.

For the orientationally drawn samples studied, the $\ln\dot{\varepsilon}(\sigma)$ dependence consists of two linear segments with different slopes (Fig. 3), which indicates that two creep mechanisms are operative. The first mechanism is related to the development of deformation in amorphous regions (specifically, fibril slip with respect to each other), and the second mechanism is related to chain slip through a crystallite (i.e., restructuring in crystallite regions). At low stresses, only the first process occurs; the second process becomes operative after reaching a certain critical stress. The linear segments in the $\ln\dot{\varepsilon}(\sigma)$ dependence correspond to two activation processes with different activation volumes. An estimation performed in [20] shows that the first process (which occurs at low loads) has a larger activation volume ($V_1 = 16.5 \times 10^{-29}$ m^3) than the second process ($V_2 = 7.4 \times 10^{-29}$ m^3), which becomes important at high strain rates. The value of the activation volume V_2 agrees well with the unit cell volume of polyethylene, which supports the assumption that the second process is related to α relaxation and chain slip through crystallites, whereas the first activation volume characterizes chain slip in less ordered interfibrillar and intercrystallite regions.

Figure 4 shows the logarithm of the lifetime of the samples produced by oriented drawing (the time to failure τ taken in seconds) as a function of the applied stress σ. This curve consists of two linear segments. Such dependences are described by the Zhurkov formula [22]

$$\tau = \tau_0 \exp[(U_0 - \gamma\sigma)/(kT)],$$

where U_0 is the fracture activation energy, γ is the fluctuation volume, and τ_0 is the preexponential factor.

The presence of two linear segments suggests two fracture mechanisms to be operative. At low stresses, one of them becomes dominant, and the second becomes substantial at high stresses. Using the slopes of the linear segments in Fig. 4, we calculated the fluctuation volume for each mechanism ($\gamma_1 = 11.6 \times 10^{-29}$ m^3, $\gamma_2 = 11.6 \times 10^{-29}$ m^3).

It should be noted that, for both the creep and lifetime of the samples, the point of transition between the two deformation modes corresponds to $\sigma = 0.3$ GPa, which is one-third of the breaking stress. This finding unambiguously indicates the presence of a relationship between the creep and fracture mechanisms. As in the case of creep tests, the activation volume in lifetime tests at high loads, γ_2, is found to be larger than γ_1. To explain this fact, we assume that the structure of highly oriented systems made of flexible-chain crystallizing fibrils consists of fibrils with alternating crystalline and amorphous zones and of unfolded, partly stressed through molecular chains located in interfibrillar gaps and mainly extended along the fibril axis [23].

According to Zhurkov's concept of a loaded sample [22], a load in certain chains is redistributed and some of the chains become overstressed and fail. This process leads to defect accumulation and, eventually, to fracture of the sample. This redistribution proceeds during a flow under a load. At small loads, fibrils and interfibrillar chains mainly slip with respect to each other, so defects are predominantly accumulated in noncrystalline regions. This process is characterized by the fluctuation volume γ_1. At high loads, chains slip through crystalline regions; there occurs slip along crystalline planes, and crystallites cleave. Therefore, fracture accumulates in the chain areas that were initially in crystallites. This behavior causes a decrease in the fluctuation volume γ_2.

Thus, the processes of both creep and fracture localize in disordered regions at low loads and in the defect zones of crystalline regions at high loads. The fact that the transition between these two modes occurs at the same load (compare Figs. 3, 4) indicates that creep processes play a significant role in the fracture of oriented samples upon loading.

5.2. Deformation of Orientationally Crystallized Samples upon Loading

The deformation behavior of orientationally crystallized samples upon loading is radically different. After initial elongation, they retain their sizes until fracture. This behavior is explained by the fact that a load applied to a sample is taken by a crystalline skeleton. The initial elongation is specified by the deformation of this skeleton. This elongation is 10–13% (which is virtually equal to the elongation at fracture of the sample). Thereafter, the sample behaves like a rigid solid and withstands a load without any creep for a time determined by the long-term strength of the sample and by the load. At loads higher than 0.12–0.15 GPa, the samples failed almost immediately, whereas lower loads were withstood by the samples for several months at a maximum elongation of 3%.

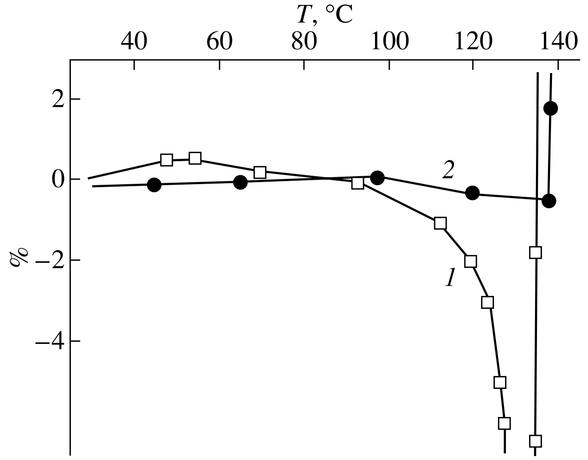


Fig. 5. Thermomechanical curves for (*1*) orientationally crystallized polyethylene samples and (*2*) fibrils forming upon fracture of a sample.

The orientationally crystallized samples also exhibit this type of behavior (characteristic of brittle low-molecular solids) upon heating under isometric conditions, where the ends of a sample are fixed. As the temperature increases up to the melting temperature of fibrillar crystallites (at which the samples failed), no increase in internal stresses was detected in the samples. These results agree with thermomechanical test data (Fig. 5). Unlike ordinary oriented samples of flexible-chain polymers, the orientationally crystallized samples demonstrate a shrinkage-free plateau upon heating to 100°C, whereas the samples produced by drawing usually shrink at 50°C. Internal stresses are retained in the samples up to the T_m of fibrillar crystallites. The fibrils (whiskers) forming upon fracture of the orientationally crystallized samples exhibit no shrinkage upon heating up to the temperature T_m of unfolded-chain crystallites, at which the stresses in a sample drops sharply.

5.3. Time Variation in the Mechanical Properties of Oriented Samples

Measurements of the mechanical properties of as-produced films show that the samples prepared using the two different methods have comparable strengths (1.15 and 1.03 GPa), whereas the elastic modulus E of the samples produced by oriented drawing (Fig. 6a, curve *2*) is substantially higher than that of the orientationally crystallized samples (Fig. 6a, curve *1*) at all degrees of orientation. Measurements of these charac-

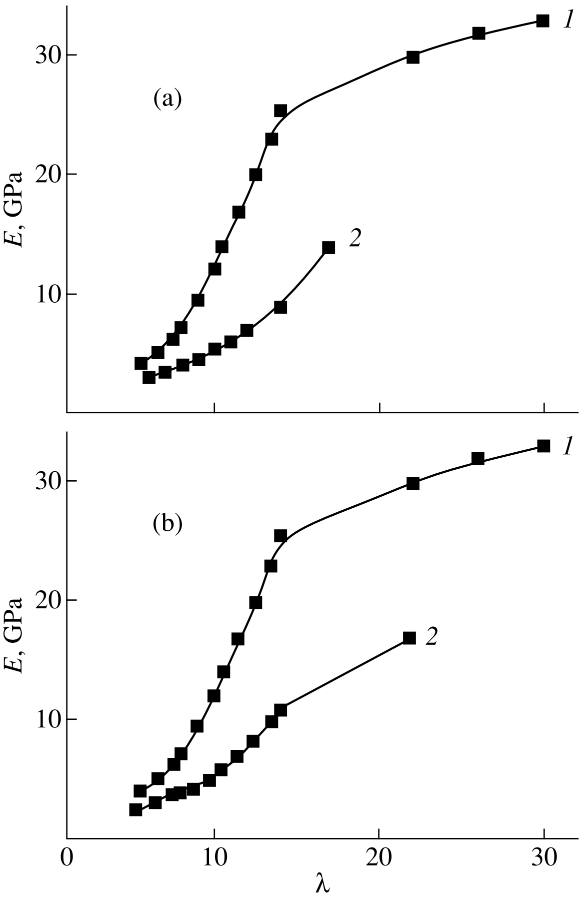


Fig. 6. (a) Dependence of the elastic modulus on the degree of orientation for (*1*) samples produced by oriented drawing and (*2*) orientationally crystallized samples. (b) Dependence of the elastic modulus on the degree of orientation for samples produced by oriented drawing: (*1*) for as-produced samples and (*2*) after one-year storage.

teristics performed after one year showed that the strengths of the samples of both types decreased by 10% and that the elastic moduli of the samples produced by oriented drawing decreased substantially (Fig. 6b).[1] However, the elastic modulus of the orientationally crystallized samples remains unchanged after one year [13].

As is seen from Fig. 6b, the decrease in the elastic modulus E after one year is greater for larger values of the draw ratio λ_d. For example, at a draw ratio $\lambda_d = 15$, the elastic modulus E decreases by a factor of 2.5 and becomes equal to the elastic modulus of the orientationally crystallized samples (Fig. 6a). The decrease in the elastic modulus indicates a decrease in the orientation and slow relaxation of oriented building blocks with time. This relaxation can be caused by a number of factors. Thermal motion can decrease the degree of orientation due to a weak linkage between the structural elements; the orientation can also be decreased because of an increase in the void dimensions between the building blocks as a result of a decrease in the transverse sizes of fibrils, which become thinner (Figs. 1a, 1b). Moreover, the appearance of discontinuities between fibrils induces noticeable fibrillation in the samples. The appearance of free (unfixed) fragments capable of relaxing, as a result of the fracture of overstressed microfibrillar building blocks, also favors a decrease in the orientation and the elastic modulus.

However, the method of multiple oriented drawing allows us to achieve higher limiting values of molecular orientation as compared to oriented crystallization, where the tension of a melt is restricted by solidification during unfolded-chain crystallization, since the effective melting temperature of unfolded chains is higher than the temperature of the process. An increase in the melt tension temperature leads to a decrease in the melt strength and, as a consequence, to a decrease in the maximum draw ratios of the melt. Higher degrees of orientation in the samples produced by multiple drawing lead to higher values of the elastic modulus, which remains rather high even after a decrease upon long-term storage.

After ten years of storage, the elastic moduli of the samples of both types remain unchanged, although the samples prepared at maximum values of λ_d become strongly fibrillar. The accumulation of defects upon storage for ten years resulted in a 15% decrease in the strengths of all samples.

6. CONCLUSIONS

Structural studies of highly oriented polyethylene samples produced using two different methods allowed us to establish the mechanism of formation of their supramolecular structures and its effect on the deformation behavior and long-term mechanical properties of the samples. Based on Zhurkov's concepts of the kinetic nature of the strength of solids, we have analyzed the relation between the structural factors and the behavior of a polymer material upon loading (in particular, practically important properties such as long-term strength and creep).

REFERENCES

1. P. J. Flory, J. Am. Chem. Soc. **84** (15), 2857 (1962).
2. A. Peterlin, Polym. Eng. Sci. **14** (9), 627 (1974).
3. A. J. Pennings, C. J. H. Schouteten, and A. M. Kiel, J. Polym. Sci. Part C **38**, 167 (1972).
4. J. H. Southern, N. Weeks, R. S. Porter, and R. G. Crystal, Makromol. Chem. **162**, 19 (1972).
5. A. Keller and M. J. Machin, J. Polym. Sci., Polym. Symp. **58**, 395 (1977).
6. Yu. K. Godovskiĭ, Vysokomol. Soedin. **11** (10), 2129 (1969).

---

[1] In 1977, Frenkel [9] theoretically predicted the phenomenon of slow relaxation of the elastic modulus by comparing the properties of high-strength polyethylene and polypropylene samples produced through oriented crystallization and oriented drawing.

7. V. A. Marikhin and L. P. Myasnikova, *Supramolecular Structure of Polymers* (Khimiya, Leningrad, 1977) [in Russian].
8. V. E. Korsukov, V. A. Marikhin, L. P. Myasnikova, and I. I. Novak, J. Polym. Sci., Polym. Symp. **42**, 847 (1973).
9. S. Frenkel, J. Polym. Sci. **58**, 195 (1977).
10. V. G. Baranov, Chem. Mater. **3**, 14 (1977).
11. G. K. El'yashevich and S. Ya. Frenkel', *Orientation Phenomena in Solutions and Melts of Polymers* (Khimiya, Moscow, 1980) [in Russian].
12. G. K. Elyashevich, E. A. Karpov, E. Yu. Rosova, V. A. Marikhin, and L. P. Myasnikova, Polym. Eng. Sci. **33** (20), 1341 (1993).
13. G. K. Elyashevich, E. A. Karpov, O. V. Kudasheva, and E. Yu. Rosova, Mechanics of Time-Dependent Mater. **3**, 319 (1999).
14. A. I. Slutsker, *Encyclopedia of Polymers* (Soviet Encyclopedia, Moscow, 1974), Vol. 2, p. 515.
15. E. A. Egorov, V. V. Zhizhenkov, V. A. Marikhin, and L. P. Myasnikova, Science B **29** (2/3), 129 (1990).
16. T. E. Sukhanova, A. V. Sidorovich, G. I. Goryainov, G. M. Mikhaĭlov, and M. Mitterpakhova, Vysokomol. Soedin. B **31** (5), 381 (1989).
17. L. P. Myasnikova, Plaste Kautsch. **33** (4), 121 (1986).
18. G. K. Elyashevich, Adv. Polym. Sci. **43**, 205 (1982).
19. E. A. Karpov, A. É. Bitskiĭ, B. V. Strel'tses, and G. K. El'yashevich, Vysokomol. Soedin. A **33** (6), 1334 (1991).
20. E. A. Karpov, E. Yu. Rozova, and G. K. El'yashevich, Vysokomol. Soedin. A **36** (6), 950 (1994).
21. H. Eiring, J. Chem. Phys. **4**, 631 (1936).
22. V. R. Regel', A. I. Slutsker, and É. E. Tomashevskiĭ, *Kinetic Nature of the Strength of Solids* (Nauka, Moscow, 1974) [in Russian].
23. W. P. Leung, C. L. Choy, K. Nagasawa, and T. Konaka, J. Polym. Sci., Polym. Symp. **25** (10), 2059 (1987).

Translated by K. Shakhlevich

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1028–1033. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 994–999.
Original Russian Text Copyright © 2005 by Pakhomov, Khizhnyak, Golikova, Galitsyn.

PROCEEDINGS OF THE CONFERENCE DEDICATED TO S. N. ZHURKOV (1905–1997)

Structural Transformations in the Course of Gel Spinning of High-Strength Polymer Fibers

P. M. Pakhomov*, S. D. Khizhnyak*, A. Yu. Golikova*, and V. P. Galitsyn**
** Tver State University, Sadovyĭ per. 35, Tver, 170002 Russia*
e-mail: Pavel.Pakhomov@tversu.ru
*** Research Institute of Synthetic Fibers, Moskovskoe sh., Tver, 170032 Russia*

Abstract—The structural transformations occurring in the course of the preparation of high-strength poly(ethylene) fibers by gel spinning are investigated using x-ray diffraction and IR and Raman spectroscopy. A model of the gel and xerogel structures is proposed. It is shown that the elastic and strength characteristics of poly(ethylene) fibers can be improved through the extension of molecular folds and an increase in the degree of crystal continuity along the fiber axis. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The design of new technologies for preparing polymer materials with unique properties, including ultrahigh-strength characteristics, is an important problem in polymer physics. At present, gel spinning has been the most efficient method for strengthening flexible-chain polymers [1, 2]. This method involves three main stages: (i) transformation of a polymer solution into a gel, (ii) gel spinning of fibers and subsequent removal of the solvent (transformation into a xerogel), and (iii) orientation drawing of spun fibers. The use of gel spinning makes it possible to improve the strength parameters of fibers based on ultrahigh-molecular-weight polyethylene (UHMWPE) by one order of magnitude. However, the strength characteristics remain far from the theoretical limit (more precisely, they account for approximately 10% of this limit). In our opinion, further improvement of the mechanical properties of these fibers can be based only on the results of complex investigations into the structure–property relations for gels and intermediate products at all stages of the technological process. In the present work, the structural transformations occurring in ultrahigh-molecular-weight polyethylene in the course of gel spinning of high-strength fibers were investigated using IR and low-frequency Raman spectroscopy, wide-angle x-ray diffraction, and rheological and mechanical measurements.

2. SAMPLE PREPARATION AND EXPERIMENTAL TECHNIQUE

We studied initial UHMWPE powders with $M_W = (1.7–14.2) \times 10^6$; solutions and gels of UHMWPE in decalin, *n*-xylene, and Vaseline (liquid paraffin) at concentrations up to 5%; xerogel films; and UHMWPE fibers (with different draw ratios) prepared at a pilot plant of the Research Institute of Synthetic Fibers (Tver, Russia) [3]. The technique used for preparing solutions and gels of ultrahigh-molecular-weight polyethylene was described in detail in [4].

The IR spectra were recorded on an Equinox 55 Fourier-transform IR spectrometer equipped with a polarizer. The molecular orientation in crystalline regions was evaluated from the formula [6]

$$\langle \cos^2\theta \rangle = (2 - R)/(2 + R), \tag{1}$$

where θ is the angle between the orientation axis and the axis of the molecular segment and R is the dichroic ratio of the IR band at a frequency of 731 cm^{-1} associated with the pendular vibrations of the CH_2 groups in crystalline regions of the polymer.

The Raman spectra of the samples in the longitudinal acoustic mode (LAM) range were measured using a DILOR XY 800 triple monochromator equipped with a $NdYO_4$ laser ($\lambda = 532$ nm; power, 100 mW). The spectra of oriented fibers were recorded in a 90° geometry for the polarization *XX*. The fiber axis was collinear with respect to the *X* direction. The Raman spectra in the LAM range were processed according to the procedure described in our earlier works [5, 6]. The length L of straight chain segments, irrespective of their state (amorphous, crystalline), and the longitudinal acoustic mode frequency ν_L are related by the expression [7]

$$\nu_L = (2cL)^{-1}(E/\rho)^{1/2}, \tag{2}$$

where ν_L corresponds to the frequency of acoustic vibrations of the elastic rod, c is the velocity of light, ρ is the density of the crystalline phase of the polymer, and E is the Young's modulus along the straight chain. The shape of the LAM band determines the length distribution function $F(L)$ of the straight chain segments according to the relationship [8]

$$F(L) \propto [1 - \exp(-hc\nu/kT)]\nu^2 I, \tag{3}$$

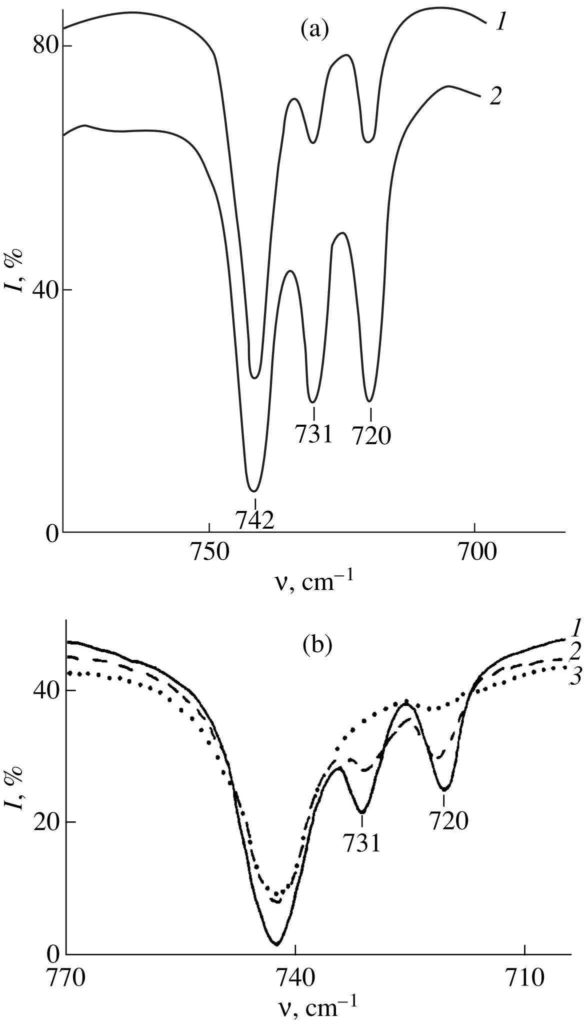


Fig. 1. IR transmission spectra of UHMWPE gels ($M_W = 1.7 \times 10^6$) prepared from decalin solutions with (a) polymer concentrations c = (*1*) 0.1 and (*2*) 1.5% and (b) the polymer concentration c = 1.0% at temperatures of (*1*) 20, (*2*) 80, and (*3*) 90°C.

where T is the absolute temperature, I is the Raman scattering intensity, k is the Boltzmann constant, and h is the Planck constant.

The wide-angle x-ray diffraction experiments were performed on a STADI P (STOE & CIE) diffractometer (monochromatic radiation at $\lambda = 0.154$ nm). The transverse size L_t of the crystallites was calculated from the Debye–Scherrer formula [9, 10]

$$L_t = K\lambda/\beta\cos\theta_{[200]}, \tag{4}$$

where $K = 0.94$ is a constant, β is the half-width of the diffraction peak (in rad), and $\theta_{[200]}$ is the diffraction angle of the (200) diffraction peak.

The rheological parameters of the solutions and gels were measured at a temperature of 60°C on a Carri-Med CSL 100 rotational viscometer. The mechanical characteristics (the strength and the elastic modulus for an extension of 2%) were determined with an Instron 1122 tensile testing machine for a gauge length of 150 mm and at an extension rate of 50 mm/s.

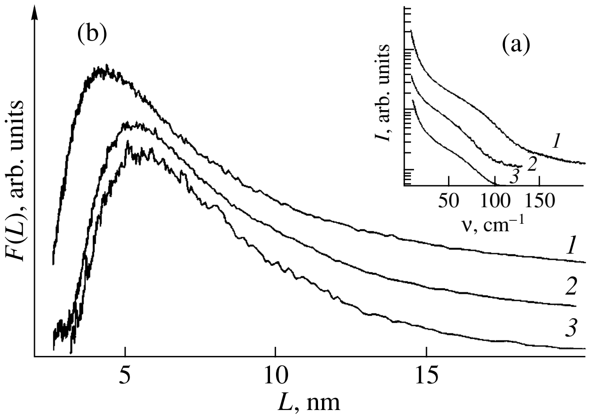


Fig. 2. (a) Raman spectra of UHMWPE gels ($M_W = 1.7 \times 10^6$, c = 2.5%) prepared from solutions in (*1*) *n*-xylene, (*2*) Vaseline, and (*3*) decalin. (b) Calculated length distribution functions of the straight chain segments.

3. UHMWPE GELS AND XEROGELS

According to the IR spectroscopic data obtained in [4, 6], poly(ethylene) gels contain a crystalline phase. This can be judged from the "crystalline" doublet at frequencies of 720 and 730 cm^{-1} in the IR spectrum of the gels (Fig. 1a). An increase in the temperature leads to a gradual decrease in the intensity of the doublet (Fig. 1b), which completely disappears at a temperature of 90°C; the gel transforms into the solution. Upon cooling, the crystal phase is again formed in the solution, which transforms into the gel, and the IR spectrum regains its initial shape. As follows from the rheological measurements, the solution $\longrightarrow$ gel transition leads to an increase in the shear modulus by three or four orders of magnitude. An increase in the elastic characteristics is also observed when the polymer concentration in the solution reaches a critical value at which a continuous spatial polymer network is formed [9]. The critical UHMWPE concentration in different solvents varies from 0.1 to 0.5%. It should be noted that an increase in the molecular weight of the polymer leads to a decrease in the critical concentration. The above data allow us to make the inference that, according to the Keller model [11], nodes of the spatial network in thermoreversible UHMWPE gels are microcrystallites.

More precise data on the structure of crystalline nodes can be obtained using Raman spectroscopy [5, 12–14] and wide-angle x-ray diffraction [9, 10, 15]. Figure 2a depicts the Raman spectra of UHMWPE gels based on different solvents. It can be seen that, in the low-frequency range, the spectrum contains a broad LAM band. The length distribution functions $F(L)$ of straight chain segments were calculated from the Raman spectra (Fig. 2b). These distribution functions

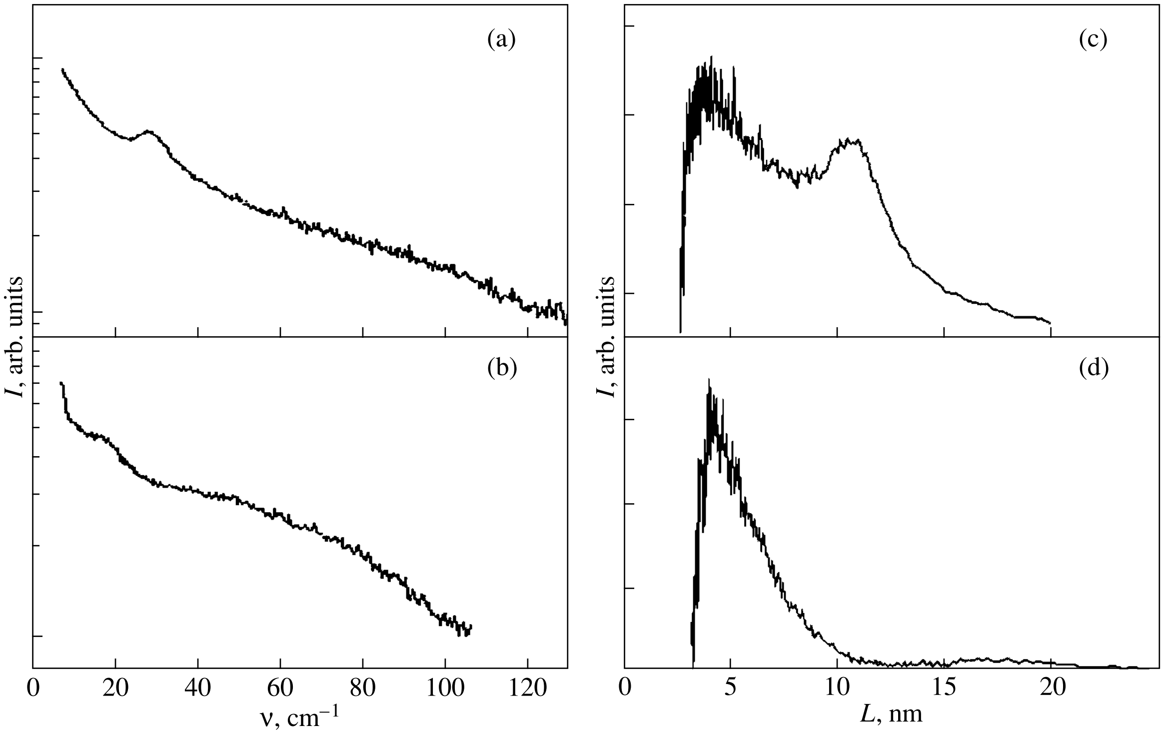


Fig. 3. Raman spectra of xerogels prepared from UHMWPE gels in (a) *n*-xylene and (b) Vaseline solutions. (c, d) Calculated length distribution functions of the straight chain segments for xerogels prepared from (c) *n*-xylene and (d) Vaseline solutions.

are represented by asymmetric dome-shaped curves in which the positions of maxima correspond to the most probable length L_p of the straight chain segments in the material. The location of the LAM band and, correspondingly, the most probable length L_p of the straight chain segments are virtually independent of both the solution concentration in the range 0.1–5.0% and the molecular weight of the polymers under investigation. The independence of the most probable length L_p of the straight chain segments from the solution concentration indicates that the straight chain segments in the gels are not isolated *trans* sequences but enter into the composition of crystallites whose sizes are determined by the thermodynamic conditions of their formation. Therefore, the most probable length L_p of the straight chain segments can be considered to be equal to the mean crystallite length L_c. The most considerable changes in the value of L_p are observed upon replacement of the solvent (Fig. 2b). In particular, the most probable lengths L_p of straight chain segments in the gels prepared from the solutions in *n*-xylene, decalin, and Vaseline are equal to 3.7, 5.0, and 5.2 nm, respectively. As was noted in our previous work [6], these differences are associated with the differences between the boiling temperature of the solvent and the crystallization temperature (90–120°C) at which the gel network is formed. Therefore, the Raman spectroscopic data confirm that the UHMWPE gels contain 3- to 6-nm-thick microcrystallites. The thickness of these microcrystallites depends on the solvent type.

The Raman spectra of solid samples (UHMWPE xerogels) and their length distribution functions of the straight chain segments are shown in Fig. 3. It can be seen from Figs. 3a and 3b that, upon removal of the solvent, the Raman spectrum of the xerogel contains a broad LAM band and a narrow band in the vicinity of the emission line of the laser. The length distribution functions of the straight chain segments for the xerogels under consideration are presented in Figs. 3c and 3d. The bimodal form of the distribution functions $F(L)$ suggests that an increase in the length of straight chain segments does not result from a gradual increase in the thickness of primary crystallites. The presence of low-intensity peaks (at ~12 and 17 nm) indicates the formation of a number of straight chain segments whose lengths exceed the thicknesses of the initial crystallites by several factors. We assume that these straight chain segments penetrate into crystallites and link adjacent crystallites together [10, 15].

The x-ray diffraction patterns of the UHMWPE gel and the corresponding xerogel in the range of the equatorial diffraction peak [200] are shown in Fig. 4. Note that the diffraction peak is asymmetric in shape and that the asymmetry is more pronounced for the xerogel. The decomposition of the diffraction peak into Gaussian components [9, 10] revealed that the experimental peak is adequately described by two components. The calculated parameters of both components are presented in

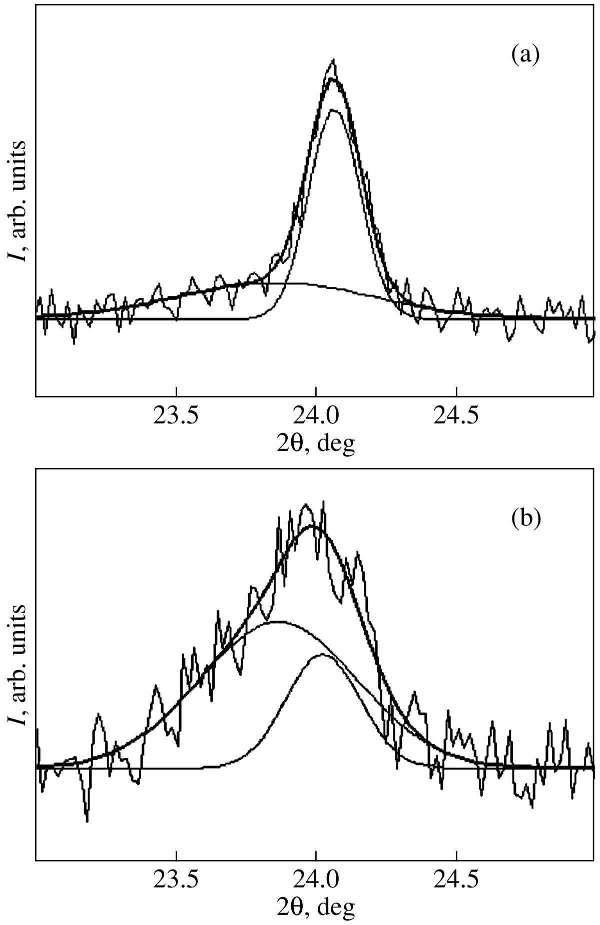


Fig. 4. X-ray diffraction patterns of (a) the UHMWPE gel and (b) the xerogel prepared from a 3.0% UHMWPE (M_W = 1.7 × 10^6) solution in Vaseline.

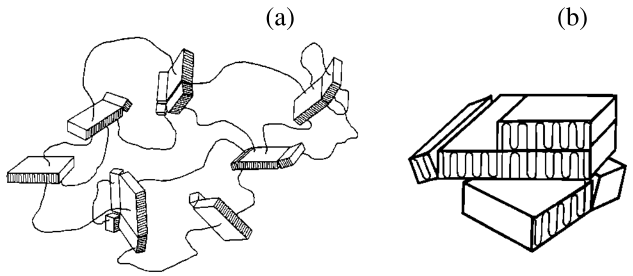


Fig. 5. Structural models of (a) the UHMWPE gel and (b) the UHMWPE xerogel.

Table 1. The angular positions θ of the diffraction peaks in the x-ray diffraction pattern, according to the Wulff–Bragg law,

$$2a_i = m\lambda/\sin\theta_i, \qquad (5)$$

determine the interplanar distances or the unit cell parameters a_i. Here, i is the number of individual components in the diffraction peak and m is the diffraction order (in our case, $m = 2$).

The unit cell parameters a_i determined from relationship (5) are listed in Table 2. The transverse crystallite sizes L_t calculated from formula (4) are also given in Table 2. The principal diffraction peak [200] ($i = 1$) at 2θ ≈ 24° corresponds to the orthorhombic crystallographic unit cell, whose parameters a_1 for the gel and xerogel are equal to 0.739 and 0.740 nm, respectively. These parameters are smaller than the standard parameter of 0.741 nm for block poly(ethylene). The peak at smaller angles ($i = 2$) can be attributed to defect unit cells with the parameter a_2 differing from the parameter a_1. Since each Gaussian component is characterized by its own values of L_t, it is believed that there are two types of crystallites with different transverse sizes. The areas under the Gaussian curves correspond to the weight functions of crystallographic unit cells of two types. The crystallite thicknesses L_p determined from the Raman spectroscopic data are also presented in Table 2.

Therefore, the analysis of the low-frequency Raman spectroscopic data demonstrates that the UHMWPE gel involves lamellar microcrystals (~5 nm thick), whose planes have a highly defective structure. The wide-angle x-ray diffraction data suggest a mosaic structure of crystallites (cracked crystallites) in the UHMWPE gel (Fig. 5a). This does not contradict the universally accepted model of the structure of a physical gel (the Keller model [11]) but only indicates a high degree of defectiveness of the crystalline nodes in the spatial gel network. According to Keller [11], the high deformability of the physical gels is associated with their sparse network and the high mobility of polymer chains. As a result of the gel $\longrightarrow$ xerogel transition, a number of straight chain segments link adjacent lamellar crystallites, thus forming clusters (a sandwich-type structure) consisting of two or more coplanar lamellar crystallites (Fig. 5b). The quasi-mosaic structure of primary crystallites in the gel and xerogel is an additional factor that facilitates the transformation of the crystalline phase upon transition to a highly oriented state during the fabrication of high-strength fibers [9, 10, 15].

Table 1. Angular positions θ and half-widths β of the diffraction peaks for the (200) plane in the UHMWPE gel and the UHMWPE-based xerogel

Sample	$2\theta_1$	$2\beta_1$	$2\theta_2$	$2\beta_2$
Gel	24.06 ± 0.01	0.69 ± 0.02	23.85 ± 0.05	0.18 ± 0.02
Xerogel	24.03 ± 0.02	0.52 ± 0.03	23.84 ± 0.04	0.27 ± 0.03

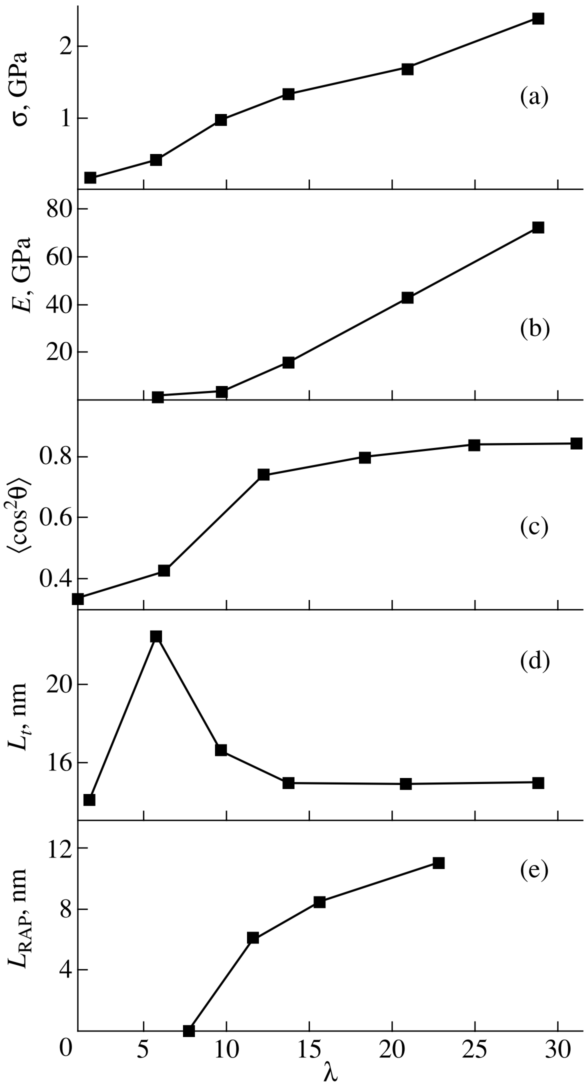


Fig. 6. Dependences of the (a) strength, (b) elastic modulus, (c) molecular orientation in crystallites, (d) transverse crystallite size, and (e) axial length of the rigid amorphous phase on the draw ratio of UHMWPE fibers.

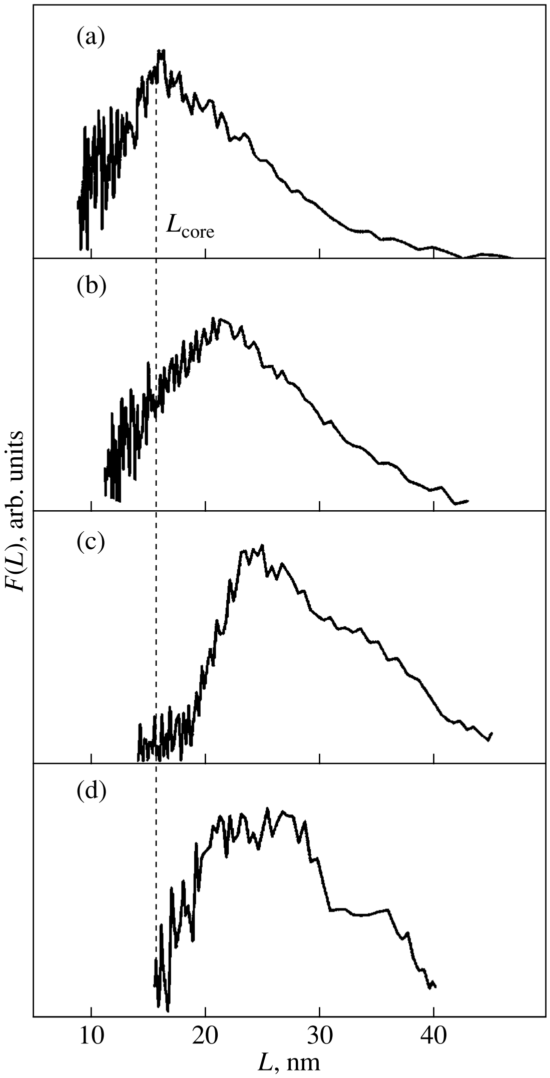


Fig. 7. Length distribution functions of the straight chain segments of UHMWPE fibers with draw ratios λ = (a) 7.7, (b) 11.6, (c) 15.6, and (d) 22.8.

4. ORIENTED UHMWPE FIBERS

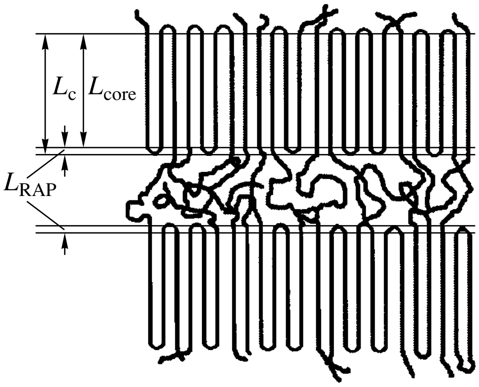
The results of complex investigations into the mechanical properties and structure of UHMWPE fibers as a function of the draw ratio λ are presented in Fig. 6. It should be noted that the fiber strength increases almost proportionally to an increase in the draw ratio λ, whereas the elastic modulus begins to increase at λ ≈ 7 (Figs. 6a, 6b). The transverse crystallite size L_t is stabilized when the molecular orientation in crystallites (according to the IR spectroscopic data) is completed (Figs. 6c, 6d). In this case, the elastic modulus does not correlate with the crystallite orientation.

The positions of the maxima of the distribution functions $F(L)$ in Fig. 7 correspond to the effective crystallite thickness L_c, which is equal to the sum of the crystalline-core thickness L_{core} and the length L_{RAP} of straight chain segments forming the so-called rigid

Table 2. Parameters (nm) of the crystalline phase in the UHMWPE gel and the UHMWPE-based xerogel

Sample	a_1	a_2	L_1	L_2	L_p
Gel	0.739 ± 0.01	0.743 ± 0.02	47.0 ± 2.0	12.0 ± 1.0	5.1 ± 0.1
Xerogel	0.740 ± 0.01	0.744 ± 0.01	31.0 ± 3.0	16.0 ± 2.0	4.3 ± 0.1

(a)
Poorly oriented



(b)
Moderately oriented

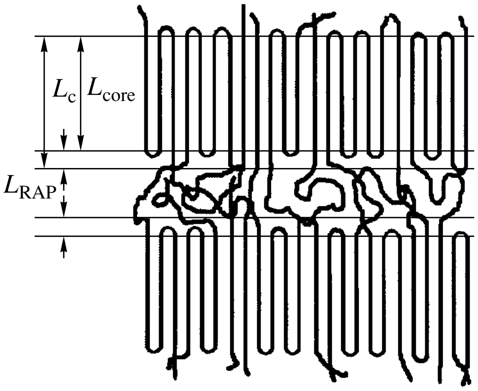


Fig. 8. Structural models of (a) poorly and (b) moderately oriented UHMWPE fibers prepared by gel spinning.

amorphous phase in the intermediate region between the crystalline and amorphous regions [5]; that is,

$$L_c = L_{\mathrm{core}} + L_{\mathrm{RAP}}. \qquad (6)$$

The inequality $L_{\mathrm{RAP}} \ll L_{\mathrm{core}}$ is satisfied for a poorly oriented sample (Fig. 8). The shift of the maximum of the distribution function $F(L)$ toward larger lengths L with an increase in the draw ratio λ (Fig. 7) is caused by the extension of chains on the surface of lamellar crystallites. The extension of chains begins simultaneously with an increase in the elastic modulus (cf. Figs. 6b, 6e). At $\lambda > 15.6$, the distribution function becomes bimodal (Fig. 7c). The bimodality implies the appearance of extended chains embedded in two adjacent crystallites. An increase in the degree of crystal continuity along the fiber axis in the course of orientation drawing is in agreement with the Ward bridging model of the structure of oriented poly(ethylene) samples [8, 16, 17].

5. CONCLUSIONS

Thus, structural investigations of thermoreversible UHMWPE gels demonstrated that the modified Keller model is valid for these materials. According to this model, lamellar crystallites ~4–5 nm in thickness and ~20–40 nm in transverse size are nodes of a spatial gel network. The crystallite planes have a defect mosaic structure. The gel $\longrightarrow$ xerogel transformation results in the formation of clusters composed of two or more coplanar lamellar crystallites. Orientation drawing leads to extension of the folds in lamellar crystallites, an increase in the degree of crystal continuity along the fiber axis, and, as a consequence, improvement of the elastic-strength characteristics of fibers.

ACKNOWLEDGMENTS

We would like to thank A.E. Tshmel for his participation in discussing and interpreting the results obtained in this study and M.D. Lechner, H. Reuter, and E. Ruhl for their assistance in performing the experiments.

This work was supported by the Deutscher Akademischer Austausch Dienst (DAAD) and the NATO Foundation (grant no. SfP 973925).

REFERENCES

1. B. Kalb and A. J. Pennings, Polym. Bull. (Berlin) **1**, 871 (1979).
2. P. Smith and P. J. Lemstra, Makromol. Chem. **180**, 2983 (1979).
3. V. P. Galitsyn, V. P. Napasnikov, and A. E. Mikushev, RU Patent No. 1,796,689 (1993).
4. P. M. Pakhomov, N. V. Larionova, and V. G. Alekseev, Vysokomol. Soedin., Ser. A **37** (5), 892 (1995).
5. P. M. Pakhomov, S. D. Khizhnyak, H. Reuter, V. P. Galitsyn, and A. E. Tshmel, Polymer **44**, 4651 (2003).
6. P. M. Pakhomov, A. E. Chmel', S. D. Khizhnyak, and V. P. Galitsin, Dokl. Akad. Nauk **386** (2), 220 (2002).
7. R. F. Schaufele and T. Shimanouchi, J. Chem. Phys. **47**, 3605 (1967).
8. G. Capaccio, M. A. Wilding, and I. M. Ward, J. Polym. Sci., Polym. Phys. Ed. **19**, 1489 (1981).
9. P. M. Pakhomov, S. D. Khizhnyak, H. Reuter, M. D. Lechner, and A. E. Tshmel, Macromolecules **36**, 4868 (2003).
10. P. Pakhomov, S. Khizhnyak, H. Reuter, and A. Tshmel, J. Appl. Polym. Sci. **89**, 373 (2003).
11. A. Keller, Faraday Discuss., Chem. Soc. **101**, 1 (1996).
12. K. Kober, S. Khizhnyak, P. Pakhomov, and A. Tshmel, J. Appl. Polym. Sci. **72**, 1795 (1999).
13. P. Pakhomov, S. Khizhnyak, V. Galitsyn, E. Ruhl, V. Vasil'eva, and A. Tshmel, J. Macromol. Sci., Phys. **B41** (2), 229 (2002).
14. P. M. Pakhomov, S. D. Khizhnyak, E. Ruhl, V. Egorov, and A. Tshmel, Eur. Polym. J. **39** (5), 1019 (2003).
15. P. M. Pakhomov, H. Reuter, S. D. Khizhnyak, and A. E. Chmel', Dokl. Akad. Nauk **391** (2), 215 (2003).
16. G. Capaccio, T. A. Crompton, and I. M. Ward, J. Polym. Sci., Polym. Phys. Ed. **14**, 1641 (1976).
17. M. Al-Hussein, G. R. Davies, and I. M. Ward, Polymer **42**, 3679 (2001).

Translated by O. Borovik-Romanova

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1034–1041. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1000–1008.
Original Russian Text Copyright © 2005 by Zavyalov.

PROCEEDINGS OF THE CONFERENCE DEDICATED TO S. N. ZHURKOV (1905–1997)

From the Kinetic Theory of Strength and Fracture Concentration Criterion to the Seismogenic Fracture Density and Earthquake Forecasting

A. D. Zavyalov

Schmidt Joint Institute of Physics of the Earth, Russian Academy of Sciences,
ul. Bol'shaya Gruzinskaya 10, Moscow, 123995 Russia
e-mail: zavyalov@ifz.ru

Abstract—The evolution of the concepts of the concentration criterion of fracture of solids, proposed by Zhurkov, in connection with its use for forecasting strong seismic events in earthquake source physics is considered. The results of long-term testing of the seismogenic fracture density as an earthquake precursor are analyzed in various seismically active regions of the world. Data on its forecasting efficiency are considered. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The author's idea to use the concentration (or crack enlargement) criterion for fracture K^* in seismology for earthquake forecasting arose for the first time in the spring of 1978 after the publication of an article by Zhurkov *et al.* [1]. That work developed kinetic concepts on the strength of solids and the thermal-fluctuation mechanism of the in-time fracture process, and the concentration criterion was formulated. This criterion was formulated as follows. If the average distance l between cracks in a loaded object is expressed in terms of the size L of cracks on a given structural scale, one can write

$$l/L = c^{-1/3}/L = K^*, \quad (1)$$

where c is the crack concentration. The parameter K^* characterizes the closeness of macroscopic cracks to each other and, hence, their capability to interact via local stress fields induced by them and to coalesce.

In 1977, researchers at the Institute of Volcanology (Far East Division, Russian Academy of Sciences, Petropavlosk-Kamchatskiĭ), with the author's participation, completed immense work on converting the regional catalog of earthquakes of the Kamchatka Peninsula and Commander Islands (several tens of thousands of events) to magnetic media and storing them on computer. Since the concentration criterion has a clear physical meaning, it was not very difficult to develop an appropriate algorithm and to calculate spatiotemporal distributions of this parameter for large volumes of rocks in seismically active regions.

Based on the analogy of fracture processes at various scale levels, the concepts of the concentration criterion of crack enlargement have been extended to large-scale processes in the Earth's crust. This was done for the first time in [2] to forecast strong earthquakes on the east coast of Kamchatka. An analogous analysis was carried out for the Nurek water basin region in [3]. The first result was rather encouraging. It was shown that strong earthquakes take place mostly in regions with decreased concentration criteria. On this basis, the seismogenic fracture (crack) concentration parameter K_{sf} [4] was implemented in seismological practice. This parameter characterizes the seismological process in a given seismically active volume V_0 of the Earth's crust at an instant of time t. An analysis of the K_{sf} values shortly before the instant of an earthquake showed them not to differ significantly from the values of the concentration criterion K^* obtained in laboratory experiments on the destruction of rocks and model materials (which differ in size by 10–12 orders of magnitude) [1, 5–8]. These values (immediately before fracture) were fitted well to the average distance between cracks plotted as a function of their average length [9] (Fig. 1).

In [1, 10, 11], it was theoretically shown that the probability of crack clustering is low in the case of small crack concentrations, where $K^* > e$ (Fig. 2). At $K^* < e$, the clustering probability sharply increases. Assuming that the crack concentration in such a cluster is sufficient for their interaction, the condition $K^* = e \approx 3$ controls the boundary of the transition from the stage of random accumulation of small cracks to the stage of their interaction, coalescence, and the transition of fracture to the next scale level.

Later, based on the dynamics of the parameter K_{sf} during a period prior to an earthquake (Fig. 3), an algorithm was developed for using the crack concentration parameter to forecast the location and time of strong earthquakes, taking into account the features of seismic processes in a seismically active region [4, 12]. The algorithm developed was tested in several seismically active regions of the world with different geodynamics

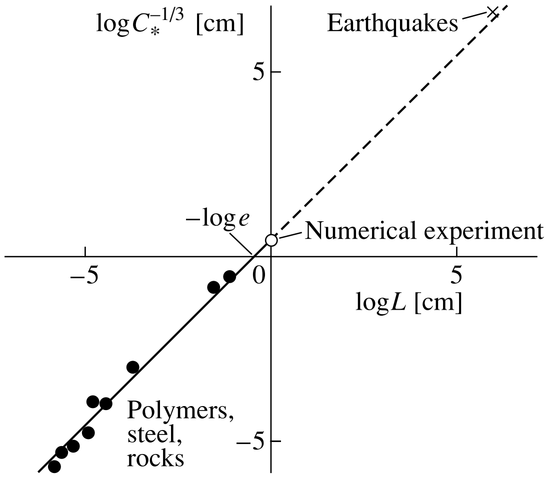


Fig. 1. Dependence of the average distance $l = c^{-1/3}$ between cracks (c is the crack concentration prior to fracture) on their characteristic length L for various-scale objects [9].

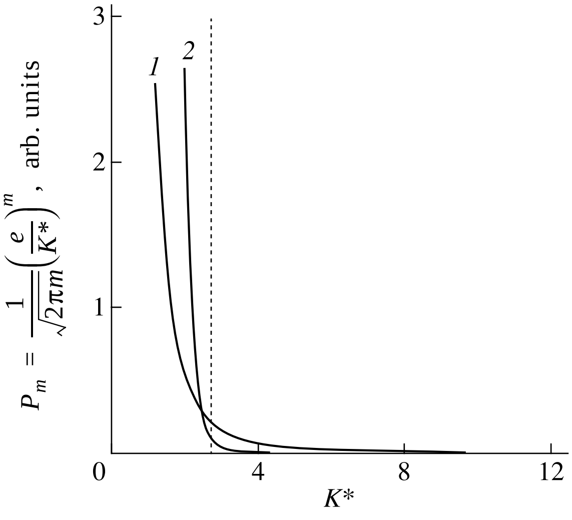


Fig. 2. Dependence of the crack-clustering probability on K^* for m equal to (*1*) 3 and (*2*) 10.

and seismic activity (the Caucasus, Turkmenistan, southern California, northeast and southwest China, Kirghizia, Greece, western Turkey, and the Kuriles). This made it possible to estimate the forecasting capabilities of the crack concentration parameter as an earthquake precursor.

Testing showed that the values of the parameter $K^*_{\rm sf}$ differ in various seismically active regions (see alarm level values in Table 1). In some regions, this difference reached a factor of 8–10. In [13–16], the effect of nonuniformity of the spatial distribution of seismogenic fractures was studied in a seismically active layer and explanations were provided for the observed dependence of the crack concentration parameter on the fracture scale (laboratory experiments with shock samples, rock impacts in mines, earthquakes) and for the differences between various seismically active regions (Table 1). To this end, the concept of the fractality of seismic processes was involved and it was shown that $K^*_{\rm sf}$ depends on the fractal dimension of the spatial distribution of seismogenic fractures. In this case, $K^*_{{\rm sf}L_2} = K^*_{{\rm sf}L_1}(L_2/L_1)^{1-\frac{d}{r}}$, where $K^*_{{\rm sf}L_2}$ and $K^*_{{\rm sf}L_1}$ are the seismogenic fracture density parameters in cells with characteristic sizes L_2 and L_1, respectively; d is the fractal dimension of the distribution of seismogenic fractures (earthquakes); and r is the spatial dimensionality (r = 1, 2, 3).

In [17, 18], when analyzing the spatiotemporal distributions of the crack concentration parameter $K_{\rm sf}$ based on data from the Caucasus earthquake catalog, the fracture transitions from the lower (n – 1)th scale level to the intermediate nth level and then to the higher (n + 1)th level were discovered to occur prior to the catastrophic Spitak earthquake of 1988. This effect was previously described for laboratory experiments [5, 19]. The result obtained was confirmed by the data from the Turkmenistan and Kamchatka earthquake catalogs. The data presented show that, based on the concentration criterion for fracture transition to the next scale level, the scheme of laboratory object fracture under a load is also applicable to seismic processes in the Earth's crust.

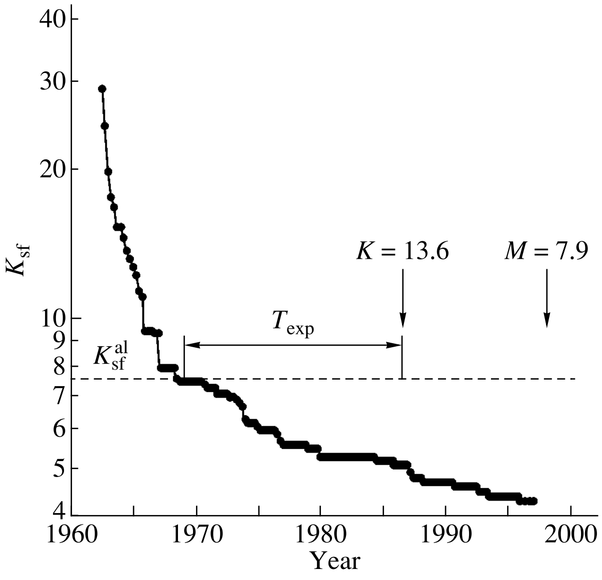


Fig. 3. Behavior of the parameter $K_{\rm sf}$ in a cell 100 × 100 km in size, whose center coincides with the Kronotsky earthquake epicenter (Kamchatka, December 5, 1997).

We note that the parameter $K_{\rm sf}$ is also useful for studying the preparation and development of the Great Fissure Tolbachik Eruption in 1975 on the Kamchatka

Table 1. Retrospective statistical characteristics of the parameter K_{sf} in various seismically active regions

	Kamchatka	Caucasus	Kirghizia	Turkmenistan	Southern California	Northeast China
Seismically active layer thickness, km	100	60	60	50	20	30
Observation period	1962–2000	1962–1990	1962–1992	1956–1991	1932–1989	1970–1996
Lower threshold of magnitude M_{tg} or class K_{tg} of forecasted earthquakes	$K_{tg} = 13.5$	$K_{tg} = 12.5$	$K_{tg} = 13.5$	$K_{tg} = 12.5$	$M_{tg} = 5.5$	$M_{tg} = 5.5$
Alarm level K_{sf}^{al}	7.5	12.9	11.4	11.8	5.5	60.2
Detection probability $P(K_{sf}\|D_1)$	0.6351	0.3711	0.5197	0.5441	0.5071	0.3738
False alarm probability $P(K_{sf}\|D_2)$	0.0526	0.0733	0.0691	0.0647	0.1069	0.1751
Average expectation time T_{exp}, years	5.3 ± 5.2	2.7 ± 2.8	4.8 ± 3.4	3.7 ± 3.4	6.6 ± 6.9	2.8 ± 1.8
Average expectation area S_{exp}, km^2	5000 ± 765	3994 ± 1249	4554 ± 1063	4421 ± 1261	4439 ± 1232	3889 ± 1314
Predictive efficiency J	9.52	4.51	5.96	6.84	4.06	2.11

	Southwest China	Greece		Western Turkey	Kuriles
		GL AU catalog	IG NOA catalog		
Seismically active layer thickness, km	30	50	50	50	100
Observation period	1970–1993	1964–1996	1964–2003	1975–2000	1962–2001
Lower threshold of magnitude M_{tg} or class K_{tg} of forecasted earthquakes	$M_{tg} = 5.5$	$M_{tg} = 5.5$	$M_{tg} = 5.5$	$M_{tg} = 5.5$	$K_{tg} = 13.5$
Alarm level K_{sf}^{al}	53.1	5.7	5.5	7.6	9.8
Detection probability $P(K_{sf}\|D_1)$	0.4812	0.3437	0.3991	0.4535	0.4970
False alarm probability $P(K_{sf}\|D_2)$	0.1345	0.0988	0.1235	0.1491	0.0898
Average expectation time T_{exp}, years	3.3 ± 2.6	4.3 ± 3.8	7.7 ± 6.5	3.9 ± 3.6	2.9 ± 3.2
Average expectation area S_{exp}, km^2	4127 ± 1217	4231 ± 1177	4161 ± 1286	4420 ± 1418	4269 ± 1238
Predictive efficiency J	3.35	3.29	2.97	2.98	5.04

Note: GL AU is the Geophysical Laboratory, Aristotle University (Thessaloniki); IG NOA is the Institute of Geodynamics, National Observatory of Athens.

Peninsula [20]. The Northern Breakthrough Eruption and the depression in the Tolbachik Crater Flat arose in regions with decreased K_{sf} values. In [21], it was shown that the fracture concentration criterion can be used to forecast rock impacts, whose energy releases are intermediate between those of earthquakes and of laboratory macroscopic fractures, and a forecasting scheme was proposed. Recently, the parameter K_{sf} has been widely used to monitor seismicity during mining in mines and shafts in the western Urals [22, 23] and the Kirovsky mine of the Apatit Company (Kola Peninsula) [24]. According to [25], K_{sf} was the most helpful parameter in forecasting large-scale seismicity in mining regions.

Thus, over the past 25–30 years, it has been demonstrated that the concepts of fracture kinetics and the concentration criterion based on them can be used to forecast earthquakes, rock impacts in mines, and volcanic eruptions. In this study, we analyze recent data on the use of the crack concentration parameter as a precursor of strong earthquakes.

2. MODEL OF THE BEHAVIOR OF THE PARAMETER K_{sf} PRIOR TO AN EARTHQUAKE

For the conditions of a seismically active region, the crack (seismogenic fracture) concentration parameter K_{sf} is the ratio of the average distance between seismogenic fractures that have taken place in a certain seismically active volume V_0 over a time ΔT to their average length:

$$K_{\mathrm{sf}} = \frac{R_{\mathrm{avr}}}{l_{\mathrm{avr}}} = \frac{\mu^{-1/3}}{l_{\mathrm{avr}}}, \tag{2}$$

where $\mu = N/V_0$ is the volume density (concentration) of fractures, which is determined by the earthquakes that have occurred; $l_{\mathrm{avr}} = (1/N)\sum_j l_j$ is the fracture length averaged over a crack ensemble; N is the total number of earthquakes belonging to energy classes in the range $[K_{\mathrm{min}}, K_{\mathrm{max}}]$ that occurred in an elementary seismically active volume V_0 over a time ΔT; and l_j is the length of a single seismogenic fracture, which can be estimated using the formula

$$\log l_j = aK_j + c. \tag{3}$$

Here, K_j characterizes the energy class (or magnitude) of the earthquake. In various seismically active regions, the coefficients a and c can take on various values. Regional relations of the type of Eq. (3) were used wherever possible. Based on experience in using the parameter K_{sf} in seismically active regions of the world, it should be noted that the most appropriate coefficients were chosen by Riznichenko [26]: $a = 0.244$ and $c = -2.266$ for the energy classes and $a = 0.440$ and $c = -1.289$ for the magnitudes.

Seismically active regions to be analyzed were partitioned into elementary seismically active volumes V_i with dimensions ΔX and ΔY (in kilometers) and depth ΔH (in kilometers). In each elementary volume, the crack concentration parameter was calculated using formula (2). To smooth possible errors in determining the earthquake hypocenters involved in K_{sf} calculations, adjacent seismically active volumes were taken to be half-overlapping [12].

It should be noted that the choice of the initial instant of time T_0 is important in calculating K_{sf}, since the seismogenic fracture concentration parameter is cumulative in terms of its physical nature and is a threshold quantity. When studying the behavior of the crack concentration parameter under laboratory conditions, it is reasonable to take T_0 as the instant of onset of sample loading. For actual conditions of a seismically active region, we choose T_0 using the concept of a seismic cycle presented in [27]. According to this concept, each seismically active volume V_i (or group of volumes) passes through its seismic cycle beginning with the stage of accumulated tectonic deformations and ending with the stage of stress release due to the occurrence of an earthquake accompanied by aftershocks (or probably without them). At the last stage, the medium returns to the initial state, at which seismicity is absent or insignificant. The seismic cycle time depends on the energy class K completing the earthquake cycle and can be estimated from the formula (see [28])

$$\log T_{\mathrm{year}} = \frac{1}{3}K - 3.5. \tag{4}$$

According to Eq. (4), the cycle duration, e.g., for earthquakes with $K \geq 12.5$, is $T \geq 5$ years. The time of the beginning of a seismic cycle for each seismically active volume V_i is taken as T_0 in calculating the spatial distribution of K_{sf}.

If in a time ΔT_i at least one strong earthquake accompanied by aftershocks occurs in a seismically active volume V_i, then further calculations of K_{sf} for this volume begin from the instant the aftershock period terminates. In this case, the influence of a strong earthquake on the seismic rate in nearby seismically active volumes is monitored. If the seismic rate exceeds its average long-term level N_L by a quantity greater than the standard deviation σ_{N_L}, then the volume under consideration is included in the aftershock region and K_{sf} for this volume is also calculated after the aftershock period has passed. The end of this period is identified from the instant the seismic rate reaches the long-term (background) level.

The results of analyzing the spatiotemporal distributions of K_{sf} and experience from retrospective forecasting [4] made it possible to propose the following scheme for practical forecasting of strong earthquakes.

(i) Distributions of K_{sf} are calculated in steps Δt for preliminarily chosen sizes of elementary seismically active volumes V_i, a range of energy classes, and depths of earthquake hypocenters.

(ii) If $K_{\mathrm{sf}} \leq K_{\mathrm{sf}}^{\mathrm{al}}$ for a volume V_i (where $K_{\mathrm{sf}}^{\mathrm{al}}$ is the alarm level determined experimentally on the basis of an analysis of retrospective data), the expectation period of a strong seismic event is announced. The critical level $K_{\mathrm{sf}}^{\mathrm{al}}$ depends on the seismic activity of the region under study, the size of the chosen seismically active volume V_i, and the earthquake energy classes used to calculate K_{sf}.

(iii) The expectation state in the volume V_i under consideration is retained for the time interval $\tilde{T}_{\mathrm{exp}}$ + $|\sigma_T|$ determined experimentally from retrospective data. This state can be relieved after the end of a strong earthquake whose hypocenter was in this seismically active volume; the same can be done for nearby volumes when the aftershock period begins.

For the volume where the earthquake has occurred, the alarm is considered actual and the time interval from the instant of the expectation state announcement (the onset of the anomaly) to the instant of the earthquake is the expectation time. In the neighbor cells, outside the hypocenter, the alarm is considered false (since the forecasted event has not formally occurred in these cells) and the time elapsed from the onset of the anomaly to the beginning of the aftershock period is the time of false alarms. For all elementary cells related to the source region of the strong earthquake, the next seismic cycle and, hence, time counting begin after the termination of the aftershock period.

(iv) If a strong earthquake does not occur in the volume V_i for a time $\tilde{T}_{\text{exp}} + |\sigma_T|$, this time is considered the time of false alarms.

The typical time dependence of K_{sf} before a strong earthquake is shown in Fig. 3. As time passes, the number of earthquakes that occur in a given volume increases and the average distance between them decreases; hence, K_{sf} is lowered. In contrast to the parameter N, which is simply indicative of the number of earthquakes, the parameter K_{sf} takes into account their energy. Therefore, if a seismic process in an elementary volume is accompanied by the generation of high-energy events, then the K_{sf} decrease rate will be higher and prerequisites for a strong earthquake will be created more rapidly.

3. RESULTS OF STUDYING K_{sf} IN SEISMICALLY ACTIVE REGIONS AND DISCUSSION

Let us define the basic parameters used in this study. We note that a detailed description of the technique of their calculation can be found in [12].

The detection probability $P(K_{\text{sf}}|D_1)$ is the conditional probability of a strong earthquake occurring when the value of the seismogenic fracture concentration parameter K_{sf} becomes anomalous in a given spatial cell. The detection probability is defined as the ratio of the number of strong earthquakes prior to which the parameter K_{sf} was anomalous (the number N_{tg} of predicted strong earthquakes) to the total number N_{tot} of strong earthquakes that occurred within the observation time T_{obs} on the observation area S_{obs}. The false-alarm probability $P(K_{\text{sf}}|D_2)$ is the conditional probability of a strong earthquake not occurring when the parameter K_{sf} was anomalous.

The expectation (alarm) time T_{exp} is the time elapsed from the onset of an anomaly (the instant the value of a prognostic factor becomes higher than a specified alarm level) to the instant of the occurrence of a strong earthquake. If a strong earthquake does not occur, this time is considered the false alarm time T_{al}. The average expectation (false-alarm) time is defined as $T_{\text{exp.avr(fa.avr)}} = (1/n)\sum_{i=1}^{n} T_{\text{exp(fa)}}$, where n is the number of anomalies of the prognostic factors.

The expectation area S_{exp} of a strong earthquake is the sum of the areas of the $\Delta X \times \Delta Y$ spatial cells in which the expectation state is announced, i.e., in which anomalous values of a prognostic factor are observed. If a strong earthquake does not occur in these cells, their area is included in the false alarm area S_{fa}. The average area of expectation (false alarm) is given by $S_{\text{exp.avr(fa.avr)}} = (1/n)\sum_{i=1}^{n} S_{\text{exp(fa)}}$.

The precursor predictive efficiency J is defined as the ratio of the flux density of strong earthquakes during the expectation time to their density during the observation time:

$$J = \frac{N_{\text{tg}}/T_{\text{exp.tot}}}{N_{\text{tot}}/T_{\text{obs}}},$$

where $T_{\text{exp.tot}}$ is the total time of expectation (alarms). In the case of random guessing, when the density of forecasted earthquakes during alarms is equal to their density over the observation time, we have $J = 1$.

The distributions of the K_{sf} values were calculated in half-overlapping rectangular cells 50 × 50 km in size. The cells covered the depth of a seismically active layer in which there occurred no less than 90% of all earthquakes detected by a network of seismic stations in the corresponding seismically active region (Table 1). The nonuniformity of the seismic-rate distribution in the seismically active layer volume was disregarded. Earthquakes with $K_{\text{tg}} \geq 12.5$ ($M_{\text{tg}} \geq 5.5$) were chosen as forecast objects. The time window increased in steps Δt = 3 months. Thus, sets of maps of the distributions of the seismogenic fracture density parameter K_{sf}, characterizing the state of a seismic process at the instant T_i, were plotted for each seismically active region. These sets were analyzed to determine the retrospective statistical characteristics: the detection probability, the false-alarm probability, and the predictive efficiency. As an example, Fig. 4 shows maps of the K_{sf} values for Kamchatka as of January 1, 1993 (Fig. 4a), and January 1, 2001 (Fig. 4b), calculated using the data from the regional earthquake catalogs for 1962–1992 and 1992–2000, respectively. The maps include the epicenters of all earthquakes with $K_{\text{sf}} \geq 13.5$ that occurred over the five subsequent years (see the average expectation time for Kamchatka in Table 1). It follows from Fig. 4 that most of the strong earthquakes occurred in regions with decreased values of $K_{\text{sf}} \leq 7.5$. From the comparison of the maps (Fig. 4), we can see that the configuration and position of anomalous zones significantly changed over the eight years, which apparently reflects changes in the seismic conditions in the region. At the same time, the succession in both the configuration and position of anomalous regions is evident.

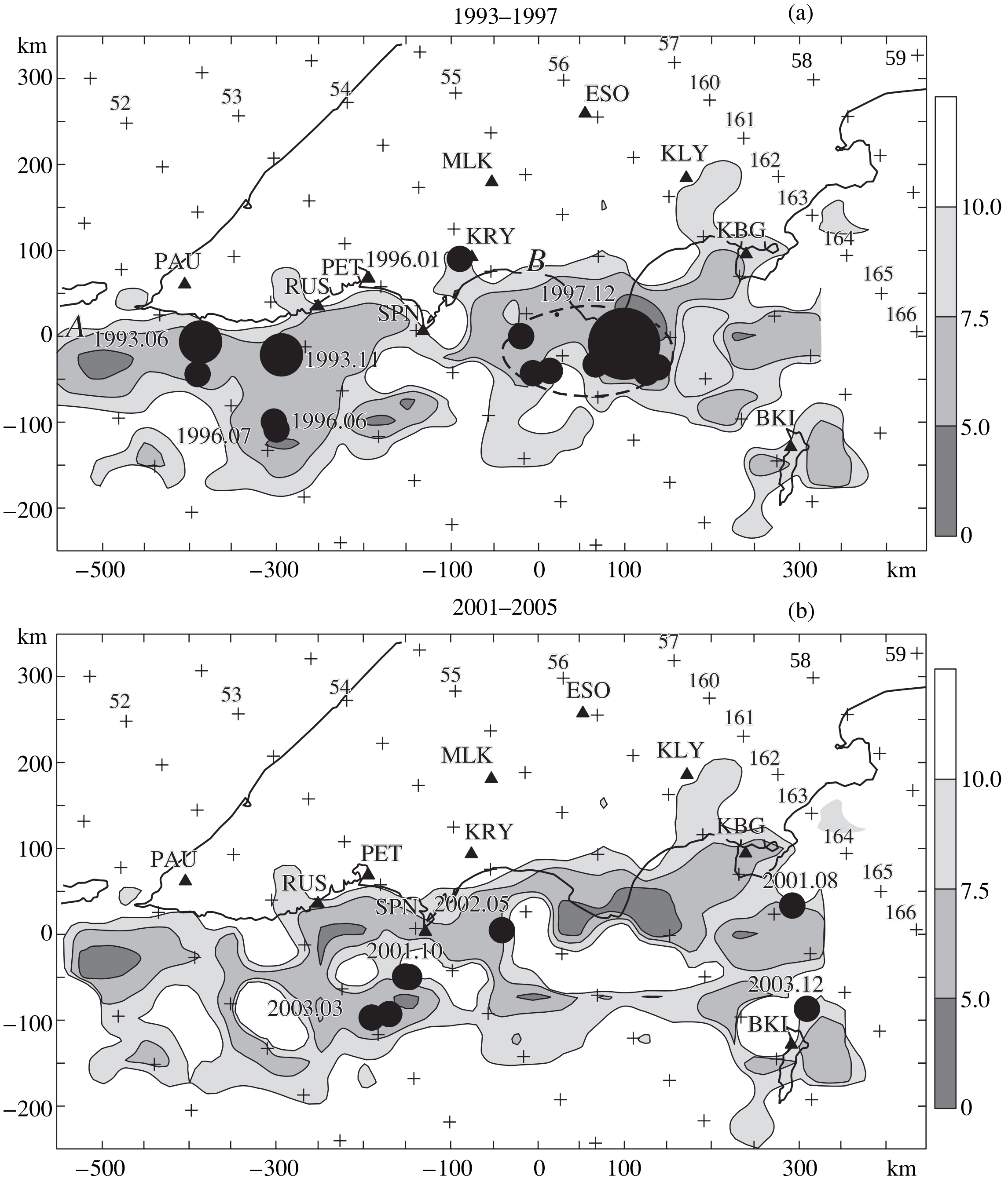


Fig. 4. Maps of the K_{sf} values for Kamchatka as of (a) January 1, 1993, and (b) January 1, 2001, calculated in cells 50 × 50 km in size according to the data on earthquakes with $K \geq 9.5$. Closed circles denote all earthquakes with $K_{tg} \geq 13.5$ that occurred in the region over the subsequent five years. The circle size is proportional to the fracture length in the source of the corresponding earthquake [26] on the map scale. The earthquake year and month are indicated above the circles. The ellipse restricts the region of aftershocks of the Kronotsky earthquake (December 5, 1997, $M = 7.9$), detected in first three days after the main shock. The geographical coordinate grid is denoted by the sign +. The scale of the K_{sf} levels is indicated on the right.

The results of analyzing K_{sf} maps for all seismically active regions are shown in Table 1. Above all, the significant deviations in the alarm levels for northeast and southwest China are worth noting. These deviations can be caused by two factors [13–16]: (i) the neglected effect of the nonuniform seismic-rate distribution over the seismically active layer thickness, with the result that aseismic regions of elementary cells V_0 are involved in the calculation of K_{sf} or (ii) inaccuracy of the correlation relations between the magnitude and the

Table 2. Retrospective statistical characteristics of the parameter K_{sf} for cells of various sizes ΔS (data for western Turkey)

	ΔS, km		
	50×50	75×75	100×100
Alarm level	7.6	7.6	8.0
Detection probability $P(K_{sf}\|D_1)$	0.4535	0.3671	0.5200
False alarm probability $P(K_{sf}\|D_2)$	0.1491	0.1526	0.1670
Expectation time T_{exp}, years	3.9 ± 3.6	4.3 ± 4.2	3.7 ± 2.7
Expectation area S_{exp}, km^2	4420 ± 1418	9141 ± 3312	17679 ± 5672
Predictive efficiency J	2.98	2.32	2.89

Table 3. Average efficiencies of predictive indications

Predictive indication		J
Seismogenic fracture density K_{sf}		4.60 ± 2.15
Reproducibility curve slope		2.21 ± 0.69
Number of earthquakes in the form of	quiescences N_q	3.05 ± 0.97
	activations N_a	4.44 ± 2.11
Seismic energy released in the form of	quiescences E_q	1.61 ± 1.04
	activations E_a	1.87 ± 0.85

fracture length in the source (especially in the range of low magnitudes), which significantly differ from the world average relations used for all other regions. We note that there are no physical reasons for these deviations. In the other regions considered, the values of the alarm level K_{sf}^{al} ranged from 5.5 to 12.9.

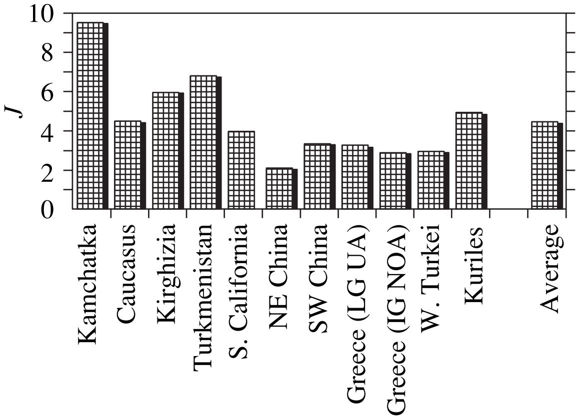


Fig. 5. Efficiency J of the precursor K_{sf} over regions.

In Table 1, the retrospective data on K_{sf} in Greece were obtained from two different versions of earthquake catalogs for this region, prepared at the Geophysical Laboratory, Aristotle University (Thessaloniki) (GL AU), and the Institute of Geodynamics, National Observatory of Athens (IG NOA). The parameters for calculating K_{sf} were identical in both cases. We can see that the values of the alarm levels, detection and false alarm probabilities, and predictive efficiencies are rather close. This counts in favor of the fact that the seismogenic fracture parameter indeed reflects the features of the regional seismic process. However, it should be noted that the expectation times differ by almost two times.

Similar results were obtained based on the data from the catalog of earthquakes in western Turkey, where the parameter K_{sf} and its retrospective statistical characteristics were calculated in cells of various sizes (Table 2). Here, the alarm levels and predictive efficiencies are also almost identical, which indicates a weak sensitivity of K_{sf} to the sizes of scanned cells. However, the detection probabilities differ by a factor of 1.5. We note that the results were obtained for only one region of western Turkey and were not examined for others.

We now turn to the last line of Table 1, i.e., the data on the predictive efficiency of K_{sf} for various seismically active regions. The average value of this parameter over all regions is $J_{K_{sf}} = 4.6 \pm 2.15$ (Fig. 5). A comparison with the average efficiency of other seismological precursors according to the data from [8] (Table 3) suggests that the parameter K_{sf} is currently the most efficient predictive criterion. In this case, 34–63% of strong earthquakes took place in regions where $K_{sf} \le \tilde{K}_{sf}^{al}$ during expectation periods of 2.7–7.7 years, which means that the parameter K_{sf} is a medium-term predictive indication. Long-term experience in using the parameter K_{sf} as an earthquake precursor and the data in Table 1 allow us to conclude that it is unlikely to significantly decrease the expectation time T_{exp} in the future or to make K_{sf} a short-term precursor.

Testing of the parameter K_{sf} in various seismically active regions showed not only its above-mentioned positive features. There are also negative properties. Let us consider one of them. In the maps shown in Fig. 4, there are several anomalous zones with different areas in which the parameter K_{sf} is below the alarm level $\tilde{K}_{sf}^{al} \le 7.5$. The total area of these anomalous regions is $S_{exp} = 20.1\%$ of the observation area where the seismic activity is higher than one earthquake per year. It is reasonable to expect a strong earthquake with probability $P(K_{sf}|D_1) = 63.5\%$ in each of these zones for an expectation time $T_{exp} = 5.3$ years (see Table 1). However, expected earthquakes did not occur in all of the anomalous regions by the end of the expectation period. They took place only in regions A (1993 and 1996) and B

(1997), i.e., in the largest areas. In smaller anomalous regions, strong earthquakes were not observed. It seems impossible to determine in advance in which region the next earthquake will occur. For this, parameters other than seismological precursor parameters also need to be used.

4. CONCLUSIONS

The concentration criterion for solid fracture developed by Zhurkov in the 1950s–1960s had a significant effect on the establishment and development of earthquake source physics and the processes resulting in it. The algorithm developed for applying the parameter K_{sf} for forecasting strong earthquakes and its testing in various seismically active regions has demonstrated its great potential to forecast the position, time, and magnitude of an earthquake (in comparison with other seismological precursors). The results obtained make it possible to recommend the use of this algorithm simultaneously with other precursors. Maps of the K_{sf} values can provide a basis for administrative decisions to prepare industrial and social infrastructures in territories within anomalous zones for strong earthquakes in order to minimize possible damage.

ACKNOWLEDGMENTS

The author would like to thank his colleagues G.A. Sobolev and V.B. Smirnov for their long-term helpful collaboration and advice during discussions of the manuscript, which culminated in significant improvements. Special gratitude is expressed to V.S. Kuksenko, whose proposal initiated the publication of this work.

This study was supported by the federal program of support for leading scientific schools (project no. NSh-1270.2003.5) and the International Science and Technology Center (project no. 1745).

REFERENCES

1. S. N. Zhurkov, V. S. Kuksenko, V. A. Petrov, V. N. Savel’ev, and U. S. Sultanov, Izv. Akad. Nauk SSSR, Ser. Geogr. Geofiz. **6**, 11 (1977).
2. G. A. Sobolev and A. D. Zavyalov, Dokl. Akad. Nauk SSSR **252** (1), 861 (1980) [Sov. Phys. Dokl. **25** (1), 69 (1980)].
3. V. S. Kuksenko, V. A. Pikulin, S. Kh. Negmatullaev, and K. M. Mirzoev, *Earthquake Forecast* (Donish, Dushanbe, 1984), Vol. 5, p. 139 [in Russian].
4. A. D. Zavyalov, Vulkanol. Seismol. **3**, 58 (1986).
5. V. S. Kuksenko, *Earthquake Forecast* (Donish, Dushanbe, 1984), Vol. 4, p. 8 [in Russian].
6. D. I. Frolov, *Earthquake Forecast* (Donish, Dushanbe, 1984), Vol. 4, p. 46 [in Russian].
7. G. A. Sobolev and A. D. Zavyalov, *Earthquake Forecast* (Donish, Dushanbe, 1984), Vol. 5, p. 160 [in Russian].
8. A. D. Zavyalov, Fiz. Zemli **4**, 16 (2002).
9. S. N. Zhurkov, V. S. Kuksenko, and V. A. Petrov, *Science’s Future: Intern. Year Book* (Znanie, Moscow, 1983), p. 100 [in Russian].
10. V. A. Petrov, Fiz. Tverd. Tela (Leningrad) **21**, 620 (1979) [Sov. Phys. Solid State **21**, 368 (1979)].
11. V. A. Petrov, *Earthquake Forecast* (Donish, Dushanbe, 1984), Vol. 5, p. 30 [in Russian].
12. A. D. Zavyalov, L. B. Slavina, V. Yu. Vasil’ev, and V. V. Myachkin, *Design Procedure of Prospective Earthquake Maps by Comples of Forecast Indications* (OIFZ RAN, Moscow, 1955) [in Russian].
13. V. B. Smirnov and A. D. Zavyalov, Vulkanol. Seismol. **4**, 75 (1996).
14. V. B. Smirnov and A. D. Zavyalov, in *Seismology in Europe. Papers Presented at the XXV General Assembly ESC* (Reykjavik, 1996), p. 276.
15. V. S. Smirnov and A. D. Zavyalov, Vulkanol. Seismol. **18**, 447 (1997).
16. A. D. Zavyalov and V. B. Smirnov, in *Proceedings of Second All-Russian Conference on Geophysics and Mathematics, Per’m, Russia, 2001* (GI UrO RAN, Perm’, 2001), p. 103.
17. A. D. Zavyalov and Yu. V. Nikitin, Vulkanol. Seismol. **1**, 65 (1997).
18. A. D. Zavyalov and Yu. V. Nikitin, Vulkanol. Seismol. **19**, 79 (1997).
19. S. N. Zhurkov, V. S. Kuksenko, V. A. Petrov, V. N. Savel’ev, and U. S. Sultanov, in *Proceedings of Physical Processes in Earthquake Centers* (Nauka, Moscow, 1980), p. 78.
20. V. I. Gorel’chik and A. D. Zavyalov, Vulkanol. Seismol. **6**, 60 (1986).
21. A. Yu. Gor, V. S. Kuksenko, N. G. Tomlin, and D. I. Frolov, FTPRPI **3**, 54 (1989).
22. A. A. Malovichko, R. A. Dyagilev, and D. Yu. Shulakov, in *Proceedings of International Coference VNIMI on Mountain Geophysics* (VNIMI, St. Petersburg, 1998), p. 147.
23. A. A. Malovichko, R. A. Dyagilev, D. Yu. Shulakov, P. G. Butyrin, and Yu. V. Baranov, in *Proceedings of Second All-Russian Coference on Geophysics and Mathematics*, Ed. by V. N. Strakhov (GI UrO RAN, Per’m, 2001), p. 367.
24. *Seismicity under Mountain Works*, Ed. by N. N. Mel’nikov (KNTS RAN, Apatity, 2002), p. 326 [in Russian].
25. R. A. Dyagilev, Abstract of Candidate’s Dissertation (Moscow, 2002).
26. Yu. V. Riznichenko, *Investigation on Earthquake Physics* (Nauka, Moscow, 1976), p. 9 [in Russian].
27. S. A. Fedotov, *Seismic Zoning of USSR* (Nauka, Moscow, 1968), p. 121 [in Russian].
28. M. A. Sadovskiĭ, L. G. Bolkhovitinov, and V. F. Pisarenko, *Deformation of Geophysical Medium and Seismic Process* (Nauka, Moscow, 1980) [in Russian].

Translated by A. Kazantsev

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1042–1047. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1009–1014.
Original Russian Text Copyright © 2005 by Steblov, Frolov, Kuksenko.

PROCEEDINGS OF THE CONFERENCE DEDICATED TO S. N. ZHURKOV (1905–1997)

Kinematics of the Earth Continental Drift

G. M. Steblov*, D. I. Frolov, and V. S. Kuksenko****
** Gamburtsev Institute of Physics of the Earth, Russian Academy of Sciences,*
ul. Bol'shaya Gruzinskaya 10, Moscow, 123810 Russia
*** Ioffe Physicotechnical Institute, Russian Academy of Sciences,*
Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia
e-mail: victor.kuksenko@mail.ioffe.ru

Abstract—The drift of regions in northeast Eurasia was studied using the precise satellite geodesy method (Global Positioning System (GPS)). Observational data were analyzed invoking the International Database. Despite the rather short observations (~6 years), the drift directions of the basic Eurasia and Northern America regions were detected and the drift velocities were estimated. The results of the study are discussed in general terms of the deformation and fracture of solids. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

This paper is devoted to analyzing the current data on the drift of certain regions of the Earth's crust. The fact that giant continents are in continuous motion can be explained using two fundamental concepts: (i) Continental drifts are caused by convective processes in deep mantle layers. (ii) Tectonic plates move with certain velocities, and these local velocities can be significantly higher at the plate boundaries. This factor makes it possible to consider this process in terms of kinetic concepts and invoke broad experience from the deformation kinetics of solids.

The first statement is actively discussed in the scientific literature, while the second concept still remains obscure, since the available data are insufficient to solve this physical problem. We interpret the available original results in terms of kinetic concepts on the fracture and deformation of solids, of which the Earth's crust is an example.

The transition from the microlevel to the macroscopic level in processes of deformation and fracture has been confirmed in many experiments [1, 2]. The experiments described in this paper, despite the short duration of observations, show that some notions are general and can initiate studies of deformation processes on larger scales.

2. STATEMENT OF THE PROBLEM

The ideas of lithospheric plate tectonics were developed in the 1960s and were presented as a fundamental concept in 1968 by Morgan, as well as by McKenzie and Le Pichon. The concept that the structure of the upper layers of the Earth's surface forms a mosaic set of solid fragments (plates) that are in continuous relative movement was an evident consequence of the study of heat convection in the mantle and of physicochemical analysis of the interaction between the crust and the mantle. These mechanisms made it possible to explain the dynamics of plate movement, since this motion requires energy sources sufficient to be a driving force of such large-scale phenomena. The current so-called "geological" models of plate movement were developed based on analysis of three types of data: stripe magnetic anomalies in oceans, azimuths of transform faults in axial regions of Mid-Oceanic ridges, and azimuths of displacement vectors over ruptures during earthquakes in subduction zones. This collection of data was used to determine the vectors of relative rotation of plates averaged over the past three million years.

There are several serious questions regarding the geological models of plates.

(i) Can the velocities averaged over three million years be considered to be contemporary?

(ii) What is the actual accuracy of geological models? This question arise from the fact that the models are based on information on only oceanic regions.

Direct measurements using modern geodetic methods allow construction of an objective model of current movements of the Earth's crust.

This study is aimed at detecting and correcting the poorly studied boundaries between the Eurasian and North American tectonic plates and estimating the velocities of their relative movement.

3. EXPERIMENT DESIGN

The development of satellite geodesy methods has made an important and often decisive contribution to modern investigations of geodynamic processes. Improvement of satellite technologies made it possible to measure long distances on the Earth's surface with a high accuracy unattainable with classical geodesy techniques (triangulation, optical range finding, leveling, etc.). These methods allow rather accurate ground-based measurements only in an extremely narrow range

1063-7834/05/4706-1042$26.00 © 2005 Pleiades Publishing, Inc.

of distances (generally, within the line-of-sight range). These circumstances have not allowed study of large-scale geodynamic processes, because errors are accumulated during unbending and leveling of geodetic networks. Determining the distances between points on the Earth's surface as long as thousands of kilometers with an accuracy of 1–2 mm made it possible for the first time to study the variation in the relative position of remote fragments of the Earth's surface with time over several years and to detect their relative movement characterized by an average velocity of several tens of millimeters per year.

The Global Positioning System (GPS) used in this study makes it possible to determine the position of an object on the Earth's surface with an accuracy of up to several millimeters. Initially (in the 1980s), GPS was developed as an off-line satellite navigation system. Its application is based on the measurement of distances from ground-based stations to active satellites emitting specially shaped signals. These distances are determined by measuring the time it takes for an emitted signal to travel from the satellite to a ground-based receiver. To determine coordinates in three-dimensional space, it is theoretically sufficient to determine the distances to three satellites; however, taking into account clock errors, in fact four satellites should be simultaneously observed.

Currently, the core of the system is a "constellation" of 24 satellites. GPS satellites function as a carrying platform for various types of equipment, including radio transmission devices, atomic frequency standard, computers, etc., that make it possible to determine coordinates both on the Earth's surface and in the air or in circumterrestrial space. Each satellite generates a signal of special shape carrying information that enables one to identify the satellite and its position at any point in time.

In the late 1980s, researchers in many geodynamic and geodetic organizations came to the conclusion that GPS technology can be efficiently employed in scientific investigations, such as the study of earthquakes, movements of faults in the Earth's crust, and deformations of the crust as a whole, as well as in many other applications. The basic idea of the organization of the International GPS Service (IGS) was to develop a global network whose points are arranged closely enough for mutual referencing and to cover all land on the Earth. With this configuration, the global network forms a geocentric coordinate system, which makes it possible to exactly calculate the geocentric coordinates of the points under study on the basis of the precise coordinates of the reference points due to simultaneous processing of the data measured at points on the areas under study and the reference points of the global network. Referencing of local and regional grounds of interest to the reference points forming the geocentric coordinate system underlay future studies of the deformation of the Earth's crust on the global scale.

Before 1997, the entire northern region of Eurasia was monitored by GPS points arranged only in its European region. This fact lead to significant difficulties for geodetic measurements over the entire territory of East -Siberia and the Far East, with the consequence that ill-founded conclusions were made from geodynamic analysis of these regions. Since 1997, on the basis of seismological stations under the Geophysical Service, Russian Academy of Sciences, we have created a reference network of GPS stations [3] over the entire territory of North Eurasia—the North Eurasia Deformation Array (NEDA), covering the continent from the East European Platform, through the Ural ridge and all Siberia, to Chukotka, Kamchatka, and Sakhalin. Recently, NEDA stations have been involved in regional GPS projects that were performed and improved with the development of the NEDA in the regions of strongest seismic activity at the boundaries of the three largest stable interacting fragments of the Earth's crust: the Eurasian (EUR), North American (NAM), and Pacific (PAC) segments. Within regional studies, we have carried out [4, 5] long-term measurements (i) at the poorly explored boundary between Eurasia and Northern America in the region of the Chersky Ridge, (ii) in the subduction zone of seismic and volcanic activity on the Kamchatka Peninsula, and (iii) in the region of complex deformation and high seismic activity caused by the convergence of Eurasia and North America on Sakhalin Island.

4. MEASUREMENT TECHNIQUE

Since 1997, the NEDA network has provided continuous GPS measurements in the vast expanse of North Eurasia. All NEDA stations are classified by the IGS service as global stations; i.e., the measurements performed at them are processed daily by no fewer than three international centers. The transmission of diurnal measurements is fully automated, with the delay not exceeding 1 h. Primary information is collected and transmitted using computers connected to receivers.

Files with measured data are transmitted from the station computers to the central server at the Geophysical Center, Russian Academy of Sciences.

The data analyzed are transmitted from the central computer of the Geophysical Center to global data centers of the International Service, where final processing of the data on the entire planet is performed.

The results of the data analysis and the solutions of problems obtained become available for regional centers (and, in particular, for the central computer of the Geophysical Center) via the Internet.

The schematic diagram of data transmission by the NEDA network is shown in Fig. 1.

An analysis of long-term series of observations over the global network, including regional segments, permits one to obtain (with sufficient confidence) the field of relative velocities of individual points on the Earth's

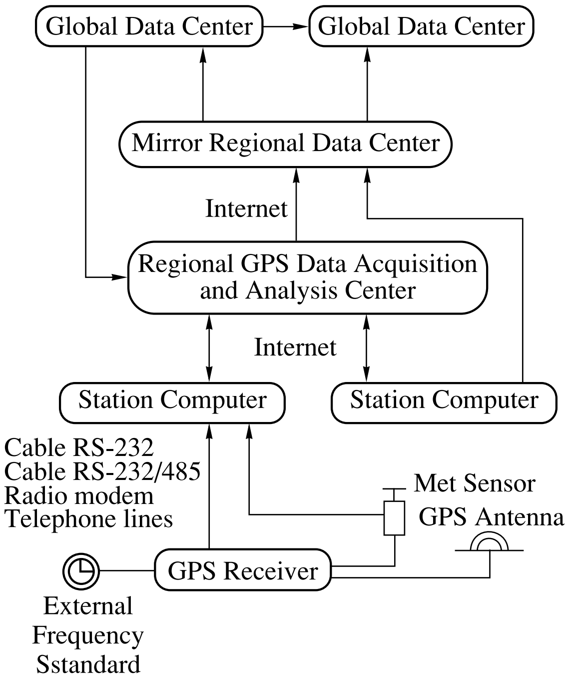


Fig. 1. Schematic diagram of transmission of the data obtained by the NEDA network.

surface. In this case, the general pattern of relative movements of the segments of the Earth's crust is controlled by the field of horizontal velocities, which are determined by projecting the relative velocities of these points onto the Earth's surface. It is very important that, by analyzing this field, one can separate stable nondeforming fragments of the Earth's surface and estimate their relative movement. In this case, it is also important to determine the degree of localization of the deformable boundary regions of these fragments, namely, to elucidate whether the fragment boundaries form narrow deformation belts or deformations extend to inner regions of assumed platforms, i.e., to determine the degree of justification of the conventional concepts on the stability of ancient platforms.

5. MEASUREMENT RESULTS

An analysis of long-time series of measurements made it possible to determine which reference stations relate to various nondeforming segments of the Earth's crust. As an example, Fig. 2 shows stable stations of the Eurasian plate. It should be noted that the velocities of relative movement of reference stations with respect to one plate do not exceed 0.5 mm/yr. For this reason,

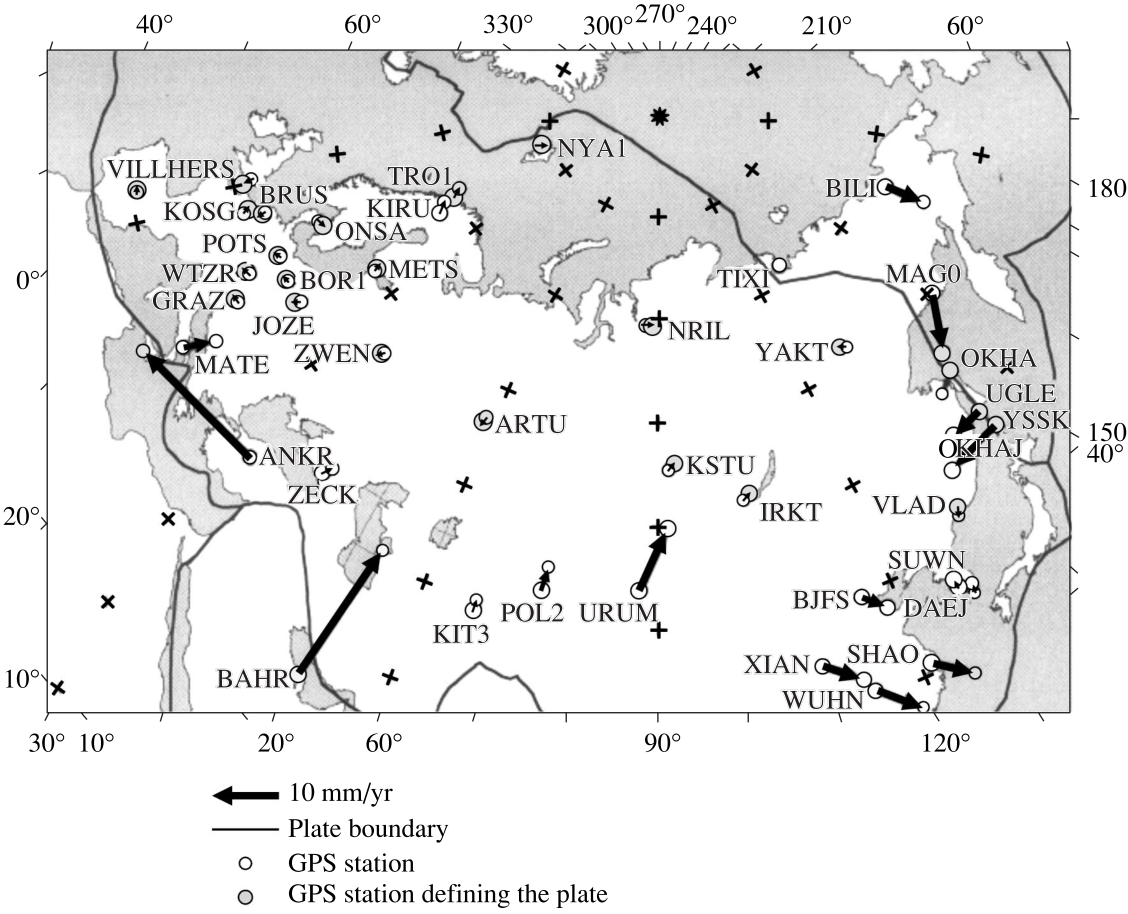


Fig. 2. GPS stations monitoring the Eurasian tectonic plate.

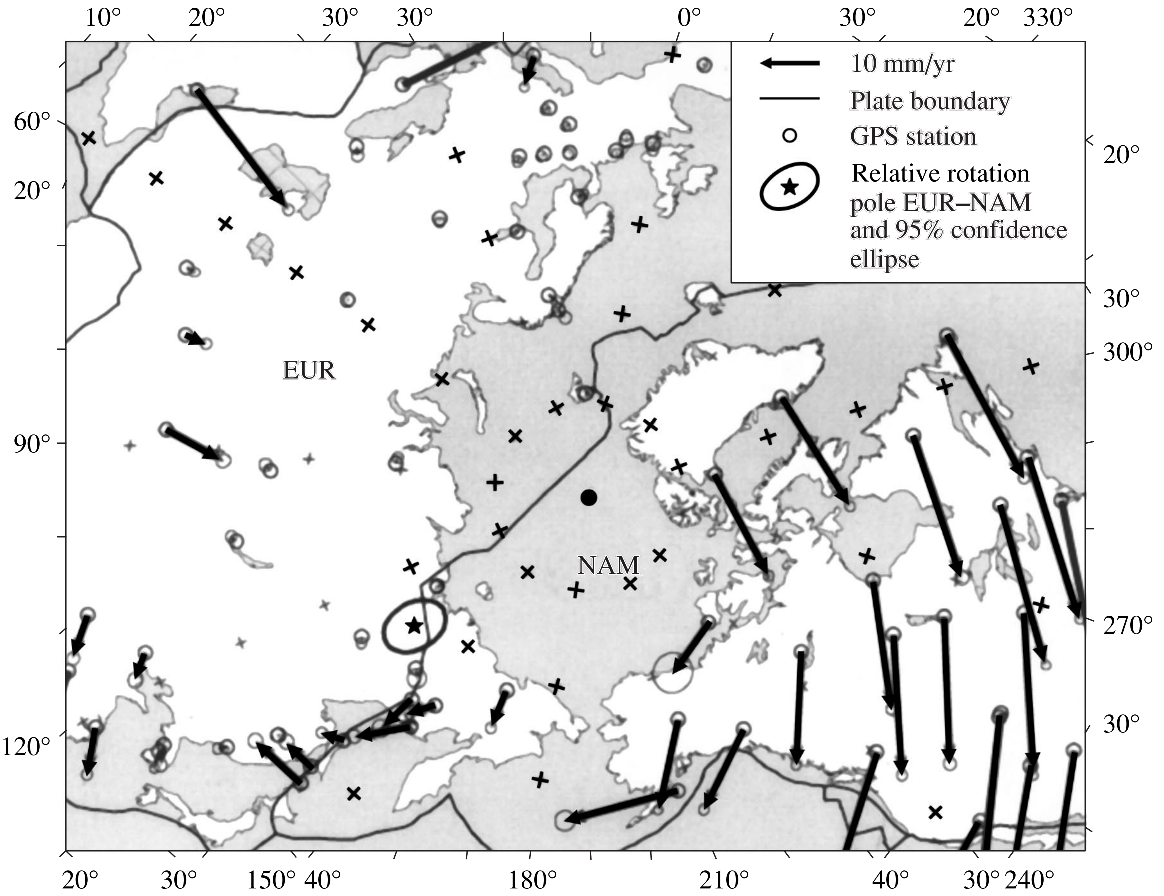


Fig. 3. Global movement of the North American tectonic plate (NAM) with respect to the Eurasian plate (EUR) according to the GPS data.

these land regions can be considered to be nondeformable segments of the Earth's crust.

The determination of the system of reference stations made it possible to construct a relative coordinate system based on the three largest plates: the Eurasian, North American, and Pacific Ocean plates. This three-plate relative coordinate system (including the vectors of relative rotation of plates as key elements) formed the basis for a geodynamic analysis of their interaction in the three regions mentioned above.

The justified choice of the relative coordinate system made it possible to estimate the position of the pole of relative rotation of the Eurasian and North American plates. This pole is indicated by an asterisk in Fig. 3. An analysis of the GPS data revealed the relative movement of these plates. As an example, Fig. 3 shows the global clockwise movement of the entire North American plate with respect to Eurasia. The angular velocity of rotation is ~0.2° per million years. Eurasia rotates counterclockwise with respect to North America. Studies of the deformation zone configuration [5–7] at the plate boundaries become possible due to detailed regional measurements. These measurements were carried out with allowance for the local conditions: either at steadily positioned stations of regional grounds or at field points of periodic observation at an interval of one year.

Regional measurements were taken into account in the combined analysis of the data obtained (i) at permanent stations of the NEDA array of the Geophysical Center (since 1997), (ii) at permanent IGS stations in this region, and (iii) from field measurements carried out repeatedly at an interval of one year (since 1995).

To analyze the kinematics of tectonic plates in northeast Asia, let us compare their velocities with respect to Eurasia and North America determined from the GPS data [5, 8].

The velocities with respect to the EUR (Fig. 4a) are less than 2 ± 1 mm/yr for stations arranged to the west of the Chersky Ridge. The velocities of stations on the east side of the Chersky Ridge, Chukotka, and northwest Kamchatka are significantly higher, 3–8 mm/yr, which indicates clockwise rotation of this region with respect to Eurasia. The drift of stations in the Chersky Ridge region indicates the possible effect of distributed deformation at the EUR–NAM boundary. However, the convergence velocity here is low due to the small distance to the pole of their mutual rotation; hence, the

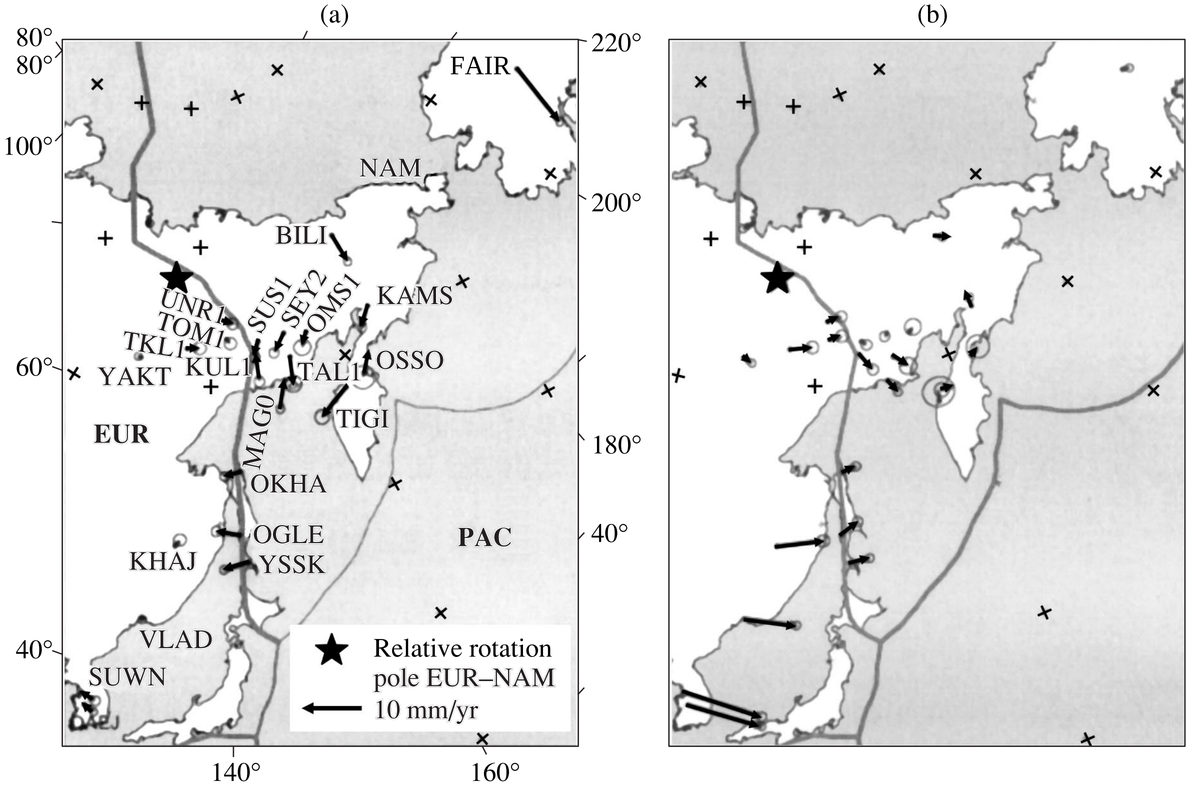


Fig. 4. Relative velocities of GPS stations in Northeast Asia, measured with respect to (a) Eurasia and (b) North America. The stations are remote from the subduction zone in Kamchatka.

effect of stress accumulation is also rather weak and can be disregarded even at distances greater than 100 km from the boundary.

The velocities with respect to North America (Fig. 4b) are less than 2 ± 1 mm/yr on Chukotka and in the northwest Kamchatka region. Among the six stations on the east side of the Chersky Ridge, the three most northern (SUS1, SEY2, OMS1) move the slowest (at a velocity of 1 mm/yr), while the three others (MAG0, KUL1, TAL1) move at a velocity of 2–3 mm/yr. The velocities on the western side of the Chersky Ridge are 2–11 mm/yr, which indicates counterclockwise rotation and corresponds to the vector of EUR–NAM relative rotation. From the comparison of these velocities, it follows that the western branch of the seismic belt of the Chersky Ridge is the northeast boundary of Eurasia. The velocity of the TIXI station (Fig. 2) near the Arctic boundary of Siberia with respect to Eurasia is lower than that with respect to North America; hence, in this region, the EUR–NAM plates move mostly to the east of the TIXI.

6. RESULTS AND DISCUSSION

The main experimental results obtained in this study are as follows.

A long-term experiment to develop a system of permanent GPS stations combined with seismic stations of the Russian Academy of Sciences has been theoretically justified and implemented for North Eurasia. The system is fully automated and makes it possible to solve a number of fundamental scientific and practical problems, such as (i) improving the accuracy of determination of GPS satellite orbits, which is very significant for precision real-time navigation, and (ii) creating a basis for geodynamic investigations over the entire territory of North Eurasia, primarily in seismically active regions, based on observational regional networks.

A reference frame has been constructed, which is based on the three largest plates: the Eurasian, North American, and Pacific Ocean plates. The pole of relative rotation of the Eurasian and North American plates was determined. This reference frame formed a basis for geodynamic analysis of the interaction of the plates in the Far East region.

Let us indicate the most important results of the analysis.

(i) The present velocities of relative movement of plates have been determined. These velocities were shown to differ significantly from those given by the geological model of plates.

(ii) It has been ascertained for the first time that the western branch of the seismic belt of the Chersky Ridge is the northeast boundary of Eurasia and that the

Chukotka and Kamchatka peninsulas make up a single whole with the North American continent.

We now attempt to interpret the results in terms of the physical kinetics of deformation and fracture of solids and consider the possibility of extending the regularities in the deformation of solids detected for small samples under laboratory conditions to global objects under natural deformation conditions.

This generalization is based on the concept of the similarity of some physical phenomena at various scale levels. For fracture processes, this similarity has been confirmed experimentally [2]. The features of the crack formation, accumulation, and propagation are similar for various scale levels. The concentration criterion for the fracture process to switch from a lower hierarchical level to a higher level is valid both for laboratory samples with initial crack sizes of ~10–100 nm and for cracks in the Earth's crust that are several kilometers long, which cause earthquakes. These problems are discussed in detail in [9, 10]. The relation between fracture and deformation processes is also well known.

The structural self-similarity of solids is also observed in a wide range of scales from nanostructures to lithospheric plates [1]. We believe that the kinetic concepts are universal for various scale levels and can be used to analyze continental plate movement. Since the plate velocities are directly measured in these studies, we can write a deformation kinetics equation and estimate the parameters of interest. Unfortunately, the short duration of these measurements and low deformation rates complicate this analysis. However, the high measurement accuracy makes this technique promising.

We also note prospects for plate movement simulation under laboratory conditions with selection of certain materials and loading conditions, which will make it possible to estimate the stresses at the boundaries of moving plates and to forecast the effect of their movement.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, projects no. 04-05-65287 and 03-05-6483.

REFERENCES

1. M. A. Sadovskiĭ, Dokl. Akad. Nauk SSSR **247** (4), 329 (1979) [Sov. Phys. Dokl. **24**, 586 (1979)].
2. S. N. Zhurkov and V. S. Kuksenko, Izv. Akad. Nauk SSSR, Ser. Geogr. Geofiz. **6**, 11 (1977).
3. M. G. Kogan, G. M. Steblov, R. W. King, T. A. Herring, D. I. Frolov, S. G. Egorov, V. Ye. Levin, A. Lerner-Lam, and A. Jones, Geophys. Res. Lett. **27** (14), 2041 (2000).
4. G. M. Steblov, M. G. Kogan, R. W. King, D. I. Frolov, S. A. Mironov, and Yu. K. Rykhalsky, Suppl. EOS. AGU Fall Meeting Abstracts **79** (45), 218 (1998).
5. G. M. Steblov, M. G. Kogan, R. W. King, C. H. Scholz, R. Bürgmann, and D. I. Frolov, Geophys. Res. Lett. **30** (18), 1924 (2003).
6. M. G. Kogan, A. Lerner-Lam, T. Koczynski, S. V. Gavrilov, O. N. Galaganov, D. I. Frolov, G. M. Steblov, and R. W. King, AGU Fall Meeting Abstracts **77** (46), 149 (1996).
7. M. G. Kogan, R. Bürgmann, N. F. Vasilenko, C. H. Scholz, R. W. King, A. I. Ivashchenko, D. I. Frolov, G. M. Steblov, Ch. U. Kim, and S. G. Egorov, Geophys. Res. Lett. **30** (3), 1102 (2003).
8. M. G. Kogan, R. W. King, G. M. Steblov, A. Lerner-Lam, and V. E. Levin, Suppl. EOS. AGU Fall Meeting Abstracts **79** (45), 218 (1998).
9. V. S. Kuksenko, Fiz. Tverd. Tela (St. Petersburg) **47** (5), 788 (2005) [Phys. Solid State **47** (5), 812 (2005)].
10. A. D. Zavyalov, Fiz. Tverd. Tela (St. Petersburg) **47** (6), 1000 (2005) [Phys. Solid State **47** (6), 1034 (2005)].

Translated by A. Kazantsev

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1048–1052. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1015–1018.
Original Russian Text Copyright © 2005 by Wolf, Chuburin.

Specific Features of Low-Energy Electron Scattering by Thin Films of Cubic Crystals

G. V. Wolf and Yu. P. Chuburin
Physicotechnical Institute, Ural Division, Russian Academy of Sciences, ul. Kirova 132, Izhevsk, 416001 Russia
e-mail: wolf@otf.fti.udmurtia.su

Received June 22, 2004; in final form, September 29, 2004

Abstract—The specific features of the electron states in thin films of cubic crystals are considered in the energy range above the vacuum zero of the crystal potential. It is demonstrated that bands of bound states embedded in the energy continuum can exist along particular directions of the two-dimensional Brillouin zone. These bands can substantially affect the intensity of low-energy electron scattering. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

In our previous works [1–3], we considered the specific features of scattering of low-energy electrons ($E <$ 10–20 eV) by a planar crystalline film. These features are associated with the existence of both bound electron states with energies somewhat lower than the boundary of the continuous spectrum and quasi-stationary states with energies higher than this boundary. The reflected intensity, in the former case, is determined by a crystal analog of the Wigner formula for low-energy scattering and, in the latter case, by the Breit–Wigner formula describing the resonance at a quasi-discrete level [1–5].

In this paper, it will be demonstrated that bound electron states can exist along particular directions of the two-dimensional Brillouin zone of a cubic-crystal film when the energy bands adjacent to the boundary of the continuous spectrum are characterized by a specific dispersion. These states are localized above the boundary of the continuous spectrum bounded by an energy paraboloid $E = \mathbf{k}^2$, i.e., above the first diffraction threshold.[1] The existence of bound electron states leads to a number of important consequences and offers strong possibilities for controlling the intensity of the reflected electrons by lowering the symmetry of the crystal under specified actions.

2. THEORETICAL TREATMENT

Let us consider a Lippmann–Schwinger equation describing the bound and quasi-stationary states of electrons of a crystalline film:

$$\Psi_{\mathbf{k}}(\mathbf{r}, E) = -\int_{\Omega} G_{\mathbf{k}}^{(0)}(\mathbf{r}, \mathbf{r}'; E) V(\mathbf{r}') \Psi_{\mathbf{k}}(\mathbf{r}', E) d\mathbf{r}', \quad (1)$$

where Ω stands for a unit cell infinite along the z direction and $V(\mathbf{r})$ is the effective single-particle potential, which is invariant with respect to the transformations of the symmetry group of the film under consideration. It is assumed that, at large distances from the surface of the film, i.e., at $|z| > z_0$, this potential becomes zero.[2] The Green's function of free electrons of the crystalline film can be represented in the form

$$\begin{aligned} G_{\mathbf{k}}^{(0)}(\mathbf{r}, \mathbf{r}'; E) &= \frac{1}{2\pi S} \sum_{\mu} \int_{-\infty}^{+\infty} \frac{\exp[i(\mathbf{k} + \mathbf{K}_{\mu}, \lambda)(\mathbf{r} - \mathbf{r}')]}{(\mathbf{k} + \mathbf{K}_{\mu})^2 + \lambda^2 - E} d\lambda \\ &= \frac{-1}{S} \sum_{\mu} \frac{\exp[i(\mathbf{k} + \mathbf{K}_{\mu}, \sqrt{E - (\mathbf{k} + \mathbf{K}_{\mu})^2})(\mathbf{u} - \mathbf{u}', |z - z'|)]}{2i\sqrt{E - (\mathbf{k} + \mathbf{K}_{\mu})^2}}. \end{aligned} \quad (2)$$

Here, S is the area of the cross section of the unit cell Ω by the surface of the crystal, $\mathbf{u}(\mathbf{u}')$ is the component of the vector $\mathbf{r}(\mathbf{r}')$ aligned parallel to the surface, $\mathbf{k}$ is the reduced two-dimensional quasi-momentum, and E is the electron energy. The summation is taken with respect to the vectors of the reciprocal lattice of the film.

---

[1] We use the atomic system of units with energy expressed in Ry.

[2] This asymptotic cutoff of the potential is quite reasonable from the physical standpoint and allows one to simplify significantly the theoretical treatment without changing the nature of the results [4].

Separating out the term with $\mathbf{K}_\mu = 0$ in relationship (2), we can rewrite expression (1) for $z > z_0$ in the form

$$\Psi_\mathbf{k}(\mathbf{r}, E) = e^{i\mathbf{ku}} \frac{\exp[-z\sqrt{k^2 - E}]}{2S\sqrt{k^2 - E}}$$

$$\times \int_\Omega e^{-i\mathbf{ku'}} e^{\sqrt{k^2 - E}z'} V(\mathbf{r'})\Psi_\mathbf{k}(\mathbf{r'}, E)d\mathbf{r'} \tag{3}$$

$$+ \sum_\mu e^{i(\mathbf{k}+\mathbf{K}_\mu)\mathbf{u}} \frac{\exp[-z\sqrt{(\mathbf{k}+\mathbf{K}_\mu)^2 - E}]}{2S\sqrt{(\mathbf{k}+\mathbf{K}_\mu)^2 - E}}$$

$$\times \int_\Omega e^{-i(\mathbf{k}+\mathbf{K}_\mu)\mathbf{u'}} e^{z\sqrt{(\mathbf{k}+\mathbf{K}_\mu)^2 - E}} V(\mathbf{r'})\Psi_\mathbf{k}(\mathbf{r'}, E)d\mathbf{r'}.$$

In essence, expression (3) is a plane-wave expansion of the function $\Psi_\mathbf{k}(\mathbf{r}, E)$. It follows from expression (3) that, in the general case, when $E > \mathbf{k}^2$, the first term is a nondecreasing function of z and the state $\Psi_\mathbf{k}(\mathbf{r}, E)$ ceases to be bound with respect to z in the discrete spectrum. However, under the condition

$$\int_\Omega e^{-i\mathbf{ku'}} e^{\pm z'\sqrt{k^2 - E}} V(\mathbf{r'})\Psi_\mathbf{k}(\mathbf{r'}, E)d\mathbf{r'} = 0, \tag{4}$$

the function $\Psi_\mathbf{k}$ remains a function quadratically integrable in Ω even in the region lying above the first threshold $\Psi_\mathbf{k}(\mathbf{r}, E)$. In this case, the states bound with respect to z can exist above the energy boundary $E = \mathbf{k}^2$ of the continuous spectrum.

This situation can occur with films of cubic crystals. Let us consider, for example, the specific case of a crystalline film with a (001) face-centered cubic lattice. For this film, the Brillouin zone is a square with sides $\sqrt{8}\pi/A$, where A is the lattice constant of the film. Next, we consider a state with wave vector $\mathbf{k}_\Delta$ aligned parallel to the direction Δ of the two-dimensional Brillouin zone ($\mathbf{k}_\Delta = \sqrt{2}\pi/A(\xi, 0, 0)$, $0 < \xi < 1$). According to the Wigner theorem [6], the eigenfunctions $\Psi_\mathbf{k}(\mathbf{r}, E)$ can be transformed by the irreducible representations of the group of wave vector $\mathbf{k}$. The irreducible representations of the group of vector $\mathbf{k}_\Delta$, which is isomorphic to the group C_{2v} [5], are presented in the table. For the function $\Psi_\mathbf{k}^{\Delta_3}(\mathbf{r}, E)$ transformed through the representation Δ_3, integral (4) for $E > \mathbf{k}^2$ can be written in the form

$$\int_\Omega e^{-i\mathbf{ku}}[\cos(z\sqrt{k^2 - E}) \pm i\sin(z\sqrt{k^2 - E})]$$

$$\times V(\mathbf{r})\Psi_\mathbf{k}^{\Delta_3}(\mathbf{r}, E)d\mathbf{r}. \tag{5}$$

Irreducible representations of the group of vector $\mathbf{k}_\Delta$

$\Delta(C_{2v})$	E	C_2	σ	σC_2
Δ_1	1	1	1	1
Δ_2	1	–1	–1	1
Δ_3	1	1	–1	–1
Δ_4	1	–1	1	–1

Note: E is the identical transformation, C_2 is the rotation through the angle π about the [100] axis, and σ is the reflection in the plane $z = 0$.

It is easy to see that the function $e^{-i\mathbf{ku}}\cos(z\sqrt{k^2 - E})$ is transformed by the representation Δ_1 and that the function $e^{-i\mathbf{ku}}\sin(z\sqrt{k^2 - E})$ is transformed through the representation Δ_2. The above integral vanishes as a consequence of the well-known selection rule for matrix elements of the operators of scalar quantities. This implies that bound states of the above symmetry can exist in the energy continuum.[3]

Let us now analyze the scattering of electrons by this film. The wave function of a scattered electron with a momentum $\mathbf{p} = (\mathbf{k}, \sqrt{E - \mathbf{k}^2})$ satisfies the Lippmann–Schwinger inhomogeneous equation

$$\Psi(\mathbf{r}, \mathbf{p}) = e^{i\mathbf{pr}} - \int_\Omega G^{(0)}(\mathbf{r}, \mathbf{r'}; \mathbf{p})V(\mathbf{r'})\Psi(\mathbf{r'}, \mathbf{p})d\mathbf{r'}. \tag{6}$$

Taking into account type (2) of the function $G^{(0)}(\mathbf{r}, \mathbf{r'}; \mathbf{p})$ for the forward-scattering amplitude [7], we derive the expression[4]

$$a^+(\mathbf{p}) = 1 - \frac{1}{2iS\sqrt{E_R(\mathbf{k}) - \mathbf{k}^2}} \int_\Omega e^{-i\mathbf{pr}} V(\mathbf{r})\Psi(\mathbf{r}, \mathbf{p})d\mathbf{r}. \tag{7}$$

On the other hand, the equation determining the state of electron scattering can be written in the form

$$\Psi(\mathbf{r}, \mathbf{p}) = e^{i\mathbf{pr}} - \int_\Omega G(\mathbf{r}, \mathbf{r'}; \mathbf{p})V(\mathbf{r'})e^{i\mathbf{pr'}}d\mathbf{r'}. \tag{8}$$

Here, $G(\mathbf{r}, \mathbf{r'}; \mathbf{p})$ is the single-electron Green's function that corresponds to the Hamiltonian of the film: $\hat{H} = -\Delta + V(\mathbf{r})$.

---

[3] The same statement is true for the states transformed through the representation Δ_4. Similarly, it can be shown that bands of bound states with energies $E > \mathbf{k}^2$ can exist along the direction Σ [$k_\Sigma = \frac{\sqrt{2}\pi}{A}(\xi, \xi, 0)$, $(1 > \xi > 0)$] of the Brillouin zone of the crystalline film with a (001) face-centered cubic lattice.

[4] In the case of low-energy states localized below the second diffraction threshold [$\mathbf{k}^2 < E < (\mathbf{k} + \mathbf{K}_\mu)^2$], there exist only one reflected electron beam and one transmitted electron beam.

The film under consideration is assumed to be thin enough for the separation between the adjacent quantum-well subbands of bound states, which undergo crowding with an increase in the film thickness, to be relatively large. Hence, if the energies of states of the discrete spectrum are close to the energy of incident electrons, it follows from the theory of self-adjoint operators that the function $G(\mathbf{r}, \mathbf{r}'; \mathbf{p})$ can be represented in the single-band form [4]:

$$G(\mathbf{r}, \mathbf{r}'; \mathbf{p}) = \frac{\Psi(\mathbf{r}, \mathbf{p}_0)\Psi^*(\mathbf{r}', \mathbf{p}_0)}{E_0(\mathbf{k}) - E} + \tilde{G}(\mathbf{r}, \mathbf{r}'; \mathbf{p}), \quad (9)$$

where $\mathbf{p}_0 = (\mathbf{k}, \sqrt{E_0(\mathbf{k}) - \mathbf{k}^2})$ and $\Psi(\mathbf{r}, \mathbf{p}_0)$ is the eigenfunction of the Hamiltonian $\hat{H}$ that corresponds to the energy $E_0(\mathbf{k})$ of the bound state. The function $\tilde{G}(\mathbf{r}, \mathbf{r}'; \mathbf{p})$ does not have poles in the vicinity of the energy of incident electrons.

In contrast to the well-known energy bands located below the boundary of the continuous spectrum, the bands of bound electron states with energies $E_n(\mathbf{k})$ higher than the energy boundary $E = \mathbf{k}^2$ of the continuum spectrum do not form surfaces in the $\mathbf{k}$ space. For $E_n(\mathbf{k}) > \mathbf{k}^2$, the wave vector $\mathbf{k}$ deviates from the direction along which there can exist states that are normalized in the cell of the film and described by a wave function exponentially decreasing away from the surface of the film toward vacuum ($|z| \longrightarrow \infty$). This deviation transfers the aforementioned states to resonance states with a finite lifetime. The wave function of these resonance states increases for $|z| \longrightarrow \infty$ [5, 8]. The existence of these resonances and their continuous transition to bound states (if they exist) follow from the analytical $\mathbf{k}$ dependence of the energies of nondegenerate electron states corresponding to the solutions of Eq. (1). Therefore, if bands of bound states localized above the energy boundary of the continuous spectrum arise along particular directions of the two-dimensional Brillouin zone, there exists a surface of quasi-stationary states,

$$E_R(\mathbf{k}) = \mathscr{E}_R(\mathbf{k}) - i\Gamma(\mathbf{k}), \quad \Gamma(\mathbf{r}) > 0, \quad (10)$$

which involves these bands; that is,

$$\begin{gathered} E_R(\mathbf{k}_\Delta) = \mathscr{E}_R(\mathbf{k}_\Delta) = E_0(\mathbf{k}_\Delta), \\ \Gamma(\mathbf{k}_\Delta) = 0. \end{gathered} \quad (11)$$

Taking into account the analyticity of the functions $E_0(\mathbf{k})$ and $\Psi(\mathbf{r}, \mathbf{p}_0)$ with respect to the wave vector $\mathbf{k}$, from relationship (9) for the resonance states in the vicinity of the energy boundary $E_0(\mathbf{k}_\Delta)$, we derive the following expression:

$$G_R(\mathbf{r}, \mathbf{r}'; \mathbf{p}) = \frac{\Psi(\mathbf{r}, \mathbf{p}_R)\Psi^*(\mathbf{r}', \mathbf{p}_R)}{E_R(\mathbf{k}) - E} + \tilde{G}_R(\mathbf{r}, \mathbf{r}'; \mathbf{p}), \quad (12)$$

where $\mathbf{p}_R = (\mathbf{k}, \sqrt{E_R(\mathbf{k}) - \mathbf{k}^2})$ and $\Psi(\mathbf{r}, \mathbf{p}_R)$ is the wave function of the resonance state, which is a solution to Eq. (1) for the complex energies $E_R(\mathbf{k})$. The function $\tilde{G}_R(\mathbf{r}, \mathbf{r}'; \mathbf{p})$ does not have poles in the vicinity of the surface $E = E_R(\mathbf{k})$.

From relationships (12), (8), and (7) without regard for the nonsingular contributions near the surface $E_R(\mathbf{k})$, we obtain

$$a^+(\mathbf{p}) \approx -\frac{\left|\int\limits_\Omega e^{i\mathbf{pr}} V(\mathbf{r})\Psi^*(\mathbf{r}, \mathbf{p}_R)d\mathbf{r}\right|^2}{2iS\sqrt{E_R(\mathbf{k}) - \mathbf{k}^2}[E_R(\mathbf{k}) - E]}. \quad (13)$$

The momentum of incident electrons can be given by

$$\mathbf{p} = (\mathbf{k}, \sqrt{E - \mathbf{k}^2}) = \mathbf{p}_{\Delta 0} + \delta\mathbf{p}, \quad (14)$$

where

$$\mathbf{p}_{\Delta 0} = (k_{\Delta 0}, 0, \sqrt{E_0(\mathbf{k}_{\Delta 0}) - \mathbf{k}_{\Delta 0}^2}), \quad (15)$$

and

$$\delta\mathbf{p} = (0, \delta k_y, 0). \quad (16)$$

The deviation δk_y is assumed to be small. In the expansion of the integrand in relationship (13), we use an approximation linear in δk_y and obtain the following expression for the forward-scattering amplitude:

$$a^+(\mathbf{p}) \approx \frac{\delta k_y^2 |W(\mathbf{p}_{\Delta 0})|^2}{2iS\sqrt{E_R(\mathbf{k}) - \mathbf{k}^2}[E_R(\mathbf{k}) - E]}, \quad (17)$$

where

$$W(\mathbf{p}_{\Delta 0}) = \int\limits_\Omega e^{-ip_{\Delta 0z}z} V(\mathbf{r}) \frac{\partial U(\mathbf{p}_{\Delta 0}, \mathbf{r})}{\partial k_y} d\mathbf{r}, \quad (18)$$

and $U(\mathbf{p}_{\Delta 0}, \mathbf{r})$ is the periodic part of the Bloch function $\Psi(\mathbf{r}, \mathbf{p}_R) = \exp(i\mathbf{kr})U(\mathbf{p}_R, \mathbf{r})$ for the momentum $\mathbf{p}_R = \mathbf{p}_{\Delta 0}$. In this case, it is necessary to retain the contributions linear in δk_y, because, in the zeroth-order approximation, as follows from condition (4), the integrals involved in expression (13) are equal to zero and the denominator has a double zero at $\delta k_y \longrightarrow 0$.

The probability of the transmission of electrons is determined by the expression [7]

$$I^+(\mathbf{p}) = (2\pi)^3 \int |C(\mathbf{p}, \mathbf{p}')|^2 |a^+(\mathbf{p}')|^2 d\mathbf{p}', \quad (19)$$

where the function $C(\mathbf{p}, \mathbf{p}')$ describes the distribution of the momentum of incident electrons in the direction of the momentum $\mathbf{p}$. According to relationship (14), the function $C(\mathbf{p}, \mathbf{p}')$ can be considered a δ function. Taking into account the symmetry of the Brillouin zone of the crystalline film with a (001) face-centered cubic lattice, conditions (10) and (11), and the smallness of δk_y, we obtain the relationship

$$I^{+}(\mathbf{p}) \approx \frac{|W(\mathbf{p}_{\Delta 0})|^4}{16S^2[(1-M_{yy}^{-1})^2+\gamma^2]\sqrt{[p_{\Delta 0z}^2-(1-M_{yy}^{-1})\delta k_y^2]^2+\gamma^2\delta k_y^4}}, \tag{20}$$

where M_{yy}^{-1} is the diagonal component of the tensor of the inverse effective mass,

$$M_{yy}^{-1} = \left(\frac{m}{m^*}\right)_{yy} = \frac{1}{2}\frac{\partial^2\mathcal{E}_R(\mathbf{k}_{\Delta 0})}{\partial k_y^2},$$
$$\gamma = \frac{1}{2}\frac{\partial^2\Gamma(\mathbf{k}_{\Delta 0})}{\partial k_y^2} = \frac{\varepsilon}{2}, \tag{21}$$

and ε is the curvature[5] of the cross section of the "resonance zone" $\Gamma = \Gamma(\mathbf{k})$ by the plane $k_x = k_{\Delta 0}$. It follows from expression (20) that, in the case when $M_{yy}^{-1} > 1$, the transmission coefficient $T(\mathbf{p}) = |a^{+}(\mathbf{p})|^2$ has a maximum at $\delta k_y = 0$. For $M_{yy}^{-1} < 1$, the transmission coefficient has a minimum at $\delta k_y = 0$ and a maximum at

$$\delta k_y = \pm p_{0z}\sqrt{(1-M_{yy}^{-1})/[(1-M_{yy}^{-1})^2+\gamma^2]}. \tag{22}$$

In the first case, we have

$$\partial^2 T(\mathbf{p}_{\Delta 0})/\partial\delta k_y^2 \sim (1-M_{yy}^{-1})/p_{\Delta 0z}^4$$

and the transmission coefficient decreases most sharply for grazing electron beams, i.e., near the point of intersection of the band of bound states with the energy boundary of the continuum spectrum. In the second case, the requirement of smallness of the quantity δk_y gives

$$p_{\Delta 0z}^2 = E_0(\mathbf{k}_{\Delta 0}) - k_{\Delta 0}^2 \ll 1 - M_{yy}^{-1}.$$

This limitation does not seem to be too severe. It can hold not only in the vicinity of the energy boundary of the continuous spectrum but also in the case of a negative effective mass and sufficiently far from it. The ratio of the transmission coefficients for the momenta $\mathbf{p}$ and $\mathbf{p}_{\Delta 0}$, which differ by the quantity δk_y determined by relationship (22), can be found from the expression

$$\frac{T(\mathbf{p})}{T(\mathbf{p}_{\Delta 0})} = \frac{\sqrt{(1-M_{yy}^{-1})^2+\gamma^2}}{\gamma} \approx \frac{2(1-M_{yy}^{-1})}{\varepsilon}. \tag{23}$$

When the curvature ε is small, this ratio can be relatively large.

Thus, the above analysis of the symmetry of electron states demonstrated that bands of bound states embedded in the energy continuum can exist along particular directions of the two-dimensional Brillouin zone in a crystalline film with a (001) face-centered cubic lattice. These states can arise in energy bands that have a specific structure and lie below the boundary of the continuum spectrum. The energy spectrum of electrons above the vacuum zero of the crystal potential should contain bands involved in the paraboloid $E = \mathbf{k}^2$ bounding the continuum spectrum. In actual fact, the existence of these bands (or at least special electron states) follows from numerous calculations of the electronic structure of crystalline films near the vacuum zero of the potential of electrons E_V. For example, according to the results obtained in [9, 10], upward directed energy bands exist in the vicinity of the vacuum zero of the potential E_V. However, what happens with these bands at $E > E_V$ has not been considered and the pattern of dispersion curves is simply cut off at some energy in the vicinity of E_V.[6]

3. CONCLUSIONS

The theoretical analysis performed above demonstrated that the bands of bound states embedded in the energy continuum can substantially affect the scattering of low-energy electrons by a very thin crystalline film. In the case where the momentum of primary electrons slightly deviates from the direction corresponding to the band of states of the discrete spectrum, the transmission coefficient reaches its maximum. For a weak dispersion of the resonance bandwidth, the above coefficient can differ significantly from the transmission coefficient in the direction of the band of bound states. This can be of interest for the modulation of electron scattering by quantum-confined film structures. By varying the quantity δk_y and measuring the ratio of the transmission coefficient $T(\mathbf{p})$ to the coefficient $T(\mathbf{p}_{\Delta 0})$, we can experimentally estimate the parameters γ and M_{yy}^{-1} of the resonance band.[7] Since within the electron–hole interaction the wave function of diffraction of slow electrons is a time-reversed wave function of the final state of the photoelectron, these parameters can be

---

[5] From symmetry considerations, $\partial\Gamma(\mathbf{k}_{\Delta 0})/\partial k_y = 0$.

[6] The dependence of the intensity of low-energy electron scattering on the dispersion of the bands of bound states localized above E_V but below the boundary of the continuum spectrum was considered in our earlier paper [2].

[7] For a particular film, the value of M_{yy}^{-1} can be calculated from the energy of quasi-stationary states according to the method proposed in [3].

used, for example, in photoelectron spectroscopy of very thin crystalline films.

ACKNOWLEDGMENTS

This work was supported by the Presidium of the Russian Academy of Sciences within the framework of the program “Low-Dimensional Quantum Structures.”

REFERENCES

1. G. V. Wolf, Yu. P. Chuburin, and L. A. Rubtsova, Poverkhnost, No. 10, 81 (1991).
2. Y. P. Chuburin and G. V. Wolf, J. Phys.: Condens. Matter **8** (6), 631 (1996).
3. G. V. Wolf, Yu. P. Chuburin, A. E. Pavlov, and L. A. Rubtsova, Poverkhnost, No. 12, 24 (1992).
4. J. C. Taylor, *Gauge Theories of Weak Interactions* (Cambridge Univ. Press, New York, 1976; Mir, Moscow, 1978).
5. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory*, 3rd ed. (Nauka, Moscow, 1974; Pergamon, Oxford, 1977).
6. H. Streitwolf, *Gruppentheorie in der Festkörperphysik* (Teubner, Leipzig, 1967; Mir, Moscow, 1971).
7. Yu. P. Chuburin, Teor. Mat. Fiz. **72** (1), 120 (1987).
8. Yu. P. Chuburin, Teor. Mat. Fiz. **110** (3), 443 (1997).
9. H. Bross and M. Kauzmann, Phys. Rev. B **51** (23), 17135 (1995).
10. H. Erschbaumer, A. J. Freeman, C. L. Fu, and R. Podloucky, Surf. Sci. **243** (1–3), 317 (1991).

Translated by O. Moskalev

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1053–1059. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1019–1025.
Original Russian Text Copyright © 2005 by Artemov.

METALS AND SUPERCONDUCTORS

Response of a Gas of Two-Dimensional Vortices in a Layered Superconductor to an AC Magnetic Field

A. N. Artemov
Donetsk Physicotechnical Institute, National Academy of Sciences of Ukraine, Donetsk, 83114 Ukraine
e-mail: artemov@kinetic.ac.donetsk.ua

Received February 4, 2004; in final form, July 1, 2004

Abstract—Experimental data concerning the response of superconducting YBCO films to a weak ac magnetic field in the vicinity of the superconducting transition point are analyzed. Earlier, these data were interpreted as evidence of either the critical behavior of magnetic vortices in a thin film or the polarization of a system of two-dimensional vortex dipoles. The frequency dependence of the response was related to the frequency dispersion of the medium. In the present paper, the effect of an ac electromagnetic field on a two-dimensional vortex gas is considered in layered superconductors above the Berezinskiĭ–Kosterlitz–Thouless transition point. Local model constitutive equations are employed in which neither frequency nor spatial dispersion is included. It is shown that the main experimentally observed features of the film response can be explained in terms of the temperature-dependent conductivity and the size effect in a finite-size sample © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The magnetic flux penetration into materials can be explored by studying the response of samples to an ac magnetic field. At present, there are a number of theories and models in which the temperature, field, and frequency dependences of the response are associated with certain physical phenomena and processes.

For normal metals, the central role is played by the normal skin effect. In [1], the response of a plane-parallel conductive plate to an electromagnetic field under symmetric boundary conditions was studied in the quasi-stationary approximation [2]. It was shown that, in this approximation, the response is determined by a single parameter d/δ, where d is the plate thickness, $\delta = c/\sqrt{2\pi\sigma\omega}$ is the skin depth, ω is the field frequency, and σ is the conductivity of the plate. The system was shown to exhibit a size effect; namely, the dissipative part of the susceptibility χ_1'' at the fundamental frequency reaches a maximum when the skin depth is about half the plate thickness, $d/\delta \approx 2$.

For type-II superconductors, one of the most useful concepts is the magnetic flux pinning, which results in the Bean [3] and Kim [4] critical-state models based on it. These models have provided a quantitative description of the magnetization and response of hard superconductors in terms of the critical current [5].

The discovery of high-temperature superconductors made it possible to conduct experiments at higher temperatures. It has been found that the metastable critical state is destroyed by thermal fluctuations. The concept of magnetic flux creep [6] was found to be useful under these conditions and made it possible to explain the characteristic frequency dependence of the response of a superconductor to an applied magnetic field [7, 8].

The ideas mentioned above lead to a qualitative understanding of the magnetization of bulk low-temperature type-II superconductors, bulk high-temperature ceramic superconductors, and high-temperature superconductor granular films at relatively low frequencies (≤1 MHz) of the applied magnetic field. In all these cases, a size effect is observed; namely, the temperature-dependent dissipative part of the susceptibility reaches a maximum at a temperature T_m where the magnetic field penetration depth becomes approximately equal to the characteristic size of the sample. The limiting frequency indicated above is rather conventional, because its actual value depends on the properties and dimensions of the superconducting samples.

At higher frequencies, it becomes observable that magnetic flux vortices penetrating a sample do not have time to reach the stationary critical state during a field cycle. In this case, it is necessary to consider not only the superconducting critical current but also normal currents, whose dissipation is related to the relaxation of the magnetic flux to the stationary critical state [9].

The theories and models mentioned above cover the main set of ideas successfully used to describe the response of a type-II superconductor to a weak ac magnetic field. However, several experiments measuring the diamagnetic response of superconducting films [10, 11] have recently been reported, the results of which do not conform with these models.

In the present paper, we briefly outline the main results of those experiments and discuss the interpretations of the results from [11] given in [11, 12] in terms of the frequency dispersion of the medium. Afterwards, we propose an alternative explanation for the experimental data based on the Berezinskiĭ–Kosterlitz–Thouless phase transition (BKT) [13, 14] in a gas of two-dimensional (2D) vortices (pancakes) in layered superconductors and on the size effect. It is assumed that the relation between the field and current is local. An argument in favor of the suggested approach is the evaluation of the frequency dependence of the temperature at which the maximum of the dissipative part of the sample response to an ac magnetic field is reached and a model calculation of the temperature dependence of the susceptibility, which agree with the experimental data.

2. DISCUSSION OF THE EXPERIMENTAL DATA

The effect of a weak ac magnetic field on an epitaxial YBCO film was studied in [10]. The variation of the temperature T_m at which the dissipative part of the response $\chi_1''(T)$ reaches a maximum at the fundamental frequency was studied as a function of the magnetic field amplitude (H_{ac} = 40–500 mOe) and frequency (ω = 1–30 kHz). The results were analyzed in terms of the critical state and magnetic flux creep. According to the Bean critical state model, the field amplitude is related to the critical current at the temperature T_m [15, 16] and the flux creep model explains the temperature T_m shift, which varies in proportion to lnω [7, 8].

Both of these features are clearly defined in the calculated dependences. However, the kink in the $T_m(H_{ac}^{2/3})$ linear dependence is left unexplained. In our opinion, this kink indicates that the nature of the dissipation of magnetic field energy in the sample changes at a temperature just below the superconducting transition temperature. The sample response is nonlinear both below and above this temperature, which is confirmed by the dependence of T_m on the field amplitude.

In [11] (see also [17]), the temperature dependence of the response of a YBCO film to a weak (<1 mOe) ac magnetic field was studied at frequencies between 30 mHz and 800 kHz. At a temperature T_c, the nondissipative part of the response exhibits a kink, and the dissipative part of the response is nonzero only at temperatures $T > T_c$. We believe that here, as in the experiment reported in [10], the nature of the response changes. The maximum of the dissipative part of the response shifts toward higher temperatures as the external field frequency increases. It was suggested in [11] that, as the temperature approaches T_c, the correlation length of the superconductor becomes equal to the sample thickness and a thickness-dependent phase transition (similar to the BKT transition in thin films [18]) takes place. The sample response at $T > T_c$ was attributed to the critical behavior of the system of free magnetic vortices, which run through the sample and are created as a result of fluctuation-induced dissociation of neutral vortex pairs. The experimental data under discussion were explained qualitatively in [12] in terms of the phenomenological linearized theory of pancake dipoles in layered superconductors in an external field [19, 20]. In [11, 12], the frequency dependence of the response observed in [11] was attributed to the frequency dispersion of the medium and the maximum of the dissipative part of the response was associated with the resonance absorption of the electromagnetic field energy.

The fact that there are two separate interpretations of a single experiment, neither supported by satisfactory arguments, makes both of them suspicious. Additional modifications of the theory are necessary in order to explain the nonlinear response observed in [10] (the authors of [11, 12] are apparently unaware of that paper). Moreover, we believe that the physical factors responsible for the sample response in the models discussed in the papers mentioned above are not actually dominant.

We suggest that the change in the nature of the sample response in the experiments reported in [10, 11] is due to the BKT phase transition [13, 14] in the pancake system in layered superconductors, that the frequency dependence of the response is due to the viscous movement of free pancakes, and that the maximum of the dissipative part of the response is due to a size effect similar to that described by Fisher and Kao [1].

The BKT transition is a reconstruction of the vortex system state in a layered superconductor occurring at a temperature T_{BKT}. At low temperatures $T < T_{BKT}$, the magnetization of the film is determined by vortices that enter the sample and run through its thickness. These vortices are a bound state of pancakes localized in different layers of the superconductor [21]. The vortices can be pinned by structural defects of the sample, thereby causing a nonzero value of the critical current. Magnetization of a superconductor in this state can be described in terms of the critical-state model.

As the temperature increases, the state of high-temperature YBCO superconductors changes. Free pancakes appear in a superconductor and significantly alter its properties. Though these superconductors are more anisotropic than layered because of the strong Josephson links between the layers, the transition between the two phases is governed by the same factors as is the BKT transition in 2D systems. The main factors are (i) the instability of the system of dipoles (bound pairs of pancakes) with respect to the dissociation into a free vortex gas that occurs at the temperature $T = T_{BKT}$ and (ii) the Debye screening of the free vortex interaction.

This model of the layered superconductors is based on experimental and theoretical studies of their behavior near T_{BKT}. We mention only a few of them, the most important for qualitative understanding of the physical processes that occur in these systems. In [22], the current-voltage characteristics of YBCO single crystals were measured both along the superconducting layers (in the *ab* plane) and across them (along the *c* axis). It was found that the transition temperature T_c^c from the superconducting state to the normal state along the *c* axis is lower than the temperature T_c^{ab} at which the transition to the normal state takes place in the *ab* plane. We believe that this difference means that, at the temperature T_c^c, thermal fluctuations destroy coherence between the adjacent layers and the sample turns into a set of disconnected superconducting layers. From analysis of current–voltage characteristics, it follows that T_c^c is close to and apparently coincides with T_{BKT}. This kind of behavior of layered superconductors is in good agreement with the results of numerical simulations based on the 3D anisotropic *XY* model [23].

The critical current, which is due to the tension of Josephson vortices, disappears when the coherence of the order parameter phase is destroyed. This fact was confirmed by measurements of the current–voltage characteristic of an epitaxial YBCO film presented in [24]. It was shown that the critical current of the film vanishes as a certain temperature $T_{\mathrm{BKT}} < T_c$ is approached from below, where the exponent of the current–voltage characteristic switches from 3 to 1. A similar result is reported in [11, inset to Fig. 1]. The absence of a critical current means the absence of vortex pinning.

Another characteristic feature of the behavior of layered superconductors in the vicinity of T_{BKT} is related to the penetration of magnetic flux into a sample. Above this point, the superconductor state with vortices is favored over the state without vortices even in a zero magnetic field. Therefore, the lower critical field H_{c1} is zero and there is no energy barrier to the penetration of a magnetic flux.

The results of the experiments reported in [10, 11] can be qualitatively explained in terms of the features of layered superconductors (YBCO in particular) discussed above. At low temperatures (below the BKT transition point), a superconductor can be either in the superconducting critical state (as in [10]) or in the Meissner state, as in [11]. At high temperatures ($T > T_{\mathrm{BKT}}$), the magnetization of a sample is determined by the gas of free pancakes. In this case, the energy of the external magnetic field is dissipated due to the viscous flow of the magnetic flux and damping of normal currents.

Furthermore, we believe that there is no need to include the frequency dispersion of the medium to explain the experimental results in question. At the very least, the frequency dependence of the susceptibility observed in [11] is in good agreement with a model that includes a linear local relation between the current and electric field in the sample and in which the temperature dependence of the conductivity is similar to that dictated by the BKT transition.

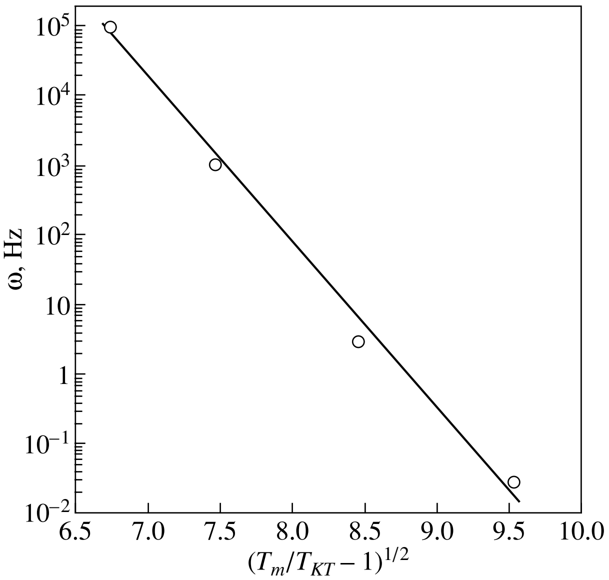


Fig. 1. Temperature T_m of the maximum in the dissipative part of the susceptibility as a function of the applied magnetic field frequency ω (the experimental data from [11]).

Let us consider a plane-parallel plate of thickness *d* subjected to electromagnetic waves incident on both surfaces of the plate in such a way that the phases of the magnetic field at both surfaces are always equal. Let the current density be related to the electric field as $j(x, t) = \sigma_f(T)E(x, t)$. According to [1], the dissipative part of the susceptibility is maximum when the skin depth $\delta = c/\sqrt{2\pi\omega\sigma_f}$ is about half the plate thickness, $d/\delta \approx 2$. Here, σ_f is the conductivity of the superconducting medium dictated by the viscous flux flow and related to the normal-state conductivity of the material σ_n and to the pancake concentration $n(T)$ as $\sigma_f(T) \sim \sigma_n/n(T)$. In order to estimate $n(T)$, we assume that the correlation length ξ^+ at $T > T_{\mathrm{BKT}}$ is proportional to the Debye screening length for the interaction between pancakes, $\xi^+ \sim \sqrt{8\pi\lambda(T)^2 T/\phi_0^2 n(T)s}$, where λ is the magnetic field penetration depth, ϕ_0 is the magnetic flux quantum, and *s* is the period of the layered structure. Using the relation $\xi^+(T) \sim \exp\{(T/T_{\mathrm{BKT}} - 1)^{-1/2}\}$ for $T \longrightarrow T_{\mathrm{BKT}}$ from above [25], we get $n(T) \sim \exp\{-2(T/T_{\mathrm{BKT}} - 1)^{-1/2}\}$. Substituting these estimations into the condition for the maximum of the dissipative part of the susceptibility, we get

$$2 \approx \frac{d}{\delta} = \frac{d\sqrt{2\pi\omega\sigma_f}}{c} \sim \frac{\sqrt{\omega}}{\exp\{-(T_m/T_{\mathrm{BKT}} - 1)^{-1/2}\}}.$$

Taking the logarithm and keeping only the relevant terms, we obtain the following relation between the temperature T_m at which the susceptibility is maximum and the magnetic field frequency ω:

$$\ln\omega \sim -\left(\frac{T_m}{T_{\mathrm{BKT}}} - 1\right)^{-\frac{1}{2}}. \tag{1}$$

The experimental results reported in [11] and replotted in [12] are shown in Fig. 1. It can be seen that there is reasonable agreement with our formula. Note that we did not use any fitting parameters when deriving Eq. (1). This is a very good reason to consider our suggestion concerning the role of the BKT transition and the local nature of the relation between the field and current to be true.

However, our estimations do not agree with the results from [10], where T_m was shown to depend on the external magnetic field amplitude. This disagreement suggests that the response is nonlinear both at low temperatures $T < T_{\mathrm{BKT}}$ and at $T > T_{\mathrm{BKT}}$. In the next section, we perform a numerical model calculation of the nonlinear response of layered superconductors to an ac electromagnetic field in the temperature range from T_{BKT} to T_c.

3. LOCAL NONLINEAR MODEL OF THE PANCAKE GAS SUSCEPTIBILITY

The main problem in calculating the susceptibility of a superconductor in this state is to find the constitutive relations between the current and electric field in the medium. At present, there is no theoretical model that provides relations between the current and field for this medium in a sufficiently wide temperature range.

In order to numerically calculate the response of a layered superconductor to an external ac magnetic field, we employ an empirical model relation between the current and electric field that qualitatively describes the current–voltage characteristic of high-temperature superconductors in the range between T_{BKT} and T_c.

Calculations are performed for a flat plate of thickness d under symmetric excitation by a magnetic field. The x axis is normal to the plate, and the plate surfaces are the $x = \pm d/2$ planes. An electric field is directed along the y axis, and a magnetic field is directed along the z axis. In this geometry, the Maxwell equations inside the plate can be written as

$$\frac{\partial e}{\partial l} = -\frac{\partial b}{\partial \tau}, \quad \frac{\partial b}{\partial l} = -j(e), \tag{2}$$

where we introduced dimensionless variables $e = E_y\sigma_n/J_{\mathrm{GL}}(0)$, $b = B_z\sigma_n/J_{\mathrm{GL}}(0)$, $l = x4\pi\sigma_n/c$, and $\tau = t(4\pi\sigma_n)$. Here, σ_n is the conductivity of the superconductor in the normal state and $J_{\mathrm{GL}}(0)$ is the decoupling current density at a zero temperature. It is convenient to write the constitutive relations in the form solved for the electric field:

$$e(j) = j\left[\left(\frac{4a}{e^{e_0}}\right)^{\frac{1}{\sqrt{4-2a}}} + \frac{j}{j_{\mathrm{GL}}}\right]^{a\left(1-\frac{j}{j_{\mathrm{GL}}}\right)}. \tag{3}$$

Here, $j_{\mathrm{GL}} = (1 - T/T_c)^{3/2}$ is the dimensionless decoupling current and e_0 is a parameter which can be related to the vortex core energy (in calculations, we set $e_0 = 6$). The value of $a = K(T_c/T - 1)$ varies from 0 at $T = T_c$ to 2 at $T = T_{\mathrm{BKT}}$. K is a parameter characterizing the anisotropy of the material; in calculations, this parameter is taken to be 100, which roughly corresponds to YBCO. The equation $a(T_{\mathrm{BKT}}) = 2$ determines the temperature T_{BKT}. The quantity $a + 1$ is usually called the exponent of the current–voltage characteristic.

In square brackets in Eq. (3), the first term describes the contribution to the resistivity of the free pancakes that appear due to thermal fluctuations above the BKT transition point and the second term describes the contribution of the vortices that appear as a result of severance of vortex dipoles by the current. Equation (3) correctly describes all current and temperature asymptotes. For intermediate values of the current and temperature, this expression qualitatively describes the behavior of real systems.

In the low-current limit, the second term in the square brackets is much less than the first and we get a linear $e(j)$ dependence. In the region of intermediate current values, the second term is greater than the first but is much less than unity and we get a high-current asymptotic behavior $e \sim j^{a+1}$. When the current value approaches the decoupling current, saturation is reached; most of the vortex dipoles are broken by the current, and further growth of the current does not have a significant effect on the free-vortex concentration. In the limit $j \longrightarrow j_{\mathrm{GL}}$, the system passes to the normal state. This type of $e(j)$ behavior can be modeled by introducing a current-dependent coefficient of the exponent a. Typically, saturation is not observed in the measured current–voltage characteristics because of the thermal instability at high current values.

In the high-temperature limit $T \longrightarrow T_c$ (or $a \longrightarrow 0$), Eq. (3) describes the transition of the system to the normal state with a linear $e(j)$ dependence. In the opposite limit $T \longrightarrow T_{\mathrm{BKT}}$ ($a \longrightarrow 2$), we get the characteristic dependence $e \sim j^3$ for all current values except in the saturation region.

By eliminating the electric field from Eqs. (2), we get a nonlinear diffusion equation for the magnetic field,

$$\frac{\partial b}{\partial \tau} = \frac{\partial e}{\partial j}\frac{\partial^2 b}{\partial l^2}, \tag{4}$$

where the diffusion constant is obtained by differentiating Eq. (3) and using the second of equations (2).

Symmetric boundary conditions lead to a symmetric field distribution inside the plate. The magnetic field is an even function of x, and the electric field is an odd function of x. Therefore, we can restrict ourselves to solving the problem only for one-half of the plate (e.g., for the right-hand half). The boundary conditions at the plate center are

$$\frac{\partial b}{\partial l}(0, \tau) = 0, \quad e(0, \tau) = 0.$$

The tangential components of the magnetic induction and of the electric field have to be continuous at the plate surfaces. Therefore, we have

$$b_{\text{in}}\left(\pm\frac{d}{2}, \tau\right) = b_{\text{out}}\left(\pm\frac{d}{2}, \tau\right),$$

$$e_{\text{in}}\left(\pm\frac{d}{2}, \tau\right) = e_{\text{out}}\left(\pm\frac{d}{2}, \tau\right).$$

According to the statement of the problem, a plane wave $b(l, \tau) = b_0\cos[\omega(\tau + l - d/2)]$ and $e(l, \tau) = b_0\cos[\omega(\tau + l - d/2)]$ with an amplitude b_0 is incident on the right-hand plate surface from $+\infty$. The reflected wave is a superposition of plane waves propagating in the positive direction of the x axis, with their amplitudes and frequency spectrum depending on the properties of the nonlinear material of the plate. The electric and magnetic fields of the reflected wave are similar in direction to those of the incident wave and are related to each other by the equation $e_r(l, \tau) = -b_r(l, \tau)$ following from the Maxwell equations in vacuum. Therefore, the boundary conditions at the right-hand surface are

$$b_{\text{in}}(d/2, \tau) = b_0\cos(\omega\tau) + b_r(d/2, \tau),$$

$$e_{\text{in}}(d/2, \tau) = b_0\cos(\omega\tau) - b_r(d/2, \tau).$$

By adding these expressions, we get boundary conditions for the field inside the plate. Therefore, the problem is completely defined by Eq. (4) and the boundary conditions for the magnetic induction

$$\begin{gathered} b(d/2, \tau) + e(d/2, \tau) = 2b_0\cos(\omega\tau), \\ \frac{\partial b}{\partial l}(0, \tau) = 0. \end{gathered} \tag{5}$$

We solved the boundary problem numerically by using the explicit absolutely stable Dufort–Frankel scheme [26]. Calculations were performed with a precision of no worse than 5%.

The temperature dependence of the susceptibility was calculated using the following procedure. For each amplitude of the magnetic field, calculations started from the normal state at T_c ($a = 0$). An expression for the magnetic field found analytically was used as the initial conditions. As the temperature was varied from $a = 0$ to 1.98, the field found at the end of the previous stage was used as the initial conditions for the next stage. The system was allowed to relax to the stationary state during one period, and the next period was used to calculate the response of the system. We checked the adequacy of this procedure by increasing the relaxation time at some temperatures by up to 20 periods and did not found a significant variation in the response.

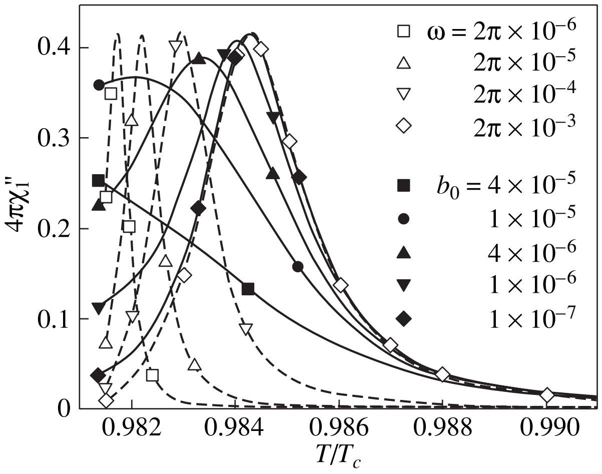


Fig. 2. Temperature dependences of the dissipative part of the susceptibility of the model system calculated in the nonlinear mode (solid curves) for the external field frequency $\omega = 2\pi \times 10^{-3}$ and various values of the external field amplitude b_0 and in the linear mode for various frequencies (dashed curves).

The dimensionless parameters of the modeled system were chosen so that (i) they satisfied the quasi-stationary condition ($\omega \ll 1$), (ii) the ratio of the sample thickness and the skin depth in the normal state was close to that typical for experiments ($d \ll \delta$), and (iii) the minimum required precision was achieved with reasonable computational effort. Accordingly, we set the plate thickness to be $d = 1$ and the field frequency, $\omega = 2\pi \times 10^{-3}$. The dimensionless skin depth was $\delta = \sqrt{2/\omega} \approx 17.8$.

The results of numerical calculations of the temperature dependence of the dissipative part of the susceptibility at the fundamental frequency are presented in Fig. 2 for values of the external magnetic field amplitude ranging from $b_0 = 10^{-7}$ to 4×10^{-5}. For comparison, the linear susceptibility of the system with the same set of parameters is also shown for various frequency values.

The temperature dependence of the dissipative part of the susceptibility $4\pi\chi_1''$ for field amplitudes $b_0 < 4 \times 10^{-5}$ has a maximum, which is situated close to T_{BKT} and whose position depends on the field amplitude. The maximum is due to the size effect, and its position T_m depends on the field amplitude because of the nonlinear conductivity of the medium.

This size effect becomes identical to the size effect considered in [1] at a sufficiently low amplitude of the external field, in the linear mode. The maximum in $4\pi\chi_1''(T)$ appears because the concentration of free vortices decreases rapidly as the temperature approaches T_{BKT} from above, which leads to an increased conductivity σ_f. As a result, the skin depth decreases and becomes approximately equal to half the sample thickness at a certain temperature T_m. In the nonlinear mode, the conductivity depends on the current and varies from point to point inside the sample. Therefore, in contrast to the linear mode, it is impossible to define the skin depth in this case. However, there is a characteristic field penetration depth, which decreases rapidly as T_{BKT} is approached; as a result, the nonlinear size effect takes place.

The calculated dependences have several prominent features. As the ac field amplitude decreases, the field-amplitude dependence of the susceptibility weakens progressively. At amplitudes $b_0 \le 10^{-7}$, the system is practically in a linear mode. The maximum magnitude of the dissipative part of the susceptibility $4\pi\chi_1'' \approx 0.417$ is close to the value observed experimentally (see [12, Fig. 1]). The peak value decreases as the temperature T_m becomes lower. The same tendency is observed in the experimental data from [11] (replotted in [12]), despite the fact that the reasons for the displacement of the maximum in the model and in the experiment are different. We believe that the variation in the peak value of the dissipative part of the susceptibility in the experiment is due to the nonlinear type of response, which becomes more and more pronounced as T_{BKT} is approached. This tendency is clearly demonstrated in Fig. 2, where all curves corresponding to the nonlinear response for various values of the external field amplitude merge into the curve of the linear response at high temperatures.

The maximum value of the temperature T_m for a given frequency is achieved at the minimum external field amplitude, in the linear mode. An increase in the field amplitude shifts T_m to lower temperatures. At $b_0 = 4 \times 10^{-5}$ and $\omega = 2\pi \times 10^{-3}$, the $4\pi\chi_1''(T)$ dependence does not exhibit a maximum in the temperature range from T_{BKT} to T_c. Experimentally, the maximum in this case is observed at a lower temperature, where the response is determined by other physical mechanisms.

The calculated $T_m(b_0)$ dependence agrees qualitatively with the experimental data from [10], so the kink in the linear $T_m(H_{\mathrm{ac}}^{2/3})$ dependence can be attributed to the BKT phase transition occurring in the sample at a temperature T_{BKT}.

4. CONCLUSIONS

It follows from the experiments performed in [10, 11] that there is a narrow temperature range (T_{BKT}, T_c) in which the response of YBCO films to an ac magnetic field differs qualitatively from that observed at low temperatures.

It was suggested in [11] that, at the temperature T_c (T_{BKT} in our notation), the film undergoes a phase transition, which depends on the film thickness and is similar to the BKT transition in 2D systems. As a result, free magnetic vortices running through the sample appear and the sample response to an ac magnetic field is determined by the behavior of these vortices in the vicinity of the transition. The authors of [12] associated the response with the frequency-dependent polarization of neutral pancake dipoles under an ac external field.

In our approach, the response of superconducting YBCO films in the temperature range in question is related to the behavior of free pancakes, which appeared because of the BKT transition and whose concentration depends both on the temperature and local current density.

In the papers cited above, the frequency dependence of the response and the presence of a maximum in the nondissipative part of the susceptibility are attributed to the frequency dispersion of the material. In our opinion, these features are due to the viscous flow of free pancakes and to the size effect in a finite-size sample, with the relation between current and field being local. Apart from providing qualitative agreement with experiment, this approach has made it possible to adequately describe, without using fitting parameters, the frequency dependence of T_m (Fig. 1) and the peak value of the dissipative part of the susceptibility.

The nonlinear character of the response in our model makes us believe that the experimental data from [10] can also be explained in this model. The decrease in the temperature T_m with an increase in the field amplitude in the model calculations is in qualitative agreement with the dependences observed in the experiments at temperatures above the kink on the $T_m(H_{\mathrm{ac}}^{2/3})$ line. The temperature at which this kink is observed can be associated with the temperature T_{BKT}. Our ability to quantitatively fit the results of calculations to particular experimental data is significantly challenged by the large volume of computations involved.

REFERENCES

1. H. Fisher and Y. H. Kao, Solid State Commun. **7** (2), 275 (1967).
2. L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Nauka, Moscow, 1982; Pergamon, Oxford, 1982).
3. C. P. Bean, Phys. Rev. Lett. **8** (7), 250 (1962).
4. Y. B. Kim, Phys. Rev. Lett. **8** (7), 256 (1962).

5. A. M. Campbell and J. E. Evetts, *Critical Currents in Superconductors* (Taylor and Francis, London, 1972; Mir, Moscow, 1975).
6. P. W. Anderson, Phys. Rev. Lett. **9** (7), 309 (1962).
7. A. N. Artemov, V. F. Drobot'ko, D. G. Emel'yanenkov, V. I. Makhov, and V. A. Khokhlov, Phys. Lett. A **157** (1), 85 (1991).
8. K.-H. Müller, Physica C **168** (5/6), 585 (1990).
9. M. W. Coffey, J. Clem. Phys. Rev. Lett. **67** (3), 386 (1991).
10. V. A. Khokhlov, V. V. Krzhizhanovskiĭ, A. Yu. Prokhorov, V. F. Drobot'ko, A. V. Klimov, and G. G. Levchenko, Fiz. Tverd. Tela (St. Petersburg) **43** (9), 1541 (2001) [Phys. Solid State **43** (9), 1603 (2001)].
11. J. Kötzler, D. Görlitz, S. Skwirblies, and A. Wriedt, Phys. Rev. Lett. **87** (12), 127005 (2001).
12. K. Medvedyeva, B. J. Kim, and P. Minnhagen, Phys. Rev. Lett. **89** (14), 149703 (2002).
13. V. L. Berezinskiĭ, Zh. Éksp. Teor. Fiz. **61** (3), 1144 (1971) [Sov. Phys. JETP **61** (3), 610 (1971)].
14. J. M. Kosterlitz and D. G. Thouless, J. Phys. C **6** (6), 1181 (1973).
15. A. N. Artemov, A. M. Grishin, V. N. Korenivskii, A. N. Ul'yanov, and V. A. Khokhlov, Int. J. Mod. Phys. B **4** (4), 591 (1990).
16. J. Kötzler, G. Nakielsky, M. Baumann, R. Behr, F. Goerke, and E. H. Brandt, Phys. Rev. B **50**, 3384 (1994).
17. J. Kötzler, D. Görlitz, and A. Wriedt, Phys. Rev. Lett. **89** (14), 149704 (2002).
18. T. Schneider, Physica C **195** (1–2), 82 (1992).
19. P. Minnhagen, Rev. Mod. Phys. **59** (4), 1001 (1987).
20. B. J. Kim, P. Minnhagen, and P. Olsson, Phys. Rev. B **59** (17), 11506 (1999).
21. J. Clem, Phys. Rev. B **43**, 7837 (1991).
22. N.-C. Yeh and C. C. Tsuei, Phys. Rev. B **39**, 9708 (1989).
23. W. Weber and H. J. Jensen, Phys. Rev. B **44**, 454 (1991).
24. A. N. Artemov, A. M. Grishin, A. V. Zinovchuk, Yu. I. Medvedev, and Yu. M. Nikolaenko, Fiz. Tekh. Vys. Davleniĭ **11** (3), 110 (2001).
25. J. M. Kosterlitz, J. Phys. C **7**, 1046 (1974).
26. D. E. Potter, *Computational Physics* (Wiley, New York, 1973; Mir, Moscow, 1975).

Translated by G. Tsydynzhapov

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1060–1065. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1026–1031.
Original Russian Text Copyright © 2005 by Borisov, Vakhrushev, Naberezhnov, Okuneva.

SEMICONDUCTORS AND DIELECTRICS

Inelastic Neutron Scattering by TA Phonons in Heavily Doped Gallium Arsenide

S. A. Borisov, S. B. Vakhrushev, A. A. Naberezhnov, and N. M. Okuneva
Ioffe Physicotechnical Institute, Russian Academy of Sciences, Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia
e-mail: alex.nabereznov@mail.ioffe.ru

Received June 28, 2004

Abstract—The temperature dependence of the shape of the inelastic neutron scattering peak from TA phonons in GaAs heavily doped by Te was studied within the temperature interval from 363 to 253 K. It was shown that doping with tellurium (to a carrier concentration $N_e \approx 2 \times 10^{18}$ cm^{-3}) gives rise to the appearance of an additional contribution to neutron scattering on the high-energy side of the TA phonon resonances at values of the reduced wave vector $q < 0.1a^*$. Below 320 K, the intensity of this additional component rises sharply, and then, below 273 K, the main TA peak and the additional shoulder merge almost completely. This additional scattering is believed to be due to a defect-induced mode, which may be responsible for the observed anomalies in the physical properties in this crystal. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The existence of a phase transition (PT) in a system of dipole centers remains an issue of undying interest in solid-state physics. Indeed, according to the Langevin–Debye theory describing the interaction between dipoles in terms of the self-consistent-field model, one would expect an unlimited rise in the polarizability of this system (“polarization catastrophe”) with decreasing temperature. In actual fact, however, no ferroelectric PT was observed even in the early experiments [1, 2] on alkali halide crystals containing dipole impurities. It was shown later in [3] that the reason for this disagreement between theory and experiment lies in the specific features of the dipole–dipole interaction; indeed, not only the magnitude but also the sign of this interaction depend on the relative spatial orientation and position of the dipoles. In systems containing impurities, this results in considerable fluctuations of the local fields acting on the dipoles, thus making the self-consistent-field approximation inapplicable. It has also been shown that spatial fluctuations substantially reduce the effective polarizability and prevent the onset of ferroelectric instability. This conclusion fully holds in the case of nonpiezoelectric crystals.

However, piezoelectric crystals exhibit a linear relation between polarization and elastic lattice strains. This linear striction interaction suppresses long-wavelength fluctuations of the ordering field, so, at a high enough dipole impurity concentration, one may expect the formation of an ordered state even in weakly polarizable crystals. It was demonstrated in [4] that, in III–V compound semiconductors doped with Group VI elements, charged donor–vacancy complexes consisting of impurity atoms and native structural defects form in the reaction $V_{\mathrm{Ga}}^{-} + \mathrm{Te}_{\mathrm{As}}^{+} + e^{-} \longrightarrow (\mathrm{Te}_{\mathrm{As}}V_{\mathrm{Ga}})^{-}$, where V_{Ga} denotes a vacancy on the Ga site. These complexes may have the properties of dipole centers [5], possess large dipole and elastic moments, and interact efficiently with elastic lattice strains.

The possible existence of a PT in such systems was predicted theoretically in [6, 7], where it was shown phenomenologically that the free energy of a doped crystal as a function of polarization can be written in the form

$$F(P) = F(0) + (1/2)\alpha' P_\beta^2 + b_1 P_\beta^4 + (1/2)(b_2 + b_d)P_\beta^2 P_\alpha^2, \quad (1)$$

where P_β is the β component of the polarization, b_1 and b_2 are the anharmonic constants of a pure crystal, b_d allows for the contribution from dipole centers to second-order anharmonicity, and α' is a renormalized parameter depending on both the temperature T and the concentrations of charge carriers (N_e) and dipole defects (N_d). As a result of renormalization of α, a critical temperature $T_0(N_c, N_e)$ at which the parameter $\alpha'(T_0)$ reverses sign appears at certain concentrations N_e and N_d; i.e., the crystal as a whole becomes unstable against transition to a state with a nonzero polarization. Note that, because the components of the polarization vector and strain tensor are linearly related, this PT should be characterized by both spontaneous polarization and spontaneous strain and, therefore, should be a ferroelectric–ferroelastic transition induced by a system of dipole impurity centers.

The possible existence of this PT has been studied experimentally on Te- and Se-doped GaAs piezoelectric crystals. The concentration and temperature dependences of the longitudinal velocity of sound $V_L(T, N_e)$

and of the specific heat C_{mol} in Se-doped GaAs were investigated in [6], and the infrared absorption coefficient $\alpha(N_e)$ in Te-doped GaAs was studied in [8]. At 300 K, anomalies were revealed in the $V_L(N_e)$ and $\alpha(N_e)$ relations at $N_e \approx 2 \times 10^{18}$ cm^{-3}. This fairly high critical concentration may be assigned to the fact that dipole defects form only for $N_e > 10^{17}$ cm^{-3} and that their concentration becomes substantial for $N_e \geq 10^{18}$ cm^{-3} [6]. Measurements of the $V_L(T)$ and $C_{\mathrm{mol}}(T)$ relations on a sample with $N_e \approx 2 \times 10^{18}$ cm^{-3} confirmed the existence of anomalies at $T_c \approx 300$ K.

Thus, these results give grounds to suggest the existence of a specific PT at the impurity concentration and temperature indicated above. However, to date, there have been almost no studies on the lattice dynamics of such compounds. If this PT does exist, it should give rise to anomalies in the low-energy part of the spectrum. For this reason, we investigated the low-energy excitation spectrum and, in particular, obtained data on the temperature-induced evolution of the shape of the neutron resonance line corresponding to inelastic neutron scattering from transverse acoustic (TA) phonons.

2. EXPERIMENTAL TECHNIQUE

Studies were conducted on a single, irregularly shaped crystal of pure GaAs, about 4 cm^3 in volume, and a GaAs single crystal doped with tellurium ($N_c \approx 2 \times 10^{18}$ cm^{-3}), $20 \times 20 \times 40$ mm in size, on a Neutron-3 neutron spectrometer at the WWR-M reactor (PNPI, Gatchina). The crystal could be rotated about the $[1\bar{1}0]$ axis, and the incident neutron wavelength was 2.485 Å. Pyrolytic graphite was used as a monochromator, and a Cu(111) single crystal was used as an analyzer. All measurements were performed in the vicinity of the (220) point, for which the structural factor for inelastic scattering from acoustic phonons is maximal for the wave vectors accessible at the given neutron wavelength. Measurements were carried out on pure gallium arsenide at room temperature only and on the doped crystal in the temperature interval from 363 to 253 K for the values of the reduced wave vector in the range $0.06a^* < q < 0.2a^*$ (a^* is the reciprocal lattice vector). For q less than $0.06a^*$, intense Bragg scattering precluded from reliable determination of the weak inelastic peak.

3. RESULTS AND DISCUSSION

In pure gallium arsenide, no line shape distortions were observed for any values of q and the phonon resonance width was determined almost entirely by the spectrometer energy resolution and could be fitted well by a Gaussian function. The results agree well with the data obtained at 12 K and reported in [9]. For this reason, the temperature dependence of the line shape was not studied further.

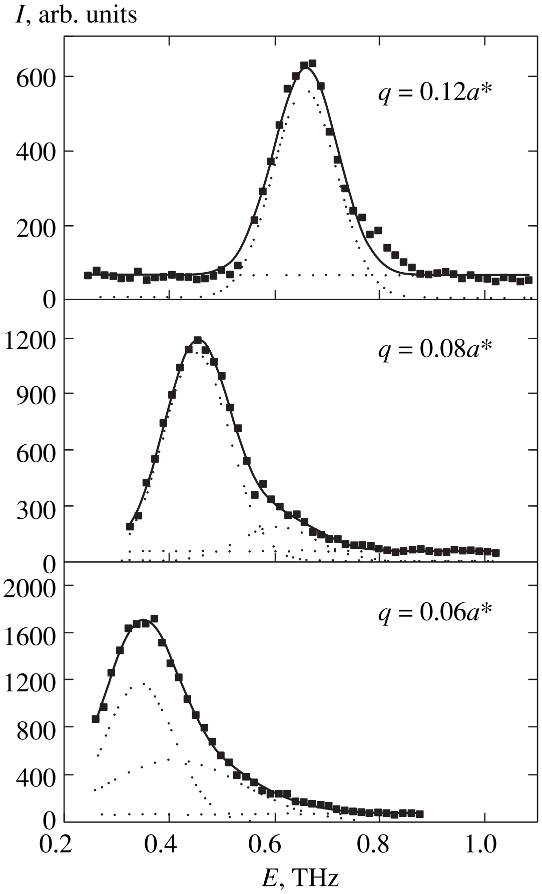


Fig. 1. TA phonon resonance line shape plotted versus reduced wave vector q in doped GaAs near the (220) point measured at 363 K. Solid lines are fitting of the experimental data with a sum of Gaussians and a linear background. Dotted lines are the contributions of individual components to the observed inelastic neutron scattering spectrum.

Measurements performed on doped GaAs at large q revealed, at all temperatures, phonon resonance with a width determined by the energy resolution and described by a Gaussian with parameters similar to those obtained for pure GaAs. The resonance peak was accompanied by a weak shoulder on the high-energy side. As seen from Fig. 1, which displays inelastic neutron scattering spectra for various values of the reduced wave vector q at a temperature of 363 K, a decrease in q results in an increase in the shoulder intensity. We fitted the inelastic scattering spectra with a sum of the Gaussians corresponding to the TA phonon, the additional shoulder, and the contribution due to elastic Bragg scattering, which was observed at small q (the Bragg tail) because of an insufficiently high resolution, and a linear background. The fitting procedure converged in several iterations to an acceptable χ^2 value.

For $q > 0.1a^*$, the shape of observed phonon resonance did not vary on cooling. For small q ($q < 0.1a^*$), a decrease in temperature caused an increase in the intensity of the additional shoulder and below 273 K

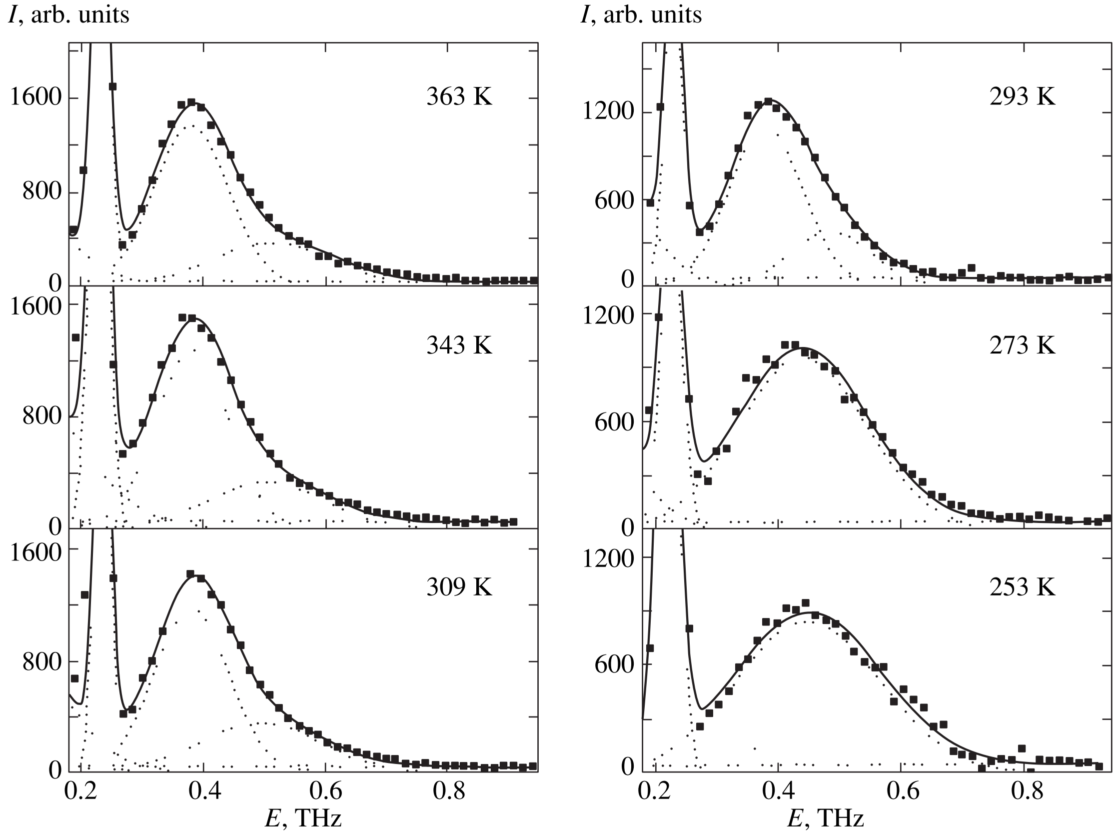


Fig. 2. Evolution of the shape of TA resonance in doped GaAs obtained at $\mathbf{Q} = (2,2,0.07)$ on cooling. The peak at $E \approx 0.2$ THz is the elastic Bragg scattering tail. The dotted line is the fit obtained by the method specified in the text.

the shoulder practically merged with the main TA peak. The experiment yielded a broad bell-shaped inelastic neutron scattering distribution, which made unambiguous separation of the TA peak from the additional scattering impossible. Figure 2 presents the temperature dependence of the TA resonance line shape near the (220) point; the solid line is the fit, and the dotted lines are contributions from the above-mentioned components. As already pointed out, starting from 273 K, the main and the additional peaks practically merged, thus making the fitting procedure ambiguous. We analyzed the temperature dependence of the TA phonon resonance parameters for $T \geq 300$ K. As is well known, in the high-temperature approximation ($kT \gg \omega$), the following scaling relation holds [10]:

$$I \sim T/\omega_j^2, \tag{2}$$

where I is the integrated peak intensity, ω_j is the frequency of the jth phonon resonance, and T is the absolute temperature. Thus, in the absence of strong mode-coupling effects, the quantity $I\omega_j^2/T$ should not depend on temperature. Figures 3 and 4 display the temperature dependences of I_{TA}/T (curve *1*, Fig. 3) and ω_{TA} (curve *1*, Fig. 4), where ω_{TA} is the TA phonon frequency and I_{TA} is the integrated intensity. We see that the above relations hold even in the range 300–320 K, where the total intensity of inelastic scattering in the vicinity of the TA resonance grows most strongly (curve *3*, Fig. 3); as a result, both ω_{TA} and I_{TA}/T do not depend on temperature. This allows us, when fitting the data obtained below 293 K, to fix the values of these parameters and vary the parameters of the additional scattering only. As is evident from Fig. 3 (curve *2*), the intensity of the additional wing grows substantially at temperatures below 320 K, i.e., exactly in the region where heavily doped GaAs revealed anomalies in the IR absorption coefficient [8], the peak in specific heat, and the minimum in the velocity of longitudinal acoustic waves [6]. It is also natural to expect a decrease in the velocity of transverse acoustic waves. Figure 5 presents dispersion curves for pure GaAs at room temperature (and at 12 K from [9]) and for doped GaAs. We see that the transverse sound velocity (which is proportional to the slope of the dispersion curve at small q) in the doped GaAs is slightly lower than that in the pure material and is 2991 ± 149 m/s for the doped GaAs and 3461 ± 173 m/s for pure gallium arsenide. The latter value is in good agreement with the value of 3345 m/s reported for 300 K in [11].

Ultrasonic studies at 12.5 MHz of the relative velocity variation ($\Delta V_l/V_l$) of longitudinal acoustic waves as a function of temperature in a doped crystal with $N_e = 3 \times 10^{18}$ cm^{-3} showed [6] that this quantity passes through a local minimum near 300 K corresponding to a change in $\Delta V_l/V_l$ of approximately 0.1% and that near this point the temperature dependence of $\Delta V_l/V_l$ changes slope. Unfortunately, it was not possible to observe this effect experimentally for the transverse acoustic branch, because the expected value is too small to be measured.

Let us consider now the possible nature of the observed shoulder in the inelastic neutron scattering spectra. As is well known, vibrations of a perfect lattice can be substantially modified by introducing, for instance, point defects. This problem was analyzed in considerable detail, in particular, in [12–14], where it was shown that the square of the characteristic frequency associated with a defect is $\omega^2 = f_d/m_d$; here, f_d is the coupling constant and m_d is the mass of the defect. Two different types of new (additional) vibrational modes can be conceived of here, depending on the actual relationship between these parameters: local modes with frequencies above the maximum frequency of the original-lattice vibrations (if $f_d > f$ and/or $m_d < m$) and resonance modes with $\omega \ll \omega_{\max}$ (if $f_d \ll f$ and/or $m_d \gg m$). In [15], the local-mode frequency in $Ta_{88}Nb_{12}$ (with the Nb atoms considered as a light impurity) was derived from an analysis of coherent inelastic neutron scattering and it was shown that this mode is observed as a shoulder on the high-energy side of the transverse $(q, 0, 0)$ mode near the Brillouin zone boundary. Later, a resonance mode of frequency $\omega_{\max}/7$ associated with libration and axial displacement of two neighboring interstitial atoms (which make up a dumbbell-shaped structure) was theoretically predicted to exist in fcc metals [16] in the case where the mass of each of the two interstitial atoms is identical to the mass of an atom of the original lattice (i.e., $m_d = m$; this situation arises, for instance, when a sample is irradiated by thermal neutrons). It was also shown in [16, 17] that these resonance modes should be observed in coherent inelastic neutron scattering at defect concentrations on the order of 10^{19} cm^{-3} and with small q.

Experimentally, a similar mode was observed by Nicklow *et al.* [18] in irradiated copper as a pronounced shoulder on the high-frequency side of the TA phonon resonance at 10 K, with the resonance frequency being displaced noticeably toward higher energies with respect to that of the transverse acoustic phonon for pure copper. This additional shoulder dropped strongly in intensity after annealing at room temperature, but only 2-h annealing at 800 K removed it completely. As for gallium arsenide, the local modes and the associated effects were earlier observed repeatedly under doping with carbon, boron, oxygen, or other light elements; the results obtained in those studies can be found, e.g., in [12, 19–21] (see also references therein). In our sample, atoms of the impurity Te (atomic number 52, atomic

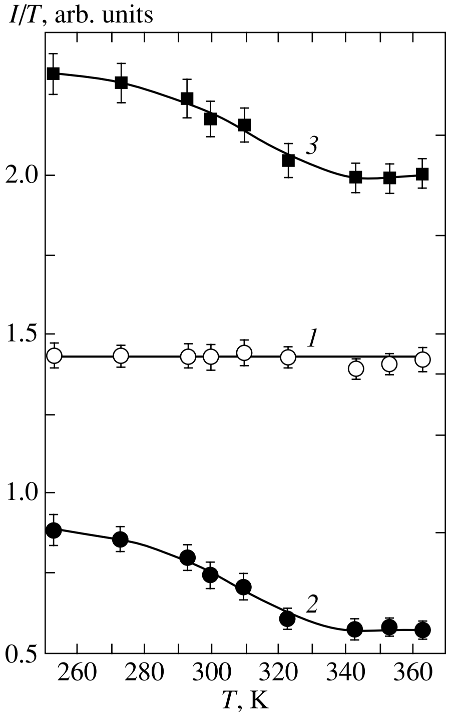


Fig. 3. Temperature dependences of I/T (I is the integrated intensity, T is the temperature) for (*1*) the TA resonance, (*2*) additional shoulder, and (*3*) total intensity of inelastic scattering in the TA resonance region obtained for doped GaAs at $q = 0.07a^*$.

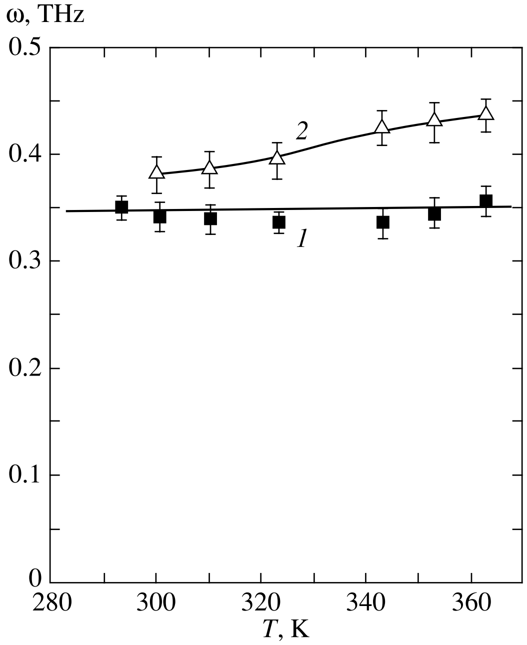


Fig. 4. Temperature dependences of the frequencies (*1*) of the TA resonance and (*2*) of the additional shoulder at $q = 0.07a^*$.

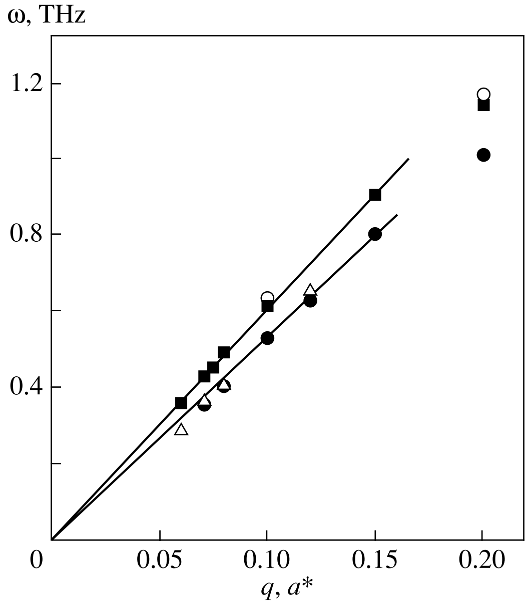


Fig. 5. TA resonance dispersion curves for pure GaAs (open circles are data from [9] obtained at 12 K, filled squares are our room-temperature results) and doped GaAs (our data: filled circles and open triangles correspond to room temperature and 310 K, respectively).

weight 127.61) are almost twice as heavy as Ga (atomic number 31, atomic weight 69.72) and As atoms (atomic number 33, atomic weight 74.91); therefore, one would expect the appearance not of a local but rather of a defect-induced resonance mode in the low-frequency excitation region. If it originated from the formation of the abovementioned dumbbell-shaped structures of Ga or As atoms, then, based on the value $\omega_{\max} \approx 9$ THz [9], the frequency of this mode should be $\omega_{\max}/7$ [16], which is about 1.28 THz. The frequency for similar structures of Te atoms should be about one-half the above value. Indeed, as seen, for instance, from Fig. 2, the maximum of the additional shoulder at $q = 0.07a^*$ for temperatures high enough for reliable determination is observed at $E \approx 0.6$ THz, which is in good agreement with the purely qualitative estimate of the resonance mode frequency indicated above. As already mentioned, this mode could account for the experimentally observed distortion of the TA resonance shape and a shift of the position of the maximum at small *q*. This mode should be dispersionless if it is truly a resonance mode. However, the above figures show clearly that this mode exhibits dispersion and that its frequency at small *q* decreases with decreasing temperature. Indeed, as the temperature is reduced from 363 to 300 K (the temperature at which we are still able to determine reliably the position of the maximum of the additional shoulder), the frequency at $q = 0.07a^*$ decreases by about 15% (curve *2* in Fig. 4). In accordance with the relationship $I \sim T/\omega_J^2$, one should expect a growth of the intensity due to scattering on this mode by a factor of about 1.3, which agrees well with the result presented in Fig. 3.

Because these defects have a dipole moment and dipole forces are of the long-range type, a decrease in temperature, as shown in [6, 7], should bring about the onset of long-range order in the dipole system and, accordingly, the appearance of a localized “soft” mode that would propagate to finite distances and result in the formation (below 320 K) of a spatially inhomogeneous state similar to the one observed in the $PbMg_{1/3}Nb_{2/3}O_3$ relaxor [22]. This consideration is persuasively supported by the observation of a strong rise in the IR absorption coefficient at the carrier concentration $N_e \approx 2 \times 10^{18}$ cm^{-3} [8]. Thus, the fact that the temperature at which the intensity of the additional shoulder exhibits a strong increase coincides with the temperature at which anomalies in the physical properties are observed finds a logical explanation, and there are grounds to believe that this mode is really responsible for the observed anomalies in physical properties.

It should also be pointed out that if this mode does indeed exist, one should expect, at small *q*, a strong dependence of the TA phonon resonance line shape on the Brillouin zone, in which the measurements are being carried out, as has been observed, for instance, for $BaTiO_3$ [23] in the case of interacting modes. Our preliminary measurements have shown that this effect apparently exists, but the insufficiently high neutron flux did not permit us to carry out a full-scale study of the TA resonance line shape for various Brillouin zones and determine the parameters of the defect-induced mode and the parameters of its interaction with other phonons.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 02-02-16695), a grant from the President of the Russian Federation (project no. NSh-2168.2003.2), a program of the Department of Physical Sciences of the RAS, and CRDF (project no. RP1-2361-ST-02).

REFERENCES

1. W. Känzig, H. R. Hart, Jr., and S. Roberts, Phys. Rev. Lett. **13** (18), 543 (1964).
2. U. T. Fiory, Phys. Rev. B **4** (2), 614 (1971).
3. B. E. Vougmeister and M. D. Glinchuk, Rev. Mod. Phys. **62** (4), 993 (1990).
4. V. T. Bublik, M. G. Mil’vidskiĭ, and V. B. Osvenskiĭ, Izv. Vyssh. Uchebn. Zaved. Fiz. **1**, 7 (1980).
5. I. Ya. Buyanova, S. S. Ostapenko, and M. K. Sheĭnkman, Fiz. Tverd. Tela (Leningrad) **27** (3), 748 (1985) [Sov. Phys. Solid State **27** (3), 461 (1985)].
6. V. V. Prudnikov and I. A. Prudnikova, Kristallografiya **37** (5), 1093 (1992) [Sov. Phys. Crystallogr. **37** (5), 585 (1992)].

7. V. V. Prudnikov and I. A. Prudnikova, Izv. Vyssh. Uchebn. Zaved. Fiz. **9**, 105 (1989).
8. E. A. Balagurova, Yu. B. Grekov, A. F. Kravchenko, I. A. Prudnikova, V. V. Prudnikov, and N. A. Semikolenova, Fiz. Tekh. Poluprovodn. (Leningrad) **19** (9), 1566 (1985) [Sov. Phys. Semicond. **19** (9), 963 (1985)].
9. D. Strauch and B. Dorner, J. Phys.: Condens. Matter **2**, 1457 (1990).
10. Yu. A. Izyumov and N. A. Chernoplekov, *Neutron Spectroscopy* (Énergoatomizdat, Moscow, 1983) [in Russian].
11. C. W. Garland and K. C. Park, J. Appl. Phys. **33** (2), 759 (1962).
12. O. Madelung, *Introduction to Solid State Theory* (Springer, Berlin, 1978; Nauka, Moscow, 1985).
13. A. A. Maradudin, *Theoretical and Experimental Aspects of the Effects of Point Defects and Disorder on the Vibration of Crystals* (Academic, New York, 1966).
14. D. A. Robbie, M. J. L. Sangster, and P. Pavone, Phys. Rev. B **51** (16), 10489 (1995).
15. J. Als-Nielsen, in *Proceedings of the Fourth IAEA Symposium on Neutron Inelastic Scattering* (Copenhagen, 1968), Vol. 1, p. 35.
16. P. H. Dederichs, C. Lehmann, and A. Scholz, Phys. Rev. Lett. **31** (18), 1130 (1973).
17. R. F. Wood and M. Mostoller, Phys. Rev. Lett. **35** (1), 45 (1975).
18. R. M. Nicklow, R. R. Coltman, F. W. Young, Jr., and R. F. Wood, Phys. Rev. Lett. **35** (21), 1444 (1975).
19. D. A. Robbie, R. S. Leigh, and M. J. L. Sangster, Phys. Rev. B **56** (3), 1381 (1997).
20. H. Ch. Alt and B. Dischler, Appl. Phys. Lett. **66** (1), 61 (1995).
21. H. Ch. Alt, H. Mussig, and H. Brugger, in *Proceedings of the 9th Conference on Semiconducting and Insulating Materials* (Tolouse, France, 1996), p. 155.
22. A. Naberezhnov, S. Vakhrushev, B. Dorner, D. Strauch, and H. Moudden, Eur. Phys. J. B **11** (1), 13 (1999).
23. J. Harada, J. D. Axe, and G. Shirane, Phys. Rev. B **4** (1), 155 (1971).

Translated by G. Skrebtsov

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1066–1072. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1032–1038.
Original Russian Text Copyright © 2005 by Bagaev, Galkina, Klokov, Klevkov, Krivobok, Martovitskiĭ, Sentyurina, Sharkov.

SEMICONDUCTORS AND DIELECTRICS

Nonequilibrium Acoustic Phonons in ZnTe at Liquid-Helium Temperatures: Effect of Impurities and Structural Defects on the Phonon Mean Free Path

V. S. Bagaev, T. I. Galkina, A. Yu. Klokov, Yu. V. Klevkov, V. S. Krivobok,
V. P. Martovitskiĭ, N. N. Sentyurina, and A. I. Sharkov
Lebedev Physical Institute, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 119991 Russia
e-mail: bagaev@lebedev.ru

Received July 8, 2004

Abstract—ZnTe crystallites isolated from a druse of coarse-grained polycrystalline ZnTe, prepared by chemical vapor synthesis at a temperature of ~650°C, were studied using x-ray diffractometry, luminescence, and the heat pulse technique. The crystallites are stacked in {110}-oriented macroscopic layers with a common twin system. The {111} twin planes separated at a distance of 50–100 μm are perpendicular to the (110) growth layers. Acoustic phonon propagation was studied using the heat pulse technique. A comparison of the responses to the arrival of differently polarized phonons in a given sample with the responses obtained on high-purity coarse-grained ZnTe with randomly distributed twin systems with a separation of 5–10 μm and on twin-free single-crystal ZnTe suggests that twins radically affect the scattering of acoustic phonons. The mean free paths of LA, FTA, and STA phonons were determined by comparing the experimental responses with Monte Carlo calculations. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Zinc telluride (ZnTe) crystals are comparatively poorly studied typical representatives of the group of wide-bandgap II–VI compounds. While certainly being a promising material with a high luminescence quantum yield in the very convenient visible region, only a fairly limited interest has been expressed in ZnTe for applications because of a number of features inherent to it (the original conduction of the *p* type characteristically observed in crystals and films prepared by any method, a tendency to self-compensation, the presently almost unattainable inversion of conduction type by doping).

Nevertheless, in addition to its well-known application as a substrate material, {110} ZnTe plates have proved convenient for use as both coherent and incoherent electrooptical sensors in the terahertz region (2–40 THz) due to the high quality and suitable electrooptical coefficients of this material [1].

With the possibility of regulating the ZnTe growth technology, which permitted substantial variation of the crystallization regimes, and by properly controlling the crystal structure of samples grown in conditions far from thermodynamic equilibrium, we carried out a study of the effect of point defects and of the extent of twinning on the scattering and propagation of acoustic phonons and on the emissive properties of the samples prepared.

This study was also motivated by the need to characterize the material, which was grown using a modification of the chemical vapor deposition method (namely, chemical vapor synthesis (CVS) from separate components), by x-ray diffractometry, measurement of the low-temperature luminescence, and the heat-pulse technique.

The growth conditions in the method chosen by us differ substantially from the quasi-equilibrium conditions governing crystallization. In particular, obtaining a few tens of grams of the material under quasi-equilibrium growth requires one day, whereas the CVS method produces the same amount of material in tens of minutes. It should be added that the total amount of the main impurities in a CVS-grown material is, as far as we know, very low and is at the level of the best materials. This method differs from the quasi-equilibrium techniques in that, due to spontaneous nucleation of the crystallites, the grown druse contains (depending on the actual growth regime parameters) only 5–10% of the total druse weight in the form of single-crystal grains 5–10 mm in size.

In publications [2, 3], the mechanism of phonon scattering in high-purity coarse-grained (ZT04) [2] and single-crystal (ZT-1) [3] zinc telluride was studied. The ZT04 sample contains an insignificant amount of impurities but consists of grains up to a few hundred microns in size, which contain randomly oriented twin systems with twin planes a few microns apart. By contrast, the ZT-1 sample does not have twin planes but contains up to 10^{17} cm^{-3} impurities. As for the ZnTe studied in this work, its impurity content is intermediate between those of the samples investigated in [2, 3], its twin system is ordered, and the distance between the twin boundaries can be as large as a few hundred microns.

2. PREPARATION OF ZNTE AND X-RAY STRUCTURAL STUDIES

A druse of polycrystalline ZnTe with a clearly pronounced surface relief at the crystallization front was obtained using CVS of component vapors interacting in the crystallizer zone at a temperature of 650–700°C. The synthesis was carried out in a quartz reactor 90 mm in diameter in dynamic vacuum. The component vapors were produced in separate cells joined with the crystallizer and effused head-on. The flux densities and the relative partial pressures of zinc and tellurium vapors were controlled by properly varying the cell temperatures.

The starting components in the process were tellurium (in which the content of the main impurities was less than 10^{16} cm^{-3}) and commercial zinc (99.999%), which we purified preliminarily.

The growth surface forming in the course of condensation at ~650°C with a vapor supersaturation of ~$(1.5–5.0) \times 10^3$ and a pressure ratio $P_{\text{Zn}}/P_{\text{Te}_2} \geq 2$ consisted of a multitude of fairly large crystallites with a [110] growth direction (Fig. 1).

For a total mass of the components of ~250 g, no more than 2 h of synthesis was required before the cells no longer contained any material for reaction. We believe that the rates of ZnTe crystallite formation obtained are characteristic of nonequilibrium crystallization involving concentration and thermal fluctuations before the crystallization front.

X-ray diffractometry showed that the crystallites shaped as irregular tetrahedral pyramids are formed by {110} macroscopic layers growing from the vertex of a pyramid towards its base. A fragment of a crystallite (ZT05) formed by one (110) macroscopic-layer system was cut out for study. As in [2], a plate thus prepared was found to contain twins represented by one system of parallel lines in the etch pit pattern shown in Fig. 2. X-ray studies revealed that the twin planes are arranged perpendicular to the (110) growth layers, whereas in the case where the direction of growth is along {111} they are parallel to the growth layers [4]. Thus, transition from the {111} layer growth characteristic of crystallization from congruently sublimed ZnTe to the {110} layer growth observed in the synthesis from the starting components is accompanied by both a change of the growth plane and rearrangement of the twin planes relative to the growth surface. The twin formation mechanisms in a CVS-grown material apparently differ from those operating in ZnTe and CdTe films grown epitaxially on various substrates (see the wealth of data reported in [5]). Figure 3 shows an atomic model of the twin plane constructed by us for the case under consideration.

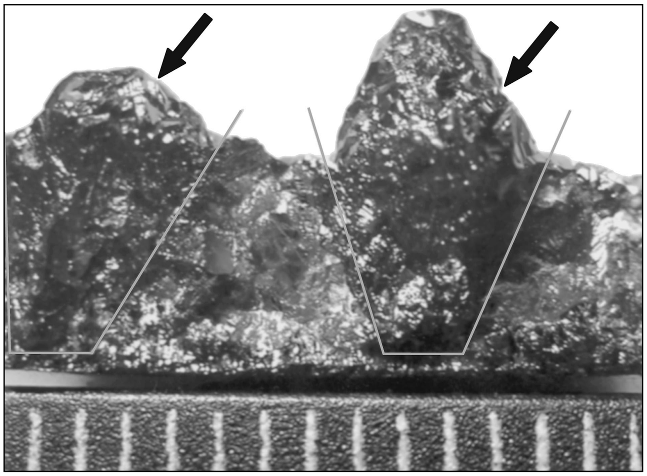


Fig. 1. Photograph of part of a ZnTe druse. Arrows identify crystallites. The ruled division, shown at the bottom, is in units of 1 mm.

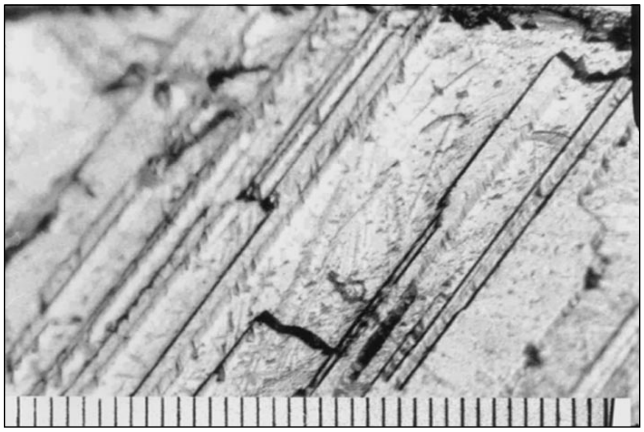


Fig. 2. Optical surface image of a ZT05 crystallite obtained with a UNION 6551 microscope. The ruled division, shown at the bottom, is in units of 10 μm.

The ZnTe samples intended for photoluminescence and heat pulse measurements were mechanically ground and polished. After this, grease was removed from the ZnTe sample by boiling it in acetone for two minutes and the damaged layer was removed by chemical wet etching in a 6 vol % Br_2 solution in methanol (dehydrated, OSCh grade) for two minutes (the rate of damaged-layer removal was 60 μm/min). Following the wet etching in Br_2/methanol, the samples were thoroughly rinsed consecutively in methanol, a 1 *N* NaOH solution in methanol for 5 min, again in methanol, in hot distilled water, and finally in cool water, after which they were dried in air.

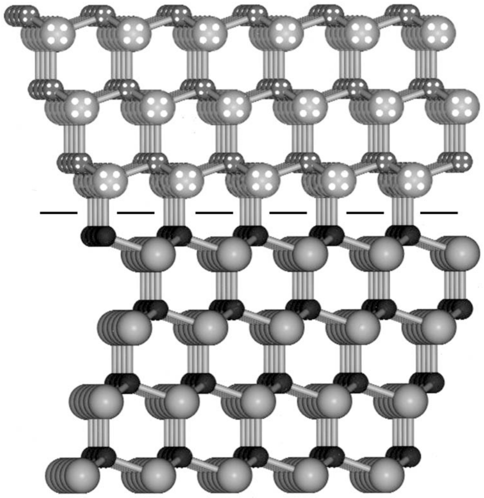


Fig. 3. Atomic model of a twin in a sphalerite-type crystal with a horizontal {111} twin plane perpendicular to the figure plane (schematic). Dashed line shows the twin plane. Larger balls denote atoms of tellurium, and smaller balls denote those of zinc.

To reveal the crystallographic structure, the surface from which the damaged layer was removed was subjected to selective etching in a 12.5 *N* NaOH solution in water at a temperature of 80°C for 15–120 s, with subsequent rinsing in hot and cool distilled water. Selective etching left clearly visible twin boundaries and dislocations on the sample surface, with the dislocation density being no greater than 10^2 cm^{-2}.

3. PHOTOLUMINESCENCE MEASUREMENTS

The photoluminescence (PL) spectra were measured in helium vapor in the temperature range 5–30 K. Optical excitation was provided by an argon and a helium–cadmium laser with pump photon energies of 2.41 eV (5145 Å) and 2.81 eV (4416 Å), respectively. Spectral measurements were performed under excitation of both the natural sample growth face and of a sample after etching. The excitation spot was about 100 μm in size. The spectrum was analyzed using a DFS-24 double-grating monochromator with a resolution of no worse than 0.1 meV. The PM tube output was measured in the photon counting mode.

Figure 4 displays an edge PL spectrum of the growth surface of a single-crystal sample obtained at 5 K. Note that the pattern of a spectrum does not depend on the actual excitation spot. The integrated PL intensities measured under excitation of the growth and etched face of the sample differ only weakly. The spectrum exhibits transitions originating from a free exciton (FE), transitions associated with exciton–impurity complexes, and their phonon (LO) replicas. The exciton-impurity complexes (the ground-state line A^0X) contained Li, Cu, or Ag acceptors. This is indicated by the fairly rich structure of the corresponding two-hole transitions denoted by *FA* in the spectra; here, *F* stands for the final state of the hole localized at the acceptor and *A*, for the acceptor impurity (for instance, $3S_A$Cu). The long-wavelength part of the spectrum (see inset to Fig. 4) exhibits fairly strong radiation associated with the isoelectronic impurity O_{Te}. The very weak transition

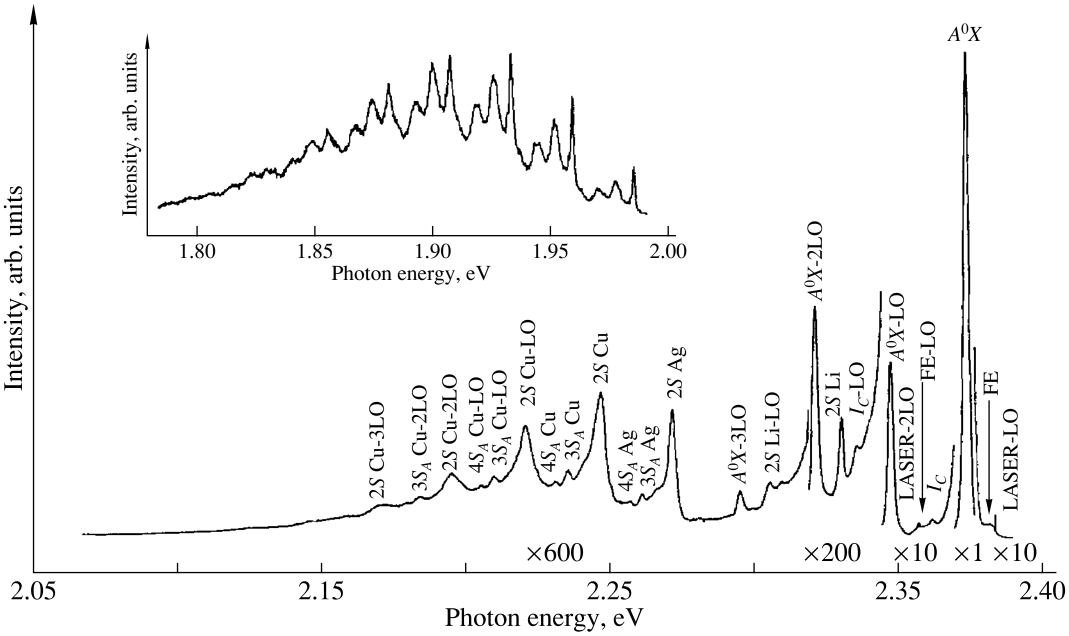


Fig. 4. PL spectrum of a single-crystal sample obtained at $T = 5$ K. Inset shows a fragment of the spectrum in the region of oxygen emission.

near 2.362 eV (I_C) apparently derives from the exciton bound to a doubly charged acceptor, whose manifestation in ZnTe is typical of the given nonequilibrium growth conditions. These doubly charged acceptors can be C_{Te}, Si_{Te}, or the zinc vacancy [6].

Let us consider in more detail the A^0X line, which is slightly broadened (the half-width is ~2.5 meV) and has a long-wavelength wing. A similar effect is observed for acceptor impurity concentrations $N_A \geq 5 \times 10^{15}$ cm^{-3} [7]. Such a concentration of a shallow acceptor impurity also affects the pattern of excitonic PL. At $T = 5$ K, there is practically no emission in the region of the exciton band bottom. As the temperature is raised to 25 K, where exciton ejection from a shallow impurity center becomes efficient, fairly bright PL associated with the polariton appears.

Thus, characterization based on PL suggests that the sample under study is essentially a single crystal of fairly high crystal lattice quality, which is indicated by the high PL quantum yield, the presence of exciton emission (which exhibits a clearly pronounced polariton structure), the distinct two-hole transition pattern, and the absence of any transitions associated with extended defects and complexes. The uniform distribution of impurities (Ag, Li, Cu, O) in substituted sites at a level of 5×10^{15} cm^{-3} is an important characteristic of the sample under study.

A comparison of the PL spectra measured by us earlier on a coarse-grained ZnTe sample (ZT04; the average acceptor concentration N_A is less than 10^{15} cm^{-3}) [2] with the spectrum of this crystallite (ZT05) reveals the following: (i) a higher total acceptor concentration in the crystallite (~5×10^{15} cm^{-3}); (ii) the absence of so-called Y spectral lines, which are usually assigned to the presence of extended defects, such as dislocations and grain boundaries; and (iii) a substantially more intense PL in the exciton region, which attests to high structural perfection of the crystallite.

It also appears instructive to compare the PL spectra of the ZT05 crystallite with the spectra of single crystals (ZT-1) studied in [3]. In addition to the radiation associated with substitutional Ag, Li, Cu, and O, single crystals also exhibit radiation deriving from a donor impurity and a complex with a zinc vacancy. The PL quantum yield in the edge region of single crystals is about one order of magnitude lower than that in the crystallites studied here, and the emission intensity of the latter in the oxygen region is several times weaker. This difference should be attributed to the fact that ZT05 crystallites have a lower oxygen concentration and do not contain the various complex defects characteristic of the single crystals (ZT-1) studied in [3].

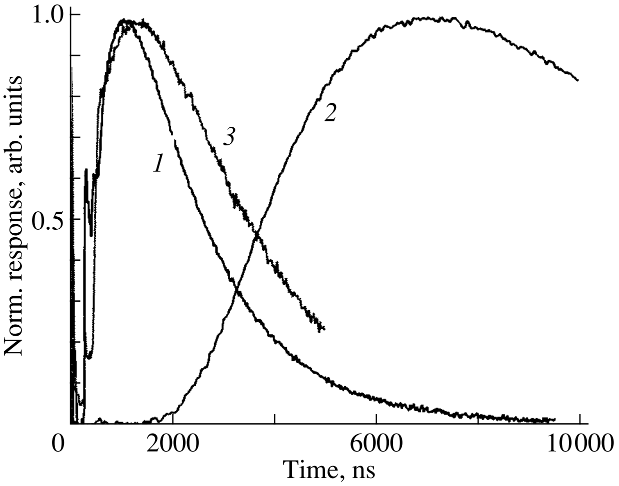


Fig. 5. Normalized bolometer response to the arrival of nonequilibrium acoustic phonons in (*1*) ZT-1 [3], (*2*) ZT04 [2], and (*3*) ZT05 obtained under photoexcitation.

4. STUDY OF NONEQUILIBRIUM HEAT TRANSPORT

Nonequilibrium heat transport was studied using the heat pulse technique in transmission geometry. Nonequilibrium phonons were produced either directly by photoexcitation of the sample surface with nitrogen laser pulses or by deposition of a gold film onto the sample with heating by these pulses. Nonequilibrium phonons were detected by a thin-film superconducting bolometer 50×70 μm in size, which was made of granular aluminum. The schematic diagram of the experiments is similar to that described in [2].

Figure 5 compares experimental bolometer responses to the arrival of nonequilibrium phonons detected in different ZnTe samples, ZT-1, ZT04, and ZT05. We note first that the response at the moment of laser pulse arrival is caused not by bolometer illumination directly by laser light (for instance, following multiple reflections from the helium chamber walls) but rather by the incidence on the bolometer of the luminescence that derives primarily from bound excitons (see Section 3) and has passed through the crystal.

As shown in [2], the ZT04 response to the arrival of nonequilibrium phonons has a clearly pronounced diffusive pattern. Indeed, its rise is smooth and starts not at the ballistic transit time of 200 ns but rather at 1500 ns; the response reaches a maximum at 7.5 μs, and its shape does not depend on the method by which the nonequilibrium phonons were produced or on the excitation energy; and there is no phonon separation in polarization. As demonstrated in [8], these features are characteristic of the case where scattering from extended rather than point defects is dominant. The mean free path in ZT04 was shown in [2] to be 14 μm.

The situation observed with ZT05 is radically different, with the response pattern approaching that revealed

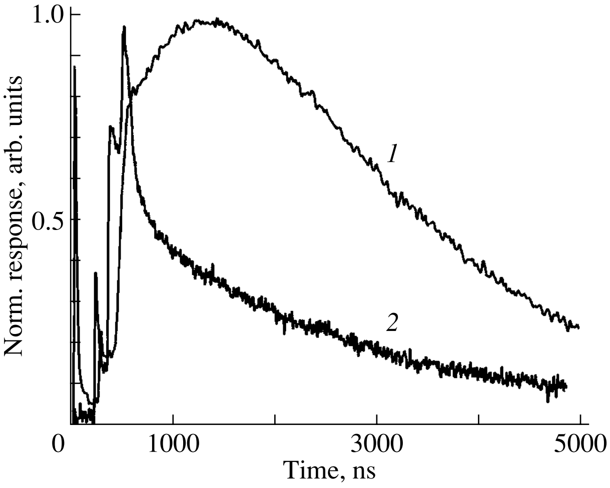


Fig. 6. Bolometer responses to the arrival of nonequilibrium acoustic phonons obtained for ZT05 under (*1*) photoexcitation and (*2*) pulsed heating of a gold film deposited on the sample. The peaks at 190, 340, and 530 ns correspond to the arrival of the LA, FTA, and STA phonons, respectively.

in single-crystal ZnTe. The response is noticeably shorter, it starts at a time consistent with ballistic transport, and differently polarized phonons are clearly detected separately (this suggests the presence of phonons with a mean free path of hundreds of microns). The responses obtained from nonequilibrium phonons generated by a gold film heated using pulsed laser radiation differ from those produced under photoexcitation (Fig. 6), and their shape depends on the excitation energy, which indicates that the frequency-dependent phonon scattering from point defects dominates over the frequency-independent scattering from extended defects.

The strong difference between the responses to nonequilibrium phonons generated by a pulse-heated gold film and those produced by photoexcitation can be understood qualitatively by conceiving these responses to consist, as it were, of two parts for which the difference between the phonon mean free paths is comparable to the sample thickness. One part derives from lower frequency phonons having a large mean free path (a few hundred microns or greater). Depending on the actual conditions, these phonons produce either peaks or a steeply rising front at the times corresponding to the ballistic transit of phonons through the sample. The other part is due to higher frequency phonons. Because of the strong frequency dependence of the phonon mean free time ($\tau = A_{\text{SCAT}}\nu^{-4}$ for elastic scattering from point defects), these phonons feature substantially shorter (on the order of a few tens of microns) mean free paths. These phonons are responsible for the diffusive part of the response. In the case of photoexcitation, the fraction of the high-frequency phonons is larger and, accordingly, the intensity of the diffusive part of the response is higher. Conversely, a heated metal film produces ballistic peaks of a higher intensity, as has been observed for germanium [8].

Calculated relaxation times of nonequilibrium acoustic phonons due to their decay and elastic scattering from isotopes and their mean free path in ZnTe

Parameter	Phonon frequency, THz					
	3.0	2.0	1.5	1.0	0.75	0.5
Decay time, ns	2.8	21.4	90	700	2900	22000
Scattering time, ns	0.14	0.72	2.3	11.4	36	180
Mean free path, μm	0.56	2.8	9.2	45	140	720

If the scattering from extended defects in ZnTe is insignificant, one can carry out an analysis of the kind conducted in [9] for CdTe and derive in this way the quantities characterizing the propagation and evolution of nonequilibrium phonons in ZnTe.

The basic assumption underlying our analysis is as follows: because ZnTe is a material rich in isotopes (zinc has five isotopes, with the main isotope, Zn^{64}, making up 48.6%; tellurium has eight isotopes, with the most abundant of them, Te^{128} and Te^{130}, present in contents of 31.7 and 33.8%, respectively) and the content of impurity atoms is, by contrast, small (and part of them can be distributed over the lattice in the form of precipitates), we assume phonon scattering from the isotopes to be a dominant process. In this case, the scattering intensity can be calculated using the technique described in [10] and it can be found that $A_{\text{SCAT}} = 8.74 \times 10^{-41}$ s^{-3}. A similar assumption was also made in [9] to analyze the responses in CdTe. Taking now the values of the third-order elasticity constants from [11] and calculating the LA phonon decay intensity using the expressions given in [12], we arrive at $A_{\text{LIFE}}(\text{LA}) = 1.46 \times 10^{-54}$ s^{-4}. The corresponding characteristic times and phonon mean free paths with respect to elastic scattering and decay are listed in the table as functions of phonon frequency.

Although these calculations are no more than rough estimates (particularly for the decay processes), they nevertheless allow us to make some conclusions. First, there occur several tens of scattering events for each decay (see table). Second, for a plate thickness of a few hundreds of microns, ballistic peaks may be produced by phonons with frequencies on the order of 0.5 THz or less. Third, most of the LA phonons arriving at times of about 500 ns have a frequency of about 1 THz.

Estimation of the frequency of FTA and STA phonons meets with some difficulties. First, we are not aware of any references in the literature to calculations of FTA and STA phonon decay rates in ZnTe; in any case, however, the intensity of these processes is substantially weaker (see, e.g., [13] for NaF and [14] for silicon). In addition, there is one more mechanism for decreasing the TA phonon frequency, namely, elastic scattering–induced conversion to LA phonons with

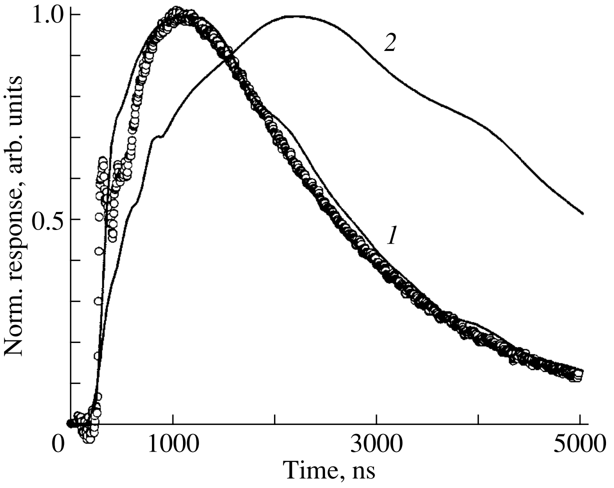


Fig. 7. Comparison of the experimental response of a ZT-1 sample (circles) with Monte Carlo calculations performed for LA phonon decay constant equal to (*1*) 1.46 × 10^{-54} and (*2*) 7 × 10^{-54} s^{-4}.

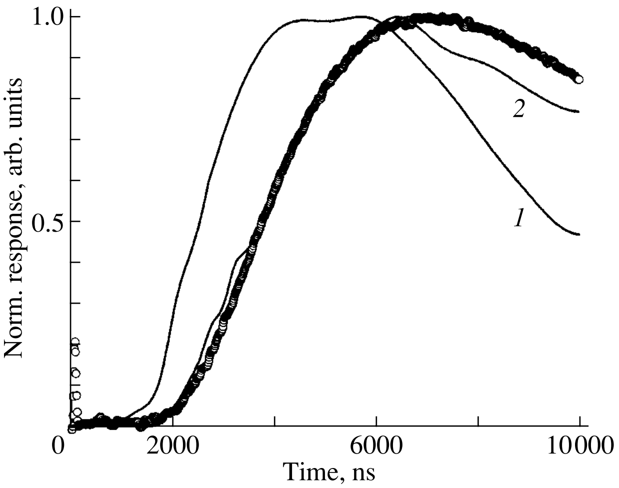


Fig. 8. Comparison of the experimental response of a ZT04 sample (circles) with Monte Carlo calculations performed for various values of the probability of phonon scattering from twin planes: (*1*) 20 and (*2*) 30%.

their subsequent decay. Given the high intensity of scattering processes, we believe that it is this mechanism that is substantial for materials with a high intensity of elastic scattering (e.g., ZnTe, CdTe). However, the density of states of LA phonons (6.3%) is considerably lower than that of STA phonons (61%). Therefore, decreasing the effective frequencies of the STA phonons is a substantially slower process and their average frequency turns out to be higher than that of the LA phonons.

The above estimation of the phonon decay constant should be considered a certain approximation, because the third-order elasticity constants governing the lattice anharmonicity were measured at room temperature and not on samples prepared by our technology. Note that the values of the LA phonon decay constants in silicon and germanium calculated using data for low temperatures were larger by a factor of 1.7–2.5 than those obtained at room temperature [12]. This stimulated Monte Carlo simulation of the nonequilibrium phonon propagation in a ZnTe plate [15], with A_{LIFE}(LA) used as a fitting parameter. It was found possible to fit well the calculated response to experimental response (Fig. 7) with $A_{\mathrm{LIFE}}(\mathrm{LA}) = 7 \times 10^{-54}$ s^{-4}, which differs from the calculated value by a factor of about 5. In the simulation, it was assumed that the STA phonons did not decay. The value of A_{LIFE}(FTA) was assumed to be one tenth of A_{LIFE}(LA). This is certainly an inadequate approach because, as a result of the low density of states of LA phonons, FTA phonon decay may appreciably affect the phonon spectrum evolution. Estimation of this effect shows that A_{LIFE}(FTA) may affect the accuracy of A_{LIFE}(LA) determination by 20–30%.

Having estimated the phonon decay intensity, it becomes possible to calculate the probability of phonon scattering from twin planes in ZT04, as was done in [9] for CdTe. The main difficulty encountered in this analysis is estimating the average plane separation. As pointed out in [2], this distance can vary from one grain to another from a few microns to a few tens of microns. In simulation, this distance was taken to be equal to 5 μm (the corresponding model was described in considerable detail in [16]). Figure 8 compares the results of the calculations with the experimental response of ZT04. The best fit is seen to be achieved for a ~30% average probability of scattering from plane boundaries. This value is substantially larger than the corresponding quantity calculated for phonon scattering from twin planes in CdTe [9] (where it was found to be 4%) and in InTl [17] (about 1%).

We carried out calculations of the reflection and transmission coefficients (including the case of mode conversion) of phonons incident on a single twin boundary in ZnTe, which is shown schematically in Fig. 3. The calculations were performed in the long-wavelength approximation. The room-temperature values of the elasticity tensor were used. The results of the calculations averaged over the angles of incidence of phonons show that the reflection probabilities amount to a fraction of a percent for the LA and FTA phonons and to 2% for STA phonons. A detailed account of these calculations will be published in a later paper.

Thus, we witness a disagreement between the high probability of phonon scattering from twin planes obtained by Monte Carlo simulation of the responses and the analysis of phonon propagation through a single plane. It is conceivable that phonons scatter not only from twins but also from grain boundaries (three to four grains may be encountered in passage across the sample), where the scattering probability is close to unity [16]. Otherwise, the models and approximations employed to estimate phonon transmission through a twin plane system are not adequate.

5. CONCLUSIONS

A study of acoustic-phonon scattering in ZnTe samples with different amounts of impurities, dislocations, and twin systems has shown that, at a concentration of impurities of 10^{15}–10^{17} cm^{-3} and a dislocation density of 10^2–10^4 cm^{-2}, the absence or presence of twin planes separated at a distance of a few tens of microns has almost no effect on the mean free paths, which for these samples are governed by isotopic scattering only. In high-purity, coarse-grained ZnTe with randomly arranged twin systems separated by 1–10 μm, however, scattering from twins is dominant, so the mean free path of acoustic phonons decreases by more than one order of magnitude as compared to that in a twin-free or weakly twinned (albeit more strongly doped) material.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project nos. 02-02-17392, 03-02-16854, and 04-02-17078), the program for support of leading scientific schools (project no. 1923.2003.2), and the program under the Presidium of the RAS “Low-Dimensional Quantum Systems.”

REFERENCES

1. Q. Wu and X.-C. Zhang, Appl. Phys. Lett. **71**, 1285 (1997); A. Leitenstorter, S. Hunshe, J. Shah, M. C. Nuss, and W. H. Knox, Appl. Phys. Lett. **74**, 1516 (1999); M. Schall and Uhd Jepsen, Appl. Phys. Lett. **77**, 2801 (2000); Ajay Nahata, James T. Yardley, and Tony F. Heinz, Appl. Phys. Lett. **81**, 963 (2002).
2. V. S. Bagaev, T. I. Galkina, A. I. Sharkov, A. Yu. Klokov, V. P. Martovitskiĭ, V. V. Zaĭtsev, and Yu. V. Klevkov, Fiz. Tverd. Tela (St. Petersburg) **45** (11), 1941 (2003) [Phys. Solid State **45** (11), 2039 (2003)].
3. T. I. Galkina, A. Yu. Klokov, A. I. Sharkov, Yu. V. Korostelin, and V. V. Zaĭtsev, Fiz. Tekh. Poluprovodn. (St. Petersburg) **37** (5), 539 (2003) [Semiconductors **37** (5), 519 (2003)].
4. Yu. V. Klevkov, V. P. Martovitskĭ, and S. A. Medvedev, Fiz. Tekh. Poluprovodn. (St. Petersburg) **37** (2), 129 (2003) [Semiconductors **37** (2), 119 (2003)].
5. Yu. Yu. Loginov, Pol D. Braun, and Ken D’yurouz, *Regularities of Structural-Defect Formation in II–VI Compound Semiconductors* (Logos, Moscow, 2003) [in Russian].
6. W. S. Kuhn, A. Lusson, B. Quhen, C. Grattepain, H. Dumont, O. Gorochov, S. Bauer, K. Wolf, M. Morz, T. Reisinger, A. Rosenauer, H. P. Wagner, H. Stanzl, and W. Gebhardt, Prog. Cryst. Growth Charact. **31**, 119 (1995).
7. H. Venghaus and P. J. Dean, Phys. Rev. B **21**, 1596 (1980).
8. V. S. Bagaev, Kh. Bel’ska-Levandovska, M. M. Bonch-Osmolovskiĭ, T. I. Galkina, S. Levandovskiĭ, G. N. Mikhaĭlova, A. G. Poyarkov, and G. Yung, Zh. Éksp. Teor. Fiz. **77**, 2117 (1979) [Sov. Phys. JETP **50**, 1013 (1979)].
9. A. I. Sharkov, T. I. Galkina, A. Yu. Klokov, and Yu. V. Klevkov, Fiz. Tverd. Tela (St. Petersburg) **45** (1), 156 (2003) [Phys. Solid State **45** (1), 163 (2003)].
10. S. Tamura, Phys. Rev. B **30** (2), 849 (1984).
11. Devinger Pal Singh, Indian J. Phys. A **62**, 205 (1988).
12. S. Tamura, Phys. Rev. B **31** (4), 2574 (1985).
13. S. Tamura and H. J. Maris, Phys. Rev. B **31** (4), 2595 (1985).
14. A. Berke, A. P. Mayer, and R. K. Wehner, Solid State Commun. **54** (5), 395 (1985).
15. M. M. Bonch-Osmolovskiĭ, T. I. Galkina, A. Yu. Klokov, A. I. Sharkov, and D. V. Kazakovtsev, Fiz. Tverd. Tela (St. Petersburg) **38** (4), 1051 (1996) [Phys. Solid State **38** (4), 582 (1996)].
16. A. I. Sharkov, T. I. Galkina, A. Yu. Klokov, R. A. Khmelnitskiĭ, V. A. Dravin, A. A. Gippius, V. G. Ralchenko, and A. V. Karabutov, Diamond Relat. Mater. **9**, 1100 (2000).
17. L. S. Mitchell and A. C. Anderson, J. Low Temp. Phys. **91** (5–6), 341 (1993).

Translated by G. Skrebtsov

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1073–1075. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1039–1041.
Original Russian Text Copyright © 2005 by Derevyanchuk, Zenkova, Kramar, Nitsovich.

SEMICONDUCTORS AND DIELECTRICS

Magnetooptical Absorption of a Molecular Crystal in the Exciton Frequency Range

A. V. Derevyanchuk, K. Yu. Zenkova, V. M. Kramar, and B. M. Nitsovich[†]
Chernivtsi National University, Chernivtsi, 58012 Ukraine
e-mail: zenkova@itf.cv.ukrtel.net

Received August 27, 2004

Abstract—Nonlinear absorption of a molecular crystal in an external magnetic field is considered. The exciton absorption shape functions under polarized laser radiation and a weak magnetic field, as well as the mechanisms responsible for the formation of a hysteresis loop in the dependence of the output light intensity on the applied magnetic field, were studied for the particular case of benzene. It was established that, for the magnetooptical response of the molecular crystal under study, the formation of bistable loops is inverse in character, which makes it possible to monitor and control the behavior of bistable elements in optical logics systems with an external magnetic field at a fixed illumination frequency. *© 2005 Pleiades Publishing, Inc.*

Optical bistability (OB) has promising application potential in the development of optical information recording and storage systems [1]. This phenomenon is observed under certain conditions of light propagation through a medium, and its essence lies in the fact that, for a fixed value of an input parameter (the intensity of light or an external magnetic field), there are two stable states differing in terms of the output signal intensity I. While the mechanism of OB formation and the medium in which it can be realized may differ [2], designing an arbitrary OB system inevitably meets with the common problem of having to search for a way to control the development of nonlinear effects and to obtain a hysteresis loop at the output of the system with preset parameters.

Molecular crystals are specific in that they can support collective excited states, Frenkel excitons [3]. The energy of such an excitation in one nondegenerate state of an isolated molecule breaks up into bands of excited states (exciton bands), whose number coincides with the number of molecules in the unit cell of the crystal. Each of these bands is characterized by a specific direction of the dipole moment of the exciton transition initiated by an external light wave of a specific polarization. Therefore, the absorption spectrum of polarized light is associated with transitions that involve the states of only one isolated exciton band [4, 5]. By contrast, illumination of a crystal with a beam of natural (unpolarized) light can give rise to excitation of states of several bands, which would be accompanied by the observation of a structured absorption band. Variations in the parameters of the exciting radiation or in other external factors should bring about a transformation of the absorption band shape, which favors the onset of optical multistability.

In this paper, we report on a study of the effect exerted on the exciton absorption band of a molecular crystal by the polarization of exciting radiation and the magnetic field **H**. We consider a molecular crystal whose unit cell contains two molecules. One nondegenerate intramolecular excitation corresponds to two exciton energy bands, $E_1(\mathbf{k})$ and $E_2(\mathbf{k})$, in this crystal. This case can be illustrated by a benzene crystal whose exciton bands are separated by the resonance (Davydov) splitting $\Delta = 40\ \text{cm}^{-1}$ [6]. Assuming a quadratic dispersion law, the energies of excitonic states in each band can be written in the form

$$E_1 = E_0 + Ly^2, \quad E_2 = E_0 + \Delta + \eta L y^2, \tag{1}$$

where E_0 is the bottom of the lower exciton band, L is its width, ηL is the width of the second band, and $0 < y = ka/\pi < 1$ is a dimensionless exciton quasimomentum ranging from 0 to 1 (a is the lattice period).

The absorption coefficient K is determined [7] by the transition probability of the electronic system of the crystal to an excitonic state:

$$K(\omega, \varphi) = 2\pi \sum_{\alpha=1}^{2} D_\alpha^2 S_\alpha(\omega, \varphi), \tag{2}$$

where $S_\alpha(\omega)$ is the shape function of the absorption band associated with the transition of the crystal to the α–e excitonic state and

$$D_\alpha = D_0(\mathbf{e}_0, \mathbf{d}_\alpha) = D_{0\alpha}\cos\varphi_\alpha \tag{3}$$

is the matrix element of electron–photon coupling; its magnitude depends on the angle φ_α the light wave polarization vector $\mathbf{e}_0$ makes with the direction of the corresponding dipole moment $\mathbf{d}_\alpha$. The exciton bands under study can be associated with a pair of different,

---

[†] Deceased.

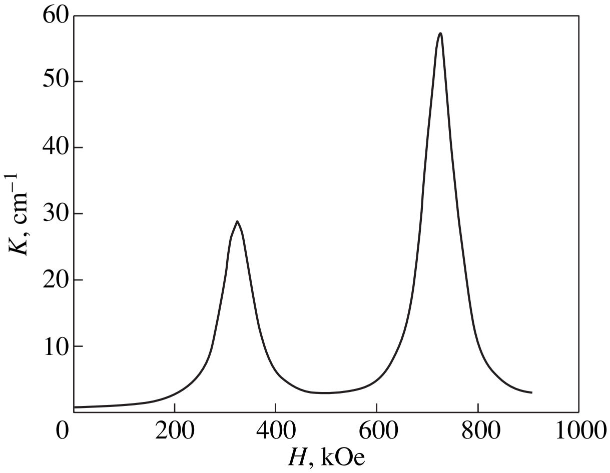


Fig. 1. Absorption coefficient versus magnetic field strength plotted for low exciton densities and the polarization vector azimuth $\varphi = 45°$.

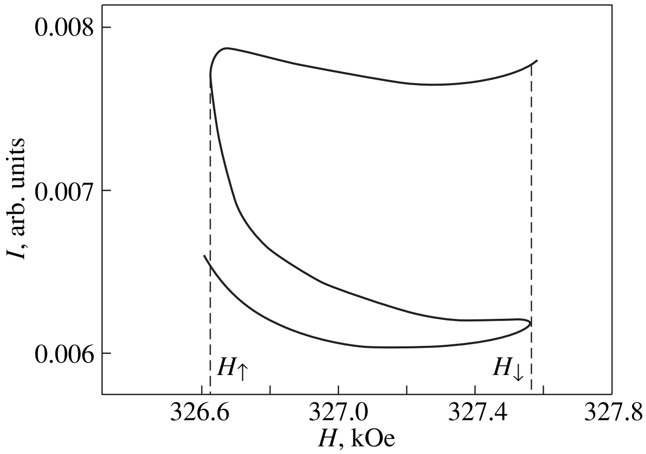


Fig. 2. Magnetooptical bistability of benzene in the exciton frequency range ($\omega = 0.06$).

mutually perpendicular transition dipole moments. Therefore, we can write

$$D_1 = D_0 \cos\varphi, \quad D_2 = \xi D_0 \sin\varphi, \tag{4}$$

where ξ is a parameter taking into account the difference between the values of the dipole moments d_1 and d_2.

It was shown in [8] that the shape function of the exciton absorption band of a crystal in a weak external magnetic field has the form

$$S_\alpha(\omega, h) \equiv \frac{\Gamma_\alpha(\omega_h)}{[\omega_h - E_{\alpha 0} - \Delta_\alpha(\omega_h)]^2 + \Gamma_\alpha^2(\omega_h)}, \tag{5}$$

where Δ_α and Γ_α are the real and imaginary parts of the mass operator of the system, respectively; $\omega = (\hbar\omega - E_0)/L$ is a dimensionless quantity characterizing the excitation frequency w of the exciton state; $\omega_h = \omega - h$; $h = (\delta H^2)/L$; and δ determines the magnetic field–induced diamagnetic shift of an exciton level.

We analyzed the dependence of the exciton absorption band on the orientation of the polarization vector of the incident light wave and an external magnetic field. With the magnetic field kept fixed, we studied the polarization properties of the crystal. When the light polarization vector coincides in direction with the dipole moment $\mathbf{d}_1$ ($\varphi = 0$), an exciton absorption band becomes clearly visible at the frequency of the lower exciton band E_1. Its shape has typical short-wavelength asymmetry. In the case of the perpendicular polarization ($\varphi = \pi/2$), an absorption band is observed at the frequency of the upper exciton band E_2, which differs in terms of both the absorption peak intensity and the shape of the long-wavelength absorption wing (the asymmetry changes sign). For the intermediate orientations of the polarization vector, the exciton band acquires a structured pattern: the spectral response of the absorption coefficient contains both components of the Davydov doublet.

An increase in the magnetic field strength causes parallel translation of the absorption band to higher energies. Therefore, the optical properties of the crystal undergo frequency selection: at a fixed frequency and arbitrary orientation of the polarization vector of the light beam ($0 < \varphi < \pi/2$), the $K(H)$ dependence exhibits a structure with a pair of clearly pronounced maxima. Thus, by properly varying the magnetic field strength, one can produce an effect equivalent to that of varying the incident light frequency. Because the exciton level shifts toward higher energies as the magnetic field increases, one can fully reconstruct the frequency dependence of the exciton absorption coefficient $K(\omega)$ of a crystal illuminated in the short-wavelength wing of the absorption band by properly varying the magnetic field strength. The results of calculations performed for crystalline benzene using the characteristic parameter values ($L = 10^3$ cm^{-1}, $\Omega_0 = 120$ cm^{-1}, $\eta = 0.7$, $\xi = 0.83$) are plotted in Fig. 1.

Intense laser pumping favors the formation of a variable-density exciton gas in a crystal. The exciton absorption coefficient (depending on the external magnetic field) and the exciton density N (governed by the input signal intensity I_0) are related by [7]

$$N = I_0 K(\omega, H, \varphi). \tag{6}$$

By fixing the input parameters (the intensity of the incident wave, its frequency and polarization) and using the Bouguer–Lambert law, one can determine the dependence of the output intensity on the external magnetic field, $I(H)$. The results of calculations performed for $\varphi = 45°$ are presented in Fig. 2. We can see a finite interval of magnetic field strength $\Delta H = H_\downarrow - H_\uparrow$ within which magnetooptical bistability (MOB) can exist. Note that optical information can be recorded and read with an MOB-based element by properly varying the external magnetic field. This is the main feature distinguishing the MOB phenomenon from ordinary optical bistability.

Note that, for the parameters chosen by us in the calculation performed here, the interval within which MOB exists is very narrow, which implies a fairly high sensitivity of MOB elements to magnetic field strength. Furthermore, the field range over which the MOB exists does not depend on the input signal frequency ω. Increasing the latter by an amount $\Delta\omega$ produces a completely identical MOB hysteresis loop, which is observed already at different values of the magnetic field, $H' = H + \Delta\omega$. In addition, the direction of motion along a MOB hysteresis loop is the reverse of that for the frequency-driven OB [2]. Thus, we obtain essentially a reversed mechanism of MOB formation, in which the crystal state is switched from strong to weak absorption by decreasing the magnetic field. In the case of ordinary OB, this is achieved by increasing the input signal frequency, which assumes the availability of a tunable laser. The MOB-based arrangement is free of this shortcoming.

REFERENCES

1. H. Gibbs, *Optical Bistability: Controlling Light with Light* (Academic, New York, 1985; Mir, Moscow, 1988).
2. B. S. Ryvkin, Fiz. Tekh. Poluprovodn. (Leningrad) **19** (1), 3 (1985) [Sov. Phys. Semicond. **19** (1), 1 (1985)].
3. A. S. Davydov, *Theory of Molecular Excitons* (Ukrainian Academy of Sciences Press, Kiev, 1951; McGraw-Hill, New York, 1962).
4. V. L. Broude, V. S. Medvedev, and A. F. Prikhot’ko, Zh. Éksp. Teor. Fiz. **2**, 379 (1957) [Sov. Phys. JETP **2**, 317 (1957)].
5. N. I. Ostapenko, V. I. Sugakov, M. P. Chernomorets, and M. T. Shpak, Pis’ma Zh. Éksp. Teor. Fiz. **27** (8), 452 (1978) [JETP Lett. **27** (8), 423 (1978)].
6. V. L. Broude and M. N. Onoprienko, Zh. Opt. Spectra **10**, 634 (1961).
7. A. S. Davydov, *Solid State Theory* (Nauka, Moscow, 1976) [in Russian].
8. K. Yu. Zenkova, A. A. Zenchenko, and B. M. Nitsovich, Fiz. Tverd. Tela (St. Petersburg) **43** (1), 19 (2001) [Phys. Solid State **43** (1), 17 (2001)].

Translated by G. Skrebtsov

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1076–1082. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1042–1047.
Original Russian Text Copyright © 2005 by Nogay, Young Huh, Yugay.

Ionic and Superionic Conduction in NASICON-Like Structures of the $Na_3Sc_2(PO_4)_3$ Type

A. S. Nogay[1], Young Huh[2], and K. N. Yugay[2, 3]

[1] *Seĭfullin Agricultural University, Astana, 473000 Kazakhstan*
e-mail: nogay@mail.kz
[2] *Korea Electrotechnology Research Institute, Seoul, 437-082 Korea*
e-mail: yhuh@keri.re.kr
[3] *Omsk State University, Omsk, 644077 Russia*
e-mail: yugay klimenty@mail.ru

Received January 15, 2004; in final form, July 7, 2004

Abstract—A model is proposed for ionic and superionic conduction in the α, β, and γ phases of the $Na_3Sc_2(PO_4)_3$ compound. This model accounts not only for the conducting properties of the aforementioned phases but also for their structural and kinetic parameters. Within the kinetic approach to the description of the electrical conduction in the $Na_3Sc_2(PO_4)_3$ compound, it is established that the ionic conduction in the α dielectric phase is governed by the localization and delocalization of ions in deep potential wells and that hoppings between the wells are caused by the interaction of ions with phonons. The superionic conduction in the β and γ phases is associated both with the increase in the contribution of the ion delocalization in potential wells due to a decrease in their depth and with the decrease in the contribution of the ion localization. These contributions are responsible for the flight mechanism of conduction in the superionic phases. The contribution of each conducting channel to the total conduction is taken into account through the relaxation time of charge carriers. The estimated relaxation times are close to the values determined from the data on the conducting and dielectric characteristics of the α, β, and γ phases of the $Na_3Sc_2(PO_4)_3$ compound.

1. INTRODUCTION

Superionic conductors have found wide application as elements in systems for data processing and as power sources. This has stimulated extensive investigations into the physical properties of superionic conductors of different structural types with the aim of elucidating the mechanism responsible for superionic conduction [1, 2].

As present, there is a wealth of information on the structural parameters and the conducting and dielectric properties of the superionic conductor $Na_3Sc_2(PO_4)_3$ and its analogs $Na_3M_2(PO_4)_3$ (M = Fe, Cr) belonging to the NASICON family [3–11]. However, the mechanism of superionic conduction in compounds of this family is still not clearly understood. In this respect, the search for adequate tools for studying the nature of this phenomenon is an important problem. The purpose of the present work was to construct a kinetic model that would be entirely appropriate for describing both ionic and superionic conduction in phosphate compounds with a rhombohedral crystal lattice of the $Na_3Sc_2(PO_4)_3$ type.

2. QUALITATIVE ANALYSIS

The ionic conductivity of the α-$Na_3Sc_2(PO_4)_3$ dielectric phase increases initially by two orders of magnitude upon the transition to the β superionic phase and then, again, by two orders of magnitude upon the transition to the γ superionic phase [4, 9]. As a result of two reversible phase transitions at $T_{\alpha \to \beta}$ = 339 K and $T_{\beta \to \gamma}$ = 439 K, the conductivity of this compound reaches an extremely high value of 2.5×10^{-2} $(\Omega\ \mathrm{cm})^{-1}$ at T = 570 K. The conducting properties and the structural characteristics of sodium scandium phosphate can be described within an energy model that accounts for the specific features of the occupation of small-sized (A) and large-sized (B) holes by compensating cations in the crystal lattice. According to Kalinin *et al.* [11], the potential relief of a conducting channel revealed for the $\{[Sc_2(PO_4)_3]^{3-}\}_{3\infty}$ rhombohedral crystal lattice in the α, β, and γ phases of the $Na_3Sc_2(PO_4)_3$ compound can be schematically drawn as profiles of one-dimensional conducting channels. Note that, to a regular alternation of the A and B holes in the crystal lattice, there corresponds an alternation of the A shallow and B deep wells in the potential relief (Fig. 1). The ferroelectric properties [4] and the low conductivity of the α-$Na_3Sc_2(PO_4)_3$ phase can be associated with the fact that the $\{[Sc_2(PO_4)_3]^{3-}\}_{3\infty}$ mobile rhombohedral crystal lattice is characterized by monoclinic distortions responsible for the nonuniform occupation of B holes by sodium cations with an asymmetric double-well potential [11]. In this case, sodium cations involved in the

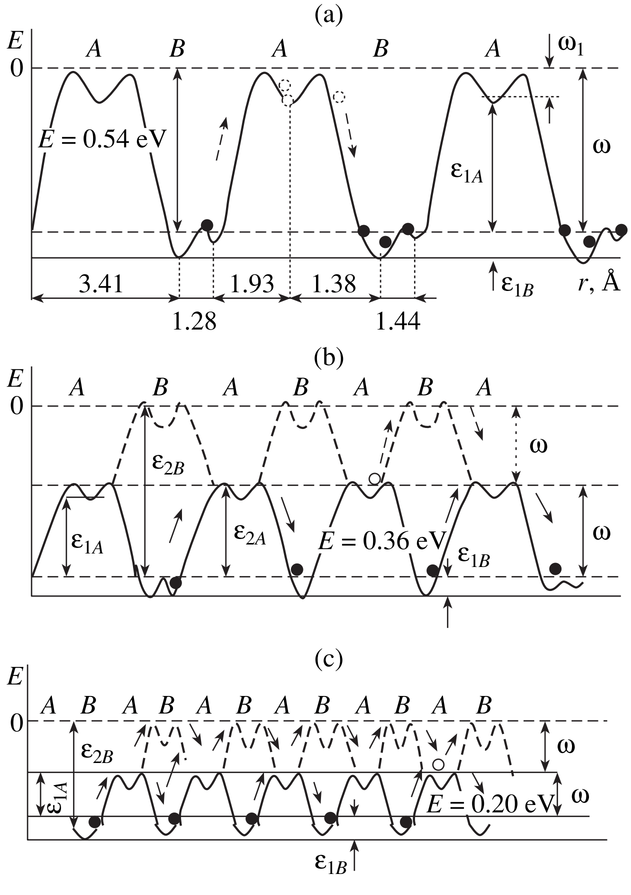


Fig. 1. Schematic representation of the one-dimensional model for the potential relief of a conducting channel in (a) the α ferroelectric phase, (b) the β superionic phase, and (c) the γ superionic phase of the $Na_3Sc_2(PO_4)_3$ compound. Closed circles indicate sodium ions that occupy deep potential wells *B* and participate in hopping conduction. Open circles represent sodium ions that occupy shallow potential wells *A* and are involved in flight conduction. Solid lines show channels of hopping conduction. Dashed lines represent potential barriers corresponding to the channels of flight conduction. Arrows indicate the mechanical trajectories of conduction ions.

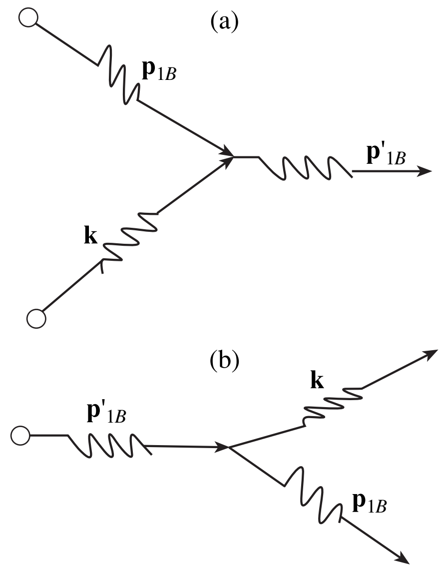


Fig. 2. Schemes illustrating the scattering of sodium ions by phonons: (a) phonon absorption by a sodium ion and (b) phonon emission by a sodium ion.

formation of random sodium dipoles can also serve as charge carriers in the presence of an external electric field. The most appropriate potential relief of the conducting channel, with due regard for the available data on the structure and conduction of the α-$Na_3Sc_2(PO_4)_3$ phase [3–11], is depicted in Fig. 1a. It can be seen from this figure that the condensation of conduction cations at the bottom of deep potential wells and the high potential barriers in the conducting channel between the *A* and *B* holes of the crystal lattice provide conditions typical of the hopping conduction mechanism, as is the case in the majority of dielectric materials.

Within the semicrystalline approach underlying our analysis of the ionic conduction, the motion of ions is accompanied by their interaction with phonons. In particular, this interaction results in the localization of ions in potential wells with their subsequent escape from these wells. The well depth should be small enough for an ion absorbing a thermal phonon to escape from the well. From this standpoint, the conduction in the α-$Na_3Sc_2(PO_4)_3$ phase can be treated as a migration of sodium ions along the conducting channel that is attended by the ion–phonon interaction leading to the absorption and emission of phonons by ions. Certainly, not any interaction of an ion with a phonon results in the localization of this ion in a hole or its escape from the hole. Nonetheless, it is these processes that are characteristic of the conduction in the α-$Na_3Sc_2(PO_4)_3$ phase [10]. This conduction will be referred to as the hopping conduction by analogy with the hopping conduction of electrons in semiconductors. However, in the case of electrons acting as charge carriers in semiconductors, hoppings occur through electron tunneling from one state to another state. The term “hopping conduction mechanism” is used to describe the electron hopping. In the case of ionic conduction, the probability of tunneling an ion from one well to another well is very low and hoppings occur according to the classical scenario: an ion localized in a well absorbs a phonon with an energy no less than the well depth, escapes from the well, and moves freely to a collision, after which the ion emits a phonon and is localized in another well. Thus, in our case, the above term is only a convenient name for a specific ion motion.

The ion–phonon interaction can be illustrated by the scheme depicted in Fig. 2. Let us assume that a sodium ion with momentum $\mathbf{p}_B$ and energy ε_B, which is located at the bottom of the *B* potential well, executes hopping from this well due to the absorption of a phonon with momentum $\mathbf{q}$ and energy ω, as is shown in Fig. 2a. The changes in the momentum $\mathbf{p}_B \longrightarrow \mathbf{p}$ and the energy $\varepsilon_B \longrightarrow \varepsilon_p$ of the interacting ion obey the laws of conservation of momentum and energy. Upon absorption of the phonon with momentum $\mathbf{q}$ and energy ω, we have

$$\mathbf{p}_B + \mathbf{q} = \mathbf{p}, \quad (1a)$$

$$\varepsilon_B + \omega = \varepsilon_p, \quad (1b)$$

where $\varepsilon_B < 0$, $\omega > 0$, and $\varepsilon_p > 0$.

For an ion with energy ε_p and momentum $\mathbf{p}$ to pass into the *B* well, the ion must emit a portion of the energy. In this case, we can write the relationships

$$\mathbf{p} = \mathbf{p}_B + \mathbf{q}, \quad (2a)$$

$$\varepsilon_p = \varepsilon_B + \omega, \quad (2b)$$

which describe the process inverse with respect to the process corresponding to expressions (1a) and (1b). Relationships similar to expressions (1a), (1b), (2a), and (2b) can also be written for *A* wells. We can assume that there occur the transitions *A*–*A*, *B*–*B*, and *A*–*B* between the wells. However, it should be taken into account that not just any interaction with the phonon will correspond to a process [described by relationships (2a) and (2b)] resulting in localization of the ion in the *A* well. In the channel associated with the *A* wells, the escape of the ion from one well and its localization in another well can be separated by many collisions until one of these collisions with the phonon will coincide with the process corresponding to relationships (2a) and (2b). Note that the *A* well with the localized ion should not necessarily be adjacent to the *B* well from which this ion has escaped. Therefore, the localization and delocalization of ions in *A* and *B* wells are random processes. For a further analysis, it is important to emphasize that, between two sequential localization events in the channels associated with the *A*–*A*, *B*–*B*, and *A*–*B* wells, the ion can experience many collisions with phonons. It should also be emphasized that ions involved in the current form a nondegenerate gas. Moreover, we should note that the number of ions in a free state due to hoppings initiated by phonons is important rather than the type of wells between which the ion executes hoppings.

Unlike the α phase of the $Na_3Sc_2(PO_4)_3$ compound, the conduction mechanism in the β and γ superionic phases changes as a result of structural transformations upon the phase transition and thus provides an anomalously high conductivity. The $\alpha \longrightarrow \beta$ phase transition leads to an almost complete disappearance of monoclinic distortions of the crystal lattice [3]. This results in a lowering of the potential relief of the conducting channel in the β-$Na_3Sc_2(PO_4)_3$ phase, the destruction of random sodium dipoles, and a more uniform occupation of *A* and *B* holes by sodium cations (Fig. 1b). According to [4, 6], three-fourths of sodium ions occupy the *B* wells and only one-fourth of the ions occupy the *A* wells. In Fig. 1b, superstructure distortions in the β-$Na_3Sc_2(PO_4)_3$ phase [3, 5] are represented by ordered states of sodium cations in *B* holes of the crystal lattice. It should be noted that sodium cations located in the *B* deep wells, as before, can participate in the hopping conduction, and a decrease in the barrier height favors this process. However, the energy depth of the *A* shallow wells is close to zero due to the disappearance of monoclinic distortions. As a result, the probability of transition of these cations to a free state due to the phonon absorption increases considerably. Furthermore, since the *A* wells almost disappear, the flight length between two sequential localization events in the *B* wells increases and a decrease in the depth of the *B* wells apparently leads to a decrease in the localization time. Therefore, the conduction in the β phase in the electric field proceeds through a mechanism that can be called the flight mechanism. This conducting channel is absent in the α-$Na_3Sc_2(PO_4)_3$ phase. Most likely, it is this flight mechanism that is responsible for the increase in the conductivity by two orders of magnitude upon the $\alpha \longrightarrow \beta$ phase transition.

A further jump in the conductivity upon transition from the β phase to the γ-$Na_3Sc_2(PO_4)_3$ phase [9] can be associated with the complete disappearance of monoclinic distortions in the crystal structure [5]. This results in a new decrease in the potential barrier height and provides a regular relief of the conducting channel and a uniform occupation of *A* and *B* holes by all sodium cations (see Fig. 1c). The ratio between the concentrations of sodium cations occupying the *A* and *B* wells in the γ phase remains identical to that in the β-$Na_3Sc_2(PO_4)_3$ phase. In this case, as in the β-$Na_3Sc_2(PO_4)_3$ phase, the conduction can proceed according to both the hopping (for sodium ions in the *B* wells) and flight (for sodium ions in the *A* wells) mechanisms. It should be noted that, in both cases, the efficiency of the conduction mechanism in the γ phase substantially increases (i) as a result of the decrease in the height of the potential barrier between the *A* and *B* wells and the disappearance of ordered states for hopping conduction and (ii) due to the decrease in the depth of the *A* wells for flight conduction.

Consequently, the conduction in superionic conductors of this type occurs through the hopping and flight mechanisms. It is not ruled out that the conduction in other superionic conductors can proceed by a similar mechanism. This is confirmed by the data presented in [1, 2] for the α-AgI compound, according to which the ratio between the flight time τ_f and the residence time τ_s of silver ions in potential wells is of the order of unity. However, in order to reveal the contribution of each mechanism to the total superionic conduction, it is necessary to analyze the kinetic and conduction parameters.

3. THE KINETIC DESCRIPTION

Now, we turn from the aforementioned qualitative analysis to the kinetic description. As is known, a system can be described by kinetic equations when it is a stochastic system. Undeniably, the system of ions under investigation is stochastic because, first, it is a many-particle system and, second, ions of this system interact with thermal phonons whose distribution obeys the Bose–Einstein statistics. In an external electric field, the mean velocity of directed motion of charged carriers is often considerably lower than their thermal velocity at reasonable temperatures. However, such a ratio between the directed and thermal velocities is not a necessary requirement for the applicability of the stochastic approach, but if this ratio holds true, we are always dealing with a stochastic system. As was noted above, the hopping mechanism of ionic conduction follows the classical scenario, according to which ions appear to be in wells through collision of free ions with phonons rather than through tunneling, as is the case with electrons in semiconductors.

Let us consider the kinetic equation with due regard for the drift of conduction ions and their collisions with phonons in the external electric field **E**; that is,

$$\frac{\partial f}{\partial t} + \frac{\mathbf{p}}{m}\frac{\partial f}{\partial \mathbf{r}} + e\mathbf{E}\frac{\partial f}{\partial \mathbf{p}} = I_c(\mathbf{p}), \qquad (3)$$

where $f \equiv f(\mathbf{r}, \mathbf{p}, \mathbf{t})$ is the distribution function of conduction ions and $I_c(\mathbf{p})$ is the ion–phonon collision integral. The collision integral will be written under the following assumptions.

(i) Conduction ions with each other are ignored, because their density is two or three orders of magnitude lower than the density of lattice ions.

(ii) It is assumed that conduction ions obey the classical statistics. This is justified taking into account their mass.

(iii) The interaction of conduction ions with other particles of the crystal lattice is considered in terms of the interaction between ions and phonons that are governed by quantum statistics.

It is important to note that, as will be shown below, disregarding the collisions between conduction ions leads to a linear collision integral. It should also be noted that the conduction in the given case is provided by free ions. The collision integral is determined by the balance of conduction ions in a phase volume element $d\mathbf{p}d\mathbf{r}$ in the vicinity of the point $(\mathbf{p}, \mathbf{r})$: (1) ions enter this volume from the *A* and *B* wells upon absorption of the corresponding phonons and upon interaction with phonons of other phase volumes, and (2) ions leave this volume upon emission of phonons either to be localized in the *A* and *B* wells or to pass into other phase volumes. The scheme illustrating this balance is shown in Fig. 3. The collision integral takes the form

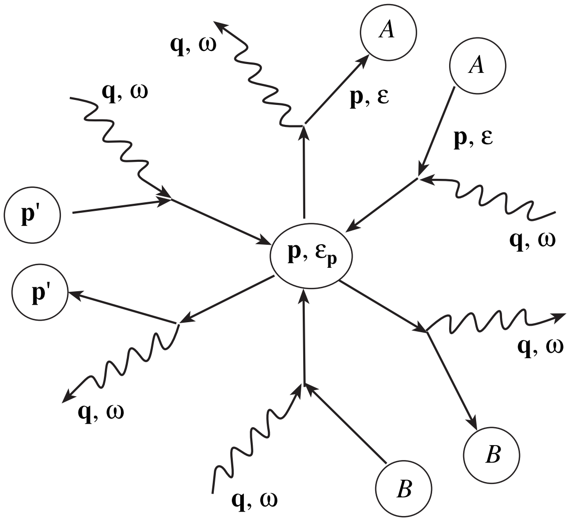


Fig. 3. Scheme illustrating the ion balance in the phase volume $d\mathbf{p}d\mathbf{r}$.

$$\begin{aligned} I_c(\mathbf{p}) = {} & \sum_{q,\, p_A}[Z(\mathbf{p}_A, \mathbf{q}; \mathbf{p})N_A n(\mathbf{q})\delta(\varepsilon_A + \omega - \varepsilon_{\mathbf{p}}) \\ & - W(\mathbf{p}; \mathbf{p}_A, \mathbf{q})f(\mathbf{p})\delta(\varepsilon_{\mathbf{p}} - \varepsilon_A - \omega)] \\ & + \sum_{q,\, p_B}[Z(\mathbf{p}_B, \mathbf{q}; \mathbf{p})N_B n(\mathbf{q})\delta(\varepsilon_B + \omega - \varepsilon_{\mathbf{p}}) \\ & - W(\mathbf{p}; \mathbf{p}_B, \mathbf{q})f(\mathbf{p})\delta(\varepsilon_{\mathbf{p}} - \varepsilon_B - \omega)] \\ & + \sum_{q}\Big[\int d\mathbf{p}' Z(\mathbf{p}', \mathbf{q}; \mathbf{p})n(\mathbf{q})f(\mathbf{p}')\delta(\varepsilon_{\mathbf{p}'} + \omega - \varepsilon_{\mathbf{p}}) \\ & - \int d\mathbf{p}' W(\mathbf{p}; \mathbf{p}', \mathbf{q})f(\mathbf{p})\delta(\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}'} - \omega)\Big]. \end{aligned} \qquad (4)$$

Here, $Z(\mathbf{p}_A, \mathbf{q}; \mathbf{p})$, $Z(\mathbf{p}_B, \mathbf{q}; \mathbf{p})$, and $Z(\mathbf{p}', \mathbf{q}; \mathbf{p})$ are the coefficients proportional to the probabilities of entering the phase volume element $d\mathbf{p}d\mathbf{r}$ in the vicinity of the point $(\mathbf{p}, \mathbf{r})$ upon absorption of phonons with momentum **q** and energy ω by an ion with momentum $\mathbf{p}_A$ and energy ε_A in the *A* well, an ion with momentum $\mathbf{p}_B$ and energy ε_B in the *B* well, and a free ion with momentum $\mathbf{p}'$ and energy $\varepsilon_{\mathbf{p}'}$, respectively; $W(\mathbf{p}; \mathbf{p}_A, \mathbf{q})$, $W(\mathbf{p}; \mathbf{p}_B, \mathbf{q})$, and $W(\mathbf{p}; \mathbf{p}', \mathbf{q})$ are the probabilities of leaving the phase volume element $d\mathbf{p}d\mathbf{r}$ in the vicinity of the point $(\mathbf{p}, \mathbf{r})$ upon emission of phonons with momentum **q** and energy ω and localization of an ion with momentum $\mathbf{p}_A$ and energy ε_A in the *A* well, localization of an ion with momentum $\mathbf{p}_B$ and energy ε_B in the *B* well, and transition of an ion to the free state with momentum $\mathbf{p}'$ and energy $\varepsilon_{\mathbf{p}'}$, respectively; N_A and N_B are the numbers of *A* and *B* wells per unit volume, respectively; and $n(\mathbf{q})$ is the phonon distribution function. This function will be treated as an equilibrium function by assuming that the current motion of ions is a weak perturbation of the system. In relationship (4), the δ functions imply the conservation of energy in the corresponding interaction processes. Hereafter, for simplicity, we will consider the stationary, spatially homogeneous case.

In the collision integral (4), the first, third, and fifth terms are associated with the transfer of the ions to the phase volume element $d\mathbf{p}d\mathbf{r}$ in the vicinity of the point $(\mathbf{p}, \mathbf{r})$ due to the phonon absorption from the A well, B well, and other phase volume elements, respectively. The second, fourth, and sixth terms describe the escape of ions due to the emission of phonons from the above phase volume element to the A well, B well, and other phase volume elements, respectively. The characteristic feature of the collision integral (4) is that it is linear with respect to the ion distribution function $f(\mathbf{p})$. This is explained by ignoring the direct interaction of conduction ions with each other. The interaction between conduction ions occurs through their interaction with phonons.

By combining the terms independent of the distribution function $f(\mathbf{p})$ and proportional to this function in relationship (4), the collision integral can be rewritten in the following form:

$$I_c(\mathbf{p}) = \sum_{\mathbf{q}}[Q_{\mathbf{q}} - R_{\mathbf{q}} f(\mathbf{p})], \tag{5}$$

where

$$\begin{aligned} Q_{\mathbf{q}} &= \sum_{\mathbf{p}_A} Z(\mathbf{p}_A, \mathbf{q}; \mathbf{p}) N_A n(\mathbf{q}) \delta(\varepsilon_A + \omega - \varepsilon_{\mathbf{p}}) \\ &+ \sum_{\mathbf{p}_B} Z(\mathbf{p}_B, \mathbf{q}; \mathbf{p}) N_B n(\mathbf{q}) \delta(\varepsilon_B + \omega - \varepsilon_{\mathbf{p}}) \\ &+ \int d\mathbf{p}' Z(\mathbf{p}', \mathbf{q}; \mathbf{p}) n(\mathbf{q}) f(\mathbf{p}') \delta(\varepsilon_{\mathbf{p}'} + \omega - \varepsilon_{\mathbf{p}}), \end{aligned} \tag{6}$$

$$\begin{aligned} R_{\mathbf{q}} &= \sum_{\mathbf{p}_A} W(\mathbf{p}; \mathbf{p}_A, \mathbf{q}) \delta(\varepsilon_{\mathbf{p}} - \varepsilon_A - \omega) \\ &+ \sum_{\mathbf{p}_B} W(\mathbf{p}; \mathbf{p}_B, \mathbf{q}) \delta(\varepsilon_{\mathbf{p}} - \varepsilon_B - \omega) \\ &+ \int d\mathbf{p}' W(\mathbf{p}; \mathbf{p}', \mathbf{q}) \delta(\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}'} - \omega). \end{aligned} \tag{7}$$

Under equilibrium conditions, the collision integral is equal to zero; that is,

$$I_c(\mathbf{p}) = 0. \tag{8}$$

Hence, we immediately obtain

$$Q_{\mathbf{q}} = f_0(\mathbf{p}) R_{\mathbf{q}}. \tag{9}$$

Substitution of expression (9) into the collision integral (5) gives the relationship

$$I_c(\mathbf{p}) = -\frac{f(\mathbf{p}) - f_0(\mathbf{p})}{\tau}, \tag{10}$$

where the quantity

$$\tau = \left(\sum_{\mathbf{q}} R_{\mathbf{q}}\right)^{-1} \tag{11}$$

has the meaning of the momentum relaxation time. Therefore, the approximation used for the collision integral (4) (disregarding collisions between conduction ions), in essence, is the relaxation time approximation (τ approximation).

In the spatially homogeneous case, the current density j in a uniform electric field within the τ approximation is determined by the quantity $f_1(\mathbf{p}) = f(\mathbf{p}) - f_0(\mathbf{p})$ (see, for example, [12]). In this case, we have

$$f_1(\mathbf{p}) = eE\tau \frac{\partial f_0(\mathbf{p})}{\partial p_x} \tag{12}$$

and $j_x = \sigma E$, where the conductivity σ (under the assumption that the electric field E is aligned with the x axis) is given by the expression

$$\sigma = \frac{n_i e^2 \tau}{m_i}. \tag{13}$$

Here, n_i is the ion concentration, m_i is the mass of conduction ions, and the relaxation time is described by formula (11). Since the conductivity is determined by the relaxation time and the concentration of conduction ions, we consider in more detail relationship (11) and the processes responsible for the change in the ion concentration n_i. We introduce the following designations:

$$\frac{1}{\tau_A} = \sum_{p_A} W(\mathbf{p}; \mathbf{p}_A, \mathbf{q}) \delta(\varepsilon_{\mathbf{p}} - \varepsilon_A - \omega), \tag{14}$$

$$\frac{1}{\tau_B} = \sum_{p_B} W(\mathbf{p}; \mathbf{p}_B, \mathbf{q}) \delta(\varepsilon_{\mathbf{p}} - \varepsilon_B - \omega), \tag{15}$$

$$\frac{1}{\tau_{\mathbf{p}'}} = \int d\mathbf{p}' W(\mathbf{p}; \mathbf{p}', \mathbf{q}) \delta(\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}'} - \omega). \tag{16}$$

It is evident that the quantities τ_A, τ_B, and $\tau_{\mathbf{p}'}$ have the meaning of the corresponding relaxation times in the case when there is, for example, only the channel associated with the A wells without the other two channels, etc. Then, as can been easily seen, the total relaxation time τ is defined by the relationship

$$\frac{1}{\tau} = \frac{1}{\tau_A} + \frac{1}{\tau_B} + \frac{1}{\tau_{\mathbf{p}'}} \tag{17a}$$

or

$$\tau = \frac{\tau_A \tau_B \tau_{\mathbf{p}'}}{\tau_A \tau_B + \tau_A \tau_{\mathbf{p}'} + \tau_B \tau_{\mathbf{p}'}}. \qquad (17b)$$

Hence, it follows that the total relaxation time is governed by the shortest time among the relaxation times τ_A, τ_B, and $\tau_{\mathbf{p}'}$. It is clear that all three phases (α, β, and γ phases) of the $Na_3Sc_2(PO_4)_3$ compound are characterized by the hierarchy of the relaxation times $\tau_{\mathbf{p}'} < \tau_A < \tau_B$. For the α-$Na_3Sc_2(PO_4)_3$ phase in which the A and B wells are sufficiently deep, the total relaxation time τ is determined by the relaxation time $\tau_{\mathbf{p}'}$. However, the ion concentration n_i is governed only by the conducting channels associated with the A and B wells. The equation for the concentration n_i of conduction ions can be written in the form

$$\frac{dn_i}{dt} = k_A N_A n_q - K_A n_i + k_B N_B n_q - K_B n_i, \qquad (18)$$

where k_A, k_B, K_A, and K_B are the rate constants of ion escape from the A and B wells upon phonon absorption and the rate constants of ion localization in the A and B wells upon phonon emission, respectively. These rate constants are determined by the corresponding probabilities Z and W of the processes. In particular, the rate constants k_A and K_A can be represented in the form

$$k_A = \frac{1}{n_i} \sum_{\mathbf{p}_A, \mathbf{q}} \int d\mathbf{p} Z(\mathbf{p}_A, \mathbf{q}; \mathbf{p}) n(\mathbf{q}) \delta(\varepsilon_A + \omega - \varepsilon_{\mathbf{p}}),$$

$$K_A = \frac{1}{n_i} \sum_{\mathbf{p}_A, \mathbf{q}} \int d\mathbf{p} W(\mathbf{p}; \mathbf{p}_A, \mathbf{q}) f(\mathbf{q}) \delta(\varepsilon_{\mathbf{p}} - \varepsilon_A - \omega).$$

Similar relationships can be written for the rate constants k_B and K_B. The stationary solution of Eq. (18) has the form

$$n_{i0} = \frac{k_A N_A + k_B N_B}{K_A + K_B}. \qquad (19)$$

Hereinafter, this stationary concentration n_{i0} [defined by expression (19)] rather than the concentration $n_i$will be used in formula (13) for the conductivity σ. According to [9], the conductivity of the α phase at $T = 290$ K is determined to be $\sigma_\alpha = 2.6 \times 10^{-5}$ $(\Omega \text{ cm})^{-1}$ ($n_{i0} \approx 10^{22}$ cm^{-3}). Then, the total relaxation time for this phase can be estimated as follows:

$$\tau = \frac{\sigma_\alpha m}{n_{i0} e^2} = 2.6 \times 10^{-7} \text{ s}. \qquad (20)$$

This estimate is close to the relaxation time $\tau = 8.5 \times 10^{-7}$ s, which was experimentally determined from the dielectric characteristics $(\tan\delta(\log\omega))$ [13]. The total relaxation time in relationship (20) is close to the relaxation time $\tau_{\mathbf{p}'}$.

In [9], the conductivity of the β phase at $T = 350$ K was found to be $\sigma_\beta = 1.9 \times 10^{-3}$ $(\Omega \text{ cm})^{-1}$. As a result, the total relaxation time τ in the β phase is given by

$$\tau = \frac{\sigma_\beta m}{n_{i0} e^2} = 1.9 \times 10^{-5} \text{ s}. \qquad (21)$$

This relaxation time is close to the relaxation time $\tau = 1.7 \times 10^{-5}$ s obtained from the dielectric measurements [13]. Consequently, the relaxation times for the α and β phases of the $Na_3Sc_2(PO_4)_3$ compound differ by two orders of magnitude. This difference can be explained by both the decrease in the contribution of the ion localization in the A wells and the decrease in the depth of the B wells in the β-$Na_3Sc_2(PO_4)_3$ phase. Therefore, we can make the inference that the relaxation processes associated with the interaction in the **p**' channel are steady-state processes.

In the β phase, the relaxation processes are controlled by the slower processes in the A and B wells. The relaxation time τ_β is defined as follows:

$$\frac{1}{\tau_\beta} = \frac{1}{\tau_A} + \frac{1}{\tau_B}. \qquad (22)$$

Since the hierarchy between the relaxation times τ_A and τ_B is obvious ($\tau_A < \tau_B$), estimate (21) corresponds to the relaxation time τ_A. From the viewpoint of the conduction mechanism in the β-$Na_3Sc_2(PO_4)_3$ phase, we can argue that the flight mechanism is initiated by the considerable decrease in the depth of the A wells.

In the γ superionic phase of the $Na_3Sc_2(PO_4)_3$ compound, the depth of the A and B wells becomes even smaller as compared to the β phase, the localization of sodium cations in the A wells is insignificant, and the relaxation process associated with the channel of B wells proceeds when the processes governed by the other two channels are in equilibrium. At $T = 570$ K, we have the conductivity $\sigma_\gamma = 2.5 \times 10^{-2}$ $(\Omega \text{ cm})^{-1}$ and the relaxation time

$$\tau = \frac{\sigma_\gamma m}{n_{i0} e^2} = 2.5 \times 10^{-3} \text{ s}. \qquad (23)$$

Most likely, this is the relaxation time τ_B. In the γ superionic phase of the $Na_3Sc_2(PO_4)_3$ compound, the flight mechanism of conduction becomes even more efficient.

It is evident that a similar mechanism of conduction should be observed in sodium iron and sodium chromium phosphates, whose composition, properties, and structure are similar to those of the $Na_3Sc_2(PO_4)_3$ compound.

REFERENCES

1. *Physics of Superionic Conductors*, Ed. by M. B. Salamon (Zinantne, Riga, 1982) [in Russian].

2. Yu. Ya. Gurevich and Yu. I. Kharkats, *Superionic Conductors* (Nauka, Moscow, 1992) [in Russian].
3. B. I. Lazoryak, V. B. Kalinin, S. Yu. Stefanovich, and V. A. Efremov, Dokl. Akad. Nauk SSSR **254** (4), 1096 (1980).
4. S. Yu. Stefanovich, V. B. Kalinin, and A. Nogai, Ferroelectrics **55** (1–4), 325 (1983).
5. M. de la Rochere, F. d’Yvoire, G. Collin, R. Comes, and J. P. Boilot, Solid State Ionics **9–10**, 825 (1983).
6. M. d’Yvoire, E. Pintard-Screpel, E. Bretey, and M. de la Rochere, Solid State Ionics **9–10**, 851 (1983).
7. E. P. Haritonova, V. I. Voronkova, V. K. Yanovskii, S. Yu. Stefanovich, and A. V. Mosunov, Ferroelectrics **221** (1), 73 (1999).
8. T. Yu. Losevskaya, E. P. Kharitonova, V. I. Voronkova, V. K. Yanovskiĭ, S. Yu. Stefanovich, N. I. Sorokina, and V. I. Simonov, Kristallografiya **44** (1), 95 (1999) [Crystallogr. Rep. **44** (1), 90 (1999)].
9. A. S. Nogay, Vestn. Omsk. Univ. **2**, 19 (2001).
10. A. S. Nogay and K. N. Yugay, Vestn. Omsk. Univ. **4**, 26 (2002).
11. V. B. Kalinin, S. Yu. Stefanovich, and A. Nogay, Neorg. Mater. **22** (1), 107 (1986).
12. A. I. Ansel’m, *Introduction to the Theory of Semiconductors* (Nauka, Moscow, 1978).
13. A. S. Nogay, D. G. Kirchev, and M. D. Chaĭka, Vestn. Nauki Akmol. Agr. Univ. **11** (5), 313 (2000).

Translated by O. Borovik-Romanova

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1083–1087. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1048–1052.
Original Russian Text Copyright © 2005 by Vorob'ev, Ekhanin, Nesmelov.

SEMICONDUCTORS AND DIELECTRICS

Electrical Breakdown in Solid Dielectrics

G. A. Vorob'ev, S. G. Ekhanin, and N. S. Nesmelov
Tomsk State University of Control Systems and Radio Electronics, Tomsk, 634050 Russia
e-mail: gemma@main.tusur.ru

Received October 12, 2004

Abstract—The mechanism of electrical breakdown in solid dielectrics is analyzed using the results of our investigations performed in this direction over a period of several decades. It is shown that the electrical breakdown in solid dielectrics involves interrelated prebreakdown processes, such as high-voltage polarization, defect formation, electron impact excitation and electron impact ionization of luminescence centers and ions in the host crystal lattice, etc. The electrical breakdown is initiated by electric-field and thermal generation of defects in the crystal. In turn, the generation of defects leads to the formation of defect regions and channels that provide an assisted transfer of charge carriers. Electron currents flow (and electrons are accelerated by the electric field to energies sufficient to induce impact ionization) in these regions of the crystal with a lattice distorted by defects. In this respect, the known approaches to the elaboration of the breakdown theory for alkali halide and other dielectric crystals on the basis of analyzing the motion and acceleration of electrons in an ideal crystal structure have appeared to be incorrect. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

There exist three types of breakdown in solid dielectrics, namely, thermal, electrical, and electrochemical breakdown. The processes occurring in the course of thermal and electrochemical breakdown are known in sufficient detail, but the same cannot be said of the electrical breakdown in solid dielectrics. Let us briefly review the main stages of research into the mechanism of electrical breakdown in solid dielectrics. In [1–3], the authors considered a number of fundamental theoretical problems associated with the electrical breakdown in solid dielectrics. In his monograph [2], Skanavi pointed out the first three well-justified classical theories of electrical breakdown in solid dielectrics. These are the Rogovskiĭ theory (breakdown of an ionic crystal in an electric field), the Ioffe theory (impact ionization by ions), and the Smurov theory (breakdown due to the detachment of electrons from atoms). However, these theories led to overestimated breakdown strengths and, thus, poorly represented the facts. Nonetheless, the Ioffe theory was of considerable importance in the further development of the concept of electrical breakdown in solid dielectrics, because the theory of impact ionization accounts for an electrical strengthening, i.e., an increase in the breakdown strength with a decrease in the electrode spacing. This effect has been confirmed by the experimental data on the breakdown in gases. The first experimental works seemingly allowed the conclusion that thin layers of solid dielectrics undergo electrical strengthening. However, in later experiments performed by Aleksandrov and colleagues [4, 5], who more correctly measured the thickness of the layer subjected to breakdown, this inference was not confirmed. Quite possibly, as was noted in [5], the breakdown developed in "weak" regions of the samples; hence, the electrical strengthening could not be revealed. The mechanism of electrical breakdown in solid dielectrics due to electron impact ionization remained more intriguing for researchers than the mechanism associated with ion impact ionization, because the mean free path of ions at a field strength of 10^6 V cm^{-1} should be greater than 100 lattice constants, which is impossible. Moreover, there have been experimental data that confirm the electron hypothesis of the breakdown mechanism. Valter and Inge [6, 7] revealed channels of incomplete breakdowns and short breakdown times (10^{-7} s) in alkali halide crystals, which lent support to the electron mechanism of electrical breakdown in solid dielectrics. Hippel [8–10] also obtained the experimental results (current instability prior to breakdown, copper ions introduced into alkali halide crystals remain fixed at their sites, development of a discharge from an anode, etc.), which counted in favor of the electron mechanism. In the Franz' opinion [1], Hippel proved that the electrical breakdown in solid dielectrics occurs through the mechanism of electron impact ionization. However, direct evidence for this mechanism was lacking.

2. DEPENDENCE OF THE ELECTRICAL BREAKDOWN IN SOLID DIELECTRICS ON THE CATHODE MATERIAL. ELECTRICAL STRENGTHENING

In [11], it was established that the breakdown strength E_{br} of rock salt and ebonite at dc and pulsed voltages does not depend on the material of metal electrodes (lead, potassium, silver) or on the presence of electronic and ionic surface charges. Hippel and Alger [12] found that the breakdown strength E_{br} of the KBr

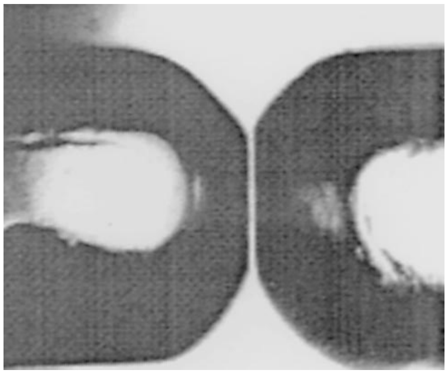


Fig. 1. Micrograph of an alkali halide crystal layer with a thin dielectric interlayer ~5 μm thick between holes.

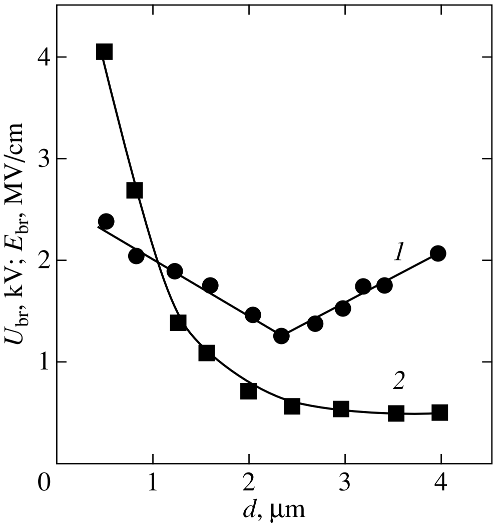


Fig. 2. Dependences of (*1*) the breakdown voltage U_{br} and (*2*) the breakdown strength E_{br} on the thickness d of the NaCl crystal layer.

crystal with gold electrodes is 60% less than the breakdown strength of the KBr crystal with electrolyte electrodes and that the KBr crystal with an electrolyte cathode has a higher value of E_{br}. Subsequent investigations demonstrated that the breakdown strengths E_{br} of solid dielectrics with metal and graphite electrodes are less than those with electrolyte electrodes. These results turned out to be very important for the development of the theory of electrical breakdown in solid dielectrics, because they made it possible to reveal new features of this phenomenon.

The experiments performed by Aleksandrov and colleagues [4, 5] suggested that the electrical breakdown could be conveniently studied with layers located inside single crystals rather than with thin films. In this case, coaxial holes in crystal plates should be mechanically drilled (to a small depth). Then, these holes should be deepened with an aqueous solution of the corresponding salt to a required thickness of the dielectric layer (Fig. 1). In more recent works [13], the layer thickness was reduced with the use of water–ethanol solutions. This made it possible to produce dielectric layers with a smoother surface. The holes thus prepared either were filled with graphite or an electrolyte or were used to evaporate metal electrodes. According to [14], the breakdown strength E_{br} of NaCl crystal samples with 0.3-μm-thick layers (this thickness corresponds to the resolution of the optical microscope) and with electrolyte electrodes amounted to approximately $\sim 10^{10}$ V m^{-1} at the breakdown probability $\psi = 90\%$. This value is considerably higher than the breakdown strength of thicker layers ($d \sim 0.10$–0.15 mm) (Fig. 2).

Thus, it was proved that electrical breakdown is accompanied by electrical strengthening. Moreover, it can be seen from Fig. 2 that the dependence of the breakdown voltage on the thickness of the dielectric layer $U_{br} = f(d)$ resembles the Paschen curve in the case of gas breakdown.

3. ULTRASTRONG ELECTRIC FIELDS

One consequence of the electrical strengthening of solid dielectrics with a decrease in their thickness is noteworthy. Thin dielectric layers at electric field strengths higher than the breakdown strength of thick layers can exhibit phenomena that cannot be observed in thicker layers due to the onset of the breakdown. Investigations have demonstrated that, in micrometer-sized layers of alkali halide crystals, it is possible to observe new phenomena, such as the generation of additional dislocations, impact ionization currents, and electroluminescence (intrinsic and activator emission). These (and other) phenomena, which can be discovered in studying thin layers of solid dielectrics, have been the subject matter of a new direction in the physics of dielectrics, namely, the physics of ultrastrong electric fields. This important problem was discussed in our recent review [15] and will not be considered in detail in the present paper. Note only that investigations into the activator electroluminescence in alkali halide crystals have proved the impact mechanism of excitation of an activator whose luminescence involves photons with energies higher than 6 eV. As the activator concentration increases, the intensity of activator luminescence initially increases and then decreases. The decrease in the luminescence intensity can be explained by the fact that an increase in the activator concentration leads to an increase in the frequency of scattering collisions of electrons with activator ions. This results in an increase in the mean free path of electrons within which they accumulate the energy necessary for the impact excitation of activator ions. Furthermore, upon deposition of a semitransparent metallic film on the anode surface, the emission of hot electrons through the anode into vacuum is observed in ultrastrong electric fields. These observations have proved that the electron impact ionization and the electron impact excitation of both ions in the host crystal lattice and impurity ions play a decisive role in the prebreakdown processes and the breakdown itself. However, it should be noted that noticeable emission of hot electrons into vacuum occurs only in the case where the crystal structure of the dielectrics undergoes progressively enhanced transformations upon application of sequential voltage pulses. Similar transformations in the crystal structure of an alkali halide dielectric have been observed when the crystal exhibits electroluminescence. These experimental data indicate that, under the given conditions, there occur intensive degradation processes. It is evident that electrons are differently accelerated in regions with destroyed and undestroyed crystal lattices. In this respect, it is necessary to investigate the defect formation in alkali halide crystals in prebreakdown electric fields and to elucidate how this process affects the electrical conductivity and elementary electro-optical events.

4. DEFECT FORMATION AND CHARGE TRANSFER IN MICROLAYERS OF ALKALI HALIDE CRYSTALS

Earlier [16], we revealed that, in a strong (prebreakdown) electric field, dislocations are generated in microlayers of alkali halide crystals prior to the onset of the breakdown. According to [17], the increase in the number of dislocations depends on the electric field strength and the exposure time. The effect of the electric field on the dislocation structure of the crystal has a threshold nature. The shorter the exposure time, the higher the field strength required to generate new dislocations. The application of an electric field of the same strength (alternately, with direct and reversed polarity) retards the dislocation generation. As the temperature decreases, the onset of intensive generation of dislocations shifts toward stronger fields. New dislocation etch pits are predominantly formed in regions where dislocation outcrops have already been observed on the crystal surface even before the application of an electric field. Therefore, the dislocation generation in response to an electric field is caused by the multiplication of dislocations that exist in the crystal prior to the application of the electric field.

In [18], it was shown that, in a strong electric field, new dislocations are generated on the cathode surface of dielectrics. The nonuniform distribution of dislocations over the bulk of the crystal layer and the absence of pronounced slip bands after exposure to the electric field indicate that an important role in the dislocation generation is played by the electrostatic interaction of the electric field with charged dislocation jogs and with the system of point defects. These defects are redistributed in the layer and soften the cathode region. The difference between the etch patterns of the cathode and anode surfaces suggests that the majority of generated dislocations are half-loops emerging on the cathode surface. In this case, the mechanism of dislocation multiplication can involve double transverse sliding of screw dislocation portions with the formation of new dislocation loops.

Investigations into the fundamental absorption spectra of thin layers of alkali halide crystals before and after exposure to an electric field have demonstrated that, in a strong field, the effective band gap of the crystal decreases [19]. This manifests itself in a shift of the fundamental absorption edge toward the low-energy range. Most likely, the additional absorption responsible for this shift is associated with both the optical generation of excitons and the absorption of light by point defects in the vicinity of dislocations induced by the strong electric field.

5. DEFECT FORMATION AND BREAKDOWN IN MICROLAYERS OF ALKALI HALIDE CRYSTALS

In ultrastrong electric fields, the breakdown in alkali halide crystals is preceded initially by the dislocation generation (a component of the dark pause) and then by the electroluminescence. As is known [20], the charge transfer along dislocation lines occurs more readily. The vacancy mobility in these channels is considerably higher than that in other crystal regions. Owing to the outflow of cation vacancies to the anode (in these channels), the local field strength in the vicinity of the cathode microapex should increase and stimulate the injection of electrons into channeling regions.

Visual examinations of the dislocation etch patterns after exposure of the layer to a pulsed ultrastrong electric field $E > (2.5\text{–}3.0) \times 10^8$ V m^{-1}, which induces an electron current and electroluminescence, revealed regions (spots) with an increased density and larger sizes of etch pits on the surface of the layer (Fig. 3). The application of additional pulses of an electric field of the same strength leads to an increase in the sizes of these regions; as a result, the regions with an increased density of etch pits can cover the entire surface of the layer.

Moreover, visual inspections of the luminescent spots and the etch patterns revealed that local currents flow in the regions with an increased density of etch pits. According to calculations, the local currents should cause an increase in the temperature and give rise to mechanical stresses in the vicinity of conducting channels [21]. Owing to the heat shock mechanism of defect formation, these effects should bring about an increase in the diameter of channels with an increased electrical conductivity.

Electron microscope examinations of the regions characterized by a local charge transfer in NaCl crystals revealed the occurrence of thermoplastic phenomena that lead not only to loosening of the material in the charge transfer channels but also to a transfer of the material to the surface (Fig. 4) [21, 22]. The cross sec-

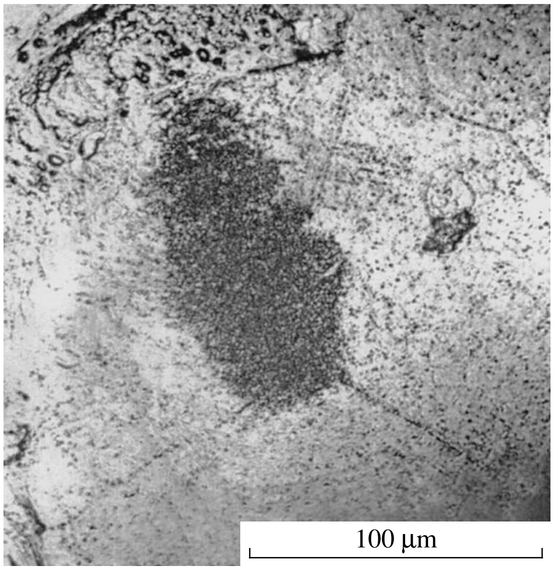


Fig. 3. Dislocation pattern on the surface of an alkali halide crystal layer after the onset of electroluminescence. The spot length is approximately equal to 100 μm.

Fig. 4. Microdamages on the anode surface of the NaCl crystal layer in the vicinity of the charge transfer channel.

tion of these channels with a modified crystal structure proves to be several orders of magnitude larger than the initial cross section. The use of electrolyte electrodes in the experiments limits the rate of rise and the magnitude of through currents in these channels of the dielectric and provides time matching of the electron and hole currents [21]. As a result, the alkali halide crystal regions with a modified structure exhibit electroluminescence due to impact excitation of luminescence centers by hot electrons. The electroluminescence is accompanied by impact ionization and is characterized by a quasi-stationary distribution of the potential throughout the layer thickness, as well as by a constant current amplitude and luminescence intensity [21]. A variation in the voltage within certain limits has no effect on this state. The electroluminescence at dc voltages and small overvoltages lasts for several tens of milliseconds at room temperature and for a few seconds at –30°C. At ac voltages and low temperatures, the electroluminescence can last for a few hours.

It should be noted that electroluminescence is not observed in alkali halide crystals with metal electrodes, because the stage of decrease in the breakdown strength during the development of the electrical breakdown immediately gives way to the stage of fracture of dielectrics. For electrolyte electrodes, the electric discharge attended by the electroluminescence can be observed over a long time. This situation is similar to electric discharges in gases. High-resistance electrolyte electrodes play a stabilizing role in the development of electrical breakdown due to the voltage redistribution in local regions of the layer with an increase in the electric current.

6. CONCLUSIONS

Thus, it has been established that the motion of charged defects and the defect formation precede and stimulate all the basic prebreakdown processes occurring in microlayers of alkali halide crystals. The electrical breakdown is caused by the electron impact ionization. However, electron currents accompanied by the impact ionization flow in local regions of the dielectric layer in which the electric field generates linear and point defects. In turn, local passage of the electron current brings about an enhanced local generation of defects due to the thermoplastic phenomena. This stimulates further development of the electrical breakdown.

Therefore, the electrical breakdown observed in microlayers of alkali halide crystals is associated with the electron impact ionization. This breakdown is characterized by a number of specific features.

(1) The discharge occurs in channels with a loose crystal structure due to the field-induced generation of linear and point defects.

(2) In samples with electrolyte electrodes, the discharge can last for a long time (the stage of catastrophic fracture of dielectrics is absent). As a result, the development of the electrical breakdown in alkali halide crystals resembles the development of a gas discharge.

We believe that, in thick layers of alkali halide crystals and other dielectrics with metal electrodes, all the processes revealed in the study of thin layers of alkali halide crystals occur simultaneously and stimulate each other. As a consequence, the fracture of the dielectric occurs at the final stage of the electrical breakdown.

It is obvious that the energy spectrum of electron–hole states of the channels providing flow of the electron current accompanied by impact ionization and the energy spectrum of the crystal prior to exposure to an electric field should differ significantly due to the very high concentration of linear and point defects, which increases with an increase in the electric-field strength.

Therefore, the known approaches to the construction of the breakdown theory for alkali halide crystals on the basis of analyzing the motion and acceleration of electrons in an ideal crystal structure cannot offer satisfactory results.

REFERENCES

1. W. Franz, *Dielektrischer Durchschlag* (Springer-Verlag, Berlin, 1956; Inostrannaya Literatura, Moscow, 1961).
2. G. I. Skanavi, *Physics of Dielectrics (Region of Strong Fields)* (Fizmatgiz, Moscow, 1958) [in Russian].
3. A. A. Vorob’ev and G. A. Vorob’ev, *Electrical Breakdown and Fracture of Solid Dielectrics* (Vysshaya Shkola, Moscow, 1966) [in Russian].
4. A. P. Aleksandrov and A. F. Ioffe, Zh. Tekh. Fiz. **3** (1), 32 (1933).
5. A. P. Aleksandrov and A. M. Zolotareva, Zh. Tekh. Fiz. **4**, 429 (1934).
6. A. F. Val’ter and L. D. Inge, Elektrichestvo **6**, 83 (1930).
7. L. D. Inge and A. F. Valter, Z. Phys. **64**, 830 (1930).
8. A. Hippel, Z. Phys. **67**, 707 (1931).
9. A. Hippel, Z. Phys. **68**, 309 (1931).
10. A. Hippel, Naturwissenschaften **14**, 79 (1935).
11. A. A. Vorob’ev and A. K. Krasin, Zh. Tekh. Fiz. **7**, 15 (1937).
12. A. Hippel and R. S. Alger, Phys. Rev. **76**, 127 (1949).
13. G. A. Vorob’ev and I. S. Pikalova, Prib. Tekh. Éksp., No. 1, 198 (1967).
14. I. S. Pikalova, Candidate’s Dissertation (Tomsk State Univ., Tomsk, 1968).
15. G. A. Vorob’ev, S. G. Ekhanin, and N. S. Nesmelov, Izv. Vyssh. Uchebn. Zaved., Fiz. **8**, 26 (2000).
16. G. A. Vorob’ev, S. G. Ekhanin, M. M. Milyushina, and N. S. Nesmeyanov, Fiz. Tverd. Tela (Leningrad) **15**, 2545 (1973) [Sov. Phys. Solid State **15**, 1695 (1973)].
17. N. I. Lebedev and N. S. Nesmelov, Fiz. Tverd. Tela (Leningrad) **14**, 1282 (1972) [Sov. Phys. Solid State **14** (4), 1106 (1972)].
18. S. G. Ekhanin, N. S. Nesmelov, and E. V. Nefedtsev, Kristallografiya **35** (1), 237 (1990) [Sov. Phys. Crystallogr. **35**, 145 (1990)].
19. S. G. Ekhanin, N. S. Nesmelov, and N. V. Nefedtsev, Izv. Vyssh. Uchebn. Zaved., Fiz. **3**, 105 (1990).
20. A. A. Vorob’ev, Available from VINITI No. 1802-81 (1981), Vol. 2.
21. S. G. Ekhanin, Doctoral Dissertation (State Pedagog. Univ., Tomsk, 2002).
22. S. G. Ekhanin, N. S. Nesmelov, and E. V. Nefedtsev, Fiz. Tverd. Tela (Leningrad) **32** (2), 409 (1990) [Sov. Phys. Solid State **32** (2), 235 (1990)].

Translated by O. Borovik-Romanova

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1088–1090. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1053–1055.
Original Russian Text Copyright © 2005 by G. Masharov, S Masharov.

DEFECTS, DISLOCATIONS, AND PHYSICS OF STRENGTH

Complexes of Impurity Atoms in Diluted Substitutional Solid Solutions

G. S. Masharov and S. I. Masharov
Ural State Technical University, ul. Mira 19, Yekaterinburg, 620002 Russia

Received July 20, 2004

Abstract—The equilibrium concentration of complexes composed of two B atoms is calculated for an A–B diluted substitutional solid solution with a body-centered cubic lattice. The conditions under which the fraction of the complexes becomes significant are determined. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

It is generally believed that, in A–B diluted substitutional solid solutions with a low concentration of one of the components, for example, B, the impurity atoms are randomly distributed over the lattice. However, by analogy with interstitial solid solutions that are characterized by the formation of complexes, i.e., aggregates consisting of several interstitial atoms [1], it can be assumed that similar complexes exist in substitutional solid solutions.

As a rule, the possibility of forming complexes is estimated by calculating the energies of states for individual atoms and atomic complexes with inclusion of the elastic lattice distortions arising from the difference between the radii of atoms of different types; i.e., the complex formation is analyzed from the viewpoint of the energy minimum [1]. In this paper, it will be shown that, even without regard for geometrical distortions of the lattice, complexes in substitutional solid solutions under specific conditions are thermodynamically equilibrium aggregates.

2. THEORETICAL ANALYSIS

We consider an A–B substitutional solid solution with a body-centered cubic lattice that involves the components A and B at atomic concentrations c_A and c_B, respectively. (In what follows, c_α will stand for the atomic concentration of the component α.) It is assumed that $c_B \ll c_A$. Let n_1 be the number of B atoms randomly distributed over the crystal and n_2 be the number of complexes formed by B atoms in the solid solution. In this case, the free energy of the solid solution can be represented in the form

$$F = E_{\mathrm{conf}} - kT\ln W_1 W_2, \tag{1}$$

where E_{conf} is the configurational energy of the system, and W_1 and W_2 are the thermodynamic probabilities of the distributions of individual atoms and atomic complexes in the crystal, respectively. Under the assumption that the complexes are noncrossing, the thermodynamic probabilities can be written as

$$W_1 = \frac{(N_A + n_1 - 2(z-1)n_2)!}{n_1!(N_A - 2(z-1)n_2)!}, \tag{2}$$

$$W_2 = \frac{\left(\frac{z}{2}N\right)!}{n_1!\left(\frac{z}{2}N - n_2\right)!}, \tag{3}$$

where $N = N_A + N_B$, N_α is the total number of atoms of the α type in the solid solution, and z is the coordination number of the lattice. It is obvious that W_2 is the number of possible variants of the distribution of n_2 complexes over $zN/2$ nearest neighbor pairs in the solid solution.

Let $\nu_{\alpha\beta}$ be the energy of interaction between α–β nearest neighbor pairs with the minus sign. Then, the configurational energy of the system E_{conf} can be written in the form

$$\begin{aligned} E_{\mathrm{conf}} = {} & -(\nu_{BB} + 2(z-1)\nu_{AB})n_2 \\ & -\left(\frac{zN}{2} - n_2 - 2(z-1)n_2\right) \\ & \times(\nu_{AA}\bar{c}_A^2 + \nu_{BB}\bar{c}_B^2 + 2\nu_{AB}\bar{c}_A\bar{c}_B). \end{aligned} \tag{4}$$

Here, we introduced the following designations:

$$\begin{aligned} \bar{c}_A &= \frac{N_A - 2(z-1)n_2}{N_A + n_1 - 2(z-1)n_2}, \\ \bar{c}_B &= \frac{n_1}{N_A + n_1 - 2(z-1)n_2}. \end{aligned} \tag{5}$$

The quantities n_1 and n_2 are taken as thermodynamic variables of the problem. These quantities are related by the expression

$$n_1 + 2n_2 = N_B. \tag{6}$$

1063-7834/05/4706-1088$26.00 © 2005 Pleiades Publishing, Inc.

The equilibrium equations can be derived from the conditions of the extremum of the function $\Phi = F + \lambda(n_1 + 2n_2 - N_B)$ with respect to the variables n_1 and n_2 (where λ is the Lagrange factor). These equations have the form

$$\frac{z}{2}(\nu_{AA}\bar{c}_A^2 + \nu_{BB}\bar{c}_{BB}^2 + 2\nu_{AB}\bar{c}_A) + kT\left\{\ln\frac{n_1}{N_A + n_1 - 2(z-1)n_2} + \frac{z}{2}\ln\frac{zN_A + zn_1 - 2(z-1)n_2}{z(N_A + n_1 + 2n_2)}\right\} + \lambda = 0, \tag{7}$$

$$\begin{aligned}&(z-1)(\nu_{AA}\bar{c}_A^2 + \nu_{BB}\bar{c}_B^2 + 2\nu_{AB}\bar{c}_A\bar{c}_B)\\ &-2(z-1)\bar{c}_B(-\nu_{AA}\bar{c}_A + \nu_{BB}\bar{c}_B + \nu_{AB}(\bar{c}_A - \bar{c}_B))\\ &-\nu_{BB} - 2(z-1)\nu_{AB}\\ &+kT\left\{\ln\frac{n_2}{1/2zN_A + 2n_1 + 2(z-1)n_2} + 2(z-1)\ln\frac{N_A - 2(z-1)n_2}{N_A + n_1 - 2(z-1)n_2} + z\ln\frac{z(N_A + n_1 + 2n_2)}{zN_A + 2n_1 + 2(z-1)n_2}\right\} + 2\lambda = 0.\end{aligned} \tag{8}$$

In order to solve Eqs. (7) and (8), we take into account the conditions $\frac{n_1}{N_A} = c_1 \ll 1$ and $\frac{2n_2}{N_A} = c_2 \ll 1$. Then, in expressions (7) and (8), we discard all (except logarithmic) terms containing c_1 and c_2. As a result, we have

$$c_1 = \exp\left\{-\frac{z}{2kT}(\nu_{AA} - 2\nu_{AB})\right\}x, \tag{9}$$

$$c_2 = \exp\left\{-\frac{(z-1)\nu_{AA} - \nu_{BB} - 2(z-1)\nu_{AB}}{kT}\right\}x^2, \tag{10}$$

where $x = \exp(-\lambda/kT)$. It follows from relationship (6) that x can be found from the equation

$$c'_B = c_1 + zc_2, \tag{11}$$

where $c'_B = n_B/N_A$. As a consequence, we obtain

$$x = \frac{-e^{\alpha_1} + \sqrt{e^{\alpha_1} + 4zc'_B e^{\alpha_2}}}{2ze^{\alpha_2}}, \tag{12}$$

$$\alpha_1 = -\frac{z}{2kT}(\nu_{AA} - 2\nu_{AB}), \qquad \alpha_2 = -\frac{(z-1)\nu_{AA} - \nu_{BB} - 2\nu_{AB}}{kT}. \tag{13}$$

By using expressions (9), (10), (12), and (13), we find

$$c_1 = \frac{e^{w/kT}}{2z}\left(\sqrt{1 + 4zc'_B e^{w/kT}} - 1\right), \tag{14}$$

$$c_2 = \frac{e^{w/kT}}{4z^2}\left(\sqrt{1 + 4zc'_B e^{w/kT}} - 1\right)^2. \tag{15}$$

Here, $w = 2\nu_{AB} - \nu_{AA} - \nu_{BB}$ is the energy of ordering of the solid solution.

Let us now consider the specific case of ordering alloys ($w > 0$) in the limit of high ($w/kT \ll 1$) and low ($w/kT \gg 1$) temperatures. Since c'_B is a small quantity, from expressions (14) and (15), we obtain identical asymptotic representations:

$$\begin{aligned}c_1 &= c'_B - zc'^2_B\exp(-w/kT),\\ c_2 &= c_B^2\exp(-w/kT).\end{aligned} \tag{16}$$

For decomposing alloys ($w < 0$), the situation is more complicated. In the high-temperature range, the atomic concentrations c_1 and c_2 are given by formulas (16). At low temperatures, these alloys contain component B at an atomic concentration $c'_B\exp(-w/kT) \gg 1$. Hence, the atomic concentrations c_1 and c_2 can be written as

$$\begin{aligned}c_1 &= \sqrt{\frac{c'_B}{2}}\exp\left(\frac{w}{2kT}\right),\\ c_2 &= \frac{1}{z}c'_B - \sqrt{\frac{c'_B}{z^3}}\exp\left(\frac{w}{2kT}\right).\end{aligned} \tag{17}$$

Since $c'_B = c_B(1 - c_B)$, from relationships (16) and (17), we find that, in solid solutions with an ordering energy $w > 0$, the number of B–B complexes is small over the entire temperature range. By contrast, in solid solutions with an ordering energy $w < 0$, the number of complexes is small at high temperatures and tends to $N_B/2$ at $T \longrightarrow 0$; i.e., virtually all B atoms are joined into pairs. From the physical standpoint, the nature of this effect is quite clear. In solid solutions with an ordering energy $w > 0$, there is a tendency toward ordering; i.e., the A atoms tend to surround the B atoms. In solid solutions with an ordering energy $w < 0$, there is a tendency toward decomposition; as a consequence, every atom tends to surround itself by like atoms.

From expressions (9) and (10), we obtain one more relationship between the atomic concentrations c_1 and c_2:

$$c_2 = c_1^2\exp\left(-\frac{w}{kT}\right). \tag{18}$$

It can easily be seen from relationship (18) that, since $c_1 \le c_B$), the concentration of complexes in solid solutions with an ordering energy $w < 0$ can be significant.

The formation of complexes in alloys can easily be revealed by measuring the residual resistance ρ_0, i.e., the resistance caused by elastic scattering of conduction electrons at high and low temperatures. In the approximation of weak scattering [2, 3], we have

$$\rho_0 \sim \left|\sum_{\mathbf{n}}\{(a_{\mathrm{A}}(\mathbf{n})-c_{\mathrm{A}})V_{\mathrm{A}}^{\mathbf{kk'}}+(a_{\mathrm{B}}(\mathbf{n})-c_{\mathrm{B}})V_{\mathrm{B}}^{\mathbf{kk'}}\}\right|^2, \qquad (19)$$

where $V_\alpha^{\mathbf{kk'}}$ is the matrix element of the potential of the α atom (α = A, B) between the electronic states **k** and **k'**, and

$$a_\alpha(\mathbf{n}) = \begin{cases} 1, & \text{if the lattice site with radius vector } \mathbf{n} \\ & \text{is occupied by the } \alpha, \\ 0, & \text{otherwise}. \end{cases}$$

The left-hand side of relationship (19) can be represented in the form

$$\left\langle \sum_{\mathbf{nn'}}\left\{(a_{\mathrm{A}}(\mathbf{n})-c_{\mathrm{A}})(a_{\mathrm{A}}(\mathbf{n'})-c_{\mathrm{A}})\left|V_{\mathrm{A}}^{\mathbf{kk'}}\right|^2\right.\right.$$
$$+(a_{\mathrm{A}}(\mathbf{n})-c_{\mathrm{A}})(a_{\mathrm{B}}(\mathbf{n})-c_{\mathrm{B}})(V_{\mathrm{A}}^{\mathbf{kk'}*}V_{\mathrm{B}}^{\mathbf{kk'}}+V_{\mathrm{A}}^{\mathbf{kk'}}V_{\mathrm{B}}^{\mathbf{kk'}*}) \qquad (20)$$
$$\left.\left.+(a_{\mathrm{B}}(\mathbf{n})-c_{\mathrm{B}})(a_{\mathrm{B}}(\mathbf{n'})-c_{\mathrm{B}})\left|V_{\mathrm{B}}^{\mathbf{kk'}}\right|^2\right\}\right\rangle.$$

The symbol $\langle\ldots\rangle$ denotes the configurational averaging.

By separating out the terms with **n** = **n'** and **n** ≠ **n'** in relationship (20) and ignoring the correlation in the solid solution, we obtain the following expressions for the alloy containing only individual B atoms:

$$\left\langle\sum_{\mathbf{n}}(a_{\mathrm{A}}(\mathbf{n})-c_{\mathrm{A}})^2\right\rangle = \left\langle\sum_{\mathbf{n}}(a_{\mathrm{B}}(\mathbf{n})-c_{\mathrm{B}})^2\right\rangle = Nc_{\mathrm{A}}c_{\mathrm{B}}, \qquad (21)$$
$$\left\langle\sum_{\mathbf{n}}(a_{\mathrm{A}}(\mathbf{n})-c_{\mathrm{A}})(a_{\mathrm{B}}(\mathbf{n})-c_{\mathrm{B}})\right\rangle = -Nc_{\mathrm{A}}c_{\mathrm{B}},$$

$$\left\langle\sum_{\substack{\mathbf{n},\,\mathbf{n'}\\(\mathbf{n}\neq\mathbf{n'})}}(a_{\mathrm{A}}(\mathbf{n})-c_{\mathrm{A}})(a_{\mathrm{A}}(\mathbf{n'})-c_{\mathrm{A}})\right\rangle$$
$$= \left\langle\sum_{\substack{\mathbf{n},\,\mathbf{n'}\\(\mathbf{n}\neq\mathbf{n'})}}(a_{\mathrm{A}}(\mathbf{n})-c_{\mathrm{A}})(a_{\mathrm{B}}(\mathbf{n'})-c_{\mathrm{B}})\right\rangle \qquad (22)$$
$$= \left\langle\sum_{\substack{\mathbf{n},\,\mathbf{n'}\\(\mathbf{n}\neq\mathbf{n'})}}(a_{\mathrm{B}}(\mathbf{n})-c_{\mathrm{B}})(a_{\mathrm{B}}(\mathbf{n'})-c_{\mathrm{B}})\right\rangle = 0.$$

From these expressions, we find

$$\rho_0 = Ac_{\mathrm{A}}c_{\mathrm{B}}, \qquad (23)$$

where A is a constant.

If the B atoms in the solid solution are joined into complexes, the means of the terms involved in relationship (23) are nonzero. The sum over **n** and **n'** (**n** ≠ **n'**) can be written as a sum over **n** and $\boldsymbol{\rho}_i$ (i = 1, 2, …), where $\boldsymbol{\rho}_i$ stands for the radius vectors of the sites of the ith coordination sphere with respect to the atom with radius vector **n**. As a result, we obtain

$$\left\langle\sum_{\mathbf{n}}\sum_{\boldsymbol{\rho}_i}(a_{\mathrm{A}}(\mathbf{n})-c_{\mathrm{A}})(a_{\mathrm{A}}(\mathbf{n}+\boldsymbol{\rho}_i)-c_{\mathrm{A}})\right\rangle$$
$$= \left\langle\sum_{\mathbf{n}}\sum_{\boldsymbol{\rho}_i}(a_{\mathrm{B}}(\mathbf{n})-c_{\mathrm{B}})(a_{\mathrm{B}}(\mathbf{n}+\boldsymbol{\rho}_i)-c_{\mathrm{B}})\right\rangle$$
$$= c_{\mathrm{B}}(1-zc_{\mathrm{B}}), \qquad (24)$$
$$\left\langle\sum_{\mathbf{n}}\sum_{\boldsymbol{\rho}_i}(a_{\mathrm{B}}(\mathbf{n})-c_{\mathrm{A}})(a_{\mathrm{B}}(\mathbf{n}+\boldsymbol{\rho}_i)-c_{\mathrm{B}})\right\rangle$$
$$= -c_{\mathrm{B}}(1-zc_{\mathrm{B}}).$$

Actually, if one of the sites with radius vector **n** is occupied by the B atom, one of the sites with radius vector $\mathbf{n} + \boldsymbol{\rho}_1$ is also occupied by the B atom. This demonstrates the validity of expression (24). As can easily be seen, the sums over **n** and $\boldsymbol{\rho}_i$ (i = 1, 2, …) become zero. It follows from here that, in the case where all B atoms are joined into complexes, we have

$$\boldsymbol{\rho}_0 = A(c_{\mathrm{A}}c_{\mathrm{B}}+c_{\mathrm{B}}(1-zc_{\mathrm{B}})). \qquad (25)$$

3. CONCLUSIONS

Thus, in solid solutions where complexes are not formed, the residual resistance $\boldsymbol{\rho}_0$ has the same value at high and low temperatures. In solid solutions containing B–B complexes, the low-temperature residual resistance proves to be higher than the high-temperature resistance.

REFERENCES

1. A. A. Smirnov, *Theory of Interstitial Alloys* (Nauka, Moscow, 1979) [in Russian].
2. A. A. Smirnov, *Theory of Electrical Resistance of Alloys* (AN USSR, Kiev, 1960) [in Russian].
3. S. I. Masharov, Fiz. Met. Metalloved. **19** (6), 820 (1965).

Translated by I. Volkov

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1091–1099. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1056–1064.
Original Russian Text Copyright © 2005 by Mikheev.

DEFECTS, DISLOCATIONS, AND PHYSICS OF STRENGTH

Simulation of Spatial Correlations of Impurity Ions in Solids Using Configurational Entropy and the Hard-Sphere Model

V. M. Mikheev

Institute of Metal Physics, Ural Division, Russian Academy of Sciences, ul. S. Kovalevskoĭ 18, Yekaterinburg, 620219 Russia

Received September 6, 2004

Abstract—Based on the hard-sphere model, the spatial correlations are considered in a system of impurities with variable valency. In a zeroth approximation, the configurational entropy of the spatially correlated system of impurity ions is identified with the configurational entropy of a system of hard spheres. The electron mobility limited by scattering on the correlated system of impurity ions at finite temperatures is found. The theory developed explains experimentally observed anomalies of the carrier mobility in an iron-doped HgSe gapless semiconductor. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The hard-sphere model with a reasonable choice of the fitting parameters allows quantitative description of the correlation effects in liquids and gases. However, there is a with the applicability of the model for calculating the correlation effects in systems that are not similar to a system of hard spheres. In particular, this concerns strongly correlated systems with Coulomb interaction.

One of these systems (an iron-doped HgSe compound) was intensively studied experimentally over the past decade, and the experimental results were interpreted in terms of the hard-sphere model. These compounds are a typical example of systems containing an impurity with variable valency (iron), with iron forming a resonance level in the conduction band. We recall that, in the HgSe matrix, iron atoms can be either in the Fe^{+2} state (neutral with respect to the lattice) or in the Fe^{+3} state (charged with respect to the lattice). At an iron concentration n_{Fe} equal to a certain critical concentration n_c, the electron Fermi level coincides with the resonance level of iron and is pinned at it. Therefore, at $n_{\mathrm{Fe}} < n_c$, all iron atoms are ionized; for a sufficiently high doping level ($n_{\mathrm{Fe}} > n_c$), the concentration of iron ions is $n_{\mathrm{Fe}}^{+} = n_c$. For a high doping level, the iron ions appear to be spatially correlated, since in this case the iron atoms are ionized selectively. Spatial correlation of iron ions has been described in terms of the hard-sphere model (see references in [1]); the minimum distance r_c between the ions was taken to be the diameter of a sphere. The concentration of hard spheres was identified with the concentration of iron ions. In the hard-sphere model, the sphere diameter was determined experimentally. In the theory of liquids, the radius of a sphere is identified with the range of the potential of an atom or a molecule and can be determined in scattering experiments. For HgSe : Fe, the authors of [2] suggested determining r_c from the measured concentration of iron ions. In [2], an equation was postulated relating the parameter r_c to the iron concentration in the HgSe : Fe compound. In [3–5], we derived an equation relating the parameter r_c to the iron concentration by using the methods of statistical physics. We started from a heuristic statement that, in equilibrium, Fe^{+3} iron ions (charged with respect to the lattice) tend to be located as far as possible from each other. It was also assumed that the iron atoms are distributed randomly.

In contrast to liquids, a correlated system of impurity ions cannot be identified with a system of hard spheres even in a zeroth approximation. Indeed, in the limit $r_c \longrightarrow 0$, a system of hard spheres is an ideal gas of point particles. In particular, the compressibility of this system is equal to the compressibility of an ideal gas. In the limit $r_c \longrightarrow 0$ ($n_{\mathrm{Fe}} \longrightarrow n_c$), the iron ions are completely ionized and frozen into the crystal. In this limit, the compressibility of the system of ions is equal to that of the crystal. Thus, the system of impurity ions is less similar to a system of hard spheres than is a liquid. Therefore, a problem arises as to how to correctly describe a system of impurity ions using the hard-sphere model.

Spatial ordering of impurity ions is accompanied by a change in the entropy of the system. The entropy related to spatial correlations in the system of ions is called the configurational entropy. We assume that the configurational entropy of the system of impurity ions (characterized by the minimum distance r_c between the ions) is similar to the configurational entropy of a system of hard spheres of diameter r_c. Therefore, in our model in a zeroth approximation, the configurational entropy of the system of impurity ions is identified with that of a system of hard spheres. We apply our model to describe correlation effects in HgSe : Fe compounds. We derive a relation between the sphere diameter r_c and the concentration of iron atoms starting from the well-known equation of state for a system of hard spheres. Based on this relation for r_c, we consider the effect of shallow donors on the mobility of electrons scattered by a correlated distribution of iron ions. Furthermore, we study the effect of correlations in the relative positions of Fe^{+2} and Fe^{+3} ions on the electron mobility and consider the effect of shallow donors on the temperature dependence of the electron mobility in HgSe : Fe.

2. EQUATION FOR THE HARD-SPHERE DIAMETER r_c

In HgSe : Fe with an iron concentration exceeding the critical value, the concentration of iron ions n_{Fe}^{+} is fixed and equal to the critical concentration n_c. Thus, the ionized iron atoms are located at distances larger than r_c from each other. The configuration of iron ions can be described by the pair distribution function. We identify this function with the pair correlation function of a system of hard spheres whose concentration and diameter are n_c and r_c, respectively [2]. To determine r_c, we write [3]

$$n_{\mathrm{Fe}}^{+} = n_{\mathrm{Fe}} W(r_c, n_{\mathrm{Fe}}^{+}), \tag{1}$$

where $W(r_c, n_{\mathrm{Fe}}^{+})$ is the probability of finding an iron atom at a distance greater than r_c from the nearest neighbor iron ion. In [3], we estimated this probability starting from a heuristic statement that, in equilibrium, iron ions tend to be located as far as possible from each other. To find this probability, we use the standard methods of statistical physics. It can easily be shown that the probability $\tilde{W}(r_c, n_{\mathrm{Fe}}^{+})$ of finding a configuration of the system of iron ions corresponding to a fixed value of the parameter r_c is given by [6]

$$\tilde{W}(r_c, n_{\mathrm{Fe}}^{+}) = \exp[\Delta S(r_c, n_{\mathrm{Fe}}^{+})/k_B], \tag{2}$$

where $\Delta S(r_c, n_{\mathrm{Fe}}^{+})$ is the change in the entropy of the system of ions produced by a change in the minimum distance between the ions from zero to a fixed value r_c. The quantity $\Delta S(r_c, n_{\mathrm{Fe}}^{+})$ is the configurational entropy of the system of ions. The configurational entropy per iron ion is $S^0(r_c, n_{\mathrm{Fe}}^{+}) = \Delta S(r_c, n_{\mathrm{Fe}}^{+})/n_{\mathrm{Fe}}^{+}$, and the expression $W(r_c, n_{\mathrm{Fe}}^{+}) = \exp[S^0(r_c, n_{\mathrm{Fe}}^{+})/k_B]$ can be treated as the probability of finding an iron atom at a distance greater than r_c from the nearest neighbor iron ion. Therefore, we can write an equation for finding the parameter r_c as

$$n_{\mathrm{Fe}}^{+} = n_{\mathrm{Fe}} \exp[S^0(r_c, n_{\mathrm{Fe}}^{+})/k_B]. \tag{3}$$

In our theory, the configurational entropy of the system of iron ions $S^0(r_c, n_{\mathrm{Fe}}^{+})$ is identified with that of a system of hard spheres. This quantity depends only on the packing factor η ($\eta = \pi r_c^3 n_{\mathrm{Fe}}^{+}/6$). We denote the configurational entropy of a system of spheres by $S^0(\eta)$. Then, the equation for the parameter r_c assumes the form

$$n_{\mathrm{Fe}}^{+} = n_{\mathrm{Fe}} \exp[S^0(\eta)/k_B]. \tag{4}$$

Thus, the problem of calculating the parameter r_c in a system of ions is reduced to the problem of calculating the configurational entropy of a system of spheres.

The configurational entropy of a system of spheres can be calculated using the equation of state of the system. In general, the equation of state of a system of spheres is

$$PV/N_{\mathrm{Fe}}^{+} k_B T = \Phi(\eta). \tag{5}$$

Here, N_{Fe}^{+} is the number of spheres occupying the volume V (we recall that the number of spheres is equal to the number of Fe^{+3} ions charged with respect to the lattice). The function $\Phi(\eta)$) describes the deviation of the system of spheres from an ideal gas. In the limit $\eta \longrightarrow 0$, we have $\Phi(\eta) \longrightarrow 0$ and Eq. (5) reduces to the equation for an ideal gas. In thermodynamics, the pressure P in the system can be expressed in terms of the derivatives of the free energy F of the system as [6]

$$P = -\left(\frac{\partial F}{\partial V}\right)_{T,\eta} - \left(\frac{\partial F}{\partial \eta}\right)_{V,T}\left(\frac{\partial \eta}{\partial V}\right)_{T}. \tag{6}$$

Integrating Eq. (6) along an isotherm under condition (5), we obtain a general expression for the free energy of the system of spheres,

$$F = N_{\mathrm{Fe}}^{+} k_B T \int \frac{d\eta}{\eta} \Phi(\eta) + C. \tag{7}$$

The entropy of the system of spheres can be calculated using the formula

$$S = -\left(\frac{\partial F}{\partial T}\right)_{V,\eta}. \tag{8}$$

In order to separate the configurational entropy $\Delta S(\eta)$ from the expression for the total entropy of the system of spheres, we determine the integration constant C from the condition that $\Delta S(\eta) \longrightarrow 0$ as $\eta \longrightarrow 0$. The function $\Phi(\eta)$ in equation of state (5) can be expressed in terms of the virial coefficients [7]

$$\Phi(\eta) = 1 + \sum_{l=1}^{\infty} b_{l+1} \eta^l. \tag{9}$$

Using formulas (7)–(9) and the condition from which the integration constant C is determined, we find the configurational entropy per sphere to be

$$S^0(\eta) = k_B \sum_{l=1}^{\infty} b_{l+1} \eta^l / l. \tag{10}$$

Substituting Eq. (10) into Eq. (4), we obtain an equation relating the minimum distance r_c between Fe^{+3} ions to the concentration of iron atoms n_{Fe} and the concentration $\overset{+}{n}_{Fe}$ of Fe^{+3} ions in the HgSe: Fe compound:

$$\overset{+}{n}_{Fe} = n_{Fe}\exp\sum_{l=1}^{\infty} b_{l+1}\eta^l/l \qquad (11)$$
$$(\eta = \pi r_c^3 \overset{+}{n}_{Fe}/6).$$

The first five virial coefficients were calculated using numerical methods [7]: $b_2 = 4$, $b_3 = 10$, $b_4 = 18.36$, $b_5 = 28.26$, and $b_6 = 39.53$. For sufficiently strongly doped HgSe : Fe ($n_{Fe} > n_c$), the condition $\overset{+}{n}_{Fe} = n_c$ is satisfied. The critical iron concentration is $n_c = 4.5 \times 10^{18}$ cm^{-3} [1].

A numerical solution to Eq. (11) is shown in Fig. 1, where the packing factor η is plotted as a function of the parameter $n_{Fe}/\overset{+}{n}_{Fe}$. Curve *1* corresponds to Eq. (9) with only the virial coefficient b_2 retained; curve *2* corresponds to Eq. (9) with two virial coefficients, b_2 and b_3; and curve *5* was calculated taking into account five virial coefficients. It follows from Fig. 1 that the solution to Eq. (11) is rather sensitive to the accuracy of the virial expansion and we need to take into account a sufficiently large number of virial coefficients. To calculate the mobility of electrons scattered by a correlated distribution of iron ions within our model, we need to know the pair correlation function for the system of spheres. Since computations of the virial expansions of the pair correlation function are extremely labor consuming, we use interpolation formulas. For this purpose, we apply the Percus–Yevick method [7], which is widely used for describing a system of spheres. The pair correlation function $\tilde{g}(r, \eta, \overset{+}{n}_{Fe})$ describing the distribution of spheres in the Percus–Yevick approximation enables us to calculate the equation of state for a system of spheres. Calculations can be performed in two ways. Starting with the pressure P, the equation of state can be found to be [7]

$$\frac{P}{\overset{+}{N}_{Fe}k_BT} = \frac{1+2\eta+3\eta^2}{(1-\eta)^3}. \qquad (12)$$

Integrating Eq. (6) with the new condition (12), we obtain an approximate expression for the configurational entropy $S_0^{(P)}(\eta)$:

$$S_0^P(\eta) = k_B\left[\ln(1-\eta) - \frac{(1+2\eta)}{(1-\eta)^2} + 1\right]. \qquad (13)$$

Based on the isothermal compressibility α, the equation of state has the form [7]

$$\frac{P}{\overset{+}{N}_{Fe}k_BT} = \frac{1+\eta+\eta^2}{(1-\eta)^3}. \qquad (14)$$

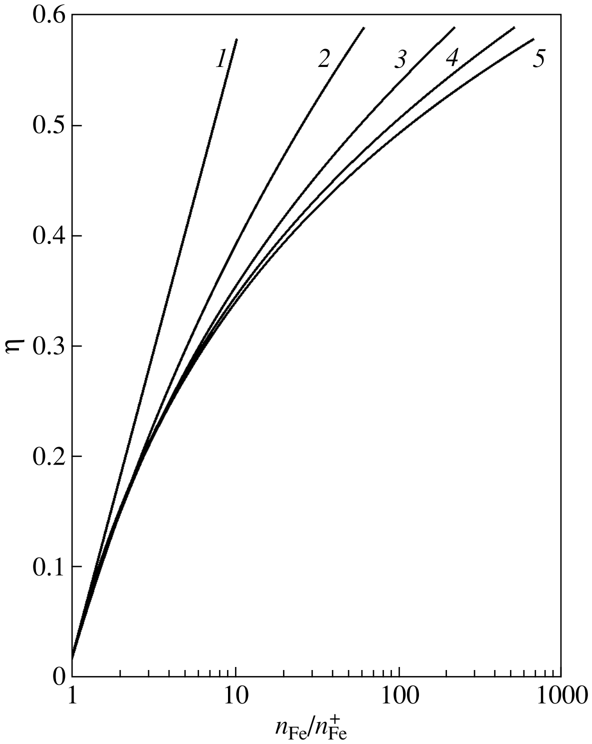


Fig. 1. Dependence of the packing factor η on the iron concentration (n_{Fe} is the concentration of iron atoms, $\overset{+}{n}_{Fe}$ is the concentration of Fe^{+3} ions). Curves *1–5* correspond to the virial expansion in which the first several terms (from one to five, respectively) are kept.

Integrating Eq. (6) with condition (14), we obtain another approximate expression for the configurational entropy $S_0^{\alpha}(\eta)$:

$$S_0^{\alpha}(\eta) = k_B\left[\ln(1-\eta) + \frac{3}{2}\frac{\eta(\eta-2)}{(1-\eta)^2}\right]. \qquad (15)$$

Thermodynamically, both methods used to derive the equations of state are equivalent. The difference between the equations of state for spheres [Eqs. (12) and (14)] derived using the Percus–Yevick approximation for the correlation function indicates that this approximation is incorrect. Nevertheless, this method is widely used in the theory of liquids to derive simple interpolation formulas that make quantitative description possible. We also use the Percus–Yevick method in our calculations. Substituting Eqs. (13) and (15) into (4), we find the dependence of the parameter η on the iron concentration in the Percus–Yevick approximation. The results are shown in Fig. 2. Curve *a* is calculated using the configurational entropy $S_0^{(\alpha)}(\eta)$, and curve *b* is calculated using the configurational entropy $S_0^{(P)}(\eta)$. Curve *c* is calculated using virial expansion (10) and can be considered a reference curve. In Fig. 2, we

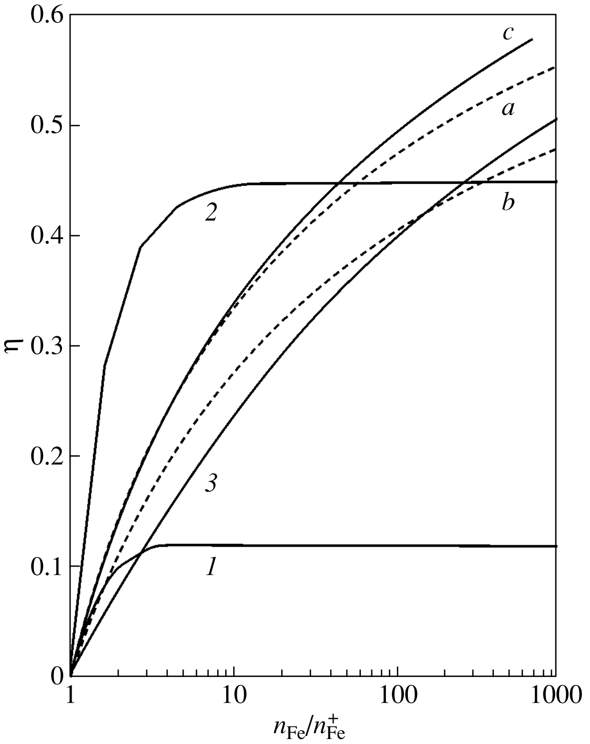


Fig. 2. Dependence of the packing factor η on the iron concentration calculated in (*1*) [2], (*2*) [8], and (*3*) [3]; curve *c* is calculated using the virial expansion. The dashed curves *a* and *b* are calculated within the Percus–Yevick approximation starting with the compressibility and pressure, respectively.

see that, in a wide range of iron concentrations ($1 < n_{\mathrm{Fe}}/\overset{+}{n}_{\mathrm{Fe}} < 50$), curve *a* is a good approximation to curve *c* obtained using a regular virial expansion procedure. For this reason, we use Eqs. (4) and (15) to calculate the parameter r_c at a given iron concentration n_{Fe}.

Figure 2 shows the results of previous calculations of the packing factor $\eta = \pi r_c^3 \overset{+}{n}_{\mathrm{Fe}}/6$. In one of the first studies of the correlations in the positions of iron ions in HgSe : Fe, the parameter r_c was assumed to satisfy the equation [2]

$$V_c \overset{+}{n}_{\mathrm{Fe}} = 1 - \exp(-V_c n_{\mathrm{Fe}}),$$

where the correlation volume is $V_c = 4\pi r_c^3/3$. The packing factor η calculated from this equation as a function of the iron concentration is shown in Fig. 2 (curve *1*). We see that this curve is close to the reference curve *c* only in a rather narrow range of iron concentrations ($1 < n_{\mathrm{Fe}}/\overset{+}{n}_{\mathrm{Fe}} < 1.5$). The authors of [2] used this equation to interpret the experimental data for HgSe : Fe in terms of the weak-correlation model and qualitatively described the anomalies in the electron mobility without any additional assumptions.

The authors of [8] analyzed the experimental data for HgSe : Fe using the strong-correlation model. Accordingly, they modified the equation from [2] by introducing a fitting parameter into it and wrote the equation in the form

$$\frac{\eta}{\eta_L} = 1 - \exp\left(-\frac{\eta}{\eta_L}\frac{n_{\mathrm{Fe}}}{\overset{+}{n}_{\mathrm{Fe}}}\right),$$

where $\eta_L = 0.45$ ($\eta = V_c \overset{+}{n}_{\mathrm{Fe}}/8$). The maximum value of the packing factor is η_L (in the limit $n_{\mathrm{Fe}}/\overset{+}{n}_{\mathrm{Fe}} \longrightarrow \infty$, we have $\eta \longrightarrow \eta_L$). At $\eta_L = 1/8$, this equation coincides with the corresponding equation in [2]. The solution to the equation postulated in [8] is plotted in Fig. 2 (curve *2*). It can be seen that the parameter η for curve *2* is several times greater than that for curve *c* at relatively low iron concentrations, namely, in the range $1 < n_{\mathrm{Fe}}/\overset{+}{n}_{\mathrm{Fe}} < 4$, where the anomalies in electron mobility were observed. Therefore, in order to attain agreement between the predictions from the strong-correlation model and the experimental data, additional assumptions and additional fitting parameters were used in [8].

In our study [3], the equations for η were obtained using the methods of mathematical statistics. We started from the heuristic statement that, in equilibrium, iron ions tend to be located as far as possible from each other:

$$\overset{+}{n}_{\mathrm{Fe}} = n_{\mathrm{Fe}}(1 - \eta/\eta_0)^{\alpha}. \tag{16}$$

Here, $\alpha = 5.92$ and the packing factor for the fcc structure is $\eta_0 = 0.74$. The solution to this equation is plotted in Fig. 2 (curve *3*). We see that curve *3* lies somewhat lower than the reference curve *c*. Therefore, the values of the mobility calculated by us in [3–5] on the basis of curve *3* at the maximum are 15 to 20% lower than the experimental values. Thus, using this intuitively clear heuristic statement, we were able to give a rough quantitative description of the effects of correlation in the positions of scatterers on the electron mobility in HgSe : Fe. The entropy approach suggested in this study allows one to give a quantitative description of spatial correlations in the system of impurity ions using regular methods.

3. ELECTRON MOBILITY AT $T = 0$

In HgSe : Fe at low temperatures, the main contribution to the electron mobility is from scattering by iron atoms and shallow donor centers. In order of magnitude, the concentration of shallow donors $n_D \sim 10^{18}$ cm^{-3} is comparable to the critical iron concentration $n_c = 4.5 \times 10^{18}$ cm^{-3}. With inclusion of shallow donors, the system of scatterers consists of three types of impurity centers: Fe^{+3} ions, Fe^{+2} atoms (neutral with respect to the lattice), and shallow donors. The mobility controlled by electron scattering on these impurity centers is [4]

$$\mu = e\tau/m,$$
$$\frac{1}{\tau} = \frac{2}{3}\frac{m}{(2\pi\hbar)^3 n_e}\int_0^{2k_F} q^3 dq\{|V_0(q)|^2 n_{\mathrm{Fe}}^0 + |V_+(q)|^2(n_{\mathrm{Fe}}^+ + n_D) \tag{17}$$
$$+\frac{1}{2}[V_+(q)(V_+^*(q) - V_0^*(q)) + V_+^*(q)(V_+(q) - V_0(q))]$$
$$\times [n_{\mathrm{Fe}}^{+2} g_{++}^{FF}(q) + 2n_{\mathrm{Fe}}^+ n_D g_{++}^{FD}]\}.$$

Here, $V_0(q)$ is the Fourier transform of the scattering potential due to neutral impurity centers, $V_+(q)$ is the Fourier transform of the screened Coulomb potential (for impurity centers charged with respect to the lattice), g_{++}^{FF} is the Fourier transform of the partial correlation function including the correlations in relative positions of the iron ions, and g_{++}^{FD} is the Fourier transform of the correlation function including the correlations in relative positions of the iron ions and shallow donors (all shallow donors are assumed to be ionized). The contribution from the interference terms, which take into account correlations between the positions of Fe^{+3} ions and "neutral" Fe^{+2} atoms, is proportional to the product $V_0(q)V_+^*(q)$ [4].

We recall that Eq. (17) for the electron mobility was obtained in [4] using the Boltzmann transport equation in the Born approximation with regard to the screening of the Coulomb potential of impurity ions described within the framework of the linear Thomas–Fermi theory. These approximations are valid if certain parameters are small. For HgSe : Fe, these parameters are (in order of magnitude) $\hbar/\tau\varepsilon_F \approx 3 \times 10^{-3}$ (for the applicability of the Boltzmann equation), $(k_F a_B)^{-1} \approx 0.1$ (for the applicability of the Born approximation in the case of Coulomb scattering potential), and $e^2/\kappa r\varepsilon_F \approx 5 \times 10^{-2}$ (for the applicability of the Thomas–Fermi theory). Here, r is the average distance between electrons, a_B is the Bohr radius, and τ is the electron momentum relaxation time. We made estimations using the parameter values characteristic of HgSe : Fe, namely, the electron concentration $n_c = 4.5 \times 10^{18}$ cm^{-3}, the Fermi energy $\varepsilon_F = 210$ meV, the electron mass $m/m_0 = 0.07$, the permittivity $\kappa = 20$, and the electron mobility $\mu = 4 \times 10^4$ cm^2/V s. The estimations show that the Born approximation, which makes it possible to obtain analytical expression (17) for the mobility, is actually the "bottleneck" for our calculations and provides a calculation accuracy s of about 10%. Therefore, to avoid an obviously excessive accuracy, we should estimate the contributions from all terms in Eq. (17) to the mobility. The corresponding calculations, which are discussed in detail in what follows, show that the terms describing the correlations between the ions (neutral and charged with respect to the lattice) have a contribution of several percent to the mobility, which obviously provides an excessive accuracy. Therefore, in order to take into account the effect of spatial correlation of iron ions on the electron mobility in HgSe : Fe, it suffices to use formula (17a) for the mobility (see below), which does not contain terms of the type $V_0(q)V_+^*(q)$ describing the correlations between the relative positions of neutral and charged scatterers:

$$\mu = e\tau/m,$$
$$\frac{1}{\tau} = \frac{2}{3}\frac{m}{(2\pi\hbar)^3 n_e}\int_0^{2k_F} q^3 dq\{|V_0(q)|^2 n_{\mathrm{Fe}}^0 + |V_+(q)|^2(n_{\mathrm{Fe}}^+ + n_D) \tag{17a}$$
$$+ |V_+(q)|^2[n_{\mathrm{Fe}}^{+2} g_{++}^{FF}(q) + 2n_{\mathrm{Fe}}^+ n_D g_{++}^{FD}]\}.$$

At sufficiently high concentrations of shallow donors, it becomes important that, because of Coulomb repulsion, an iron ion cannot approach a shallow donor ion to a distance less than a certain length $\tilde{r}_c$. To allow for this effect in Eq. (11), the concentration of iron atoms n_{Fe} is replaced by the effective concentration $\tilde{n}_{\mathrm{Fe}} = n_{\mathrm{Fe}}(1 - \tilde{V}_c n_D)$. Thus, instead of Eq. (11), we write [4]

$$n_{\mathrm{Fe}}^+ = n_{\mathrm{Fe}}(1 - \tilde{V}_c n_D)\exp\sum_{l=1}^{\infty} b_{l+1}\eta^l/l, \tag{18}$$

where $\tilde{V}_c = 4\pi\tilde{r}_c^3/3$ is determined from the equation

$$\tilde{V}_c n_c = 1 - n_{\mathrm{Fe}}^+/n_{\mathrm{Fe}}. \tag{19}$$

We note that in [4], we used the following equation for the parameter η obtained from heuristic assumptions:

$$n_{\mathrm{Fe}}^+ = n_{\mathrm{Fe}}(1 - \tilde{V}_c n_D)(1 - \eta/\eta_0)^{\alpha}. \tag{20}$$

Here, $\alpha = 5.92$ and the packing factor of the fcc structure is $\eta_0 = 0.74$. When calculating the correlation function g_{++}^{FF}, the Percus–Yevick formula [7] was used, and when calculating the correlation function g_{++}^{FD} we made use of an expression obtained by substituting the virial expansion to lowest order in η:

$$g_{++}^{FD}(r) = \begin{cases} 1, & r \le \tilde{r}_c \\ 0, & r > \tilde{r}_c. \end{cases} \tag{21}$$

This approximation it is quite sufficient, since the correlations between the positions of iron ions with respect to shallow donors frozen into the lattice are much weaker than the correlations between the relative positions of mobile iron ions. The potential $V_0(q)$ of scattering by a neutral atom was estimated using the formula for electron mobility in the case of scattering by neutral

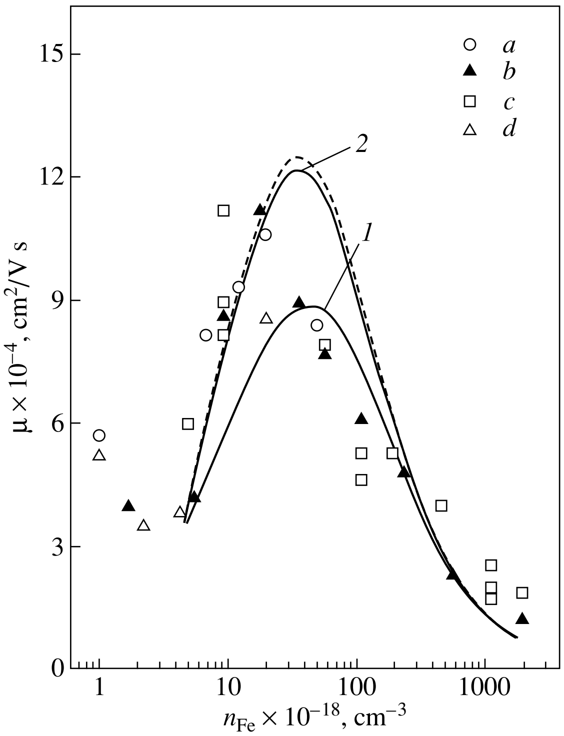


Fig. 3. Calculated electron mobility μ as a function of the iron concentration in pure HgSe : Fe ($n_D = 0$). Experimental points are taken from (*a*) [10], (*b*) [9], (*c*) [11], and (*d*) [12]. Curve *2* does not include correlations in the relative positions of the Fe^{+2} and Fe^{+3} ions. The dashed curve takes into account the correlations in the relative positions of the Fe^{+2} and Fe^{+3} ions. Curve *1* was calculated in [4].

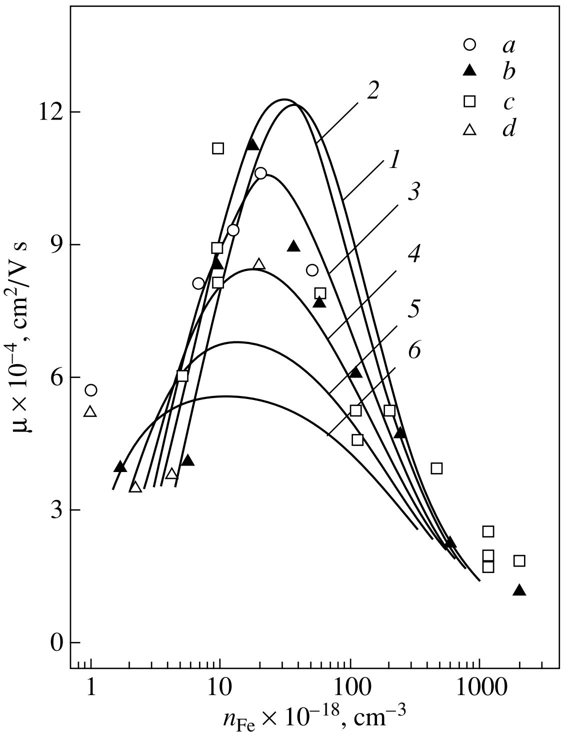


Fig. 4. Electron mobility μ calculated as a function of the iron concentration in HgSe : Fe for various values of the concentration of shallow donors n_D (10^{18} cm^{-3}): (*1*) 0, (*2*) 1, (*3*) 1.5, (*4*) 2, (*5*) 2.5, and (*6*) 3. The experimental points are the same as those in Fig. 3.

centers $\mu = e\tau_0/m$ (formula (17) with $n_D = 0$, $\overset{+}{n}_{Fe} = 0$). When calculating the mobility, we set $V_0(q)$ = const and identified the experimental value of the electron mobility ($\mu = 2.24 \times 10^4$ cm^2/Vs in HgSe : Fe with an iron concentration $n_{Fe} = 5.8 \times 10^{20}$ cm^{-3} [9]) with μ_0, since at such iron concentrations the concentration of iron atoms in the Fe^{+2} state (neutral with respect to the lattice) is two orders of magnitude greater than the concentration of Fe^{+3} atoms charged with respect to the lattice.

Based on Eqs. (17)–(20), we calculated the electron mobility (we recall that the correlation function g_{++}^{FF} was calculated from the Percus–Yevick formula [7]) using the following parameter values: $n_c = 4.5 \times 10^{18}$ cm^{-3}, ε_F = 210 meV, electron mass $m/m_0 = 0.07$, and permittivity of HgSe : Fe, $\kappa = 20$.

The results are shown in Figs. 3 and 4. Figure 3 presents the calculated electron mobility μ as a function of the concentration n_{Fe} of iron atoms in pure HgSe : Fe ($n_D = 0$). All curves have a characteristic bell-shaped form. At low iron concentrations, the main contribution is from scattering by Fe^{+3} ions charged with respect to the lattice. As the iron concentration increases, the system of scatterers becomes more ordered and the electron mobility grows. For high iron concentrations, the scattering by Fe^{+2} ions, neutral with respect to the lattice, is dominant. The concentration of these scatterers increases with iron concentration, thereby decreasing the electron mobility. Curve *1* represents the calculations performed in [4], where the parameter η was calculated from Eq. (20). In [4], the interference contribution from correlations in relative positions of Fe^{+3} and Fe^{+2} ions was disregarded. Therefore, in Eq. (17) for τ, we disregarded the terms proportional to the product $V_+(q)V_0^*(q)$. In Fig. 3, we see that the mobility calculated in [4] is smaller than the experimental values by 20–25%. Curve *2* in Fig. 3 represents the calculations in which the parameter η was determined from Eqs. (18) and (19). As before, we disregarded the contribution from correlations in the relative positions of Fe^{+3} and Fe^{+2} ions to the mobility. To a good accuracy, the experimental points fall on calculated curve *2*. The results of the calculations with regard to the correlations in the relative positions of Fe^{+3} and Fe^{+2} ions are shown in Fig. 3 by dashed curves. In Fig. 3, we see that a more complete account of the correlations in the positions of scatterers increases the mobility only slightly (by several percent) near the mobility maximum in a small interval of iron concentrations. These results justify the above replacement of formula (17) for the mobility by a simpler formula (17a), which was obtained from formula (17) by omitting the terms proportional to the factor $V_+(q)V_0^*(q)$ and taking into account the correlations in the relative positions of iron ions, both neutral and charged with respect to the lattice.

Let us compare the results of our calculations with those obtained in [13], where the effect of correlations in the positions of iron ions on the electron mobility in HgSe : Fe was considered. The latter calculations were based on two assumptions that were also made in a previous study [8]. The authors of [8] postulated an equation for the mobility that takes into account the effect of correlations in the positions of the scatterers. In an expression for the electron mobility [8, Eq. (3)], the term describing the correlations in the relative positions of Fe^{+3} and Fe^{+2} ions was assumed to have the form $2[(N_{Fe^{+3}}N_{Fe^{+2}})^{1/2}/N_{Fe}]|V_+(q)V_0(q)|S_{+0}(q)$. For randomly distributed scatterers (with the structure factor $S_{+0}(q) = 1$), this interference term assumes the maximum value, whereas the standard expression describing the interference [14] in this case is equal to zero. Thus, the authors of [13] made an unphysical assumption that the introduction of additional correlations in the positions of Fe^{+3} and Fe^{+2} scatterers increases the scattering cross section and, therefore, decreases the electron mobility. When calculating the interference effects in the hard-sphere model, the authors of [13] determined the packing factor η from the equation postulated in [8]. This equation overestimates the parameter η (curve *2* in Fig. 2). Therefore, according to [13], the electron mobility controlled by their scattering on the correlated system of Fe^{+3} ions (in the absence of correlations in relative positions of Fe^{+3} and Fe^{+2} ions) is two times greater than the experimental mobility at an iron concentration $n_{Fe} = 2 \times 10^{19}$ cm^{-3}. To reconcile the results of the theoretical calculations with the experimental results, the authors of [13] had to use the artificial constructions made in [8], according to which the additional correlations in the positions of the scatterers can increase the electron scattering cross section. According to these constructions, the correlations in relative positions of Fe^{+3} and Fe^{+2} ions reduce the scattering cross section by a factor of 2 and compensate for the disagreement between the experimental data and the values of the mobility calculated in [13]. These constructions raise doubts. We hold the generally accepted view that any correlations in the distribution of scatterers decrease the scattering cross section and, hence, increase the carrier mobility (the more ordered the positions of scatterers, the smaller the corresponding scattering cross section). Therefore, we believe that the calculations in [13] are incorrect and, in what follows, we refrain from commenting on those studies. Moreover, the allowance for the contribution from correlations in the positions of Fe^{+3} and Fe^{+2} ions to the carrier mobility is obviously in excess of the accuracy of the formula for the mobility. Indeed, the application of the Born approximation to calculating the mobility using formula (17) limits the accuracy to about 10%. However, the contribution from the correlations in the relative positions of Fe^{+3} and Fe^{+2} ions to the carrier mobility calculated using standard formulas is only several percent.

In Fig. 4, the calculated dependence of μ on n_{Fe} is plotted at different concentrations of shallow donors n_D. Curve *1* corresponds to the electron mobility of pure HgSe : Fe ($n_D = 0$), and curves *2–6* correspond to the mobility for the concentration of shallow donors varied from 10^{18} to 3×10^{18} cm^{-3} in steps of 5×10^{17} cm^{-3}. It can be seen that, for high iron concentrations, where scattering by Fe^{+2} ions (neutral with respect to the lattice) dominates, the electron mobility decreases with increasing concentration of shallow donors (additional scatterers). However, for iron concentrations at which an anomalous increase in the mobility is observed, the presence of shallow donors gives rise to new anomalies. Namely, in the region of iron concentrations $4.5 \times 10^{18} < n_{Fe} < 3 \times 10^{19}$ cm^{-3}, the electron mobility increases with the concentration of shallow donors. This effect was observed and studied in [15], and a consistent theory was developed in our study [4]. The effect has a simple explanation. Indeed, according to Eq. (11), the packing factor η increases with the parameter n_{Fe}/n_{Fe}^{+}. In turn, the parameter $n_{Fe}/n_{Fe}^{+} = n_{Fe}/(n_c - n_D)$ increases with n_D. The increase in the correlations in the system of Fe^{+3} ions results in an anomalous increase in the electron mobility with the concentration of shallow donors.

We see in Fig. 4 that, at low iron concentrations, the experimental points lie in the region where there are curves corresponding to low concentrations of shallow donors. On the contrary, at high iron concentrations, the experimental points are grouped near the curves corresponding to higher concentrations of shallow donors. Thus, it follows from our calculations that the content of uncontrollable shallow donors in HgSe : Fe samples correlates with the iron concentration in the samples. In the HgSe : Fe samples with low iron content, the concentration of shallow donors is low. With increasing the iron concentration from 2×10^{19} to 2×10^{20} cm^{-3}, the concentration of shallow donors grows from 1.5×10^{18} to 2.5×10^{18} cm^{-3}.

4. THE EFFECT OF TEMPERATURE ON SPATIAL CORRELATIONS OF IMPURITY IONS IN HgSe : Fe

The disorder in the system of impurity ions increases with temperature. In our one-parameter model, the order in the system is characterized by the parameter r_c. Large values of the parameter r_c corre-

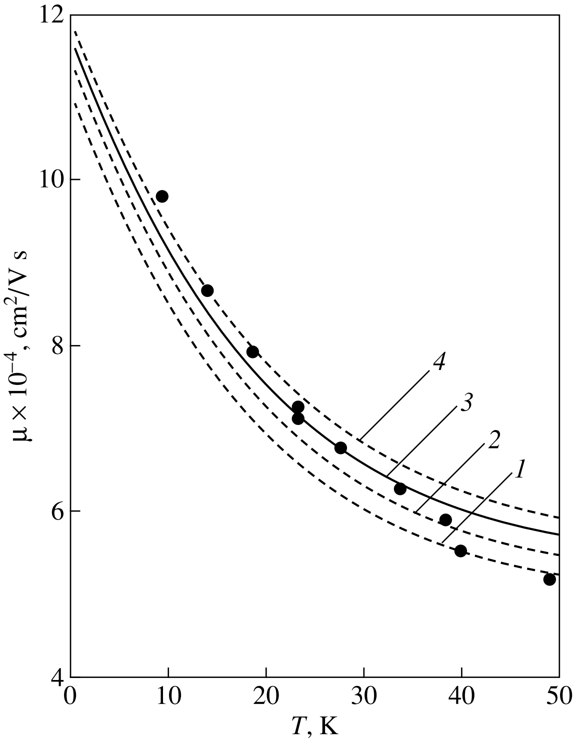


Fig. 5. Electron mobility μ calculated as a function of temperature in HgSe : Fe ($n_{\mathrm{Fe}} = 1.78 \times 10^{19}$ cm^{-3}) for various values of the concentration of shallow donors n_D (10^{18} cm^{-3}): (*1*) 0, (*2*) 0.2, (*3*) 0.4, and (*4*) 0.6. Experimental points are taken from [10].

spond to a greater degree of order in the system of impurity ions. In the limiting case $r_c = 0$, the impurity ions are randomly distributed in space. In a correlated system of impurity ions at finite temperature, any ion configuration described by the pair correlation function $g(r_c, r)$ with the parameter $r_c \le r_c^0$ can occur (r_c^0 is the minimum distance between the ions at $T = 0$). In standard textbooks on statistical physics [6], it is proved that the probability of finding a system in a configuration with a given value of the parameter r_c is proportional to $\exp[-\Delta E + T\Delta S - P\Delta V]/T$. Here, ΔE, ΔS, and ΔV are the variations in the internal energy, entropy, and volume of the system, respectively, as the parameter r_c changes from zero to a given value. In our case, ΔE is the correlation energy of the system of impurity ions and ΔS is the configurational entropy of this system. We disregard the variation in the volume of the system of impurity ions during ordering. The thermodynamic average of a physical quantity dependent on the parameter r_c over all possible configurations of the system of ions is given by [5]

$$\langle F(r_c)\rangle = \int_0^{r_c^0} F(r_c)W(r_c)dr_c \Big/ \int_0^{r_c^0} W(r_c)dr_c, \qquad (22)$$

where

$$W(r_c) = \exp\left[\frac{S_0(r_c)}{k_B} - \frac{\varepsilon_c(r_c)}{k_B T}\right].$$

$\varepsilon(r_c)$ is the correlation energy per particle, and $S_0(r_c)$ is the configurational entropy per particle. In [5], we calculated $S_0(r_c)$ using the Boltzmann formula $S_0(r_c) = k_B \ln W(r_c)$, where the probability $W(r_c)$ of finding a nearest neighbor ion at a distance from the impurity atom exceeding r_c is taken from [3]. In this study, the configurational entropy is calculated from Eq. (10). To calculate the correlation energy, we used the formulas [5]

$$\varepsilon_c(r_c) = 2\pi n_{\mathrm{Fe}}^{+} \int_0^{\infty} r^2 dr V(r)[\tilde{g}(r) - 1], \qquad (23)$$
$$V(r) = \frac{e^2}{\kappa r} e^{r/r_{TF}}.$$

We applied our theory to study the effect of temperature disordering in the system of iron ions in HgSe : Fe on the electron mobility. The temperature dependence of the electron mobility μ is calculated from Eq. (17), in which the parameter r_c was replaced by its thermodynamic average $r_c(T) = \langle r_c\rangle$. The results are shown in Fig. 5. The dependence of μ on T is calculated for a HgSe : Fe sample with an iron concentration $n_{\mathrm{Fe}} = 1.78 \times 10^{19}$ cm^{-3}. Since the concentration of shallow donors in HgSe : Fe is uncontrollable and there are no straightforward methods for measuring it, Fig. 5 shows calculated curves for the mobility for several values of the concentration of shallow donors. Curves *1*–*4* correspond to the donor concentration varying from $n_D = 0$ to $n_D = 6 \times 10^{17}$ cm^{-3} in steps of $\Delta n_D = 2 \times 10^{17}$ cm^{-3}. Experimental points in Fig. 5 are taken from [16] for a $\mathrm{Hg}_{1-x}\mathrm{Fe}_x\mathrm{Se}$ sample with $x = 0.001$. All calculated curves demonstrate a sharp decrease in electron mobility (approximately by a factor of 2) as the lattice temperature changes from 1 to 50 K. This decrease in electron mobility is caused solely by a weakening of the correlations in the positions of scatterers (iron ions) with increasing lattice temperature. In Fig. 5, we see that, to a good accuracy, the experimental points fall on the calculated curves corresponding to the concentrations of shallow donors in the range 4×10^{18} cm^{-3} $< n_D < 6 \times 10^{18}$ cm^{-3}. In the temperature range 35–50 K, the scattering by acoustic phonons appears. As a result, the calculated values of the mobility in this temperature range (where the electron scattering by acoustic phonons was disregarded) are approximately 10% greater than the experimental values of the mobility. Figure 5 clearly demonstrates an anomalous increase in the electron mobility with the concentration of shallow donors (the curves corresponding to higher donor concentrations lie higher). We mentioned this anomaly when discussing the concentration dependence of the mobility (Fig. 4); it is described in detail in [4, 15]. We note that the theory developed here allows us to quantitatively describe the temperature dependence of the electron mobility in HgSe : Fe without using any fitting parameters and without any additional assumptions.

5. CONCLUSIONS

We have suggested a modified version of the hard-sphere model. In the standard model, in a zeroth approximation, the particles are identified with hard spheres (the radius of a sphere is assumed to be equal to the range of the particle potential). In our model, there are no explicit restrictions on the particle potentials; instead, the configurational entropy of the system of particles is identified with the configurational entropy of the system of hard spheres and the diameter of a sphere is assumed to be equal to the minimum distance between the particles. This weakly restrictive version of the hard-sphere model has allowed us to formally extend it to systems of particles with a long-range potential. We have applied this model to calculate spatial correlations of iron ions in HgSe : Fe. The calculated temperature and concentration dependences of the electron mobility agree well with the experimental data.

Thus, estimations made within the suggested modification of the hard-sphere model can be used in experimental studies of spatial correlations of charged impurities in semiconductors.

REFERENCES

1. Isaak M. Tsidilkovski, *Electron Spectrum of Gapless Semiconductros* (Springer Series in Solid State Sciences, 1997), Vol. 116.
2. Z. Wilamowski, K. Swatek, T. Dietl, and J. Kossut, Solid State Commun. **74** (8), 833 (1990).
3. V. M. Mikheev, Fiz. Tverd. Tela (St. Petersburg) **41** (11), 1994 (1999) [Phys. Solid State **41** (11), 1830 (1999)].
4. V. M. Mikheev, Fiz. Tverd. Tela (St. Petersburg) **43** (3), 414 (2001) [Phys. Solid State **43** (3), 430 (2001)].
5. V. M. Mikheev, Fiz. Tverd. Tela (St. Petersburg) **43** (10), 1785 (2001) [Phys. Solid State **43** (10), 1860 (2001)].
6. L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 3rd ed. (Nauka, Moscow, 1976; Pergamon, Oxford, 1980), Parts 1 and 2.
7. R. Balescu, *Equilibrium and Nonequilibrium Statistical Mechanics* (Wiley, New York, 1975; Mir, Moscow, 1978), Vol. 1.
8. I. G. Kuleev, I. Lyapilin, and I. M. Tsidil'kovskiĭ, Zh. Éksp. Teor. Fiz. **102** (11), 1652 (1992) [Sov. Phys. JETP **75** (5), 893 (1992)].
9. W. Dobrowolski, K. Dybko, S. Skierbiszewski, T. Suski, E. Litvin-Staszewska, S. Motkowska, J. Kossut, and A. Mycielski, in *Proceedings of 19th International Conference on Physics of Semiconductors, Warszawa, 1988*, Vol. 2, p. 1247.
10. F. Pool, J. Kossut, U. Debska, and R. Reifenberger, Phys. Rev. B **35** (8), 3900 (1987).
11. N. G. Gluzman, L. D. Sabirzyanova, I. M. Tsidil'kovskiĭ, L. D. Paranchich, and S. Yu. Paranchich, Fiz. Tekh. Poluprovodn. (Leningrad) **20** (1), 94 (1986) [Sov. Phys. Semicond. **20** (1), 55 (1986)].
12. C. Skierbiszewski, T. Suski, W. Dobrowolski, K. Dybko, and A. Mycielski, Semicond. Sci. Technol. **4** (4), 293 (1989).
13. I. G. Kuleev and I. Yu. Arapova, Fiz. Met. Metalloved. **88** (3), 43 (1999) [Phys. Met. Metallogr. **88** (3), 249 (1999)]; Fiz. Tverd. Tela (St. Petersburg) **43** (3), 403 (2001) [Phys. Solid State **43** (3), 420 (2001)].
14. J. M. Ziman, *Models of Disorder* (Cambridge Univ. Press, Cambridge, 1979; Mir, Moscow, 1982).
15. C. Skierbiszewski, Z. Wilamowski, T. Suski, and J. Kossut, Semicond. Sci. Technol. **8** (3), S40 (1993).

Translated by I. Zvyagin

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1100–1106. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1065–1070.
Original Russian Text Copyright © 2005 by Anikeenok.

DEFECTS, DISLOCATIONS,
AND PHYSICS OF STRENGTH

Crystal Field at the Impurity Center Sites in Ionic Crystals

O. A. Anikeenok
Kazan State University, ul. Kremlevskaya 18, Kazan, 420008 Tatarstan, Russia
e-mail: anikeenok@rambler.ru

Received October 22, 2004

Abstract—Perturbation theory is developed for second-quantized operators in a basis of partly nonorthogonal orbitals. This method may be helpful in carrying out *ab initio* calculations of the parameters of the crystal field at the impurity center sites. As an illustration, when estimating the crystal field parameters for Yb^{3+} : KZnF_3, some fitting parameters are calculated using this method. The results agree well with experimental data, which indicates that this theory shows considerable promise. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The concept of the crystal field is widely used in interpreting experimental data on impurity centers in crystals. However, *ab initio* calculations of its parameters are severely complicated even for high-symmetry centers and even in the cluster approximation, where only the nearest neighbor environment is taken into account. For this reason, efforts have been concentrated on developing semiempirical models that include phenomenological fitting parameters. These models can explain the behavior of the physical characteristics of an impurity ion by varying the fitting parameters [1, 2]. However, only quantum-mechanical calculations enable one to determine the mechanisms of the interaction of an impurity ion with its environment and can validate the introduction of the fitting parameters. The most efficient way to prove the validity of the theory is to derive analytical expressions for the crystal field parameters from first principles and compare them with experimental data. In [3, 4], the crystal field operator was found in terms of a superposition model taking into account the overlapping effects and the transfer of an electron from a ligand to the central ion and a comparison with experiment was made for iron group ions. In [5, 6], using the same method, the crystal field parameters of ions in LiYF_4 : Nd^{3+} and KZnF_3 : Er^{3+} were calculated. However, certain quantities involved in expressions for the crystal field parameters were treated as parameters in those studies. For example, the covalence parameters were taken from [7, 8], where these parameters were adjusted when interpreting data on electron–nuclear double resonance. In all the studies mentioned above, there was good agreement with experimental data. It should be noted that these parameters can be introduced only if we know the one-electron orbitals of the ground state of the central ion and its neighbor ions or molecular complexes. Therefore, these orbitals can be a fairly good zeroth approximation in the problems in question. From the above, it follows that perturbation theory can be developed in terms of a basis consisting of partly nonorthogonal orbitals. Certain steps in this direction were made in [9, 10]. In [9], expressions for an arbitrary second-quantized operator were obtained in such a basis. In [9, 10], the electron transition amplitudes from the 2*s* and 2*p* shells of a ligand to the 4*f* shell of a rare-earth ion were calculated. The values of the covalence parameters calculated with these amplitudes agree well with experimental data. In this study, based on the results obtained in [9], we derive an expression for the crystal field operator without using the superposition model. This operator includes the contributions from excited configurations that correspond to electron transitions to empty and partially occupied shells. Under the assumption of the existence of the matrix $(I + S)^{-1}$, the effects of nonorthogonality are included exactly to all orders of perturbation expansion considered. As an example, we calculate the crystal field parameters for Yb^{3+} : KZnF_3. Because calculated electron transition amplitudes from a ligand to the empty 5*d*, 6*s*, and 6*p* shells and certain matrix elements are not currently available, we make estimates taking into account only the terms proportional to the overlap integrals squared. Nevertheless, we show that certain terms are absent from the expressions derived in [5, 6]. When estimating the contributions due to the electron transition from a ligand to the 4*f* shell, we use the transition amplitudes calculated in [9, 10].

2. PARTICLE NUMBER OPERATOR

In [9], we derived an expression for an arbitrary second-quantized operator in terms of a basis of partly nonorthogonal orbitals. In what follows, the representation of an operator in this basis will be referred to as the Ψ representation. Let us find the particle number operator in this representation. According to [9], for an arbitrary one-particle operator, we can write

$$H_\Psi = \exp\left(\frac{1}{2}Q\right)\tilde{H}\exp\left(-\frac{1}{2}Q\right), \qquad (1)$$

$$\tilde{H} = \sum a_{\xi}^{+} a_{\xi'} \langle \xi | (I+S)^{-1} | \theta \rangle \langle \theta | h | \xi' \rangle, \qquad (2)$$

where h is this operator in the coordinate representation, $Q = \sum a_{\xi}^{+} a_{\xi'} \langle \xi | q | \xi' \rangle$, I is the identity operator, and S is the overlap matrix for one-electron orbitals. For the particle number operator in the coordinate representation, we have $h \equiv 1$. Substituting this value into Eq. (2), we obtain

$$\tilde{N} = \sum a_{\xi}^{+} a_{\xi} \qquad (3)$$

and, hence,

$$N_{\Psi} = \sum a_{\xi}^{+} a_{\xi}. \qquad (4)$$

The representation defined by Eq. (2) will be referred to as the non-Hermitian representation. It can be seen that the particle number operator in both the non-Hermitian and Ψ representations has the conventional form. Furthermore, since the operators in these representations are related by a similarity transformation, their eigenvalue spectra are identical. Therefore, there are systems for which the non-Hermitian representation can be used to perform calculations, e.g., systems consisting of a small number of particles or systems for which it is advantageous to use temperature Green's functions [11] (for these functions, the exponential factors in Eq. (1) cancel each other under the operation of trace).

In concluding this section, the following general remark should be made. It is well known that the form of second-quantized operators in an orthonormal one-particle basis is the same for bosons and fermions. In much the same manner as that followed in [9, 12], it can be shown that all expressions obtained in the section Theory in [9] and Eqs. (3) and (4) in this paper are also valid for bosons. This will occur if the creation and annihilation operators satisfy the commutation relations for bosons, i.e.,

$$\begin{aligned} a_{\xi} a_{\xi'}^{+} - a_{\xi'}^{+} a_{\xi} &= \delta_{\xi\xi'}, \\ a_{\xi} a_{\xi'} - a_{\xi'} a_{\xi} &= a_{\xi}^{+} a_{\xi'}^{+} - a_{\xi'}^{+} a_{\xi}^{+} = 0, \end{aligned} \qquad (5)$$

and if all calculations are performed for the so-called permanents (products of one-particle orbitals symmetrized with respect to pair permutations). Then, in the notation used in [9], the functions $|\{\eta\}\rangle$ will have the form

$$|\{\eta\}\rangle = \Pi \frac{(a_{\xi}^{+})^{n_{\xi}}}{\sqrt{n_{\xi}!}} |0\rangle, \qquad (6)$$

where n_{ξ} are the occupation numbers of the orbitals. This approach may prove helpful if the one-particle approximation is a good zeroth approximation for a system of interacting bosons.

3. THE HAMILTONIAN OF A SYSTEM OF IONS

Let us consider a system consisting of an arbitrary number of ions. The ion positions and the quantum numbers of the ion orbitals are indicated by indices ξ, ξ', η, η', and so on; i.e., $\xi = (\mathbf{R}_i, nlm_l m_s)$. A distribution of electrons over ion orbitals will be referred to as a configuration. The Hamiltonian of the system can be written as

$$\begin{aligned} H = \sum_{i} h_k(\mathbf{r}_i) + \frac{1}{2} \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,j} \frac{Z_j}{|\mathbf{r}_i - \mathbf{R}_j|} \\ + \frac{1}{2} \sum_{i \neq j} \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|}, \end{aligned} \qquad (7)$$

where $h_k(\mathbf{r})$ is the kinetic energy and Z_i is the nuclear charge of an ion. We write the nuclear charge as the sum $Z_i = q_i + n_i + m_i$, where q_i is the ionic charge in the crystal without impurities, n_i is the number of electrons in a certain configuration of an ion, and m_i is a deviation from the number of electrons in this configuration in the impurity-free crystal. A deficiency and excess of electrons correspond to $m_i > 0$ and $m_i < 0$, respectively. We will refer to m_i as a charge defect. Substituting the above expression for the nuclear charge into Eq. (7), we obtain

$$\begin{aligned} H = \frac{1}{2} \sum_{i \neq j} \frac{q_i q_j}{|\mathbf{R}_i - \mathbf{R}_j|} + \sum_{i \neq j} \frac{m_i q_j}{|\mathbf{R}_i - \mathbf{R}_j|} + \frac{1}{2} \sum_{i \neq j} \frac{m_i m_j}{|\mathbf{R}_i - \mathbf{R}_j|} \\ + \sum_{i} h_k(\mathbf{r}_i) - \sum_{i,j} \frac{Z_j}{|\mathbf{r}_i - \mathbf{R}_j|} + \sum_{i \neq j} \frac{n_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|} \\ + \frac{1}{2} \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \frac{1}{2} \sum_{i \neq j} \frac{n_i n_j}{|\mathbf{R}_i - \mathbf{R}_j|}. \end{aligned} \qquad (8)$$

In Eq. (8), the first term is the electrostatic energy of the crystal in the ionic approximation and the second term is the interaction energy between the charge defects and the crystal lattice. It can be seen that the energy of the crystal decreases when an electron either is removed from a site with a positive Madelung energy or is added to a site with a negative energy. The third term in Eq. (8) is the interaction energy between the charge defects. If we separate the intraionic interactions from the other terms, then the interionic interactions will be the difference of quantities of the same order of magnitude and, therefore, can be considered a perturbation. This partition of the Hamiltonian can be conveniently used to consider virtual transitions of an electron between ions. Passing over to second quantization in this case does not pose complications, because the particle number operator in the formalism we develop here was already defined above.

4. PERTURBATION THEORY

In order to solve our problem, we apply perturbation theory for the degenerate case in the same form as that used in [13] and described in detail in [14]. As shown in [9, 10], the ratio of the electron transition amplitude to the transition energy is on the order of the overlap integral corresponding to the transition in question. Therefore, using the results obtained in [9], we can write the Hamiltonian to third order in the transition energy as

$$H = \bar{H} - \frac{1}{8}[Q, \bar{H}]^{(2)} + \frac{1}{2}[\bar{H}, S_1] + \frac{1}{2}[\bar{H}, S_2], \tag{9}$$

$$\begin{aligned}\bar{H} &= \sum a^{+}_{\xi} a_{\xi'} \langle \xi | \bar{h} | \xi' \rangle \\ &+ \frac{1}{2} \sum a^{+}_{\xi} a^{+}_{\eta} a_{\eta'} a_{\xi'} \langle \xi\eta | \bar{g} | \xi'\eta' \rangle,\end{aligned} \tag{10}$$

$$\begin{aligned}&\langle \xi | (I+S)^{-1} | \theta \rangle \equiv \langle \xi || \theta \rangle, \\ &\langle \xi | (I+S)^{-1} | \theta \rangle \langle \eta | (I+S)^{-1} | \zeta \rangle \equiv \langle \xi\eta || \theta\zeta \rangle,\end{aligned} \tag{11}$$

$$\langle \xi | \bar{h} | \xi' \rangle = \frac{1}{2} \sum \langle \xi || \theta \rangle \langle \theta | h | \xi' \rangle + \frac{1}{2} \sum \langle \xi | h | \theta \rangle \langle \theta || \xi' \rangle, \tag{12}$$

$$\begin{aligned}\langle \xi\eta | \bar{g} | \xi'\eta' \rangle &= \frac{1}{2} \sum \langle \xi\eta || \theta\zeta \rangle \langle \theta\zeta | g | \xi'\eta' \rangle \\ &+ \frac{1}{2} \sum \langle \xi\eta | g | \theta\zeta \rangle \langle \theta\zeta || \xi'\eta' \rangle.\end{aligned} \tag{13}$$

The matrices S_1 and S_2 can be found in [13, 14]. In the expression for the operator Q written above (see the text below Eq. (2)), $q = \ln(I + S)$. Note that the number of orbitals included in the partly nonorthogonal basis is determined by the matrix $(I + S)^{-1}$ alone; indeed, if this matrix exists, so does the matrix q. The commutator $[Q, \bar{H}]^{(2)}$, involved in Eq. (9), can be found to be

$$\begin{aligned}[Q, \bar{H}]^{(2)} &= \sum a^{+}_{\xi} a_{\xi'} \langle \xi | [q[q, \bar{h}]] | \xi' \rangle \\ &+ \frac{1}{2} \sum a^{+}_{\xi} a^{+}_{\eta} a_{\eta'} a_{\xi'} \langle \xi\eta | [q[q, \bar{g}]] | \xi'\eta' \rangle,\end{aligned} \tag{14}$$

$$\langle \xi | [q[q, \bar{h}]] | \xi' \rangle = \langle \xi | q^2 \bar{h} - 2q\bar{h}q + \bar{h}q^2 | \xi' \rangle, \tag{15}$$

$$\begin{aligned}\langle \xi\eta | [q[q, \bar{g}]] | \xi'\eta' \rangle &= \langle \xi_1\eta_2 | (q_1^2 + q_2^2)\bar{g}_{12} + 2q_1q_2\bar{g}_{12} \\ &- 4q_1\bar{g}_{12}q_1 - 4q_1\bar{g}_{12}q_2 + 2\bar{g}_{12}q_1q_2 \\ &+ \bar{g}_{12}(q_1^2 + q_2^2) | \xi'_1\eta'_2 \rangle,\end{aligned} \tag{16}$$

where the matrix elements in the right-hand sides of Eqs. (15) and (16) can be written as

$$\langle \xi | q^2 \bar{h} | \xi' \rangle = \sum \langle \xi | q | \theta \rangle \langle \theta | q | \zeta \rangle \langle \zeta | \bar{h} | \xi' \rangle, \tag{17}$$

$$\begin{aligned}&\langle \xi_1\eta_2 | q_1\bar{g}_{12}q_2 | \xi'_1\eta'_2 \rangle \\ &= \sum \langle \xi | q | \theta \rangle \langle \theta\eta | \bar{g} | \xi'\zeta \rangle \langle \zeta | q | \eta' \rangle.\end{aligned} \tag{18}$$

The other terms can be transformed analogously. It can be seen that, in our theory, the one-particle interaction operators have a one-particle form and the two-particle interaction operators have a two-particle form; i.e., so-called *n*-particle interactions do not appear.

5. CRYSTAL FIELD OPERATOR

In order to find the crystal field operator, let us analyze each of the terms in Eq. (9). It was shown in [5, 6] that, when including the excited configurations, primarily the electron transitions from a ligand to the partly filled and empty shells of the central ion should be taken into account. The second-quantized operator that involves these excited configurations and operates in the space of the ground-state configuration is constructed in the usual way [15]. For this purpose, we denote the electron creation (annihilation) operators of the filled and partly filled shells as a^{+}_{ξ} ($a_{\xi'}$), those of the empty shells as d^{+}_{η} ($d_{\eta'}$), and those of a ligand as b^{+}_{θ} ($b_{\theta'}$) and use the commutation relations that are valid for the ground-state configuration, namely, $d_{\eta'} d^{+}_{\eta} = \delta_{\eta\eta'}$ and $b^{+}_{\theta} b_{\theta'} = \delta_{\theta\theta'}$. In what follows, $|\xi_a\rangle$ are the orbitals for electrons of the central ion and $|\xi_b\rangle$ are the orbitals for electrons of the ligands. For the first term in Eq. (9), we obtain

$$\begin{aligned}\bar{H} &= \frac{1}{2} \sum a^{+}_{\xi} a_{\xi'} \Bigg[\delta_{\xi\xi'} \sum_b \frac{q_b + m_b}{|\mathbf{R}_a - \mathbf{R}_b|} \\ &+ \langle \xi || \theta \rangle \left\langle \theta \left| h_k - \frac{Z_a}{|\mathbf{r}|} - \sum_b \frac{q_b + n_b + m_b}{|\mathbf{r} - \mathbf{R}_b|} \right| \xi' \right\rangle \Bigg] \\ &+ \frac{1}{4} \sum a^{+}_{\xi} a^{+}_{\eta} a_{\eta'} a_{\xi'} \langle \xi\eta || \theta\zeta \rangle \langle \theta\zeta | g | \xi'\eta' \rangle \\ &+ \frac{1}{2} \sum a^{+}_{\xi} a_{\xi'} \langle \xi\dot{\theta}_b || \theta\zeta \rangle \langle \theta\zeta | g(1 - P) | \xi'\dot{\theta}_b \rangle + \text{H.c.},\end{aligned} \tag{19}$$

where summation over $\dot{\theta}_b$ is performed only over the filled shells of the ground-state configuration and P is the permutation operator. The values of q_b, n_b, and m_b are also related to the ground-state configuration; and H.c. stands for the Hermitian conjugate. Equation (19) does not contain the first three terms of Eq. (8), because these terms cause a shift in energy and will enter into the denominators of the perturbation series.

Now, let us consider the second term in Eq. (9). If the matrix elements of the operator Q are not small, then the Hamiltonian of the system should be taken in the form of Eq. (8) in order to separate the zeroth-approximation Hamiltonian and the perturbation. In ionic crystals, however, these matrix elements are likely to be sufficiently small for the commutator in question to be treated as a perturbation. In this case, the interaction Hamiltonian can be taken in the form of Eq. (7). To within third-order terms, we can write Eq. (14) in the form

$$\begin{aligned}[Q, \bar{H}]^{(2)} = {} & \sum \overset{+}{a}_{\xi} a_{\xi'} \langle\xi|q|\theta_b\rangle[\langle\theta_b|q|\eta_a\rangle\langle\eta_a|\bar{h}|\xi'\rangle \\ & - \langle\theta_b|\bar{h}|\zeta_b\rangle\langle\zeta_b|q|\xi'\rangle] + \sum \overset{+}{a}_{\xi} \overset{+}{a}_{\eta} a_{\eta'} a_{\xi'} \langle\xi|q|\theta_b\rangle \\ & \times [\langle\theta_b|q|\alpha_a\rangle\langle\alpha_a\eta|\bar{g}|\xi'\eta'\rangle - \langle\theta_b\eta|\bar{g}|\zeta_b\eta'\rangle\langle\zeta_b|q|\xi'\rangle] \\ & + \sum \overset{+}{a}_{\xi} a_{\xi'} \langle\dot{\theta}_b|q|\alpha_a\rangle[\langle\alpha_a|q|\zeta_b\rangle\langle\zeta_b\xi|\bar{g}|\dot{\theta}_b\xi'\rangle \\ & - 2\langle\xi|q|\zeta_b\rangle\langle\alpha_a\zeta_b|\bar{g}|\xi'\dot{\theta}_b\rangle] \\ & + \sum \overset{+}{a}_{\xi} a_{\xi'} \langle\xi|q|\zeta_b\rangle[\langle\zeta_b|q|\alpha_a\rangle\langle\dot{\theta}_b\alpha_a|\bar{g}|\dot{\theta}_b\xi'\rangle \\ & - \langle\dot{\theta}_b\zeta_b|\bar{g}(1-P)|\dot{\theta}_b\beta_b\rangle\langle\beta_b|q|\xi'\rangle] \\ & - \sum \overset{+}{a}_{\xi} a_{\xi'} \langle\dot{\theta}_b|q|\alpha_a\rangle\langle\alpha_a\xi|\bar{g}(1-P)|\beta_a\xi'\rangle\langle\beta_a|q|\dot{\theta}_b\rangle + \text{H.c.}\end{aligned} \tag{20}$$

It can be seen that, in both Eqs. (19) and (20), the nonorthogonality effects are included exactly if the matrix $(I + S)^{-1}$ exists. Therefore, these expressions are easy to analyze. By separating the one-particle and two-particle single-center interactions in the first two terms in Eq. (20), it can be found that the largest terms are the Hartree–Fock energy, the Madelung energy, and the interaction energy between a charge defect and the crystal lattice, which are renormalized by the matrix elements of Q and $(I + S)^{-1}$. For example, for rare-earth elements, the largest contribution to the last term in Eq. (20) comes from expressions of the type $S_{b,\,5d}\langle 5d, 4f|g(1 - P)|5d, 4f\rangle S_{5d,\,b}$ (which are absent from the expressions for the crystal field parameters derived in [5, 6]). Note that these contributions from the filled shells are canceled by the corresponding contributions to the second term. This is immediately clear if the operator $\overset{+}{a}_{\xi} a_{\xi'}$ in this term is related to the electrons of a filled shell. The other terms describe interactions of the density–density type, which were discussed, e.g., in [16] in terms of a phenomenological approach.

Let us turn to the third term in Eq. (9). The electron transition amplitude between states $|\{\xi\}|$ and $|\{\xi'\}|$ that differ in terms of only one of their quantum numbers was determined in [9]. In what follows, this amplitude will be designated as $\langle\{\xi\}|H_\Psi|\{\xi'\}\rangle = \langle\xi|G|\zeta\rangle$. Here, $|\zeta\rangle$ is the state where an electron is annihilated and $|\xi\rangle$ is the state where the electron is created. Using this notation, we can write

$$\begin{aligned}[\bar{H}, S_1] = {} & \sum \overset{+}{a}_{\xi} a_{\xi'} \frac{\langle\xi|G|\dot{\theta}_b\rangle\langle\dot{\theta}_b|G|\xi'\rangle}{|\Delta_{\dot{\theta}_b,\,\xi}|} \\ & - \sum \frac{\langle\xi_a|G|\dot{\theta}_b\rangle\langle\dot{\theta}_b|G|\xi_a\rangle}{|\Delta_{\dot{\theta}_b,\,\xi_a}|} \\ & - \sum \overset{+}{a}_{\xi} a_{\xi'} \frac{\langle\dot{\theta}_b\xi|\bar{g}(1-P)|\eta_d\xi'\rangle\langle\eta_d|G|\dot{\theta}_b\rangle}{|\Delta_{\dot{\theta}_b,\,\eta_d}|} + \text{H.c.},\end{aligned} \tag{21}$$

where $|\Delta_{\dot{\theta}_b,\,\eta_d}|$ is the energy of the electron transition from a ligand to the central ion. It can be seen that, in Eq. (21), as in Eq. (20), the contributions from the filled shells to the first two terms cancel each other. Due to summation over the empty shells in the second term, as would be expected, the energy of the entire configuration decreases. Most likely, the third term can make a noticeable contribution only for rare-earth elements.

In the last term in Eq. (9), we retain only the contributions from the electron transitions to the empty shells of the central ion. As follows from the studies cited above, these contributions have to be taken into account for rare-earth elements. In this case, we obtain

$$\begin{aligned}& \frac{1}{2}[\bar{H}, S_2]_d \\ & = \sum \overset{+}{a}_{\xi} a_{\xi'} \frac{\langle\dot{\theta}_b|G|\eta_d\rangle\langle\xi\eta_d|\bar{g}(1-P)|\xi'\alpha_d\rangle\langle\alpha_d|G|\dot{\theta}_b\rangle}{|\Delta_{\dot{\theta}_b,\,\eta_d}||\Delta_{\alpha_d,\,\dot{\theta}_b}|}.\end{aligned} \tag{22}$$

This term can also be significant for iron group ions if some of the nearest neighbor ions are at a considerable distance from the central ion [17]. Note that the expressions written above can also be used to include the covalency effects in impurity-free ionic crystals.

6. ESTIMATION OF THE CRYSTAL FIELD PARAMETERS FOR Yb^{3+} : KZnF$_3$

The derivation of analytical expressions for the operators discussed in the preceding section is very labor consuming, because we need to know not only the matrices $(I + S)^{-1}$ and Q but also the matrix elements of the relevant operators for arbitrary ion positions. Nevertheless, modern computers make it possible to solve this problem. To begin with, let us consider the simplest operator (which we designate as h_c^a)

$$\begin{aligned}h_c^a = {} & \frac{1}{2}\sum \overset{+}{a}_{\xi} a_{\xi'} \Bigg[\langle\xi|(I+S)^{-1}|\xi_a\rangle \\ & \times \left\langle \xi_a \left| -\sum_b \frac{q_b}{|\mathbf{r}-\mathbf{R}_b|} \right| \xi' \right\rangle\Bigg] + \text{H.c.}\end{aligned} \tag{23}$$

and, following [15], pass over from the second-quantized operators to irreducible tensor operators. As a result, we obtain

$$\sum \overset{+}{a}_{\xi} a_{\xi'} \left\langle \xi \left| -\sum_b \frac{q_b}{|\mathbf{r}-\mathbf{R}_b|} \right| \xi' \right\rangle = \sum_{k,\,q} a_q^k V_q^k, \tag{24}$$

$$\left\langle \xi \left| -\sum_b \frac{q_b}{|\mathbf{r}-\mathbf{R}_b|} \right| \xi' \right\rangle = \sum_{k,q} (-1)^{l-m_l^{\xi}} \begin{pmatrix} l & k & l \\ -m_l^{\xi} & q & m_l^{\xi'} \end{pmatrix} a_q^k, \tag{25}$$

$$W_{0q}^{0k} = 2^{-\frac{1}{2}}(2k+1)^{\frac{1}{2}} v_q^k, \quad k \neq 0, \qquad W_{00}^{00} = N[2(2l+1)]^{-\frac{1}{2}}, \tag{26}$$

$$\langle \xi | (I+S)^{-1} | \xi' \rangle = \sum_{k,q} (-1)^{l-m_l^{\xi}} \begin{pmatrix} l & k & l \\ -m_l^{\xi} & q & m_l^{\xi'} \end{pmatrix} s_q^k, \tag{27}$$

where s_q^k are analogs of a_q^k in Eq. (24). Substituting Eqs. (25) and (27) into Eq. (23) and passing over to irreducible tensor operators, we obtain

$$h_c^a = \sum_{j_1 j_2 m_1 m_2 k q} s_{m_1}^{j_1} a_{m_2}^{j_2} \begin{pmatrix} j_1 & j_2 & k \\ m_1 & m_2 & -q \end{pmatrix} \begin{Bmatrix} j_1 & j_2 & k \\ l & l & l \end{Bmatrix} \times \frac{[k](-1)^{l+q}}{[l]\begin{pmatrix} l & k & l \\ 0 & 0 & 0 \end{pmatrix}} C_q^k = \sum_{kq} B_q^k(a,c) C_q^k, \tag{28}$$

where (: : :) and {: : :} are 3j and 6j symbols, respectively; $[l] = 2l + 1$; and C_q^k is the spherical tensor operator. The quantities $s_{m_1}^{j_1}$ and $a_{m_2}^{j_2}$ are determined by the electron density distribution over the entire crystal. In the case of cubic symmetry, b_4 and b_6 are experimentally measured quantities and are defined as follows:

$$H_{\text{cr}} = \sum_{kq} B_q^k C_q^k, \quad b_4 = \frac{1}{8} B_0^4, \quad b_6 = \frac{1}{16} B_0^6. \tag{29}$$

Using the Gaussian expansion of Hartree–Fock functions [18] (with expansion coefficients a_i, α_i), we obtain the following results for the 4f shell:

$$a_0^0 = -\frac{\sqrt{7}}{16} \sum_b q_b \sum_{i,j} a_i a_j \left(\frac{1}{\alpha_i \alpha_j}\right)^4 \int_0^1 dx [105(1-x^2)^3 + 210\alpha_{ij} R_{ab}^2 x^4 (1-x)^2 + 84\alpha_{ij}^2 R_{ab}^4 x^8 (1-x^2) + 8\alpha_{ij}^3 R_{ab}^6 x^{12}] \exp(-\alpha_{ij} R_{ab}^2 x^2) C_0^0(\theta_b, \varphi_b), \tag{30}$$

$$a_0^4 = -\frac{9}{4}\sqrt{\frac{2\times 7}{11}} \sum_b q_b R_{ab}^4 \sum_{ij} a_i a_j \left(\frac{1}{\alpha_i \alpha_j}\right)^2 \times \int_0^1 dx [11x^8(1-x^2) + 2\alpha_{ij} R_{ab}^2 x^{12}] \times \exp(-\alpha_{ij} R_{ab}^2 x^2) C_0^4(\theta_b, \varphi_b), \tag{31}$$

$$a_0^6 = 5\sqrt{\frac{7\times 13}{3\times 11}} \sum_b q_b R_{ab}^6 \sum_{ij} a_i a_j \left(\frac{1}{\alpha_i \alpha_j}\right) \times \int_0^1 dx [x^{12} \exp(-\alpha_{ij} R_{ab}^2 x^2)] C_0^6(\theta_b, \varphi_b), \tag{32}$$

where $\alpha_{ij} = \alpha_i + \alpha_j$, R_{ab} is the distance between the ionic nuclei, and $C_{ij}^k(\theta_b, \varphi_b)$ is the spherical tensor function of the angles θ_b and φ_b defining the orientation of the a–b pair axis in a chosen coordinate frame. Expressions (30)–(32) are valid for any distances R_{ab}. These expressions are obtained using the transformations

$$\frac{1}{|\mathbf{r}-\mathbf{R}|} = \frac{2}{\sqrt{\pi}} \int_0^{\infty} dy \exp[-(\mathbf{r}-\mathbf{R})^2 y^2], \qquad y^2 = \frac{\alpha_{ij} x^2}{1-x^2} \tag{33}$$

which differ from the transformations used in [18]. In Eq. (33), the integrand is a superposition of polynomials like $x^{2n}(1 - x^2)^m$ rather than x^n (as in [18]), which makes mathematical manipulations much simpler. Note that, using Eq. (33), the matrix elements of the Coulomb interaction between electrons are also brought to the form of Eqs. (30)–(32) and, hence, are analytic functions of the ion position vectors.

The expressions given above for the matrix elements of the operators involved in Eq. (9) are exact. However, only the electron transition amplitudes to the 4f shell, namely, $\langle 4f0|G|2s\rangle = G_{4fs}$, $\langle 4f0|G|2p0\rangle = G_{4f\sigma}$, and $\langle 4f1|G|2p1\rangle = G_{4f\pi}$, have been calculated to date [9, 10]. The transition amplitudes to 5d, 6s, and 6p shells have not yet been estimated numerically, and the matrix elements of the operators $(I + S)^{-1}$ and Q have not yet been calculated. Therefore, we estimate the contributions from the operators considered in Section 5 to within the terms quadratic in the overlap integrals. We take into account the contributions missed in [5, 6] and use the calculated transition amplitudes currently available. The other quantities are estimated using the commonly accepted approximations. As a result, we obtain the following expressions for the crystal field parameters B_0^4 and B_0^6 for a rare-earth ion in an octahedral environment:

$$\begin{aligned} B_0^4 = {} & \frac{9}{2}\Bigg[\left(\frac{G_{4fs}}{|\Delta_{4fs}|} - S_{4fs}\right)G_{s4f} + \left(\frac{G_{4f\sigma}}{|\Delta_{4f\sigma}|} - S_{4f\sigma}\right)G_{\sigma 4f} \\ & + \frac{1}{3}\left(\frac{G_{4f\pi}}{|\Delta_{4f\pi}|} - S_{4f\pi}\right)G_{\pi 4f}\Bigg] + B_0^4(a, c) + h_\rho^4 \\ & + \frac{9}{8}(\varepsilon_{\mathrm{XF}}^{\mathrm{Yb}^{3+}} + h_M^a - h_M^b - h_m - 4\alpha - 2\langle 4f, 4f|g|4f, 4f\rangle) \\ & \times\left(S_{4fs}^2 + S_{4f\sigma}^2 + \frac{1}{3}S_{4f\pi}^2\right) \\ & - \frac{9}{8}\left[\varepsilon_{\mathrm{XF}}^{2s}S_{4fs}^2 + \varepsilon_{\mathrm{XF}}^{2p}\left(S_{4f\sigma}^2 + \frac{1}{3}S_{4f\pi}^2\right)\right] \\ & + \Bigg[F^{(4)}(f, d) - \frac{99}{70}G^{(1)}(f, d) - \frac{22}{35}G^{(3)}(f, d) \\ & - \frac{5}{154}G^{(5)}(f, d)\Bigg]\Bigg[\frac{1}{4}\left(S_{5ds}^2 + S_{5d\sigma}^2 - \frac{4}{3}S_{5d\pi}^2\right) \\ & + \frac{G_{5ds}^2}{|\Delta_{5ds}|^2} + \frac{G_{5d\sigma}^2}{|\Delta_{5d\sigma}|^2} - \frac{4}{3}\frac{G_{5d\pi}^2}{|\Delta_{5d\pi}|^2}\Bigg], \end{aligned} \tag{34}$$

$$\begin{aligned} B_0^6 = {} & \frac{39}{28}\Bigg[\left(\frac{G_{4fs}}{|\Delta_{4fs}|} - S_{4fs}\right)G_{s4f} + \left(\frac{G_{4f\sigma}}{|\Delta_{4f\sigma}|} - S_{4f\sigma}\right)G_{\sigma 4f} \\ & - \frac{3}{2}\left(\frac{G_{4f\pi}}{|\Delta_{4f\pi}|} - S_{4f\pi}\right)G_{\pi 4f}\Bigg] + B_0^6(a, c) + h_\rho^6 \\ & + \frac{39}{112}(\varepsilon_{\mathrm{XF}}^{\mathrm{Yb}^{3+}} + h_M^a - h_M^b - h_m - 4\alpha \\ & - 2\langle 4f, 4f|g|4f, 4f\rangle)\left(S_{4fs}^2 + S_{4f\sigma}^2 - \frac{3}{2}S_{4f\pi}^2\right) \\ & - \frac{39}{112}\left[\varepsilon_{\mathrm{XF}}^{2s}S_{4fs}^2 + \varepsilon_{\mathrm{XF}}^{2p}\left(S_{4f\sigma}^2 - \frac{3}{2}S_{4f\pi}^2\right)\right]. \end{aligned} \tag{35}$$

The parameters B_0^4 and B_0^6 are estimated for the distance $R = 4.16$ au corresponding to the $\mathrm{KZnF_3 : Yb^{3+}}$ compound. We obtain $B_0^4(a, c) = 592\ \mathrm{cm}^{-1}$ and $B_0^6(a, c) = 23\ \mathrm{cm}^{-1}$. The quantities $S_{4fs} = 0.009019$, $S_{4f\sigma} = -0.013558$, and $S_{4f\pi} = 0.008142$ are calculated using the functions from [19]. The quantities $G_{4fs} = -0.01297$ au, $G_{4f\sigma} = 0.01530$ au, and $G_{4f\pi} = -0.01013$ au are determined in [9, 10], and the quantities $|\Delta_{4fs}| \approx 1$ au and $|\Delta_{4f\sigma}| \approx |\Delta_{4f\pi}| \approx 0.3$ au are taken from [7]. The Hartree–Fock energies are $\varepsilon_{\mathrm{HF}}^{\mathrm{Yb}^{3+}} = -2.01$ au, $\varepsilon_{\mathrm{HF}}^{2s} = -1.07$ au, and $\varepsilon_{\mathrm{HF}}^{2p} = -0.18$ au, the Madelung energies are $h_M^a = 0.82$ au and $h_M^b = -0.42$ au, and the hole energy is $h_m = -0.23$ au. The other needed parameters are taken from [6]. The Coulomb f–d interaction parameters are $F^{(4)}(f, d) = 0.05129$ au, $G^{(1)}(f, d) = 0.04495$ au, $G^{(3)}(f, d) = 0.03728$ au, $G^{(5)}(f, d) = 0.02856$ au, $\alpha = 0.43$, $S_{5ds} = 0.19$, $S_{5d\sigma} = -0.18$, and $S_{5d\pi} = 0.12$. The covalency parameters $-G_{5di}/|\Delta_{5di}| = \gamma_{5di}$ are $\gamma_{5ds} = 0.02$, $\gamma_{5d\sigma} = -0.13$, and $\gamma_{5d\pi} = 0.09$. Finally, the density–density overlaps are $h_\rho^4 = 441\ \mathrm{cm}^{-1}$ and $h_\rho^6 = -19.2\ \mathrm{cm}^{-1}$. Substituting these values into Eqs. (34) and (35) and using Eq. (29), we obtain $b_4 = 130\ \mathrm{cm}^{-1}$ and $b_6 = 8.4\ \mathrm{cm}^{-1}$. The experimental values of these parameters for $\mathrm{KZnF_3 : Yb^{3+}}$ [20] are $b_4 = 302\ \mathrm{cm}^{-1}$ and $b_6 = 13\ \mathrm{cm}^{-1}$. Thus, there is fairly good agreement between the experimental data and the tentative estimates, which we made with the aim of testing the validity of the perturbation theory developed here. However, it is clear that the predicted values of the crystal field parameters will be refined if more correct expressions are used for the operators $(I + S)^{-1}$ and Q. For example, the overlap integrals between the orbitals of a ligand and the 5d, 6s, and 6p orbitals of the central ion are fairly large. Therefore, it is insufficient to restrict *ab initio* calculations to the terms quadratic in the overlap integrals. Furthermore, even the possibility of including these orbitals in the basis is determined be the existence of the matrix $(I + S)^{-1}$. Therefore, the next natural step forward would be to calculate this matrix with inclusion of these orbitals and the electron transition amplitude from a ligand to these orbitals.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, project no. 02-02-16648.

REFERENCES

1. D. J. Newman, Adv. Phys. **20**, 84 (1971); Adv. Phys. **20**, 197 (1971).
2. B. Z. Malkin, *Spectroscopy of Solids Containing Rare-Earth Ions* (North-Holland, Amsterdam, 1987).
3. M. V. Eremin and A. A. Kornienko, Fiz. Tverd. Tela (Leningrad) **19** (3), 775 (1977) [Sov. Phys. Solid State **19** (3), 449 (1977)].
4. M. V. Eremin, Paramagn. Reson. **20**, 84 (1984).
5. M. V. Eremin, Fiz. Tverd. Tela (Leningrad) **29** (1), 254 (1987) [Sov. Phys. Solid State **29** (1), 145 (1987)].
6. M. L. Falin, M. V. Eremin, M. M. Zaripov, I. R. Ibragimov, A. M. Leushin, R. Yu. Abdulsabirov, and S. L. Korbaeva, J. Phys.: Condens. Matter **1** (13), 2331 (1989).
7. O. A. Anikeenok, M. V. Eremin, M. L. Falin, A. L. Konkin, and V. P. Meiklyar, J. Phys. C: Solid State Phys. **17** (15), 2813 (1984).

8. O. A. Anikeenok, M. V. Eremin, and O. G. Khutsishvilli, Fiz. Tverd. Tela (Leningrad) **28** (6), 1690 (1986) [Sov. Phys. Solid State **28** (6), 935 (1986)].
9. O. A. Anikeenok, Fiz. Tverd. Tela (St. Petersburg) **45** (5), 812 (2003) [Phys. Solid State **45** (5), 854 (2003)].
10. O. A. Anikeenok *et al.*, Phys. Rev. B (in press).
11. A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics* (Fizmatgiz, Moscow, 1962; Prentice-Hall, Englewood Cliffs, N.J., 1963).
12. O. A. Anikeenok, Available from VINITI No. 2442-V87 (April 6, 1987).
13. O. A. Anikeenok and M. V. Eremin, Fiz. Tverd. Tela (Leningrad) **23** (3), 706 (1981) [Sov. Phys. Solid State **23** (3), 401 (1981)].
14. G. L. Bir and G. E. Pikus, *Symmetry and Strain-Induced Effects in Semiconductors* (Nauka, Moscow, 1972; Wiley, New York, 1975).
15. B. R. Judd, *Second Quantization and Atomic Spectroscopy* (Johns Hopkins, Baltimore, 1967).
16. D. Garcia and M. Faucher, Phys. Rev. B **30** (4), 1703 (1984).
17. V. A. Ulanov, O. A. Anikeenok, M. M. Zaripov, and I. I. Fazlizhanov, Fiz. Tverd. Tela (St. Petersburg) **45** (10), 1814 (2003) [Phys. Solid State **45** (10), 1906 (2003)].
18. H. Taketa, S. Huzinaga, and K. J. O-ohata, J. Phys. Soc. Jpn. **21** (11), 2313 (1966).
19. E. Clementi and R. Roetti, At. Data Nucl. Data Tables **14**, 177 (1974).
20. A. A. Antipin, A. A. Fediĭ, and E. A. Tsvetkov, Paramagn. Reson. **12**, 133 (1976).

Translated by Yu. Epifanov

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1107–1110. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1071–1074.

MAGNETISM AND FERROELECTRICITY

Structure and Properties of Deposited Yttrium Iron Garnet Films

V. F. Shkar'*, E. I. Nikolaev*, V. N. Sayapin*, A. I. Linnik*,
V. P. Denysenkov, A. M. Grishin**, and S. I. Khartsev****
**Galkin Physicotechnical Institute, National Academy of Sciences of Ukraine,*
ul. R. Lyuksemburg 72, Donetsk, 83114 Ukraine
e-mail: shkar@donapex.net.ua, nikol@donapex.net.ua
***Department of Condensed Matter Physics, Royal Institute of Technology, Stockholm–Kista, S-164 40 Sweden*
Received June 25, 2004

Abstract—Ferromagnetic resonance (FMR) and vibrating-sample magnetometer techniques were used to study the nature of the structural characteristics of yttrium iron garnet films deposited through either liquid phase epitaxy or laser evaporation on a (111)-oriented gallium gadolinium garnet substrate. It was proved that, based on the experimentally observed cubic magnetic anisotropy, deposited films should be considered to be single crystals. However, the absence of the FMR domain branch in a nonsaturated film and the shape of the magnetization curve indicate that a deposited film when demagnetized does not have a domain structure, as would be expected for a single-crystal film. According to the model proposed, a deposited film consists of close-packed single-crystal fragments with equal crystallographic orientation, the boundaries between which are in a partially atomically disordered state. As a result, such a film is both locally and macroscopically anisotropic, like a continuous single crystal. This film can split into domains only within a fragment (as is the case in a magnetic granular polycrystal); however, this does not happen, because the linear dimensions of a submicroscopic fragment are smaller than the equilibrium domain width. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Garnet ferrite films, such as yttrium iron garnet (YIG) films, fabricated using liquid phase epitaxy (LPE), have been intensively studied for many years. The interest is also drawn to garnet ferrite films deposited through laser evaporation (LE) and electron beam evaporation (EBE) [1, 2]. In particular, the LE method has been used to deposit films of bismuth iron garnet $Bi_3Fe_5O_{12}$ that exhibit a record-high Faraday effect, which was not possible with the LPE method. Analysis of FMR spectra made it possible to observe presumably bismuth magnetism in them [3]. When studied with a polarizing light microscope, LE-deposited $Bi_3Fe_5O_{12}$ films show an interesting property: no domain structure is observed, even though the film thickness (~ 1 μm) is more than sufficient to show a distinctly high magnetooptical contrast. However, x-ray structural analysis shows that these films are almost single-crystal [4] and, consequently, must have a domain structure. A similar pattern was observed in LE-deposited YIG films, whereas LPE-deposited YIG films did not exhibit a domain structure.

In previous studies of ultrathin YIG films LPE-deposited on an (111)-oriented gallium gadolinium garnet (GGG) substrate, it was shown that some information on the film domain structure that was not visible with a polarizing light microscope could be obtained by analyzing FMR spectra and studying magnetization reversal of the films in a quasistatic magnetic field [5].

In this study of LE (EBE)-evaporated YIG films and LPE-deposited YIG films, we compared their magnetization values and used the FMR technique to elucidate the nature of the structural features of LE (EBE)-evaporated YIG films.

2. EXPERIMENTAL

2.1. Specimens and Experimental Technique

Two types of specimens were studied: (i) ~1-μm-thick single-crystal YIG films grown on a (111)-oriented GGG substrate through liquid phase epitaxy from an oxide solution in melt [6] and (ii) ~1-μm-thick YIG films deposited on similar substrates using the LE and EBE evaporation methods. To obtain magnetization curves of the specimens studied, the vibrating-sample magnetometer technique was used. To determine the FMR parameters, a setup operating in the reflection mode was employed, which made it possible to carry out measurements in the range 1–10 GHz with magnetic-field modulation. A copper microstrip 0.1 mm in thickness and 6 mm in length was used as a measuring cell (sensitive element). A specimen 6 mm in diameter was stuck to the end of the brass axis of a rotating device and was tightly pressed to the microstrip with a spring. The rotation axis of the specimen was normal to that of the electromagnet rotating device, which allowed the external magnetic field to be oriented along the direction required in the experiment.

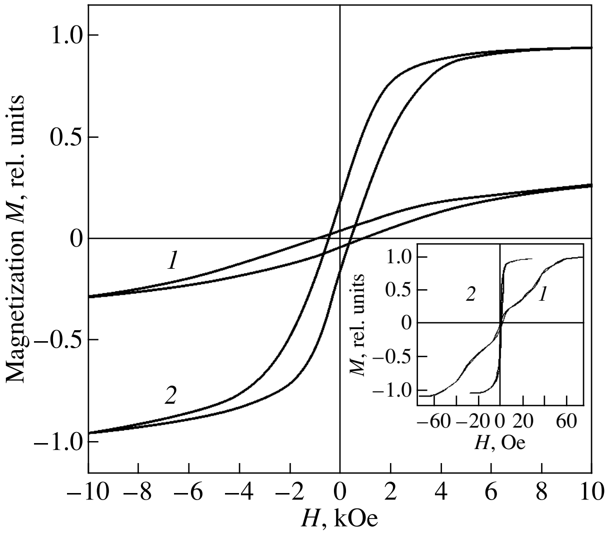


Fig. 1. Hysteresis loops in an in-plane magnetic field for films deposited using (*1*) liquid phase epitaxy and (*2*) laser evaporation. The inset shows the full form of these loops.

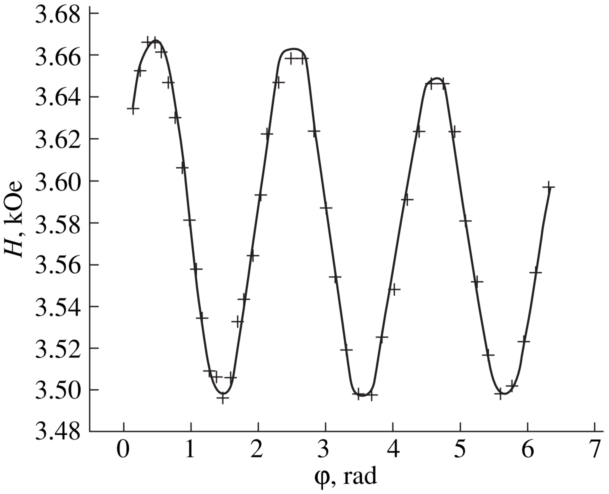


Fig. 2. Dependence of the FMR field on the azimuth angle for a laser-evaporated YIG film.

The minimum FMR linewidth at a frequency of 9.3 GHz was approximately 5 Oe for evaporated films. When the external field was oriented arbitrarily with respect to the film surface, two or three closely spaced FMR lines were excited. When the external magnetic field was oriented at 31° to the film normal, a single spectral line was observed and its intensity was maximum. The same orientation of the field (31°) was used to measure the cubic magnetic anisotropy by recording the azimuth dependence of the resonance field (the specimen was rotated around the film normal). In this case, a single FMR line was also observed in the spectrum.

The FMR linewidth for LPE-deposited YIG films was approximately 0.5 Oe. The gyromagnetic ratio, saturation magnetization, and cubic anisotropy were the same as those for the evaporated film.

2.2. Experimental Results

Figure 1 shows in-plane magnetization curves for LPE- and LE-deposited YIG films recorded with a vibrating-sample magnetometer. As seen from Fig. 1, the saturation field H_s for the LE film (film 2) is one order of magnitude lower than that for the LPE film (film 1). Furthermore, curve *2* does not have features that are seen in curve *1*, such as a chain of specific hysteresis loops. It is interesting to compare this result with the data from [7], where magnetization reversal of LPE-deposited YIG films was studied using a ferromagnetometer and the domain structure was simultaneously examined using a polarizing light microscope. The domain structure was observed due to the fairly high magnetooptical contrast of a 7-μm-thick film. The external magnetic field was applied in the film plane, and the saturation was reached in a field H_s ~ 50 Oe, which is close to our result for an LPE-deposited YIG film. The shape of the hysteresis loops is also the same for both specimens, which indicates that the magnetization processes are similar. It was shown that the features of the magnetization curve of the film mentioned above are due to domain walls moving stepwise with increasing external field. This conclusion is supported by jumplike variations in the high-frequency magnetic susceptibility [7]. Based on these data and taking into account the shape of the curves in Fig. 1, we can assume that film 2 is not split into domains and that its magnetization reversal proceeds only through the magnetic moment vector rotating towards the external field.

Figure 2 shows the azimuth dependence of the FMR field measured for a fixed value of the polar angle (31°) at a frequency of 9.3 GHz for a YIG specimen fabricated using the LE method. An analogous dependence of the resonance field is also observed for LPE-deposited YIG films (earlier, this dependence was measured for $Bi_3Fe_5O_{12}$ bismuth iron garnet films produced using the LE method [3]). As can be seen in Fig. 2, the axes of cubic magnetic anisotropy are fairly distinct in the studied specimen. This confirms the conclusion that, in both cases (evaporated or LPE-deposited YIG films), we deal with magnetic single crystals.

Figure 3 shows the frequency versus field (*F*–*H*) curve of FMR for an LPE-deposited YIG film and for a film fabricated using magnetron sputtering. The field is normal to the film plane. It follows from Fig. 3 that, for the former film, FMR is observed both in the saturation region and in the region where domain structure exists (domain FMR). This dependence is fairly well described theoretically by taking into account the cubic magnetic anisotropy in the single-crystal YIG film having a domain structure [5]. For the latter film, FMR is not observed in the nonsaturated state. As the field increases, a signal appears only close to the saturation field of the specimen. For both films in the saturated state, the *F*–*H* curves of FMR almost coincide in shape. The *F*–*H* curve for this film in the vicinity of H_s is similar to the theoretical curve [8] for a film that has neither domain structure nor cubic anisotropy.

3. DISCUSSION OF THE RESULTS

The experimental results obtained for the evaporated YIG films are obviously contradictory in terms of the conventional concepts concerning the structure of single-crystal YIG films. The results of x-ray structural analysis and the presence of cubic crystallographic anisotropy discovered using the FMR technique in the saturated state (Fig. 2) indicate that the evaporated films are single crystals. In this case, the films should have a domain structure in the demagnetized (nonsaturated) state and domain FMR should be excited in them, as is the case in the LPE-deposited YIG film. However, as mentioned above, no domain FMR is observed in the evaporated films, and the shape of the magnetization curve in Fig. 1 indicates that, in all probability, a domain structure (domain walls) does not exist.

What structure of the evaporated film could be consistent with the results obtained? In our opinion, we can assume that the single-crystal film is divided into fragments by a network of submicroscopic “cracks” with linear dimensions smaller than the equilibrium domain width. In this case, the film does not split into domains. In other words, the film is not continuous but rather consists of close-packed fragments each of which is in the saturated state and is characterized by a magnetic-moment vector $\mathbf{M}_F$. All fragments of this film retain the same crystallographic orientation of the lattice. Therefore, the film is anisotropic not only microscopically (like, for example, a finely dispersed polycrystalline system) but also macroscopically (like a single crystal). This conclusion is confirmed by the cubic magnetic anisotropy that was experimentally detected by measuring the azimuthal dependence of the FMR field for the film in the saturated state (Fig. 2).

This structure could form as a result of island film growth during deposition. When growing on a common single-crystal substrate, all islands were subjected to the influence of the crystallographic orientation of the substrate, so they reproduced the same crystallographic orientation of the lattice. Consequently, there are no abrupt structural breaks at the island boundaries, contrary to the case of grain boundaries in a polycrystalline material. For this reason and according to x-ray structural analysis data, the fragmented film does not differ from a continuous single-crystal film.

However, an ideal matching of islands was not, apparently, achieved in the growing film, so the matching areas of the fragments are in a partially atomically disordered state. This has an effect on structurally sensitive parameters, such as the FMR linewidth, which in the evaporated film is ten times larger than that in the LPE-deposited film. Due to the atomically disordered boundaries, the exchange coupling between the atoms of the neighbor fragments is broken and the coordination in the mutual arrangement (orientation) of the magnetic moments of the single-crystal fragments is destroyed. Consequently, the magnetic moments $\mathbf{M}_F$ inside the fragments can be oriented in any of the six easy magnetization directions in the (111) plane. An additional contribution to misorientation of the vectors $\mathbf{M}_F$ can be made by long-range dipole interaction of the entire ensemble of magnetic fragments of the film, which minimize the magnetostatic energy.

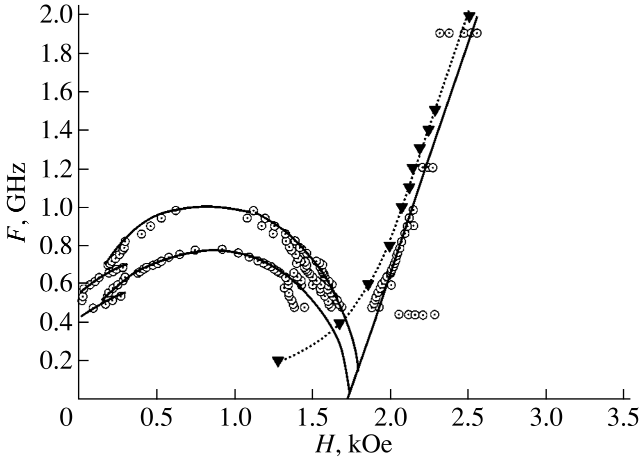


Fig. 3. Frequency versus field curve of FMR for YIG films deposited using liquid phase epitaxy (open dots) and magnetron sputtering (solid triangles).

We are unaware of any FMR theory for such a fragmented nonsaturated crystalline magnetic film. Apparently, the ensemble of variously oriented magnetic moments of film fragments not bonded by exchange interaction cannot produce a uniform in-phase spin wave excitation. This is the reason why domain FMR is not observed in the evaporated YIG film. However, FMR is observed in the saturated film, when the entire ensemble of vectors $\mathbf{M}_F$ is uniformly oriented along the external magnetic field (Fig. 3).

Based on this model for the structure of the evaporated film, let us again discuss Fig. 1. According to the conventional concepts, magnetization reversal of a ferromagnetic crystal through motion of domain walls requires less energy than in the case of rotation of the magnetic moments. Therefore, saturation of the LPE-grown film should be achieved in lower fields than saturation of an evaporated film, which, according to our model, does not have magnetic domains. However, the experimental data in Fig. 1 demonstrate the opposite. We believe that the high value of H_s in the LPE-grown

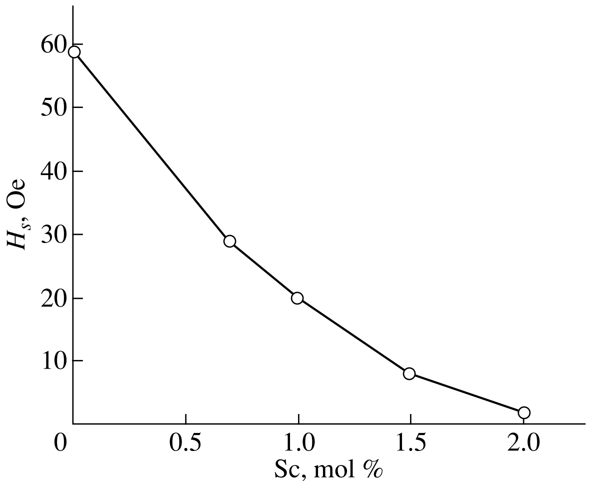


Fig. 4. Saturation field of an in-plane magnetized YIG specimen as a function of the proportion of scandium substituting for iron.

film is mostly due to tensile stresses that exist in this film and cause uniaxial magnetic anisotropy along the normal to the (111)-oriented film (so called stress-induced anisotropy). This anisotropy should be surmounted during magnetization of the film in its plane, which causes an increase in H_s. The stresses arise due to the difference in lattice parameters between the crystal lattices of the YIG film (a_0 = 1.2376 nm) and GGG substrate (a_0 = 1.2382 nm). Therefore, removal of this difference will cause a decrease in the saturation field of the film. With this aim, YIG films were LPE-grown in which Sc^{3+} ions having a larger radius were partly substituted for Fe^{3+} ions. The measured values of H_s for these films are shown in Fig. 4. It can be seen that the introduction of scandium caused the stress to decrease and, as a result, the saturation field of YIG film decreased to $H_s = 2$ Oe. It can thus be concluded that the low value H_s of the evaporated YIG film in Fig. 1 indicates that this film is almost free of stresses, which relaxed due to the fragmented structure of the evaporated film with atomically disordered fragment boundaries. It should be noted that the stresses in LPE-deposited YIG films also decrease abruptly after the formation of a net of cracks caused by mismatching stresses.

The validity of the suggested model is also supported by the results of electron microscopic studies of a laser-evaporated $Bi_3Fe_5O_{12}$ film reported in [9]. The microphotograph of a cross section of the approximately 1.3-μm-thick film demonstrates that there is a columnar structure consisting of columns with submicron cross-sectional dimensions oriented along the normal to the film surface (that is, along the direction of film growth). We believe that the results of [9] additionally validate the concept of the fragmented structure of the deposited YIG films.

Thus, the fragmented-structure model proposed by us gives a consistent explanation of the experimental data on YIG deposited films presented above, although it does not clarify the structure of the boundary that breaks the exchange coupling between the fragments of the magnetic crystal.

REFERENCES

1. T. Okuda, N. Koshizuka, H. Kobayashi, and N. Kobayashi, J. Appl. Phys. **67**, 4944 (1990).
2. Travendrarajah, M. Pardavi–Horvath, P. E. Wigen, and M. Gomi, IEEE Trans. Magn. **25**, 4015 (1989).
3. V. F. Shkar', V. P. Denisenkov, A. V. Grishin, A. A. Yalali, S. I. Khartsev, E. I. Nikolaev, and V. N. Sayapin, Fiz. Tverd. Tela (St. Petersburg) **45** (12), 2227 (2003) [Phys. Solid State **45** (12), 2334 (2003)].
4. N. Adachi, V. P. Denysenkov, S. I. Khartsev, and A. M. Grishin, J. Appl. Phys. **88** (5), 2734 (2000).
5. V. F. Shkar', E. I. Nikolaev, V. N. Sayapin, and V. D. Poĭmanov, Fiz. Tverd. Tela (St. Petersburg) **46** (6), 1043 (2004) [Phys. Solid State **46** (6), 1073 (2004)].
6. E. I. Nikolaev and I. A. Krasin, Kristallografiya **33** (2), 478 (1988) [Sov. Phys. Crystallogr. **33** (2), 282 (1988)].
7. B. A. Belyaev, S. N. Kulinich, and V. V. Tyurnev, Preprint No. 556F (Kirensky Inst. Phys., Krasnoyarsk, 1989).
8. *Ferromagnetic Resonance*, Ed. by S. V. Vonsovskiĭ (Fizmatlit, Moscow, 1961) [in Russian].
9. N. Adachi, T. Okuda, V. P. Denysenkov, S. I. Khartsev, and A. M. Grishin, in *Proceedings of 8th International Conference on Ferrites (ICF-8), Kyoto, Japan, 2000* (Kioto, 2000).

Translated by E. Borisenko

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1111–1117. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1075–1081.
Original Russian Text Copyright © 2005 by Mironov.

MAGNETISM AND FERROELECTRICITY

Magnetic Susceptibility of the Two-Dimensional Two-Sublattice Hubbard Model Calculated in the Static-Fluctuation Approximation

G. I. Mironov
Mari State Pedagogical Institute, Ĭoshkar Ola, 424002 Russia
e-mail: mir@mgpi.mari.ru

Received July 29, 2004

Abstract—The transverse dynamic susceptibility of the 2D two-sublattice Hubbard model is calculated in the static-fluctuation approximation. The static magnetic susceptibility is studied as a function of various parameters of the system. The results for the special case of the one-dimensional Hubbard model are compared to the exact solution. *© 2005 Pleiades Publishing, Inc.*

In [1, 2], a method for solving the Hubbard model [3] in the static fluctuation approximation was developed. In [4], the ground-state energy of the two-dimensional (2D) two-sublattice Hubbard model [5] was calculated and investigated. A comparison of the results from [4] with the exact solution of the one-dimensional (1D) Hubbard model [6] proved that the static-fluctuation approximation gives a rather faithful description of a Hubbard system both in the weak and strong correlation regimes. It was shown in [4] that the ground state energy obtained in the approximation of static fluctuations [1, 2] coincides with that for the exact solution in the limits of $U = 0$ and $U = \infty$ and that there is good agreement for intermediate values of U. Thus, the static-fluctuation approximation is applicable not only in the case of weak correlations but also for intermediate and strong correlations, the latter being of special importance for application to layered cuprates [5].

The objective of the present work was to calculate and study the transverse dynamic susceptibility of the 2D Hubbard model in the static-fluctuation approximation.

The magnetism in the Hubbard model has been studied in a number of papers (see, for example, [7–15]). The theory of ferromagnetism of itinerant electrons in the molecular-field approximation was developed by Stoner [13]. Unfortunately, it has been found that, at nonzero temperatures, the Stoner model does not provide a consistent description of all physical properties of magnetic transitional metals. The Stoner model was enhanced by Hubbard [3], who used splitting of retarded Green's functions and the random-phase approximation [7]. However, these methods also failed to remove the main drawbacks of the Stoner model that appear when dealing with magnetic and thermodynamic properties at nonzero temperatures.

Further progress in methods for calculating the susceptibility was achieved when diagram techniques for the Hubbard operators were used [14, 15]. This approach is a generalization of the diagram technique [16–19] used for spin operators of localized magnetic moments.

While it is easy to substantiate the diagram analysis for localized electrons (for example, the transverse ac susceptibility of a system consisting of coupled itinerant and localized electrons can be calculated, including the Kondo anomalies [20], up to the third order in the *s–d* exchange constant) [21, 22], the diagram technique for the Hubbard model has not been developed as a clear geometric algorithm, and the structure of the diagram series is unknown [8]. For example, "no method is known to obtain *a priori* graphic series not connected to any particular precedence system" [8]. Recently, the quantum Monte Carlo method has been widely used to study the susceptibility of the Hubbard model [11].

However, often analytical solutions for the susceptibility that are required, so the problem of calculating the ac susceptibility is highly topical, especially since new experimental data appear that require explanation [23, 24].

In the 2D two-sublattice *B*–*B*'–*U* Hubbard model, in contrast to the standard Hubbard model, it is assumed, following [5], that the lattice consists of two sublattices of atoms of different kinds and, furthermore, that electron transitions to the next-to-nearest neighbors in the crystal lattice are taken into account. The Hamiltonian of this model has the form

$$H = H_0 + V, \tag{1}$$

$$H_0 = \sum_{\sigma, f\in A}\left(\varepsilon_1 + \frac{1}{2}\sigma\omega_{e1}\right)n_{f\sigma} + \sum_{\sigma, l\in C}\left(\varepsilon_2 + \frac{1}{2}\sigma\omega_{e2}\right)n_{l\sigma}$$
$$+ \sum_{\sigma, f, l} B_{fl}(a^+_{f\sigma}a_{l\sigma} + a^+_{l\sigma}a_{f\sigma}) + \sum_{\sigma, l', l} B_{l'l}a^+_{l'\sigma}a_{l\sigma}, \tag{2}$$

$$V = \frac{U_1}{2}\sum_{\sigma, f\in A} n_{f\sigma}n_{f\bar{\sigma}} + \frac{U_2}{2}\sum_{\sigma, l\in C} n_{l\sigma}n_{l\bar{\sigma}}, \tag{3}$$

where $a^+_{j\sigma}$, $a_{j\sigma}$ are the Fermi creation and annihilation operators for an electron with spin σ on site j ($j = f, l$); $n_{f\sigma} = a^+_{f\sigma}a_{f\sigma}$; $\varepsilon_1(\varepsilon_2)$ is the self-energy of an electron on a site belonging to the sublattice A (C); $B_{fl} = B(f - l)$ and $B_{l'l} = B(l' - l)$ are the transfer integrals describing electron jumps from atom to atom that occur at the expense of the kinetic energy and crystalline field to the nearest sites and to the next-to-nearest sites situated on the square diagonal, respectively; $\bar{\sigma} = -\sigma$; ω_{e1} and ω_{e2} are the Zeeman frequencies of electrons belonging to different sublattices; and $U_1(U_2)$ is the Coulomb repulsion energy of two electrons situated on one site of the sublattice A (C). In order to draw the system described by the Hamiltonian in Eq. (1) closer to the case of holes moving in CuO_2 planes of HTSCs, it is assumed that jumps along diagonals can be performed by electrons belonging to only one sublattice (which models oxygen atoms in CuO_2 planes) and only within the same sublattice. We should stress that, for the sake of simplicity, we consider a hypothetical square lattice.

Magnetic excitations of electrons are described by the operators

$$S^-_a(p, q) = a^+_{p\downarrow}a_{p+q\uparrow}, \quad S^-_b(p, q) = b^+_{p\downarrow}b_{p+q\uparrow}.$$

Here, an electron with wave vector $p + q$ and spin up ($\uparrow$) is excited into the state with wave vector p and spin down ($\downarrow$) and the subscript a (b) is related to the sublattice A (C). By converting the spin density operators to the Heisenberg representation

$$S^-_a(p, q, \tau) = \exp(H\tau)a^+_{p\downarrow}a_{p+q\uparrow}\exp(-H\tau),$$

we get the following equation of motion:

$$\frac{d}{d\tau}a^+_{p\downarrow}a_{p+q\uparrow} = -\omega_{e1}a^+_{p\downarrow}a_{p+q\uparrow} + B(p)b^+_{p\downarrow}a_{p+q\uparrow}$$
$$-B(p+q)a^+_{p\downarrow}b_{p+q\uparrow} + \frac{2U_1}{N}\sum_{k_2, k}a^+_{k+p-k_2\downarrow}a_{p+q\uparrow}a^+_{k_2\uparrow}a_{k\uparrow} \tag{4}$$
$$-\frac{2U_1}{N}\sum_{k_1, k}a^+_{k+k_1-p-q\downarrow}a_{k_1\downarrow}a^+_{p\downarrow}a_{k\uparrow},$$

where $B(p) = -2B(\cos(p_x a) + \cos(p_y a))$ and $B(p + q)$ can be expressed in a similar way. The equations of motion for operators $b^+_{p\downarrow}a_{p+q\uparrow}$, $a^+_{p\downarrow}b_{p+q\uparrow}$, and $b^+_{p\downarrow}b_{p+q\uparrow}$ are obtained similarly.

Let us consider the last terms in differential equation (4). The Stoner theory [13], which is based on the Hartree–Fock approximation, is limited to the contribution [7]

$$\frac{2U_1}{N}\sum_{k_2, k}a^+_{k+p-k_2\downarrow}a_{p+q\uparrow}a^+_{k_2\uparrow}a_{k\uparrow}$$
$$-\frac{2U_1}{N}\sum_{k_1, k}a^+_{k+k_1-p-q\downarrow}a_{k_1\downarrow}a^+_{p\downarrow}a_{k\uparrow} \tag{5}$$
$$= \frac{2U_1}{N}\sum_{k}(\langle n_{k\uparrow}\rangle - \langle n_{k\downarrow}\rangle)a^+_{p\downarrow}a_{p+q\uparrow}$$
$$= 2U_1 S a^+_{p\downarrow}a_{p+q\uparrow} = 2U_1 S S^-_a(p, q).$$

The spin S in Eq. (5) is defined as

$$2S = 2\langle S_z\rangle = \frac{2}{N}\sum_k(\langle n_{k\uparrow}\rangle - \langle n_{k\downarrow}\rangle).$$

Following [1, 2], we isolate the spin projection fluctuation operator in the last two terms in Eq. (4). We get

$$\frac{d}{d\tau}a^+_{p\downarrow}a_{p+q\uparrow}$$
$$= (-2SU_1 - \omega_{e1})a^+_{p\downarrow}a_{p+1\uparrow} + B(p)b^+_{p\downarrow}a_{p+q\uparrow} \tag{6}$$
$$-B(p+q)a^+_{p\downarrow}b_{p+q\uparrow} - 2U_1\Delta S a^+_{p\downarrow}a_{p+q\uparrow},$$

where ΔS is the spin fluctuation operator

$$\Delta S = \frac{1}{N}\sum_k(n_{k\uparrow} - n_{k\downarrow}) - \frac{1}{N}\sum_k(\langle n_{k\uparrow}\rangle - \langle n_{k\downarrow}\rangle).$$

The Heisenberg operators can be written as [1, 2]

$$S^-(p, q, \tau) = \exp(H_0\tau)\tilde{S}^-(p, q, \tau)\exp(-H_0\tau), \tag{7}$$

where

$$\tilde{S}^-(p, q, \tau) = \exp(-H_0\tau)\exp(H\tau)S^-(p, q, 0)$$
$$\times\exp(-H\tau)\exp(H_0\tau). \tag{8}$$

We will call the representation of the Heisenberg operators in the form of Eq. (7) the interaction-like representation. Notice that the Hamiltonian H_0 in Eq. (7) coincides with the Hamiltonian H_0 in Eq. (1) to within the renormalization

$$\varepsilon \longrightarrow \varepsilon + U\left(\frac{1}{2} - \sigma S\right).$$

We get the following set of two differential equations for the unknown operator defined in Eq. (8) and the operator $\Delta S\tilde{S}^-(p, q, \tau)$:

$$\frac{d}{d\tau}\tilde{S}_a^-(p, q, \tau) = -2U_1\Delta\tilde{S}\tilde{S}_a^-(p, q, \tau),$$

$$\frac{d}{d\tau}\Delta\tilde{S}\tilde{S}_a^-(p, q, \tau) = 4SU_1\Delta\tilde{S}\tilde{S}_a^-(p, q, \tau) - 2U_1\Phi^2\tilde{S}_a^-(p, q, \tau), \tag{9}$$

where we denoted $\Phi^2 = 1/4 - S^2$ and made use of the fact that the spin projection fluctuation operator $\Delta\tilde{S}(\tau) = \exp(-H_0\tau)\Delta S(\tau)\exp(H_0\tau)$ satisfies the equation

$$\frac{d}{d\tau}\Delta\tilde{S}(\tau) = 0.$$

Therefore, $\Delta\tilde{S}(\tau) = \Delta S(0) = \Delta S$ and $(\Delta S)^2 = \Phi^2 - 2S\Delta S$. The last equality can be easily proved by expressing the particle number operator through the spin fluctuation operator and making use of the equality $n_{f\sigma}^2 = n_{f\sigma}$.

The solution to the set of differential equations (8) for the operator $\tilde{S}_a^-(p, q, \tau)$ is given by

$$\begin{aligned}\tilde{S}_a^-(p, q, \tau) &= \exp(2U_1S\tau)\\ &\times\{S_a^-(p, q, 0)[\cosh(U_1\tau) - 2S\sinh(U_1\tau)]\\ &- 2\Delta SS_a^-(p, q, 0)\sinh(U_1\tau)\}.\end{aligned} \tag{10}$$

Consequently, the general solution to differential equation (4) is given by

$$\begin{aligned}S_a^-(p, q, \tau) &= \exp(H_0\tau)\tilde{S}_a^-(p, q, \tau)\exp(-H_0\tau)\\ &= \exp(2U_1S\tau)\\ &\times\{\bar{S}_a^-(p, q, \tau)[\cosh(U_1\tau) - 2S\sinh(U_1\tau)]\\ &- 2\Delta S(0)\bar{S}_a^-(p, q, \tau)\sinh(U_1\tau)\},\end{aligned} \tag{11}$$

where $\bar{S}_a^-(p, q, \tau) = \exp(H_0\tau)S_a^-(p, q, 0)\exp(-H_0\tau)$.

Solving the set of four differential equations, we get solutions for the spin density operators:

$$\begin{aligned}&\bar{S}_a^-(k, q, \tau)\\
&= \frac{1}{4}\Bigg\{\Bigg[a^{+}_{p\downarrow}a_{p+q\uparrow}(0)\left(1 - \frac{\varepsilon_{2\downarrow} - \varepsilon_{1\downarrow}}{2t_p}\right)\left(1 + \frac{\varepsilon_{2\uparrow} - \varepsilon_{1\uparrow}}{2t_{p+q}}\right)\\
&- a^{+}_{p\downarrow}b_{p+q\uparrow}(0)\left(1 - \frac{\varepsilon_{2\downarrow} - \varepsilon_{1\downarrow}}{2t_p}\right)\frac{B_{p+q}}{t_{p+q}}\\
&+ b^{+}_{p\downarrow}a_{p+q\uparrow}(0)\frac{B_p}{t_p}\left(1 + \frac{\varepsilon_{2\uparrow} - \varepsilon_{1\uparrow}}{2t_{p+q}}\right)\\
&- b^{+}_{p\downarrow}b_{p+q\uparrow}(0)\frac{B_p}{t_p}\frac{B_{p+q}}{t_{p+q}}\Bigg]e^{(t_p + t_{p+q})\tau}\\
&+ \Bigg[a^{+}_{p\downarrow}a_{p+q\uparrow}(0)\left(1 - \frac{\varepsilon_{2\downarrow} - \varepsilon_{1\downarrow}}{2t_p}\right)\left(1 - \frac{\varepsilon_{2\uparrow} - \varepsilon_{1\uparrow}}{2t_{p+q}}\right)\\
&+ a^{+}_{p\downarrow}b_{p+q\uparrow}(0)\frac{B_{p+q}}{t_{p+q}}\left(1 - \frac{\varepsilon_{2\downarrow} - \varepsilon_{1\downarrow}}{2t_p}\right)\\
&+ b^{+}_{p\downarrow}a_{p+q\uparrow}(0)\frac{B_p}{t_p}\left(1 - \frac{\varepsilon_{2\uparrow} - \varepsilon_{1\uparrow}}{2t_{p+q}}\right)\\
&+ b^{+}_{p\downarrow}b_{p+q\uparrow}(0)\frac{B_p}{t_p}\frac{B_{p+q}}{t_{p+q}}\Bigg]e^{(t_p - t_{p+q})\tau}\\
&+ \Bigg[a^{+}_{p\downarrow}a_{p+q\uparrow}(0)\left(1 + \frac{\varepsilon_{2\downarrow} - \varepsilon_{1\downarrow}}{2t_p}\right)\left(1 + \frac{\varepsilon_{2\uparrow} - \varepsilon_{1\uparrow}}{2t_{p+q}}\right)\\
&- a^{+}_{p\downarrow}b_{p+q\uparrow}(0)\left(1 + \frac{\varepsilon_{2\downarrow} - \varepsilon_{1\downarrow}}{2t_p}\right)\frac{B_{p+q}}{t_{p+q}}\\
&- b^{+}_{p\downarrow}a_{p+q\uparrow}(0)\frac{B_p}{t_p}\left(1 + \frac{\varepsilon_{2\uparrow} - \varepsilon_{1\uparrow}}{2t_{p+q}}\right)\\
&- b^{+}_{p\downarrow}b_{p+q\uparrow}(0)\frac{B_p}{t_p}\frac{B_{p+q}}{t_{p+q}}\Bigg]e^{(-t_p + t_{p+q})\tau}\\
&+ \Bigg[a^{+}_{p\downarrow}a_{p+q\uparrow}(0)\left(1 + \frac{\varepsilon_{2\downarrow} - \varepsilon_{1\downarrow}}{2t_p}\right)\left(1 - \frac{\varepsilon_{2\uparrow} - \varepsilon_{1\uparrow}}{2t_{p+q}}\right)\\
&+ a^{+}_{p\downarrow}b_{p+q\uparrow}(0)\left(1 + \frac{\varepsilon_{2\downarrow} - \varepsilon_{1\downarrow}}{2t_p}\right)\frac{B_{p+q}}{t_{p+q}}\\
&- b^{+}_{p\downarrow}a_{p+q\uparrow}(0)\frac{B_p}{t_p}\left(1 - \frac{\varepsilon_{2\uparrow} - \varepsilon_{1\uparrow}}{2t_{p+q}}\right)\\
&- b^{+}_{p\downarrow}b_{p+q\uparrow}(0)\frac{B_p}{t_p}\frac{B_{p+q}}{t_{p+q}}\Bigg]\\
&\times e^{(-t_p - t_{p+q})\tau}e^{\frac{1}{2}(-\omega_{e1} - \omega_{e2} + J_p - J_{p+q} + 2S(U_2 - U_1))\tau}\Bigg\},\end{aligned} \tag{12}$$

where

$$\varepsilon_{1\sigma} = \varepsilon_1 + \frac{1}{2}\omega_{e1} + \left(\frac{1}{2} + \sigma S\right)U_1,$$

$$\varepsilon_{2\sigma} = \varepsilon_2 + \frac{1}{2}\omega_{e2} + \left(\frac{1}{2} - \sigma S\right)U_2 + J_{p+q}\delta_{\uparrow,\sigma} + J_P\delta_{\downarrow,\sigma},$$

$$t_p = \sqrt{\left(\frac{\varepsilon_{2\downarrow} - \varepsilon_{1\downarrow}}{2}\right) + B_p^2},$$

$$t_{p+q} = \sqrt{\left(\frac{\varepsilon_{2\uparrow}-\varepsilon_{1\uparrow}}{2}\right) + B^2_{p+q}},$$

$$J_p = -4B'\cos(p_x a)\cos(p_y a),$$

$$J_{p+q} = -4B'\cos((p_x+q_x)a)\cos((p_y+q_y)a).$$

A similar procedure can also be carried out for the spin density operator $\bar{S}_b^-(k, q)$. The expression obtained for $\bar{S}_a^-(k, q)$ is substituted into Eq. (11), and we obtain the general solution for the operator $S_a^-(p, q, \tau)$. An analogous procedure is also carried out for the operator $S_b^-(p, q, \tau)$. Given these solutions, we can calculate the transverse ac susceptibility of electrons belonging to the sublattices *A* and *C*.

The transverse ac susceptibility is defined as

$$\chi^{+-}(q, \omega) = i\int_0^\infty dt e^{-i\omega t}\sum_{p,\,p'}\langle[S_a^-(p, q, t) + S_b^-(p, q, t), \tag{13}$$
$$S_a^+(p', -q, 0) + S_b^+(p', -q, 0)]\rangle.$$

Using Eq. (11) and the solution for $S_b^-(p, q, \tau)$, we can obtain a rather cumbersome expression for the transverse ac susceptibility. For high-temperature superconductivity, the strong correlation limit is of most interest. As demonstrated in [2], in this case, we have $S = 1/2$. In the special case of strong correlations and $n = 1$, the total transverse susceptibility of the system is given by

$$\chi^{+-}(q, \omega) = \chi^{+-}_{aa}(q, \omega) + \chi^{+-}_{ab}(q, \omega) + \chi^{+-}_{ba}(q, \omega) + \chi^{+-}_{bb}(q, \omega)$$
$$= \sum_p\sum_{\beta,\gamma}\frac{1}{4}\left\{\frac{(\langle n^a_{p\downarrow}\rangle - \langle n^a_{p+q\uparrow}\rangle)\left[\left(1-\frac{\varepsilon_{2\downarrow}-\varepsilon_{1\downarrow}}{2\beta t_p}\right)\left(1+\frac{\varepsilon_{2\uparrow}-\varepsilon_{1\uparrow}}{2\gamma t_{p+q}}\right)-\beta\gamma\frac{B_p}{t_p}\frac{B_{p+q}}{t_{p+q}}\right]}{\omega-\beta t_p-\gamma t_{p+q}+\frac{1}{2}(\omega_{e1}+\omega_{e2}+U_1-U_2-J_p+J_{p+q})}\right.$$
$$+\frac{(\langle a^+_{p+q\uparrow}b_{p+q\uparrow}\rangle - \langle a^+_{p\downarrow}b_{p\downarrow}\rangle)\left[\left(1-\frac{\varepsilon_{2\downarrow}-\varepsilon_{1\downarrow}}{2\beta t_p}\right)\frac{B_{p+q}}{\gamma t_{p+q}}-\left(1-\frac{\varepsilon_{2\uparrow}-\varepsilon_{1\uparrow}}{2\gamma t_{p+q}}\right)\frac{B_p}{\beta t_p}\right]}{\omega-\beta t_p-\gamma t_{p+q}+\frac{1}{2}(\omega_{e1}+\omega_{e2}+U_1-U_2-J_p+J_{p+q})} \tag{14}$$
$$+\frac{(\langle b^+_{p+q\uparrow}a_{p+q\uparrow}\rangle - \langle b^+_{p\downarrow}a_{p\downarrow}\rangle)\left[\left(1+\frac{\varepsilon_{2\downarrow}-\varepsilon_{1\downarrow}}{2\beta t_p}\right)\frac{B_{p+q}}{\gamma t_{p+q}}-\left(1+\frac{\varepsilon_{2\uparrow}-\varepsilon_{1\uparrow}}{2\gamma t_{p+q}}\right)\frac{B_p}{\beta t_p}\right]}{\omega-\beta t_p-\gamma t_{p+q}+\frac{1}{2}(\omega_{e1}+\omega_{e2}+U_1-U_2-J_p+J_{p+q})}$$
$$\left.+\frac{(\langle n^b_{p\downarrow}\rangle - \langle n^b_{p+q\uparrow}\rangle)\left[\left(1+\frac{\varepsilon_{2\downarrow}-\varepsilon_{1\downarrow}}{2\beta t_p}\right)\left(1-\frac{\varepsilon_{2\uparrow}-\varepsilon_{1\uparrow}}{2\gamma t_{p+q}}\right)-\beta\gamma\frac{B_p}{t_p}\frac{B_{p+q}}{t_{p+q}}\right]}{\omega-\beta t_p-\gamma t_{p+q}+\frac{1}{2}(\omega_{e1}+\omega_{e2}+U_1-U_2-J_p+J_{p+q})}\right\}.$$

In Eq. (14), α and β take on two values, $\alpha = \pm 1$ and $\beta = \pm 1$; $\chi^{+-}_{aa}(q, \omega)$ and $\chi^{+-}_{bb}(q, \omega)$ are the dynamical responses of the electron subsystems *A* and *C*, respectively; and $\chi^{+-}_{ab}(q, \omega)$ and $\chi^{+-}_{ba}(q, \omega)$ describe magnetization transfer between the electron subsystems.

The correlation functions in the numerator of Eq. (14) for the transverse ac susceptibility are calculated in the static-fluctuation approximation in a similar way. In the end, we get ($n = 1$, $S = 1/2$; see, e.g., [1, 2, 4])

$$\langle n^a_{p\downarrow}\rangle = \frac{1}{2}\left[\left(1-\frac{\varepsilon_{2\downarrow}-\varepsilon_{1\downarrow}}{2t_p}\right)f^+\left(\frac{\varepsilon_{1\downarrow}+\varepsilon_{2\downarrow}}{2}+t_p\right)\right.$$
$$\left.+\left(1+\frac{\varepsilon_{2\downarrow}-\varepsilon_{1\downarrow}}{2t_p}\right)f^+\left(\frac{\varepsilon_{1\downarrow}+\varepsilon_{2\downarrow}}{2}-t_p\right)\right],$$
$$\langle n^a_{p+q\uparrow}\rangle = \frac{1}{2}\left[\left(1-\frac{\varepsilon_{2\uparrow}-\varepsilon_{1\uparrow}}{2t_{p+q}}\right)f^+\left(\frac{\varepsilon_{1\uparrow}+\varepsilon_{2\uparrow}}{2}+t_{p+q}\right)\right. \tag{15}$$
$$\left.+\left(1+\frac{\varepsilon_{2\uparrow}-\varepsilon_{1\uparrow}}{2t_{p+q}}\right)f^+\left(\frac{\varepsilon_{1\uparrow}+\varepsilon_{2\uparrow}}{2}-t_{p+q}\right)\right],$$

$$\langle n^b_{p\downarrow}\rangle = \frac{1}{2}\left[\left(1+\frac{\varepsilon_{2\downarrow}-\varepsilon_{1\downarrow}}{2t_p}\right)f^+\left(\frac{\varepsilon_{1\downarrow}+\varepsilon_{2\downarrow}}{2}+t_p\right)\right.$$

$$\left.+\left(1-\frac{\varepsilon_{2\downarrow}-\varepsilon_{1\downarrow}}{2t_p}\right)f^+\left(\frac{\varepsilon_{1\downarrow}+\varepsilon_{2\downarrow}}{2}-t_p\right)\right], \quad (16)$$

$$\langle n^b_{p+q\uparrow}\rangle = \frac{1}{2}\left[\left(1+\frac{\varepsilon_{2\uparrow}-\varepsilon_{1\uparrow}}{2t_{p+q}}\right)f^+\left(\frac{\varepsilon_{1\uparrow}+\varepsilon_{2\uparrow}}{2}+t_{p+q}\right)\right.$$

$$\left.+\left(1-\frac{\varepsilon_{2\uparrow}-\varepsilon_{1\uparrow}}{2t_{p+q}}\right)f^+\left(\frac{\varepsilon_{1\uparrow}+\varepsilon_{2\uparrow}}{2}-t_{p+q}\right)\right],$$

$$\langle \overset{+}{a}_{p\downarrow} b_{p\downarrow}\rangle = \langle \overset{+}{b}_{p\downarrow} a_{p\downarrow}\rangle$$

$$= \frac{B_p}{2t_p}\left[f^+\left(\frac{\varepsilon_{1\downarrow}+\varepsilon_{2\downarrow}}{2}+t_p\right)-f^+\left(\frac{\varepsilon_{1\downarrow}+\varepsilon_{2\downarrow}}{2}-t_p\right)\right], \quad (17)$$

$$\langle \overset{+}{a}_{p+q\uparrow} b_{p+q\uparrow}\rangle = \langle \overset{+}{b}_{p+q\uparrow} a_{p+q\uparrow}\rangle$$

$$= \frac{B_{p+q}}{2t_{p+q}}\left[f^+\left(\frac{\varepsilon_{1\uparrow}+\varepsilon_{2\uparrow}}{2}+t_{p+q}\right)-f^+\left(\frac{\varepsilon_{1\uparrow}+\varepsilon_{2\uparrow}}{2}-t_{p+q}\right)\right],$$

where $f^+(x) = 1/(1+\exp(\beta x))$ is the Fermi distribution function.

It follows from Eq. (15) that $S = 1/2$ in the case where the Coulomb potential is large as compared to the transfer energy. Substituting Eq. (15)–(17) into Eq. (14), we get the final expression for the transverse ac susceptibility of a system described by the Hubbard Hamiltonian.

It is instructive to study the poles of the ac susceptibility (the collective-excitation spectrum of the Hubbard model) given by Eq. (14). This spectrum is shown in Fig. 1. For the parameter values indicated, the spectrum is a band consisting of four subbands. Two central subbands cross each other and form a single subband, so there are, in fact, only three subbands. The spectrum shown in Fig. 1 apparently indicates an antiferromagnetic ordering in the 2D Hubbard model.

It is also interesting to consider the static magnetic susceptibility as a function of q. In [11], the static magnetic susceptibility $\chi(q, 0)$ was calculated numerically as a function of q in the case of an exactly half-filled band for an 8 × 8 lattice. The calculations showed that there is a sharp peak in $\chi(q, 0)$ at $q = (\pi, \pi)$. As the temperature decreases, the peak increases in magnitude. The dependence of the static magnetic susceptibility on the antiferromagnetic wave vector q in the first Brillouin zone is shown in Fig. 2 for various temperatures. According to numerical calculations, the susceptibility has a maximum at $q = (\pi, \pi)$. As the temperature decreases, the peak increases and its width decreases. In contrast to the results from [11], the dependence under discussion is nonmonotonic. Local extrema can be identified (cf. [11, Fig. 3.2a] at temperature $T = 0.33B$).

It was shown in [23] that the static magnetic susceptibility of systems similar to Li-doped La_2CuO_4

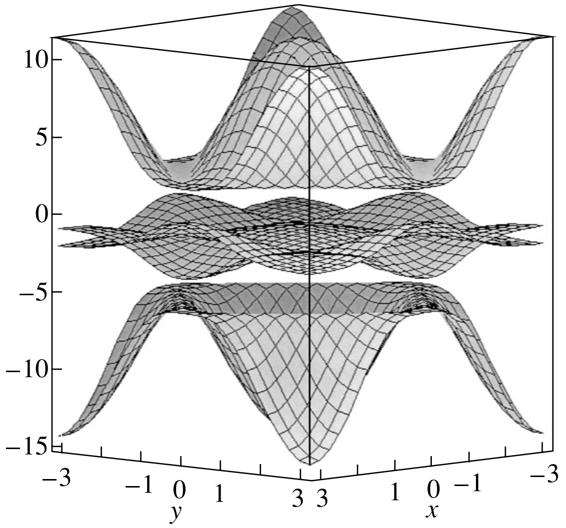


Fig. 1. Energy spectrum (the denominator of the transverse susceptibility) calculated for $S = 1/2$, $q = (\pi, \pi)$, $U_1 = 6$ eV, $U_2 = 2$ eV, $B = 1.5$ eV, $B' = -0.3B$, and $n = 1$.

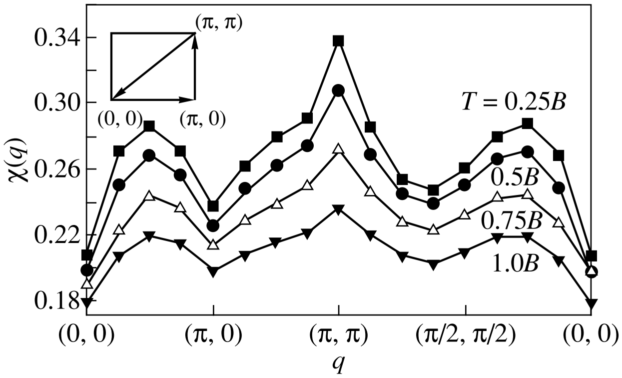


Fig. 2. Static magnetic susceptibility $\chi(q, 0)$ as a function of q for various values of T. $S = 1/2$, $U_1 = 6$ eV, $U_2 = 2$ eV, $B = 1.5$ eV, and $B' = -0.3B$.

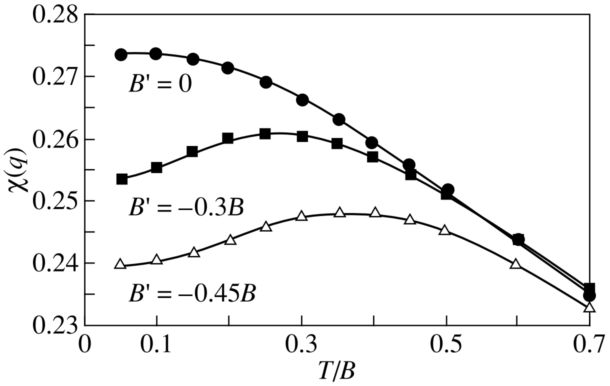


Fig. 3. Static magnetic susceptibility $\chi(q, 0)$ as a function of temperature calculated for $q = (\pi/2, \pi/2)$, $U_1 = 6$ eV, $U_2 = 2$ eV, $B = 1.5$ eV, $S = 1/2$, $\varepsilon_1 = -3$ eV, $\varepsilon_2 = -1$ eV, and various values of B'.

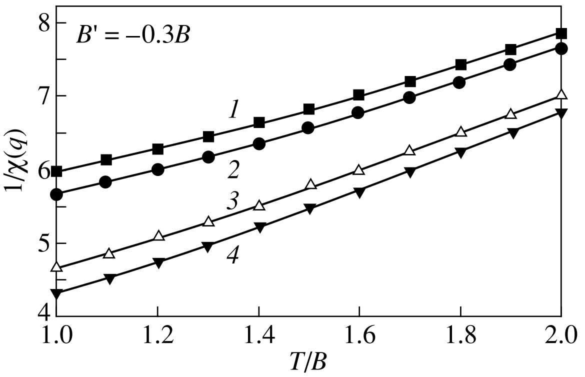


Fig. 4. Inverse susceptibility as a function of temperature for various values of the Coulomb potentials and wave vector: (*1*) $U_1 = 8$ eV, $U_2 = 4$ eV, $q = (0, 0)$; (*2*) $U_1 = 6$ eV, $U_2 = 3$ eV, $q = (0, 0)$; (*3*) $U_1 = 8$ eV, $U_2 = 4$ eV, $q = (\pi, \pi)$; and (*4*) $U_1 = 6$ eV, $U_2 = 3$ eV, $q = (\pi, \pi)$.

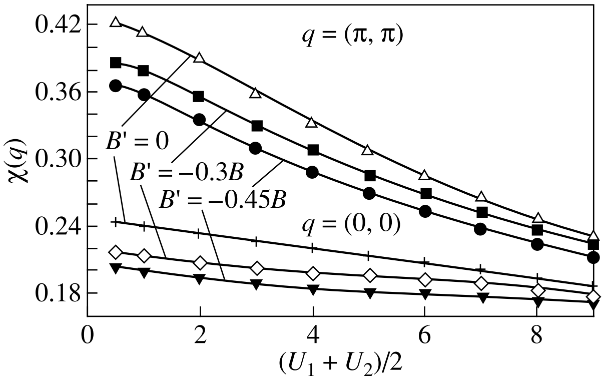


Fig. 5. Magnetic susceptibility as a function of the Coulomb potential for various values of the wave vector q and of the transfer integral B' at $T = 0.5B$.

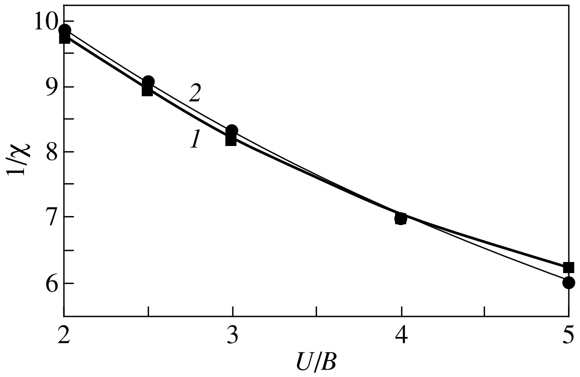


Fig. 6. Inverse susceptibility of the 1D Hubbard model as a function of the Coulomb potential for $S = 1/2$, $\omega_e/B = 0.01$, $T = 0.02B$, and $q = (\pi, \pi)$. (*1*) Exact solution taken from [25] and (*2*) the solution in the random-phase approximation according to Eq. (14).

strongly deviates from the Curie–Weiss law at low temperatures. In particular, the susceptibility increases with temperature at low temperatures and then passes through a maximum and decreases slowly (see, e.g., [23, Fig. 3]). This behavior can be explained in terms of the spin glass model [23]. Let us demonstrate that a similar behavior is possible in the Hubbard model at low temperatures under certain circumstances. The dependence of the static susceptibility on temperature (T/B) for various values of the diagonal transfer integral is shown in Fig. 3 for the wave vector $q = (\pi/2, \pi/2)$. It follows from Fig. 3 that the inclusion of the diagonal transfer integral B' causes a drastic change in the low-temperature behavior of $\chi(q)$. A maximum appears on the $\chi(q)$ dependence when B' is taken into account, and the larger the magnitude of B', the higher the temperature of the maximum. If the system is in the strongly correlated state, the inclusion of the transfer integral B' with values $B' < 0$ ($B > 0$) promotes delocalization of electrons and, consequently, decreases the magnetization in comparison with the case of $B' = 0$. Assuming that there is competition between the localization and delocalization in the Hubbard model, the fraction of the Pauli susceptibility becomes larger with inclusion of B' as compared to the case where $B' = 0$. Therefore, the peak observed experimentally in [23] can be explained not only in terms of the spin glass state but also in terms of the contribution from the diagonal transfer integral to the susceptibility at a certain value of q at low temperatures.

In [24], the temperature dependence of the static susceptibility is presented for various values of the Coulomb potential U. In order to compare our results to the data from [24], we also plotted the inverse susceptibility as a function of temperature for various values of the Coulomb potential (Fig. 4). The results are clearly in qualitative agreement. As for the quantitative comparison, the results agree in order of magnitude; as the temperature increases, the susceptibility decreases (the inverse susceptibility increases) and the curves approach straight lines. This kind of behavior corresponds to the Curie–Weiss law $\chi \sim 1/(T + \Theta)$ with $\Theta > 0$, so there is no ferromagnetic transition (at least in the given range of U). It follows from Fig. 4 that the inverse susceptibility depends on q (cf. Fig. 1).

Another interesting property is the behavior of the static magnetic susceptibility as a function of the Coulomb potential $(U_1 + U_2)/2$ for various values of the wave vector and the transfer integral B' at a constant temperature (Fig. 5). It follows from Fig. 5 that the static susceptibility depends on the Coulomb potential, the transfer integral B', and q.

As mentioned above, the static fluctuation approximation made it possible to calculate the ground-state energy of the Hubbard model. A comparison of the results for the special case of the 1D Hubbard model with the exact solution [6] shows that, in the limit of weak or strong correlations, the solution obtained in the static-fluctuation approximation coincides with the exact solution [6]. For intermediate correlations, the ground-state energy values obtained by these methods are quite close to each other [4]. In order to judge how well Eq. (14) describes the susceptibility behavior, a comparison with exact calculations is necessary. The 1D Hubbard model in a magnetic field was exactly solved in [25]. The study in [25] is a logical extension of the solution from both [6] and [26], i.e., the exact solution of the 1D Hubbard model at temperature $T \neq 0$. Figure 6 presents the dependences of the inverse static susceptibility obtained using the exact solution and calculated in the random phase approximation according to Eq. (14) (accurate to within a constant factor) [7]:

$$\chi_{\mathrm{rpa}}(q) = \chi(q, 0)/(1 - 0.5U\chi(q, 0)).$$

It follows that, qualitatively, the behaviors of the inverse static susceptibility as calculated exactly and in the static-fluctuation approximation are the same. Note that we assumed that $S = 1/2$ in the calculations. If we include the dependence of the spin on the Coulomb potential [2, Fig. 4]), the curves coincide even better.

To conclude, the method of solving the Hubbard model in the static-fluctuation approximation developed in [1, 2, 4] enables one to calculate not only the Green's functions and the ground-state energy but also the magnetic susceptibility and to study the behavior of the susceptibility as a function of various parameters of the system. Comparison of the results obtained for the special case of the 1D Hubbard model with the exact solution of the 1D Hubbard model in a magnetic field has shown that the static-fluctuation approximation gives a faithful description of the properties of the Hubbard model. Preliminary results of this work were presented in the Winter Physics Theory School "Kourovka" [27].

ACKNOWLEDGMENTS

The author is grateful to R.O. Zaĭtsev and V.V. Val'kov for their regard for my work and discussion of the results and to R.R. Nigmatullin for his regard and helpful remarks.

REFERENCES

1. G. I. Mironov, Fiz. Tverd. Tela (St. Petersburg) **39** (9), 1594 (1997) [Phys. Solid State **39** (9), 1420 (1997)].
2. G. I. Mironov, Fiz. Tverd. Tela (St. Petersburg) **41** (6), 951 (1999) [Phys. Solid State **41** (9), 864 (1999)].
3. J. Hubbard, Proc. R. Soc. A **276** (1365), 238 (1963).
4. G. I. Mironov, Fiz. Tverd. Tela (St. Petersburg) **44** (2), 209 (2002) [Phys. Solid State **44** (2), 216 (2002)].
5. V. J. Emery, Phys. Rev. Lett. **58** (26), 2794 (1987).
6. E. H. Lieb and F. Y. Wu, Phys. Rev. Lett. **20** (25), 1445 (1968).
7. T. Moriya, *Spin Fluctuations in Magnetics* (Moscow, 1988) [in Russian].
8. Yu. A. Izyumov and Yu. N. Skryabin, *Statistical Mechanics of Magnetically Ordered Media* (Moscow, 1987) [in Russian].
9. Yu. A. Izyumov, Usp. Fiz. Nauk **165** (4), 403 (1995) [Phys. Usp. **38** (4), 385 (1995)].
10. Yu. A. Izyumov, Usp. Fiz. Nauk **167** (5), 465 (1997) [Phys. Usp. **40** (5), 445 (1997)].
11. N. Bulut, Adv. Phys. **51**, 6 (2002); cond-mat/0207186 (2002).
12. E. W. Carlson, V. J. Emery, S. A. Kivelson, and D. Orgad, cond-mat/0206217 (2002).
13. E. C. Stoner, Proc. R. Soc. A **165**, 372 (1963).
14. R. O. Zaĭtsev, Fiz. Tverd. Tela (Leningrad) **19** (11), 3204 (1977) [Sov. Phys. Solid State. **19** (11), 1874 (1977)].
15. R. O. Zaĭtsev, Zh. Éksp. Teor. Fiz. **75**, 2362 (1978) [Sov. Phys. JETP **48**, 1193 (1978)].
16. A. A. Abrikosov, Physics **2** (1), 5 (1965).
17. S. V. Tyablikov, *Methods in the Quantum Theory of Magnetism* (Moscow, 1975) [in Russian].
18. A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics* (Fizmatgiz, Moscow, 1962; Prentice-Hall, Englewood Cliffs, N.J., 1963).
19. S. E. Barnes and J. Zitkova-Wilcox, Phys. Rev. B **7** (5), 2163 (1973).
20. J. Kondo, Prog. Theor. Phys. **32** (1), 38 (1964).
21. A. A. Kosov and G. I. Mironov, Fiz. Tverd. Tela (Leningrad) **24** (2), 583 (1982) [Sov. Phys. Solid State **24** (2), 327 (1982)].
22. G. I. Mironov and N. G. Fazleev, Fiz. Nizk. Temp. **14** (9), 950 (1988).
23. R. Arita *et al.*, cond-mat/0002441 (2001).
24. T. Sasagawa *et al.*, cond-mat/0208014 (2003).
25. C. Yang, A. N. Kocharian, and Y. L. Chiang, J. Phys.: Condens . Matter **12**, 7433 (2000).
26. M. Takahashi, Prog. Theor. Phys. **47**, 1 (1972).
27. G. I. Mironov and R. R. Nigmatullin, in *Proceedings of XXX International Winter School Kourovka-2004 for Physics Theorists, Yekaterinburg, Russia, 2004* (Yekaterinburg, 2004), p. 190.

Translated by G. Tsydynzhapov

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1118–1121. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1082–1084.
Original Russian Text Copyright © 2005 by Filippov.

MAGNETISM AND FERROELECTRICITY

Theory of the Magnetoelectric Effect in Ferromagnetic–Piezoelectric Heterostructures

D. A. Filippov

Novgorod State University, Sankt-Peterburgskaya ul. 41, Novgorod, 173003 Russia

Received June 4, 2004; in final form, August 27, 2004

Abstract—A theory of the magnetoelectric effect in ferromagnetic–piezoelectric bilayer structures is considered for platelike samples. The magnetoelectric voltage coefficient is expressed through the parameters characterizing the magnetic and piezoelectric phases. It is shown that the magnetoelectric voltage coefficient considerably increases in the region of electromechanical resonance. The thickness ratio between the ferromagnetic and piezoelectric phases at which the magnetoelectric voltage coefficient is maximum is determined. The calculated magnetoelectric voltage coefficients for Permendur–PZT (lead zirconate titanate) structures are presented and compared with the experimental data. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Magnetoelectric composites are mechanically interacting mixtures of magnetostrictive and piezoelectric components. The magnetoelectric effect in composites is caused by a mechanical coupling between the magnetic and piezoelectric subsystems. In a magnetic field, magnetostriction induces mechanical vibrations in a magnetic material. These vibrations are transferred to the piezoelectric phase and bring about polarization due to the piezoelectric effect. Since the magnetoelectric effect in magnetostrictive piezoelectric composites is associated with the mechanical coupling of the subsystems, the magnetoelectric coefficient considerably increases in the region of electromechanical resonance [1–6]. If the scale of variations in external effects is much larger than the characteristic sizes of the composites, these materials can be considered homogenous media with effective parameters [7]. The theory of the magnetoelectric effect for ferrite–piezoelectric composites, which is founded on the method of effective material parameters, and the experimental results obtained for disklike and platelike samples based on nickel spinel ferrite–lead zirconate titanate (PZT) are given in our previous works [3–5]. The use of a ferromagnet with magnetostriction constants larger than those of ferrite offers possibilities for increasing the magnetoelectric coefficient. Laletsin *et al.* [6] experimentally studied Permendur–PZT–Permendur structures. In this case, the composite cannot be considered a homogenous medium and the above theory is not applicable. In the present work, we obtained an expression for the magnetoelectric coefficient of heterogeneous composites.

2. THEORETICAL BACKGROUND

As a model, we consider a sample in the form of a plate prepared from a bilayer composite, i.e., the mechanical joint of a ferromagnet (*m*) and a piezoelectric (*p*) (Fig. 1). The outer surfaces of the piezoelectric and the ferromagnet are coated with metallic contacts. The thicknesses of these contacts are assumed to be negligible. Let the sample be polarized along the normal to the contact plane (the X_3 axis). Static (magnetizing) and alternating magnetic fields can be aligned both along the normal to the contact plane and with the contact plane along the X_1 axis. In accordance with these orientations of the magnetic fields, we will distinguish between the longitudinal and transverse magnetoelectric effects.

An alternating magnetic field induces vibrations in the ferromagnet due to magnetostriction. These vibrations propagate both across the thickness of the sample and in the contact plane. We will restrict our consideration to the specific case of volume vibrations propagating along the plate, because these vibrations occur at the lowest frequencies.

It is assumed that the thickness ${}^mh + {}^ph$ and the width W of the plate are considerably smaller than its length L. The faces of the plate are considered to be free; hence, the stresses at the surfaces of the plate are zero. Since the plate is thin and narrow, we can assume that

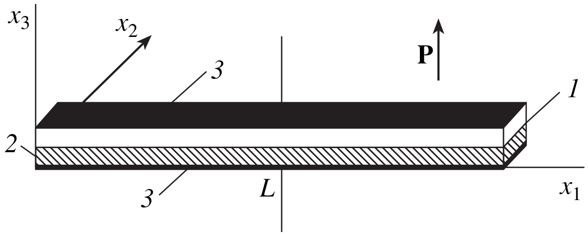


Fig. 1. Schematic drawing of the heterogeneous structure: (*1*) piezoelectric, (*2*) ferromagnetic, and (*3*) metallic contacts. The arrow indicates the direction of polarization **P**.

1063-7834/05/4706-1118$26.00 © 2005 Pleiades Publishing, Inc.

the stress components T_2 and T_3 are equal to zero not only at the surfaces but also over the entire volume of the plate and that only the component T_1 of the stress tensor is nonzero. In our case, the upper and lower faces of the plate are equipotential. Therefore, only the component E_3 of the electric field vector is nonzero. The equations for the strain tensor ${}^mS_i$ in the magnet and for the strain tensor ${}^pS_i$ and the electric field induction D_i in the piezoelectric for the transverse orientation of the fields have the form

$$ {}^mS_1 = {}^ms_{11}{}^mT_1 + {}^mq_{11}H_1, \tag{1} $$

$$ {}^pS_1 = {}^ps_{11}{}^pT_1 + {}^pd_{31}E_3, \tag{2} $$

$$ D_3 = {}^p\varepsilon_{33}E_3 + {}^pd_{31}{}^pT_1, \tag{3} $$

where ${}^ms_{11}$ and ${}^ps_{11}$ are the components of the magnetic compliance tensor and the piezoelectric compliance tensor, respectively; ${}^p\varepsilon_{33}$ is the component of the permittivity tensor for the piezoelectric; and ${}^pd_{31}$ and ${}^mq_{11}$ are the piezoelectric and piezomagnetic coefficients, respectively. For the longitudinal orientation of the electric and magnetic fields, relationship (1) contains ${}^mq_{11}H_1$ instead of ${}^mq_{31}H_3$.

The stress components involved in relationship (1) can be expressed through the strain components. By substituting these components into the equation of motion for the medium, we derive the differential equation with respect to the projection x of the displacement vector of the magnetic medium ${}^mu_x$. The solution to this equation can be written in the form

$$ {}^mu_x(x) = A_1\cos({}^mkx) + B_1\sin({}^mkx), \tag{4} $$

where ${}^mk = \omega({}^m\rho{}^ms_{11})^{1/2}$ and ${}^m\rho$ is the density of the ferromagnet.

Vibrations of the magnetic medium due to the mechanical coupling induce vibrations of the piezoelectric. The latter vibrations can be represented as a superposition of vibrations caused by the coupling with the magnet and natural vibrations initiated in the piezoelectric by the induced electric field. Since the contact between the layers is not ideal, the displacement of the piezoelectric medium ${}^pu_x(x)$ can be written as

$$ {}^pu_x(x) = \beta{}^mu_x(x) + (1-\beta){}^pu_x^{(0)}(x), \tag{5} $$

where $\beta \in (0.1)$ is the coupling coefficient between the phases and ${}^pu_x^{(0)}(x)$ stands for the displacements of the piezoelectric medium in the absence of coupling with the ferromagnet. By solving the equation of motion of the medium for the piezoelectric plate not coupled with the ferromagnet, we derive the following relationship for the displacements ${}^pu_x^{(0)}(x)$ of the piezoelectric medium:

$$ {}^pu_x^{(0)}(x) = \frac{{}^pd_{31}E_3}{{}^pk\cos({}^p\kappa)}\sin({}^pkx), \tag{6} $$

where ${}^pk = \omega({}^p\rho{}^ps_{11})^{1/2}$, ${}^p\rho$ is the density of the piezoelectric, and ${}^p\kappa = {}^pkL$ is the dimensionless parameter.

The integration constants A_1 and B_1 can be determined from the boundary conditions. Since the left and right faces of the sample are free, the resultant force affecting them is equal to zero. Consequently, we obtain the following relationships for the left and right boundaries:

$$ {}^mh{}^mT_1(-L/2) + {}^ph{}^pT_1(-L/2) = 0, \tag{7} $$

$$ {}^mh{}^mT_1(L/2) + {}^ph{}^pT_1(L/2) = 0. \tag{8} $$

Taking into account relationship (5) and boundary conditions (7) and (8) for displacements ${}^pu_x(x)$ of the piezoelectric medium, we derive the expression

$$ \begin{aligned} {}^px_x(x) = {} & \frac{\beta}{(\gamma+\beta)}\frac{\sin({}^mkx)}{{}^mk\cos({}^m\kappa)}(\gamma{}^mq_{11}H_1 - \beta{}^pd_{31}E_3) \\ & + (1-\beta)\frac{{}^pd_{31}E_3}{{}^pk\cos({}^p\kappa)}\sin({}^pkx), \end{aligned} \tag{9} $$

where ${}^m\kappa = {}^mkL/2$ and $\gamma = ({}^ps_{11}/{}^ms_{11})({}^mh/{}^ph)$ are dimensionless parameters.

The electric field strength appearing in the piezoelectric under strains can be found from relationship (3) using the open-circuit condition; that is,

$$ \int_0^W dy \int_{-L/2}^{L/2} D_3(x)dx = 0. \tag{10} $$

After substituting the obtained expressions into relationship (10) and calculating the integral for the electric field strength E_3 in the piezoelectric, we derive the equation

$$ E_3 = -\frac{\gamma\beta}{(\gamma+\beta)}\frac{{}^pd_{31}{}^mq_{11}}{{}^p\varepsilon_{33}{}^ps_{11}\Delta_a}\frac{\tan({}^m\kappa)}{{}^m\kappa}H_1. \tag{11} $$

Here, we introduced the notation

$$ \Delta_a = 1 - \left(1 - (1-\beta)\frac{\tan({}^p\kappa)}{{}^p\kappa} - \frac{\beta^2}{(\gamma+\beta)}\frac{\tan({}^m\kappa)}{{}^m\kappa}\right)K_{31}^2, \tag{12} $$

where $K_{31}^2 = {}^pd_{31}^2/({}^p\varepsilon_{33}{}^ps_{11})$ is the square of the electromechanical coupling coefficient of the piezoelectric for in-plane vibrations.

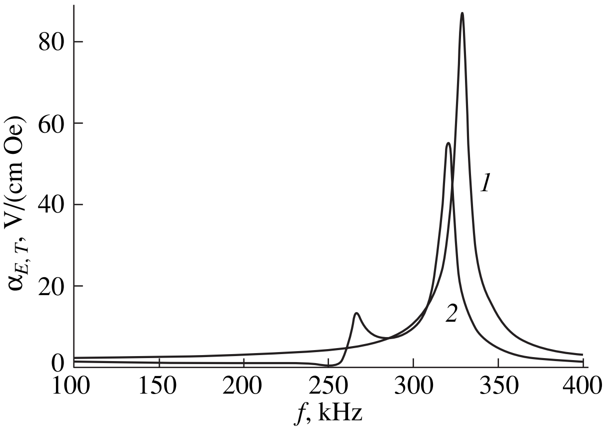


Fig. 2. Frequency dependences of the magnetoelectric voltage coefficient for the structure based on Permendur and lead zirconate titanate with coupling coefficients β = (*1*) 1.0 and (*2*) 0.4.

Let us determine the magnetoelectric voltage coefficient for a bilayer structure from the relationship

$$\alpha_E = E_{\text{av}}/H, \tag{13}$$

where $E_{\text{av}} = U({}^{m}h + {}^{p}h)$ is the average value of the electric field strength and U is the voltage difference across the electrodes. By assuming that the total voltage drop occurs across the piezoelectric, we obtain the following relationship for the magnetoelectric voltage coefficient at the transverse orientation of the electric and magnetic fields:

$$\alpha_{E,T} = -\frac{\gamma\beta}{(\gamma+\beta)}\frac{{}^{p}d_{31}\,{}^{m}q_{11}}{{}^{p}\varepsilon_{33}\,{}^{p}s_{11}\Delta_a}\frac{\tan({}^{m}\kappa)}{{}^{m}\kappa}\frac{{}^{p}h}{({}^{m}h+{}^{p}h)}. \tag{14}$$

For the longitudinal orientation of the electric and magnetic fields along the X_3 axis, the piezomagnetic coefficient ${}^{m}q_{31}$ replaces ${}^{m}q_{11}$ in the expression for the magnetoelectric coefficient. Since the quantity ${}^{m}q_{31}$ is usually smaller than the quantity ${}^{m}q_{11}$ due to the demagnetizing fields, the effect for the longitudinal orientation is usually one order of magnitude weaker than that for the transverse orientation.

It follows from relationship (14) for the magnetoelectric coefficient that, at frequencies when $\Delta_a = 0$, there occurs a resonant increase in the magnetoelectric coefficient. The losses in the structure can be taken into account through the damping coefficient by substituting the circular frequency in the form $\omega = \omega' + i\chi$ [8], where χ is the parameter characterizing the damping. Figure 2 shows the frequency dependences of the magnetoelectric voltage coefficient calculated for Permendur–PZT structures from formula (15) with different coupling coefficients β. The calculations were carried out using the following parameters: ${}^{m}s_{11} = 5.5 \times 10^{-12}$ m^2/N, ${}^{m}q_{11} = 63.75 \times 10^{-10}$ m/A, ${}^{m}h = 0.36$ mm, ${}^{p}s_{11} = 15 \times 10^{-12}$ m^2/N, ${}^{p}d_{31} = -175 \times 10^{-12}$ m/V, ${}^{p}\varepsilon_{33}/\varepsilon_0 = 1750$, ${}^{p}h = 0.36$ mm, damping coefficient χ = 20 000 rad/s, sample length $L = 7.5$ mm, and phase-coupling coefficients $\beta = 1.0$ and 0.4.

It can be seen from Fig. 2 that, in the case of rigid coupling between the ferromagnet and the piezoelectric ($\beta = 1$), a resonant increase in the magnetoelectric coefficient is observed at a frequency of approximately 330 kHz for a sample with the above dimensions. This increase in the magnetoelectric coefficient is associated with the resonance occurring in the ferromagnet. Under the conditions where the coupling coefficient is less than unity, an additional resonance peak appears in the frequency dependence. In the case of rigid coupling, the magnetic field induces mechanical vibrations in the ferromagnet and the vibrations of the piezoelectric repeat them. When the coupling coefficient is less than unity, the vibrations of the piezoelectric medium with the parameter ${}^{m}k$ are accompanied by vibrations characterized by the parameter ${}^{p}k$. This leads to the appearance of an additional peak in the frequency dependence of the magnetoelectric voltage coefficient.

In the low-frequency range, the magnetoelectric coefficient is almost independent of the frequency and its value is determined by the relationship

$$\alpha_{E,T}^{\text{low}} = -\frac{\gamma\beta}{(\gamma+\beta)}\frac{{}^{p}d_{31}\,{}^{m}q_{11}}{{}^{p}\varepsilon_{33}\,{}^{p}s_{11}(1-K_{31}^{2}\gamma\beta/(\gamma+\beta))}\frac{{}^{p}h}{({}^{m}h+{}^{p}h)}. \tag{15}$$

It follows from relationship (14) that the magnetoelectric coefficient depends not only on the parameters of the magnetic and piezoelectric phases but also on the percentage of the composite and on the phase-coupling coefficient. For a small coupling coefficient β, the magnetoelectric coefficient is directly proportional to it. When the coupling coefficient β tends to unity, the dependence becomes weaker. According to relationships (14) and (15), the magnetoelectric coefficient reaches a maximum when the thickness ratio between the ferromagnetic and piezoelectric layers is given by

$${}^{m}h/{}^{p}h = (\beta\,{}^{m}s_{11}/{}^{p}s_{11})^{1/2}. \tag{16}$$

As was shown in the experiments [6], the maximum magnetoelectric coefficient is observed for a piezoelectric layer of thickness ${}^{p}h = 0.6$ mm and a magnetic layer of thickness ${}^{m}h = 0.36$ mm. After substituting the compliances for the Permendur and PZT alloys, we found that the experimental results are in good agreement with theory for a coupling coefficient $\beta \approx 1$.

3. CONCLUSIONS

Thus, the mechanical coupling between the magnetostrictive and piezoelectric subsystems in ferromagnetic–piezoelectric composites gives rise to a magnetoelectric effect. The maximum magnetoelectric coefficient is observed for a specific thickness ratio between the magnetic and piezoelectric phases. This ratio depends on the phase-coupling coefficient and on the ratio between the compliance moduli for the ferromagnetic and piezoelectric phases.

ACKNOWLEDGMENTS

This work was supported in part by the Ministry of Education of the Russian Federation (project no. E02-3.4-278) and the program "Russian Universities" (project no. UR 01.01.007).

REFERENCES

1. M. I. Bichurin, D. A. Filippov, V. M. Petrov, and G. Srinivasan, in *Proceedings of the International Conference on Physics of Electronic Materials* (Kaluga, 2002), p. 309.
2. D. A. Filippov, M. I. Bichurin, V. M. Petrov, and G. Srinivasan, Bull. Am. Phys. Soc. **48**, 214 (2003).
3. M. I. Bichurin, D. A. Filippov, V. M. Petrov, V. M. Laletsin, N. Paddubnaya, and G. Srinivasan, Phys. Rev. B **68**, 132 408 (2003).
4. D. A. Filippov, M. I. Bichurin, V. M. Petrov, V. M. Laletin, N. N. Paddubnaya, and G. Srinivasan, Pis'ma Zh. Tekh. Fiz. **30** (1), 15 (2004) [Tech. Phys. Lett. **30** (1), 6 (2004)].
5. D. A. Filippov, M. I. Bichurin, V. M. Petrov, V. M. Laletin, and G. Srinivasan, Fiz. Tverd. Tela (St. Petersburg) **46** (9), 1621 (2004) [Phys. Solid State **46** (9), 1674 (2004)].
6. V. Laletsin, N. Padubnaya, G. Srinivasan, and C. P. De Vreugd, Appl. Phys. A **78**, 33 (2004).
7. M. I. Bichurin, V. M. Petrov, and G. Srinivasan, J. Appl. Phys. **92** (12), 7681 (2002).
8. R. Truell, C. Elbaum, and B. B. Chick, *Ultrasonic Methods in Solid State Physics* (Academic, New York, 1969; Mir, Moscow, 1972).

Translated by N. Korovin

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1122–1128. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1085–1091.

MAGNETISM AND FERROELECTRICITY

Generalized Mean-Field Theory for Lattice Magnetic Systems and Ferromagnetism of Semiconductors with Magnetic Impurities

E. Z. Meĭlikhov and R. M. Farzetdinova

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

e-mail: meilikhov@imp.kiae.ru

Received June 8, 2004; in final form, September 27, 2004

Abstract—Mean-field theory is generalized to lattice systems of interacting magnetic moments by introducing a distribution function of random local magnetic fields to take into account the nonequivalence of various lattice sites. Analytical and numerical methods developed in this approach are used to describe ferromagnetism of nonmagnetic semiconductors with magnetic impurities. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

This study deals with a generalization of mean-field theory to lattice systems of paramagnetic ions with exchange interaction. The idea behind this generalization is simple, and generalization can be made to various systems. For the sake of definiteness, however, we consider a specific system, namely, a nonmagnetic semiconductor with magnetic impurities.

Recently, ferromagnetism in semiconductors with magnetic impurities has been intensively studied [1] because these materials are candidates for electronic technology. However, the mechanisms of this phenomenon are not clearly understood. Ferromagnetic ordering of the spins of magnetic impurities can be due to various indirect interactions mediated by free carriers: RKKY exchange [2], kinematic exchange [1, 3], etc. [1]. The universal Bloembergen–Rowland (BR) mechanism [4] can also be operative; this mechanism does not require the presence of carriers (or a high density of carriers, which leads to degeneracy) and can be responsible for ferromagnetism in systems like GaAs(Mn) and GaN(Mn) [5]. In this mechanism, the spins of impurities interact via virtual holes that arise in the valence band when electrons are promoted from this band to acceptor impurity levels. Since there is a threshold for this process due to the finiteness of the acceptor ionization energy Δ, the energy of this indirect interaction decays exponentially with the distance between impurity atoms. The characteristic energy of this decay is determined by the de Broglie wavelength $\lambda_\Delta = \hbar(2m\Delta)^{-1/2}$ of a hole with energy Δ and mass m. The energy W of the indirect interaction between the spins of two impurity atoms separated by a distance r is given by [5]

$$W(r) = -\frac{J_{pd}^2 m^2 \Delta}{\pi^3 \hbar^4 N^2 r^2} K_2(2r/\lambda_\Delta), \tag{1}$$

where J_{pd} is the contact interaction energy between an impurity spin and a hole, N is the concentration of the host lattice atoms, and K_2 is the MacDonald function. At large distances ($r \gtrsim r_0$), we have $K_2(2r/\lambda_\Delta) \approx (4r/\pi\lambda_\Delta)^{-1/2}\exp(-2r/\lambda_\Delta)$. Keeping only the exponential part of the spatial dependence of the interaction energy in Eq. (1) in this case, we write the dependence in the form

$$W(r) = -J_0 \exp(-r/r_0), \tag{2}$$

where $r_0 = \lambda_\Delta/2$ and $J_0 \sim J_{pd}^2 m^2\Delta/\pi^3\hbar^4 N^2 r_0^2$ or

$$J_0 \sim \frac{1}{64\pi^3 N r_0^3}\left(\frac{J_{pd}^2}{\Delta}\right). \tag{3}$$

A similar expression for the interaction energy between magnetic impurity atoms is used in the kinematic-exchange model [1]:

$$\begin{aligned} W(r) &= -J_0 \exp(-r^2/2r_0^2), \\ J_0 &= J_{pd}/(2\pi)^{3/2} N r_0^3, \end{aligned} \tag{4}$$

where r_0 is the characteristic interaction length.

2. GENERALIZED MEAN-FIELD THEORY

In this paper, we study the magnetic state of a system of magnetic moments which interact with one another via exchange interaction described by Eq. (2) or (4). We take into account that, in real systems, magnetic impurity atoms are located at host lattice sites (at least when the impurity concentration is sufficiently low). In this respect, the system under study closely resembles the Ising model of magnetic dipoles. This model is a lattice system of magnetic moments that can point in one of two possible (opposite) directions and interact only with their nearest neighbors.

In our case, however, we cannot restrict our consideration to only nearest neighbor interactions. For this reason, the properties of the magnetic systems in question are analyzed using a generalized mean-field theory and the required distribution function of local fields (see below) is found using numerical methods.

The need to generalize the conventional mean-field theory arises from the fact that this theory not only neglects correlation between magnetic moments but also assumes all lattice sites to be equivalent. Indeed, the mean field is assumed to be the same at all sites, whereas in actuality this field varies randomly from site to site. It is natural to take into account these random variations. A method for solving this problem was proposed for the first time in [6], where a system of randomly arranged magnetic dipoles was considered. Later on, this method was analyzed in detail in a series of papers [7], where an *ad hoc* energy distribution function of pairwise interactions was used. In this generalized model, the standard equation from the mean-field theory

$$j = \tanh\left(\frac{\lambda j}{kT}\right) \tag{5}$$

(where j is the reduced magnetization of the system, λ is the mean-field constant) is replaced by a generalized relation,

$$j = \int_{-\infty}^{\infty} \tanh\left(\frac{\mu H}{kT}\right) F(j;\, H)\, dH. \tag{6}$$

Here, $F(j; H)$ is the distribution function of local fields H exerted on each magnetic moment of the system (which is not necessarily regular) with magnetization j by the other magnetic moments (equal to μ). Clearly, when calculating this distribution function, it is more consistent to use the spatial dependence of the pairwise interaction energy W of magnetic moments (e.g., of the form of Eq. (2)) than to employ an *ad hoc* distribution of these interaction energies (as done in [7]). Note that Eq. (5) given by the mean-field theory is equivalent to

$$j = \tanh\left[\frac{\mu \bar{H}(j)}{kT}\right], \quad \bar{H}(j) = \int_{-\infty}^{\infty} H F(j;\, H)\, dH. \tag{7}$$

For square or simple cubic lattices (considered in the Ising model), the distribution function $F(j, H)$ can easily be found in an explicit form. Thereafter, the magnetic properties of the system can be studied using Eq. (6).[1] In our model, the function $F(j, H)$ can be found analytically only in the (less interesting) case of a high impurity concentration. For this reason, we perform numerical calculations.

3. NUMERICAL CALCULATIONS OF THE DISTRIBUTION FUNCTIONS

In order to calculate the random-field distribution function, spins were randomly distributed over the lattice sites within a fairly large spherical region (the total number of spins was greater than 10^4) and the magnetic field H was calculated at the center of this region. The function $F(j; H)$ was calculated for a large number (about 10^4) of such samples. Figures 1–3 show numerically calculated distribution functions $F(j; H)$ for three groups of three-dimensional (3D) systems. In one group, the parameter Nr_0^3 was varied at fixed values of $j = 1$ and $p = 0.1$; in another group, the parameter p was varied at $j = 0.1$ and $Nr_0^3 = 0.5$; and in the third group, the magnetization j was varied at $Nr_0^3 = 0.1$ and $p = 0.1$. It can be seen that the functions $F(j; H)$ are close to Gaussians for $Nr_0^3 \gtrsim 1$ and deviate from Gaussians outside this concentration range. The largest deviations are observed for the function $F(1; H)$: at $Nr_0^3 \ll 1$, this function breaks up into separate peaks corresponding to individual lattice sites.

4. ANALYTICAL CALCULATION OF THE DISTRIBUTION FUNCTIONS

According to computer simulations, the distribution function $F(j; H)$ is close to a Gaussian if $Nr_0^D \gtrsim 1$, i.e., if the number of lattice sites in the effective-interaction region is much greater than the coordination number z. This is not surprising, because the local field in this case is a sum of a large number of independent terms with a large variance and, according to the limit theorem from probability theory, is described by a Gaussian distribution function.

In order to find the function $F(j; H)$ analytically, we neglect the discreteness of the lattice and only take into account that the minimum spacing between magnetic moments cannot be less than the lattice parameter a.

In this model, the local effective fields $h(r) = -W(r)/\mu$ are produced by magnetic moments randomly distributed in space. These fields vary randomly from point to point. Their sum H is a random variable and is characterized by a distribution function $F(j; H)$. This function, in general, depends on the magnetization j of the system, which is determined by the average fraction $\eta = (1/2)(1 + j)$ of the magnetic moments directed along the magnetization of the system. Note that $F(1; H)$ is the

---

[1] Other methods for solving this problem and an extensive bibliography can be found in reviews [8–10].

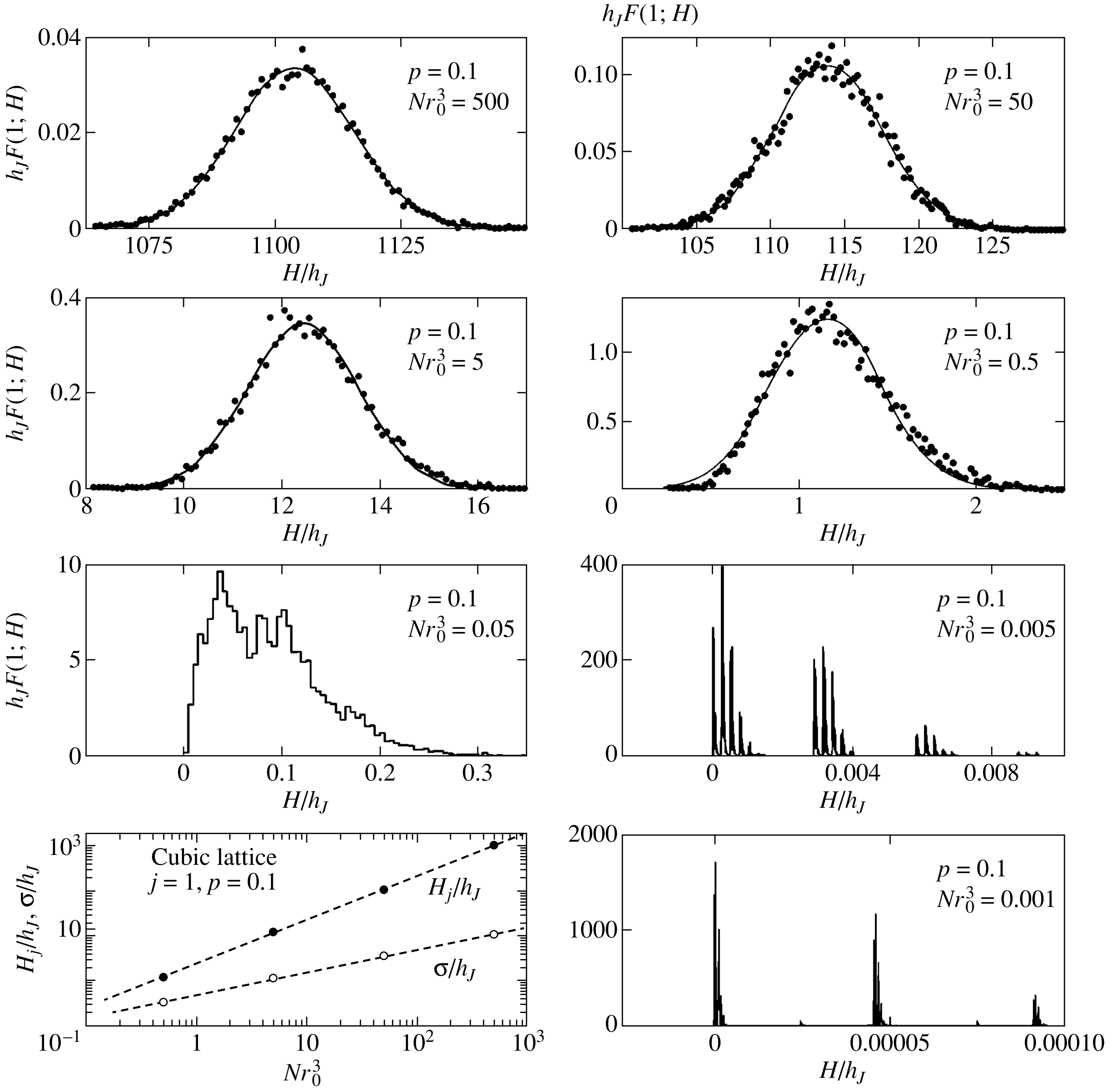


Fig. 1. Distribution functions $F(1; H)$ for completely magnetized ($j = 1$), partly occupied ($p = 0.1$) cubic lattices of Ising magnetic moments interacting through the Bloembergen–Rowland mechanism for various values of the parameter Nr_0^3. At $Nr_0^3 \gtrsim 1$, these functions are close to Gaussians with parameters H_j and σ, whose concentration dependences are shown in the bottom left-hand panel.

distribution function in the case where all magnetic moments are parallel to one another.

We find the distribution function using the Markov method [11], according to which

$$F(j;\,H) = \frac{1}{2\pi}\int_{-\infty}^{\infty} A(q)\exp(-iqH)dq,$$

$$A(q) \tag{8}$$

$$= \lim_{M_{\max}\to\infty}\left[\int_{(\zeta)}\int_{r=a}^{r_{\max}} \exp[iqh_{\zeta}(r,\zeta)]\tau(r,\zeta)drd\zeta\right]^{M_{\max}},$$

where

$$h_{\zeta}(r,\zeta) = \zeta h(r), \quad h(r) = h_J\exp(-r/r_0) \tag{9}$$

is the effective magnetic field produced by a magnetic moment situated at a random distance r from the coordinate origin and $h_J = J/\mu$. The random variable ζ takes on values ±1 with probabilities η and $(1 - \eta)$, respectively, and defines the direction of the magnetic moment; $\tau(r, \zeta)$ is the distribution function for the random variables r and ζ; and $M_{\max}$ is the number of magnetic moments within the sphere of radius $r_{\max}$ over

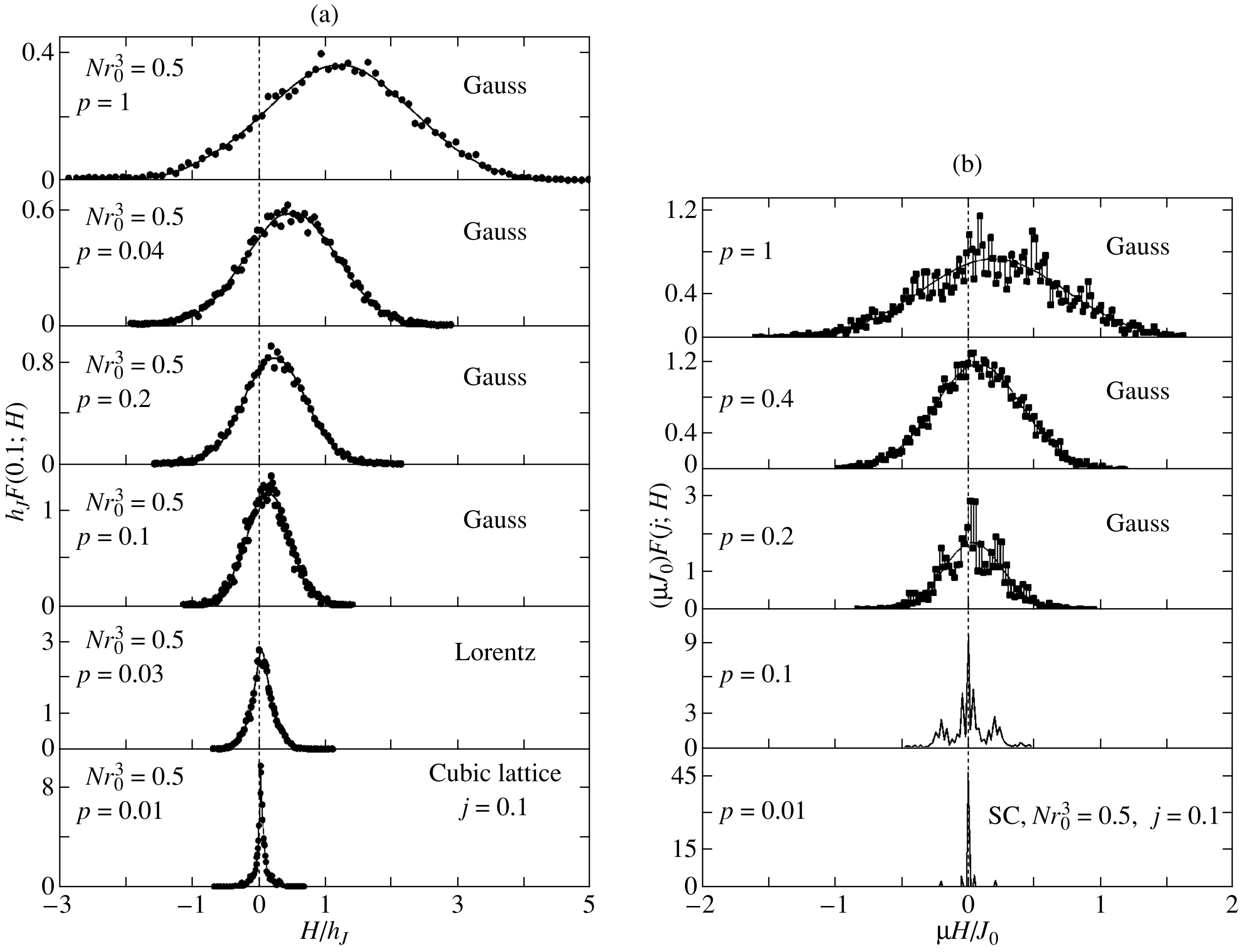


Fig. 2. Distribution functions $F(0.1; H)$ for weakly magnetized ($j = 0.1$), partly occupied ($p = 0.01$–1) simple cubic (SC) lattices of Ising magnetic moments interacting through (a) the Bloembergen–Rowland mechanism or (b) the kinematic-exchange mechanism. $Nr_0^3 = 0.5$. At $p \gtrsim 0.1$, these functions are close to Gaussians.

which integration is performed. If the distributions of r and ζ are uniform and not correlated, then

$$\tau(r, \zeta)drd\zeta = [3r^2/(r_{\max}^3 - a^3)]dr[(1-\eta)\delta(\zeta+1) + \eta\delta(\zeta-1)]d\zeta \quad (10)$$

for a 3D system and

$$\tau(r, \zeta)drd\zeta = [r/(r_{\max}^2 - a^2)]dr[(1-\eta)\delta(\zeta+1) + \eta\delta(\zeta-1)]d\zeta \quad (11)$$

for a 2D system.

Substituting Eq. (10) into Eq. (8) gives

$$A(q) = \exp[-pNC(q)], \quad (12)$$

where

$$C(q) = 4\pi\int_a^\infty \{1 - \cos[qh(r)] - ij\sin[qh(r)]\}r^2 dr \quad (13)$$

for a 3D system and

$$C(q) = 2\pi\int_a^\infty \{1 - \cos[qh(r)] - ij\sin[qh(r)]\}r dr \quad (14)$$

for a 2D system.

Equations (12)–(14) are too complicated to be used to find an analytical expression for the distribution function $F(j; H)$. For this reason, we use a small-q approximation, based on the fact that large values of q make an insignificant contribution to the inverse Fourier transform of Eq. (8). In this approximation, the functions $\cos[qh(\rho, \alpha)]$ and $\sin[qh(\rho, \alpha)]$ involved in the integrals in Eqs. (13) and (14) are replaced by the first nonvanishing term in their expansion in a power

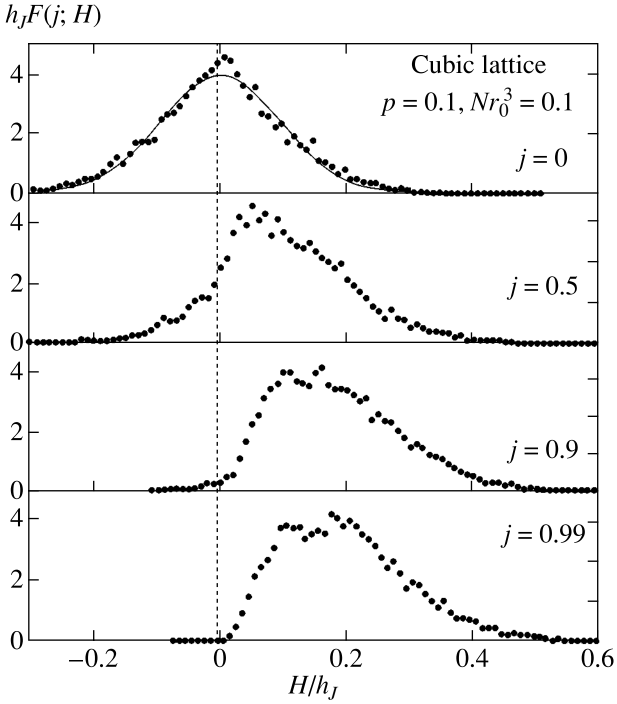


Fig. 3. Distribution functions $F(j;\ H)$ for partly occupied ($p = 0.1$) cubic lattices of Ising magnetic moments interacting through the Bloembergen–Rowland mechanism [described by Eq. (2)] for various values of the magnetization. $Nr_0^3 = 0.1$. At $j \ll 1$, the distribution function is close to a Gaussian. As j increases, the left-hand wing gradually becomes shorter. At $j = 1$, the distribution function is nonzero only for $H > 0$.

series in qh and the integrals can easily be calculated. For $Nr_0^D \gg 1$ (i.e., $a \ll r_0$), we thus obtain

$$C(q) = 8\pi r_0^3\left[\frac{1}{16}(qh_J)^2 - ijqh_J\right] \quad (3\mathrm{D}),$$
$$C(q) = 2\pi r_0^2\left[\frac{1}{8}(qh_J)^2 - ijqh_J\right] \quad (2\mathrm{D}). \tag{15}$$

The distribution function $F(j;\ H)$ in this case is found to be a Gaussian,

$$F(j;\ H) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(H - jH_j)^2}{2\sigma^2}\right], \tag{16}$$

where

$$H_j/h_J = 8\pi pNr_0^3,$$
$$\sigma/h_J = \sqrt{\pi pNr_0^3} \tag{17}$$

for a 3D system with interaction (2);

$$H_j/h_J = (4\pi)^{3/2}pNr_0^3,$$
$$\sigma/h_J = (\pi/2)^{3/4}\sqrt{pNr_0^3} \tag{18}$$

for a 3D system with interaction (4);

$$H_j/h_J = 2\pi pNr_0^2,$$
$$\sigma/h_J = (\pi/2)^{1/2}\sqrt{pNr_0^2} \tag{19}$$

for a 2D system with interaction (2); and

$$H_j/h_J = 4\pi pNr_0^2,$$
$$\sigma/h_J = (\pi/2)^{1/2}\sqrt{pNr_0^2} \tag{20}$$

for a 2D system with interaction (4).

As the magnetization decreases, the maximum of the function $F(j,\ H)$ (at $H = jH_j$) shifts to lower fields, while its width remains unchanged. The concentration dependences of the parameters H_j and σ of the Gaussian distribution derived above are supported by the results of computer simulations shown in Fig. 1 (the bottom left-hand panel in Fig. 1), which agree well with Eqs. (17). There is also good agreement in the other cases (for interaction (4) and for 2D systems).

The approximate expressions derived above are valid if the effective-interaction region contains a large number of magnetic moments, i.e., if $pNr_0^D \gtrsim 1$. However, this criterion depends on the magnetization j of the system. For example, at $j = 1$, the effective field cannot be negative, i.e., $F(1;\ H < 0) = 0$. This condition is not satisfied for Gaussian function (16); however, this function is exponentially small in the range $H < 0$ if $H_j \gtrsim \sigma$ or, according to Eqs. (17)–(20), $pNr_0^3 \gtrsim 1$. In the case of $j = 0$, the distribution function is symmetric relative to $H = 0$ and can be a Gaussian even for much lower values of the parameter pNr_0^D.

5. MAGNETIC PROPERTIES OF LATTICE SYSTEMS

The magnetic properties of the systems in question are determined by the generalized mean-field equation (6). If the distribution function $F(j;\ H)$ is close to a Gaussian, this equation reduces to the conventional mean-field equation (5). Indeed, we have $\overline{H} = jH_j$ in this case and the mean-field constant involved in Eq. (5) is $\lambda = \mu H_j \propto J_0pNr_0^D$. This leads to the conventional concentration dependence of the Curie temperature $T_C \propto J_0pNr_0^D$. If the distribution function is not a Gaussian (which is the case, as shown above, for $pNr_0^3 \ll 1$), we

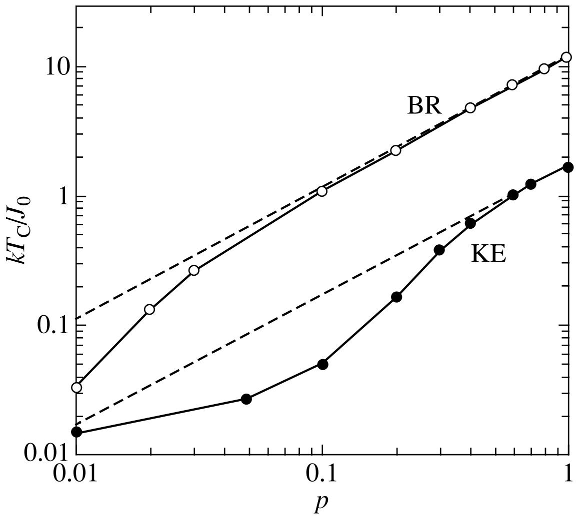


Fig. 4. Curie temperature T_C plotted as a function of the lattice site occupation probability p for a cubic lattice of Ising magnetic moments interacting through the Bloembergen–Rowland (BR) or kinematic-exchange (KE) mechanism. $Nr_0^3 = 0.5$.

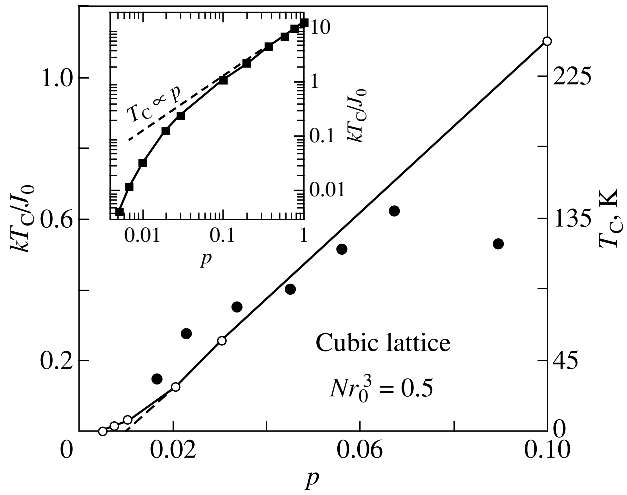


Fig. 5. Curie temperature T_C plotted as a function of the lattice site occupation probability p for a cubic lattice of Ising magnetic moments interacting through the mechanism described by Eq. (2). $Nr_0^3 = 0.5$. The right-hand scale corresponds to $J_0 = 20$ meV. The inset shows the same in logarithmic coordinates.

may expect deviations from the predictions from the conventional theory.

The results obtained by solving Eq. (6) with the numerically calculated distribution functions support this conclusion (Figs. 4, 5). For $Nr_0^D \sim 1$, the deviations from the concentration dependence of the Curie temperature predicted from the mean-field theory are noticeable at small values of the concentration p. As might be expected, these deviations are observed for smaller values of p in the case of interaction (2), which decays more slowly with distance (Fig. 4). In both cases, the linear portion of the $T_C(p)$ dependence is shifted to larger values of p, which can be interpreted in experiments as the existence of a threshold concentration $p_c \sim 0.01$ (in the 3D case) below which the system remains paramagnetic at $T = 0$. In actuality, the ground state is always ferromagnetic ($J_0 > 0$!).

As an illustration of the results obtained let us consider the GaAs(Mn) compound. This nonmagnetic semiconductor with magnetic impurities has been best studied experimentally [1, 2, 12]. The contact-interaction energy J_{pd} in GaAs(Mn) has not been determined exactly; estimates range in value from 1 to 3 eV [2]. The ionization energy of Mn acceptors is $\Delta \approx 0.11$ eV [13]. With this energy and the hole effective mass $m \approx 0.5m_0$, the effective interaction length is calculated to be $r_0 \approx 4.3$ Å, which is close to the lattice parameter $a \approx 5.6$ Å for GaAs (therefore, the shortest spacing between Mn atoms substituting for Ga atoms is $a/\sqrt{2} \approx 4$ Å). For $Nr_0^3 \sim 1$ in the 3D case, the characteristic indirect-interaction energy between Mn spins is found from Eq. (3) to be $J_0 \sim (1/64\pi^3)(J_{pd}^2/\Delta) = 5$–$40$ meV. The absolute values of the Curie temperature calculated from the average value of $J_0 = 20$ meV (for the range indicated above) are shown in Fig. 5 on the right-hand vertical scale. It can be seen that the calculations agree well with the experimental data on the GaAs(Mn) compound [12].[2]

Similar results were also obtained for a system in which impurities interact through the kinematic-exchange (KE) mechanism. In this case, Eq. (4) gives $J_0 \sim 200$ meV, which is one order of magnitude greater than the value for the BR mechanism. It follows that the predicted absolute values of the Curie temperature are similar in both cases, because, as seen from Fig. 4, the values of the dimensionless parameter kT_C/J_0 for the KE mechanism are one order of magnitude smaller than those for the BR mechanism.

6. CONCLUSIONS

The conventional mean-field theory has been significantly refined by introducing the distribution function $F(H)$ of random local magnetic fields for both regular and nonregular (random) systems with magnetic interaction. This function possesses a clear physical interpretation and enables one to determine not only the

---

[2] It should be noted that low-temperature annealing for a prolonged period brings about a considerably increased Curie temperature as compared to that of the initial samples. It is likely that the nonmonotonic $T_C(p)$ dependence shown in Fig. 5 is due to the insufficiently long duration of annealing treatment of samples in [13].

magnetization but also the energy $E = -\int \mu H F(H) dH$ of the system and, hence, other characteristics, such as the free energy and heat capacity. In systems with long-range interaction (where the local field is produced by a large number of interacting magnetic moments), the distribution function $F(H)$ is Gaussian (regardless of the magnetization j of the system) and all information on the interaction and the mean concentration of interacting magnetic moments is contained in two parameters of this distribution, namely, the position of the maximum $H = jH_j$ and its width σ. The method under discussion (neglecting the correlation between the magnetic moments) is convenient because these parameters can be determined using a known simple procedure [11]. The generalized model refines the conventional mean-field theory, which can be verified by comparing their results with known exact solutions for simple cases.

The most noticeable changes are brought about by the generalized mean-field theory in the results for a low-density system, where the spacing between magnetic moments is comparable to the interaction length. In this case, the distribution function becomes significantly non-Gaussian, analytical calculations are no longer simple, and numerical model calculations have to be made to find the function $F(H)$.

Although the idea behind the method under discussion was described long ago [7], this method has only recently been used to study the magnetic properties of various specific 2D and 3D systems [14]. In this paper, we have applied this method to study the properties of a lattice system of magnetic moments between which the interaction decreases exponentially with distance. One example of such systems is a nonmagnetic semiconductor with magnetic impurities. The magnetic ordering that occurs in this case can be adequately described in terms of the generalized mean-field theory.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, project nos. 03-02-17029 and 05-02-17021.

REFERENCES

1. J. Konig, J. Schliemann, T. Jungwirth, and A. H. MacDonald, cond-mat/0111314; in *Electronic Structure and Magnetism of Complex Materials*, Ed. by D. J. Singh and D. A. Papaconstantopoulos, (Springer, 2002).
2. T. Dietl, H. Ohno, F. Matsukara, J. Cibert, and D. Ferrand, Science **287**, 1019 (2000).
3. P. M. Krstajic, V. A. Ivanov, F. M. Peeters, V. Fleurov, and K. Kikoin, Europhys. Lett. **61**, 235 (2003).
4. N. Bloembergen and T. J. Rowland, Phys. Rev. **97**, 1679 (1955).
5. V. I. Litvinov and V. K. Dugaev, Phys. Rev. Lett. **86**, 5593 (2001).
6. M. W. Klein and R. Brout, Phys. Rev. **132**, 2412 (1963).
7. M. Thomsen, M. F. Thorpe, T. C. Choy, and D. Sherrington, Phys. Rev. B **30**, 250 (1984); T. C. Choy, D. Sherrington, M. Thomsen, and Thorpe, Phys. Rev. B **31**, 7355 (1985); M. Thomsen, M. F. Thorpe, T. C. Choy, and D. Sherrington, Phys. Rev. B **33**, 1931 (1986).
8. B. E. Vugmeister and M. D. Glinchuk, Rev. Mod. Phys. **62**, 993 (1990).
9. A. P. Young, J. D. Reger, and K. Binder, in *Monte Carlo Methods on Statistical Mechanics*, Ed. by K. Binder (Springer, 1992).
10. Vik. S. Dotsenko, Usp. Fiz. Nauk **165**, 481 (1995) [Phys. Usp. **38**, 457 (1995)].
11. S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943).
12. K. W. Edmonds, K. Y. Wang, R. P. Campion, A. C. Neumann, C. T. Foxon, B. L. Gallagher, and P. C. Main, Appl. Phys. Lett. **81**, 3010 (2002).
13. S. Sonoda, S. Shimizu, T. Sasaki, Y. Yamamoto, and H. Hori, J. Appl. Phys. **91**, 7911 (2002); H. Hori, S. Sonoda, T. Sasaki, Y. Yamamoto, S. Shimizu, K. Suga, and K. Kindo, Physica B **324**, 142 (2002).
14. E. Z. Meĭlikhov, Zh. Éksp. Teor. Fiz. **124**, 650 (2003) [JETP **97**, 587 (2003)]; E. Z. Meĭlikhov and R. M. Farzetdinova, Zh. Éksp. Teor. Fiz. **124**, 656 (2003) [JETP **97** (3), 593 (2003)]; E. Z. Meĭlikhov, Zh. Éksp. Teor. Fiz. **125**, 1329 (2004) [JETP **98** (6), 1164 (2004)]; E. Z. Meĭlikhov, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 675 (2003) [JETP Lett. **77**, 571 (2003)]; E. Z. Meĭlikhov and R. M. Farzetdinova, J. Magn. Magn. Mater. **268** (1–2), 237 (2004).

Translated by Yu. Epifanov

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1129–1133. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1092–1096.
Original Russian Text Copyright © 2005 by Zhigalov, Ivantsov, Édelman, Seredkin, Frolov, Bondarenko.

MAGNETISM AND FERROELECTRICITY

Faraday Effect in Nanogranular Co–Sm–O Films

V. S. Zhigalov, R. D. Ivantsov, I. S. Édelman, V. A. Seredkin, G. I. Frolov, and G. V. Bondarenko

Kirensky Institute of Physics, Siberian Division, Russian Academy of Sciences, Akademgorodok, Krasnoyarsk, 660036 Russia

e-mail: zhigalov@iph.krasn.ru

Received July 14, 2004; in final form, November 2, 2004

Abstract—The Faraday effect (FE) was studied in Co–Sm–O composite films consisting of nanoparticles of metallic cobalt embedded in a samarium oxide dielectric matrix. The volume of the magnetic phase was ~60%. The FE spectral dependence for the condensates studied revealed a substantial change as compared to that for bulk cobalt samples, as well as for the films of nanocrystalline Co and CoSm films prepared in this study. An enhancement of the FE was also observed in the short-wavelength part of the optical spectrum. © *2005 Pleiades Publishing, Inc.*

Composite materials made up of nanosized magnetic particles separated by a dielectric layer have considerable potential for both basic research and practical applications. The broad interest expressed in magnetic nanocomposites stems primarily from their exhibiting a number of unusual phenomena, such as giant magnetoresistance, magnetically soft and high-resistivity properties, optical transparency over a broad spectral region, and significant magnetooptical effects [1–3]. Belonging to this group is also resonant enhancement of magnetooptical effects (MOEs), which is revealed by these materials in various optical regions and was predicted by theorists (see, e.g., [4]). Because the resonance MOE originates from surface plasma vibrations in particles of the magnetic material, the MOE should depend on the particle size, the content of the magnetic phase in the composite, and the optical parameters of the metal and the dielectric. The situation is complicated by the fact that the optical and magnetooptical parameters of nanosized metal particles may differ substantially from those of the bulk materials and that the properties of composites with a high magnetic-phase content are governed both by individual characteristics of the nanoparticles and by the effects originating from their interaction.

Progress in miniaturization places a particular emphasis on film nanocomposites. The materials that have been studied thus far are primarily Co–SiO_2 film composites (see, e.g., [5, 6]). In [7], Co–Sm–O films were prepared for the first time in which a composition structure in the form of ferromagnetic nanoparticles of metallic Co embedded in a Sm oxide matrix arose under certain conditions of condensation and additional annealing. Direct observations of nanoparticles were carried out using electron microscopy and revealed some characteristic features of the magnetic properties of these composites. The results of those studies suggest that various distinctive features and magnetooptical properties could be expected in these composites, including variations in the spectral dependence of the Faraday effect (FE).

Here, we report on a study of the spectral and field dependences of the FE in films of the Co–Sm–O system, whose structure and magnetic properties are described in [7]. The films were prepared by pulsed plasma sputtering (PPS) of a $SmCo_5$ target in vacuum under an initial pressure of 10^{-6} Torr. X-ray fluorescence analysis showed that the Sm concentration varied from one sample to another within 13–17 at. %. The samples were ~100-nm-thick and did not contain the $SmCo_5$ phase, because samarium has high chemical reactivity and, in the vacuum conditions chosen, oxidizes directly in the vacuum chamber in the course of film deposition. Auger electron spectroscopy detected certain amounts of carbon in the films. Electron microscopy and x-ray diffraction measurements established that the film structure is made up of magnetic particles of a Co(C) solid solution, $D \sim 1.5$ nm in size, embedded in samarium oxide (Sm_2O_3), with the volume of the magnetic phase f being ~60% [7]. This morphology of the film nanocomposite is obviously responsible for the high electrical resistivity of the samples ($\rho \sim 0.5\ \Omega$ cm). Starting films are superparamagnetic. At room temperature, the magnetization curve has no hysteresis, but when samples are cooled, magnetic hysteresis appears, with a coercive force $H_c \cong 50$ Oe at $T = 77$ K. The temperature dependence of H_c was used to determine the blocking temperature (the temperature of transition to the superparamagnetic state), which was found to be $T_B \cong 81$ K. Our studies involved subjecting the films to heat treatment, namely, annealing in vacuum (10^{-6} Torr) at various temperatures, which initiated hysteresis at room temperature. The value of H_c depended on the actual annealing regime; the films subjected to short annealings at $T \le 500$ K exhibited a weak coercive force ($H_c \le$ 3 Oe). For the purpose of comparison, we also prepared samples using the same technology of pulsed plasma

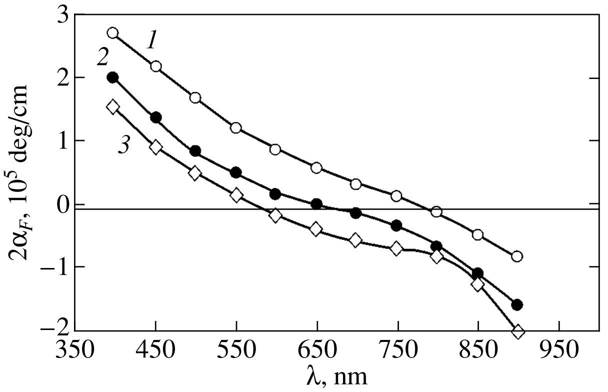


Fig. 1. Dispersion graphs of the FE ($2\alpha_F$) measured in a magnetic field $H = 4.5$ kOe for starting Co–Sm_2O_3 films on various substrates: (*1*) glass plates, (*2*) single-crystal quartz, and (*3*) single-crystal MgO. The effective cobalt thickness used in calculations of the specific FE was determined by x-ray fluorescence to within 5%.

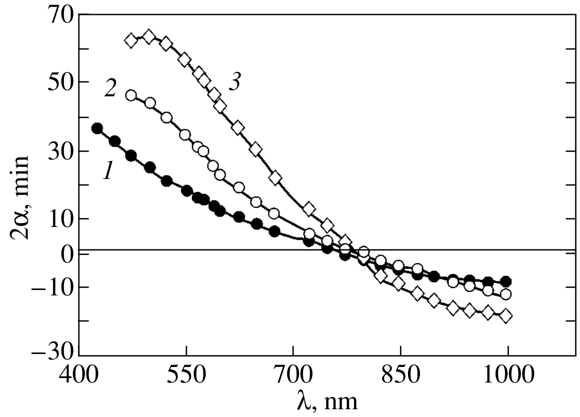


Fig. 2. Dispersion graphs of the FE ($2\alpha_F$) measured in a magnetic field $H = 4.5$ kOe for a Co–Sm_2O_3 film prepared on a glass substrate: (*1*) starting film and (*2, 3*) after annealing at (*2*) 520 and (*3*) 620 K.

sputtering of targets of pure cobalt and Co_5Sm targets in ultrahigh vacuum (10^{-9} Torr). In this case, Sm did not oxidize, as a result of which these films differed strongly from the above samples in terms of their structure and magnetic properties.

The FE spectral dependences were obtained with the null technique by modulating the plane of polarization of light in the 400- to 1000-nm spectral interval. The accuracy of measuring the rotation of the plane of polarization was ±0.2′ (Faraday rotation). A magnetic field of up to 4.5 kOe was directed along the light beam normal to the sample plane. The field was measured to within ±20 Oe.

Figure 1 presents the FE spectral curves ($2\alpha_F$) for the starting Co–Sm_2O_3 films (thickness ~100 nm) deposited on different substrates. Three features immediately attract attention. (i) Unlike the decrease in the FE with decreasing wavelength of light λ in pure cobalt films reported in the literature [8], here it is seen to grow appreciably. A similar pattern was observed with the FE in layered Co/SiO_2 films [9]. (ii) the FE reverses sign in the wavelength interval from 550 to 800 nm. (iii) Single-crystal substrates are conducive to displacement of the $2\alpha_F(\lambda)$ curves toward shorter wavelengths.

Figure 2 presents $2\alpha_F(\lambda)$ graphs for films prepared on glass substrates: for the starting film (curve *1*) and for those annealed at different temperatures for 30 min (curves *2*, *3*). We see that annealing at $T_{\text{ann}} = 620$ K brings about the appearance of a maximum near $\lambda \sim 550$ nm, with the FE increasing approximately threefold in this region. The wavelength at which the FE reverses sign does not shift. Thus, these graphs demonstrate the following points, which require explanation. First, there is a rearrangement of the FE spectrum as compared to solid Co films, which is accompanied by an enhancement of the effect in the short-wavelength region of the spectrum. Second, there is a dependence of the FE value and of the position of the characteristic points in the spectrum on the substrate type and annealing regime.

Let us compare the FE dispersion curves obtained for a Co–Sm_2O_3 composite film with those for films prepared by ablation of metallic Co and $SmCo_5$ targets in ultrahigh vacuum (Fig. 3). All the films were deposited on glass substrates using the PPS technology, have the same effective thickness of the metallic component, and were annealed at 620 K. The FE spectral curves for the films prepared by ablation of a pure cobalt target or of a Co_5Sm target in a vacuum of ~10^{-9} Torr differ in terms of their pattern from the FE graph obtained for the nanogranular Co–Sm_2O_3 film and are closer to the literature data quoted for solid Co films [8]. One can judiciously assume that the difference in the FE spectral dependences between films prepared in a higher

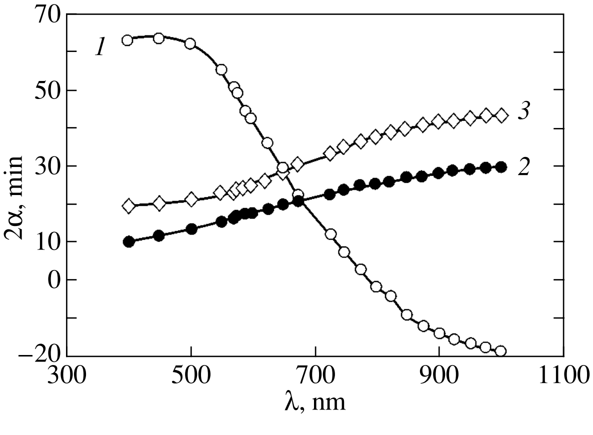


Fig. 3. Spectral curves of the FE ($2\alpha_F$) measured in a magnetic field $H = 4.5$ kOe for (*1*) a Co–Sm_2O_3 film and (*2, 3*) films prepared by ablation of (*2*) Co and (*3*) Co_5Sm targets in ultrahigh vacuum. All targets were annealed at 620 K.

and a lower vacuum originates from the formation of the Sm_2O_3 dielectric component in the latter case, which finds ready explanation in the high chemical reactivity of Sm.

As already mentioned, the observed changes in the FE wavelength dependence resemble those described to occur in multilayered Co/SiO_2 films [9], which have been interpreted satisfactorily in terms of the effective permittivity tensor [10]. As shown in [7], cobalt particles in starting films are almost spherical in shape and are distributed more or less uniformly over the bulk of the film and the linear dimension of each particle is 1.5–2.0 nm, which is about two orders of magnitude less than the film thickness (~100 nm). Therefore, in a first approximation, the effective permittivity tensor of a composite film, ε_{eff}, can be calculated using the model of spherical particles distributed uniformly over an infinite matrix [10]. In accordance with this model, the effective diagonal (ε_{eff}) and off-diagonal (γ_{eff}) components of the permittivity tensor ε

$$\hat{\varepsilon} = \begin{bmatrix} \varepsilon_{\text{eff}} & i\gamma_{\text{eff}} & 0 \\ -i\gamma_{\text{eff}} & \varepsilon_{\text{eff}} & 0 \\ 0 & 0 & \varepsilon^{\parallel}_{\text{eff}} \end{bmatrix} \tag{1}$$

are related to the parameters of the composite components through the relations

$$\varepsilon_{\text{eff}} = \varepsilon_0 \frac{2f(\varepsilon - \varepsilon_0) + \varepsilon + 2\varepsilon_0}{\varepsilon(1-f) + \varepsilon_0(2+f)}, \tag{2}$$

$$\varepsilon^{\parallel}_{\text{eff}} = \varepsilon_0 \frac{2f(\varepsilon^{\parallel} - \varepsilon_0) + \varepsilon^{\parallel} + 2\varepsilon_0}{\varepsilon^{\parallel}(1-f) + \varepsilon_0(2+f)}, \tag{3}$$

$$\gamma_{\text{eff}} = \frac{9f\varepsilon_0^2\gamma}{[\varepsilon(1-f) + \varepsilon_0(2+f)]^2}, \tag{4}$$

where ε_0 is the permittivity of the dielectric matrix (of samarium dioxide in the case considered); ε and γ are the diagonal and off-diagonal components of the permittivity tensor for the particle material (cobalt), respectively; f is the magnetic-phase filling coefficient; and $N = 1/3$. We calculated the FE spectra by substituting Eqs. (2) and (4) into the well-known relation

$$\alpha_F - i\psi_F = \frac{\pi}{\lambda}\frac{\gamma}{\sqrt{\hat{\varepsilon}}}. \tag{5}$$

The values of ε and γ were taken from [11] and [12], respectively, and it was assumed that $\varepsilon_{\text{eff}} = \varepsilon^{\parallel}_{\text{eff}}$. The calculation is complicated by the lack of literature data on ε_0 of Sm_2O_3. Furthermore, the presence of a certain amount of carbon in the film and the small dimensions of the cobalt particles may make the optical parameters of the material of the particles substantially different from those of pure cobalt films [11]. A study of the parameters ε and γ of Co films as a function of their thickness [12] showed that variations in the film thickness have a noticeable effect on these parameters.

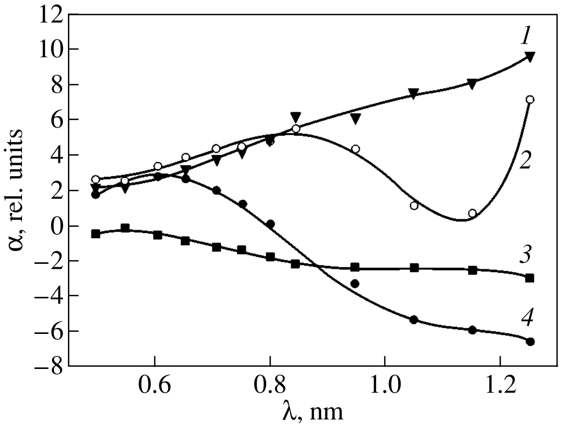


Fig. 4. FE spectral dependences calculated from Eqs. (2)–(5) for various values of the magnetic phase filling coefficient f: 1.0, 0.8, 0.6, and 0.2 (curves *1–4*, respectively). The values of ε were taken from [11], and the values of γ, from [12] for a 26-nm-thick film [12, Table 2]; $\varepsilon_0 = 4$.

Figure 4 presents α(λ) graphs plotted for various values of the magnetic phase filling coefficient f. We used the values of γ' and γ'' from [12] for a film which had a value of ε' closest in magnitude to that given in [11]. Figure 4 demonstrates a substantial change in the FE wavelength dependence as f is varied: there appears a maximum at shorter wavelengths and the FE reverses sign. The peak value of the FE, the energy position of the maximum, and the point of reversal of the FE in sign depend nonmonotonically on f. For certain values of f, the FE remains negative throughout the spectral range covered. The FE spectral dependence calculated for $f = 0.2$ fits best to the experimental curves in Figs. 1–3. This value of f is substantially smaller than the magnetic phase filling coefficient estimated on the basis of the technological conditions. This discrepancy may be assigned to the strong misfit between the optical and magnetooptical characteristics of the particle material and the characteristics accepted in the calculation. Therefore, refining model calculations were carried out. Figure 5 plots the FE spectra calculated for $f = 0.6$ from Eqs. (2), (4), and (5) using the values of γ' and γ'' quoted in [12] for all the cobalt films with different thicknesses studied therein. Figure 6 shows variations in the FE spectral dependence with the permittivity of Sm_2O_3. Both of these parameters are seen to have a significant effect on the FE spectra.

Another possible reason for the discrepancy between the experimental and calculated FE spectra may be the deviation of the shape of the particles from a sphere. Indeed, it was established in [13] that the magnetic moment of similar films prepared under identical conditions lies in the film plane. This is conceivable if the particles are shaped like oblate ellipsoids

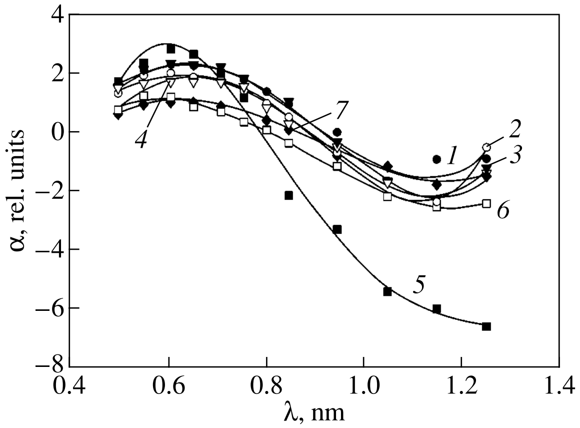


Fig. 5. FE spectral dependences calculated from Eqs. (2)–(5) for $f = 0.6$ and $\varepsilon_0 = 0.4$; ε and γ were taken from [12, Tables 1, 2] for samples with thicknesses of 108, 58, 46, 36, 26, 22, and 13 nm (curves *1*–*7*, respectively).

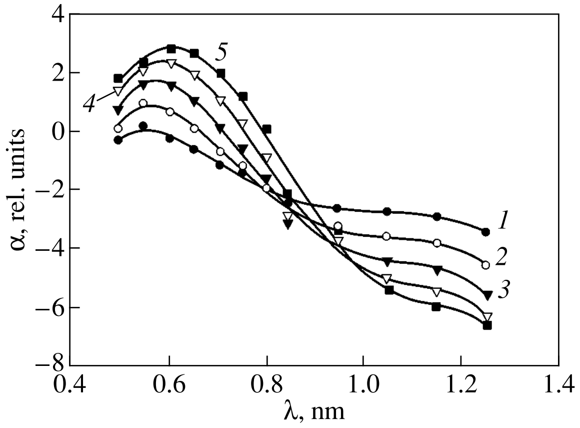


Fig. 6. FE spectral dependences calculated from Eqs. (2)–(5) for ε_0 = 2.0, 2.5, 3.0, 3.5, and 4.0 (curves *1*–*5*, respectively) and $f = 0.6$. The values of ε and γ were taken from [12] for a 26-nm-thick film.

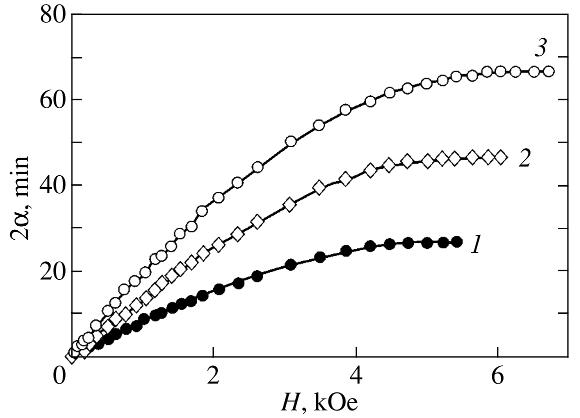


Fig. 7. Field dependences of the FE for (*1*) starting Co–Sm_2O_3 film and (*2*, *3*) after annealing at (*2*) 520 and (*3*) 620 K.

with their long axes lying in the film plane. Figure 7 presents the FE field dependences plotted for the starting sample (curve *1*) and samples annealed at T_{ann} = 520 K (curve *2*) and 620 K (curve *3*). Annealing increases the magnitude of the FE and its sensitivity to high magnetic fields. Note that the field dependence of magnetization for the samples described in [7] changes pattern already at a heat treatment temperature of 573 K. Assuming that the above-mentioned plane anisotropy occurs in the samples and that the size of the Co particles depends only weakly on annealing at the given temperatures [13], one can derive the effective saturation magnetization M_{eff} from the field dependence of the Faraday rotation, $\alpha = f(H_\perp)$, by invoking the expression for the shape anisotropy field $H_S = 4\pi M_{eff}$. The values of the effective magnetization were found to be ~330 G for the starting film and ~440 and ~520 G after annealing for curves *2* and *3*, respectively. Thus, the increase in the FE observed in annealed films can be attributed both to the growth of the effective magnetization and to a change in the optical and magnetooptical parameters of the material of the particles and/or of the matrix.

Thus, the magnetooptical properties of Co–Sm_2O_3 films with granular composite morphology, both starting and heat treated, differ strongly from those of the films of an alloy of the same elements but in the metallic state. Analogous differences were observed for these films earlier [7] in a study of the magnetic and electrical properties: the coercive force in composite films is substantially lower and the electrical resistivity is considerably higher than those for uniform metallic films prepared from the same components. One should add here the large magnitude of the FE in the short-wavelength spectral region. The possibility of fabricating nanogranular films with a large volume fraction of the magnetic phase (~60%) is presently attracting considerable attention for potential application.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 04-02-16099-a) and the Presidium of the RAS (project no. 9.1-2004).

REFERENCES

1. F. Parent, Phys. Rev. B **55** (6), 3683 (1997).
2. D. E. Lood, J. Appl. Phys. **38** (13), 5087 (1967).
3. P. H. Lissberger and P. W. Saunders, Thin Solid Films **34** (2), 333 (1976).
4. P. M. Hui and D. Stroud, Appl. Phys. Lett. **50**, 950 (1987).
5. Yu. A. Dynnik, I. S. Édel'man, T. P. Morozova, P. D. Kim, I. A. Turpanov, and A. Ya. Beten'kova, Zh. Nauchn. Prikl. Fotogr. **43** (5), 18 (1998).
6. E. Gan'shina, A. Granovsky, B. Dieny, M. Kumaritova, and A. Yurasov, Physica B **299** (3–4), 260 (2001).
7. G. I. Frolov, V. S. Zhigalov, S. M. Zharkov, A. I. Pol'skiĭ, and V. V. Kirgizov, Fiz. Tverd. Tela (St. Petersburg) **45**

(12), 2198 (2003) [Phys. Solid State **45** (12), 2303 (2003)].

8. K. H. Clemens and J. Jaumann, Z. Phys. **173**, 135 (1963).
9. I. S. Édel'man, T. P. Morozova, V. N. Zabluda, P. D. Kim, I. A. Turpanov, A. Ya. Beten'kova, and Yu. A. Dynnik, Pis'ma Zh. Éksp. Teor. Fiz. **63** (4), 256 (1996) [JETP Lett. **63**, 273 (1996)].
10. T. K. Xia, P. M. Hui, and D. Stroud, J. Appl. Phys. **67**, 2736 (1990).
11. P. B. Jonson and R. W. Christy, Phys. Rev. B **9**, 5056 (1974).
12. R. Carey and B. J. Thomas, Thin Solid Films **67**, L35 (1980).
13. V. S. Zhigalov, G. I. Frolov, and L. I. Kveglis, Fiz. Tverd. Tela (St. Petersburg) **40** (11), 2074 (1998) [Phys. Solid State **40** (11), 1878 (1998)].

Translated by G. Skrebtsov

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1134–1142. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1097–1104.
Original Russian Text Copyright © 2005 by Gufan.

MAGNETISM AND FERROELECTRICITY

Theory of Nonstoichiometric Ordering of Pb-Containing Relaxors with Perovskite Structure

A. Yu. Gufan

Research Institute of Physics, Rostov State University, pr. Stachki 194, Rostov-on-Don, 344090 Russia
e-mail: Gufan_gufan@mail.ru

Received May 11, 2004

Abstract—Conditions that are imposed on the interatomic interaction and under which a certain ordered state of $AB'_xB''_{1-x}O_3$ perovskites arises are determined with allowance for effectively pairwise interactions and the configuration entropy of interchanges of B' and B'' ions. It is shown that, for the interaction potential $u(R) = u_0/R^6$, the highest temperature of ordering of the 1 : 2 type (T_{ord} (1 : 2)) corresponds to a structure observed at $A \equiv$ Ba. The highest temperature of ordering of the 1 : 1 type (T_{ord} (1 : 1)) corresponds to a structure that typically occurs in the case of $A \equiv$ Pb. Within the approximation used, it is found that T_{ord} (1 : 1) > T_{ord} (1 : 2) for all compositions and that the 1 : 1-ordered phase is most stable. For models with $u(R) = u_0/R^n$ (n = 1–6) including the interaction in the first m coordination shells (m = 3, 6, 8, 11), it is shown that the ground state of $AB'_xB''_{1-x}O_3$ corresponds either to a decomposed solid solution or to an ordered state similar to that observed in $PbMg_{1/3}Nb_{2/5}O_3$. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Ternary oxides $AB'_{1/3}B''_{2/3}O_3$ with perovskite structure, due to their unusual properties, attract attention as a basis for developing a new generation of active materials for piezoelectric devices (lead-containing magnesium niobate and magnesium tantalate), microelectronic devices (barium zirconate tantalate), reading devices in computer engineering (lanthanum calcium manganate), etc. From the viewpoint of fundamental physics, an unexpected feature of ternary oxides $AB'_{1/3}B''_{2/3}O_3$ is the ordered distribution of ions at sites B in the perovskite structure, which arises as the temperature decreases. In itself, ion ordering over sites B at low temperatures is not surprising. According to the Nernst theorem [1], the entropy decreases and causes an ordered arrangement of B' and B'' ions. However, in the theory of solid-solution ordering [2–5], it is assumed that ordering should correspond to a chemical composition of the initial material. In the case of $AB'_{1/3}B''_{2/3}O_3$ compositions, this means that the perovskite cubic phase should be ordered with cell multiplication, which allows 1 : 2 ordering. Hence, the number of unit cells of the unordered phase forming one primitive cell in the ordered state should be a multiple of three. For example, this ordering is observed in Ba-based complex oxides [6–10]. However, it turns out that, energetically, the 1 : 2 ordering in complex oxides $AB'_{1/3}B''_{2/3}O_3$ differs only slightly from the nonstoichiometric 1 : 1 ordered state [6–14]. Small additives (2–3%) of a “third” element distributed over the perovskite B sublattice (sites 1(a) in space group O_h^1) change the structure of the ordered phase and replace stoichiometric 1 : 2 ordering by nonstoichiometric 1 : 1 ordering [10, 11].

Furthermore, ordering of ternary oxides $PbB'_{1/3}B''_{2/3}O_3$ is always of the 1 : 1 type [6–14].

2. STATEMENT OF THE PROBLEM

There can be several types of ordered structures with ordering of the 1 : 2 or 1 : 1 type [1, 15]. The three simplest types of possible stoichiometric 1 : 2 ordering of $B'B''$ ions in $AB'_{1/3}B''_{2/3}O_3$ compounds are shown in Figs. 1a–1c. Figures 2a–2c show three ordered structures of the 1 : 1 type that (i) satisfy the Lifshitz condition [1, 15] and (ii) are predicted as possible stable structures [16] within the theory based on the Landau potential [17] in the Gorsky–Bragg–Williams (GBW) approximation [2–5, 17].

Figure 1a corresponds to the case where the equilibrium value of the order parameter in the ordered phase is defined by two rays of the eight-rayed star of the wave vector $\mathbf{k}_9^{(1,2)} \equiv \mathbf{k}_\lambda^{(1,2)} \equiv \pm(\mathbf{b}_1 + \mathbf{b}_2 + \mathbf{b}_3)/3$. Here, the subscript indicates the star of vector $\mathbf{k}$; the numerical and Greek subscripts correspond to the Kovalev [18] and the Bouckaert–Smoluchowski–Wigner [19] notations, respectively. The superscript is the number of a ray of the star according to [18]. Figures 1b and 1c show the distribution of B' and B'' cations in the case where translation symmetry violation is described by rays of the stars of $\mathbf{k}_\Sigma$ and $\mathbf{k}_\Delta$: $\mathbf{k}_\Sigma^{(1,2)} \equiv \pm(\mathbf{b}_1 + \mathbf{b}_2)/3$ and

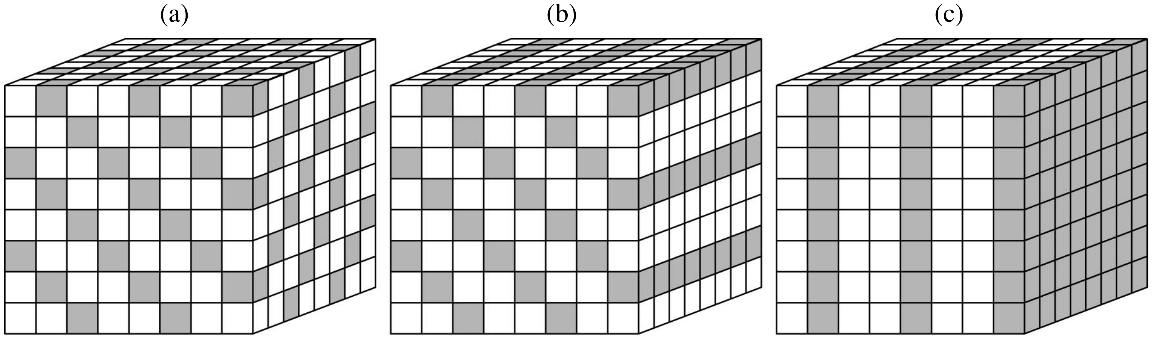


Fig. 1. Three structures of $AB'_{1/3}B''_{2/3}O_3$ compounds with a 1 : 2 ordered arrangement of B' and B'' ions. Cells filled with B' ions are shaded.

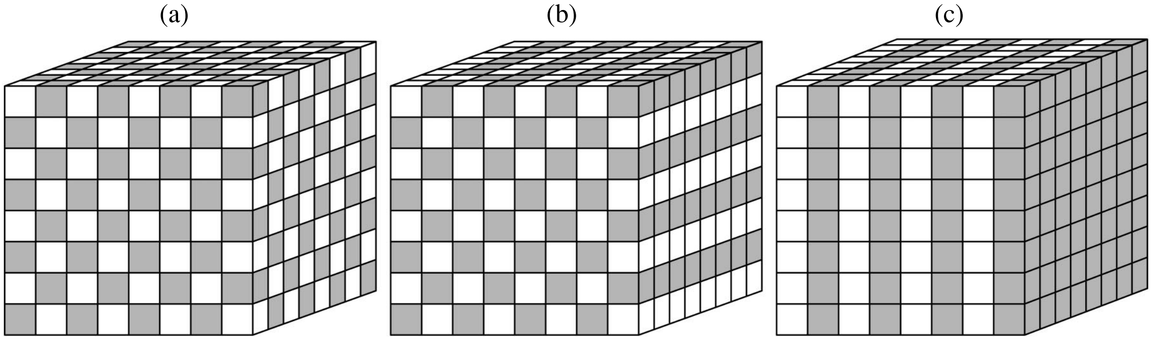


Fig. 2. Same as in Fig. 1, but for 1 : 1 ordering.

$\mathbf{k}_{\Delta}^{(1,2)} \equiv \mathbf{b}_1/3$. There are many types of structures corresponding to 1 : 2 ordering of a quasi-binary solid solution [15]. However, they are either unlikely from the viewpoint of the general Landau theory [1, 15] or are unstable at low temperatures [20]. In what follows, we restrict our consideration to competition between the structures shown in Figs. 1 and 2.

At this point, several questions arise. Why does the structure shown in Fig. 1a arise in all the ternary Ba-based oxides $AB'_{1/3}B''_{2/3}O_3$ ($A \equiv$ Ba)? This obvious question should be complemented with other ones. Why is nonstoichiometric 1 : 1 ordering shown in Fig. 2a observed in lead-containing ternary oxides $PbB'_{1/3}B''_{2/3}O_3$? (Figures 2b, 2c show two more types of the simplest perovskite structures with 1 : 1 ordering, which are also allowed by the Landau theory of phase transitions.) Why is 1 : 2 ordering in a ternary oxide so unstable with respect to the addition of a small amount of one more (fourth) cation and why does it switch to an ordered state designated as 1 : 1 (see Fig. 2a)?

All these questions should be answered using microscopic theory, which will relate the effective interaction potentials of cations in complex oxides to the ordering types observed in these compounds. In our opinion, the problem is to determine a criterion for selecting (within a simplified ordering model) an analytical form of interatomic interaction potentials that stabilize a given ordered structure. In this paper, the problem is solved for the perovskite structure within a model considering only effectively pairwise interactions and only the configuration entropy.

3. REVIEW OF THEORIES OF THE GROUND STATE OF COMPLEX PEROVSKITES

Several theoretical papers have been published [7–14] in which the objective was to explain the nonstoichiometric ordering stability in $PbB'_{1/3}B''_{2/3}O_3$ compounds. In contrast to our statement of the problem, the studies in [7–14] were mainly devoted to calculating the energies that correspond to stable ordered states of specific compositions by using various models of interatomic interactions [7–9, 12, 14]. In this case, to provide reliable results, we considered a large number of competing ordered structures selected on a rather arbitrary basis [9, 14].

In [9], the energy of the ground state of a structure was calculated using the orthogonalized-plane-wave method within the pseudopotential approximation. In [8], the Vienna *ab initio* simulated program and the one-electron approximation were used for the same purposes to characterize the state of atoms. The size and shape of ordered phase cells were chosen randomly. According to [8, 9], the lowest energy corresponds to an ordered structure having a tetragonal unit cell with parameters $a_T = a_C\sqrt{2}$, $b_T = a_C\sqrt{2}$, and $c = 3a_C$. The energy corresponding to this stable structure differs by 1% from the energy of a rhombohedral ordered structure having a unit cell with parameters $a_R = a_C\sqrt{2}$, $b_R = a_C\sqrt{2}$, and $c_R = a_C\sqrt{3}$. Thus, the structures that could exist at low temperatures according to [8, 9] are not among the ordered structures determined from x-ray diffraction patterns [6–14].

In [7], the Monte Carlo method in the cluster approximation was applied to calculate the ground state of a binary (with a component content ratio of 1 : 2) solid solution whose atoms are arranged at sites of a simple cubic lattice $30 \times 30 \times 30$ in size. In those calculations, only pair interactions between ions arranged at the vertices of cubic unit cells and three-particle interactions of the nearest neighbors arranged along a straight line were taken into account. According to [7], the inclusion of the latter interactions is necessary and sufficient to stabilize the 1 : 2 ordered state. In this case, it remains unclear how the conclusions made in [7] will change if we include pair interactions or any other three-particle interactions more completely. It is also unclear how a change in the analytical form of the pair and three-particle interactions used in the calculations in [7] will alter the results.

The authors of [12] included the pair Coulomb interaction between cations screened due to the effective permittivity, which should be attributed to the primitive cell. On this basis, the energy of $6 \times 6 \times 6$ clusters was calculated using the Monte Carlo method [12]. Based on the calculated energies for many types of structures, the authors of [12] concluded that only stoichiometric ordered structures correspond to the minimum total energy of a cluster of heterovalent ions with pure Coulomb interactions.

In [14], the interatomic interaction energy was considered most completely, including the Coulomb, dipole, and quadrupole interactions of ions, whose state was determined in the one-electron approximation. The stabilization energy of nine types of ordered structures (seven with $x = 1/2$ and two with $x = 1/3$) was estimated. According to [14], the lowest energy of the ordered state corresponds to a structure with quadrupling of one of the periods along the fourfold axis of the perovskite cubic cell. The difference in energy between this ordered state and the next applicant for the ground state (rhombohedral structure with $a = b = a_C\sqrt{2}$ and $c = a_C\sqrt{3}$) is 0.4% [14].

To our knowledge, none of the known studies provides answers the questions posed in Section 2. This fact suggests that the energy $V(R)$ of the effectively pairwise interactions of ions accepted in the models used in [7–9, 12, 14] seems to differ from the actual energy in the crystal. The need to elucidate the restrictions imposed on the form of the $V(R)$ interaction is one of the arguments in favor of the statement of the problem solved in this study.

4. PHENOMENOLOGICAL MODEL OF TERNARY Ba-CONTAINING OXIDES WITH PEROVSKITE STRUCTURE

Let us characterize the ideal perovskite structure by a reduced cubic cell containing a single formula unit ABO_3. The symmetry of the corresponding virtual crystal (or praphase) is described by space group O_h^1. We choose a reference point so that A ions in the praphase occupy the regular system of points (RSP) 1(b) with coordinates (1/2, 1/2, 1/2) [21] and that B ions are arranged over the RSP 1(a) with coordinates (0, 0, 0). Stoichiometric 1 : 2 ordering, which is observed in ternary Ba-containing oxides with perovskite structure, requires consideration of an extended primitive cell [1] containing 27 sites. We label these sites by choosing a reference point in one of them:

1. (0, 0, 0); 2. (1, 0, 0); 3. (0, 1, 0); 4. (0, 0, 1);
5. (1, 1, 0); 6. (1, 0, 1); 7. (0, 1, 1); 8. (–1, 0, 0);
9. (0, –1, 0); 10. (0, 0, –1); 11. (–1, –1, 0);
12. (–1, 0, 0); 13. (1, –1, 0); 14. (–1, –1, 1);
15. (–1, –1, –1); 16. (1, –1, –1); 17. (1, 0, –1); (1)
18. (0, –1, 1); 19. (0, 1, –1); 20. (–1, –1, 0);
21. (–1, 0, –1); 22. (0, –1, –1); 23. (1, 1, 1);
24. (–1, 1, 1); 25. (1, –1, 1); 26. (1, 1, –1);
27. (–1, –1, –1).

The parenthesized site coordinates are given in units of length of the unit cell edges of the cubic praphase. Coordinate axes are directed along the fourfold axes of crystal space group O_h^1. In order to adequately describe ordering in terms of a phenomenological model, we have to consider at least up to four-particle effective interactions between ordering cations [22]. However, we restrict our consideration to only effectively pairwise interactions, because it turns out that, if we accept certain empirically determined regularities (which characterize the solutions to the equations of state found within the self-consistent field approximation [15, 16, 23]), some of the posed problems can be solved even in this approximation of the phenomenological theory.

The model of B' and B'' cation ordering over the sites indicated in Eq. (1) contains 27 fitting parameters $P_j(B')$, which are the probabilities of B' ions occupying the jth sublattice characterized by Eq. (1). Let us consider the energies $V_{B'B'}(R)$, $V_{B''B''}(R)$, and $V_{B'B''}(R)$ of effectively pairwise interactions of B' and B'' ions, which depend on the distance R between ions and on the ion type [2–6]. In general, the nonequilibrium internal energy as a function of pair interactions is given by (see [22])

$$\begin{aligned}\frac{27}{N}E = {} & \sum_j \mu_{B'} P_j(B') + \sum_j \mu^{(0)}_{B''} P_j(B'') \\ & + \sum_i \sum_j \{ V_{B'B'}(|R_i - R_j|) P_i(B') P_j(B') \\ & + V_{B''B''}(|R_i - R_j|) P_i(B'') P_j(B'') \\ & + V_{B'B''}(|R_i - R_j|)(P_i(B') P_j(B'') + P_j(B') P_i(B''))\},\end{aligned} \qquad (2)$$

where N is the total number of sites allowed for B' and B'' atoms in the perovskite structure.

Passing in Eq. (2) to a system of 27 independent variables $P_j = P_j(B')$ and $P_i(B'') = 1 - P_i$, where $i, j = 1$–27, we obtain

$$\frac{27}{N}E = \varepsilon_0 - \mu_1 3\sqrt{3}x + \frac{1}{N}\sum_{i,j=1}^{N} u(|R_i - R_j|) P_i P_j, \qquad (3)$$

where $u(|R_i - R_j|)P_iP_j \equiv V_{AA}(|R_i - R_j|) + V_{BB}(|R_i - R_j|) - 2V_{AB}(|R_i - R_j|)$ are the ordering energies, which are dependent only on the distance between sites i and j in the approximation used. In Eq. (3), we introduced the nonequilibrium collective generalized coordinate x (the Frenkel order parameter [24])

$$x = \frac{1}{N}\sum_{i=1}^{N} P_i \qquad (4)$$

with the aim to analyze not only one-phase but also two-phase ordered states as possible equilibrium states of the quasi-binary solid solution of B' and B'' cations. In a one-phase homogeneous equilibrium state, we have $x = 3\sqrt{3}\,c$.

The third term in Eq. (3) contains 378 types of products P_iP_j. By considering the symmetry of the perovskite crystal structure and introducing new notation for the infinite sums of ordering energies, determined according to the lattice structure, we can formally rewrite the nonequilibrium internal energy (3) in terms of only four phenomenological parameters W_i:

$$\begin{aligned}\frac{27}{N}E = {} & \varepsilon_0 - \mu_1 3\sqrt{3}x \\ & + \Big\{ W_1(P_1P_2 + P_1P_3 + \ldots + P_7P_{16} + \ldots + P_{19}P_{27}) \\ & + W_2(P_1P_5 + P_1P_6 + \ldots + P_{16}P_{19} + \ldots + P_{25}P_{27}) \\ & + W_3(P_1P_{14} + P_1P_{15} + \ldots + P_{12}P_{16} + \ldots + P_{23}P_{27}) \\ & + W_8 \sum_{i=1}^{27} P_i^2 \Big\},\end{aligned} \qquad (5)$$

W_i are the sums of infinite series of ordering energies for various coordination shells,

$$\begin{aligned} W_1 &= u(a) + u(2a) + 4u(\sqrt{10}a) + \ldots, \\ W_2 &= u(\sqrt{2}a) + 2u(\sqrt{5}a) + u(\sqrt{8}a) + \ldots, \\ W_3 &= u(\sqrt{3}a) + 3u(\sqrt{6}a) + 3u(3a) + \ldots, \\ W_8 &= 6u(3a) + \ldots. \end{aligned} \qquad (6)$$

In what follows, the distance between ions i and j is measured in units of the edge length a of the reduced cubic cell of the praphase.

We note that the sum of products P_iP_j in Eq. (5) associated with the phenomenological parameter W_1 contains 81 terms. The sums in Eq. (5) associated with the parameters W_2 and W_3 contain 162 and 108 terms, respectively.

In order to determine the types of ordered states that can be predicted based on Eq. (5), it is convenient to switch from the quasi-local generalized coordinates of the crystal (the probabilities of sublattices 1–27 in Eq. (1) being filled with ions B') to symmetric collective generalized coordinates—the components of the Landau [1, 15] and Frenkel [24] order parameters.

The total number of Landau order parameters is 26. The symmetry of eight of them with respect to translations by the praphase lattice constants is defined by the star of the vector $\mathbf{k}_\lambda^{(1)} \equiv \mathbf{k}_9^{(1)}$ [18]. Similarly, the translation properties of the other twelve and six order parameters are defined by the stars of the vectors $\mathbf{k}_\Sigma^{(1)} \equiv \mathbf{k}_4^{(1)}$ and $\mathbf{k}_\Delta^{(1)} \equiv \mathbf{k}_8^{(1)}$, respectively, at $\mu = 1/3$ [18]. Thus, the probability density ρ of the B' ion distribution over the sublattices described by Eq. (1) can be expanded in terms of four irreducible representations of group O_h^1, one of which is constructed using the basis function in Eq. (4) and belongs to the star of the vector $\mathbf{k}_\Gamma^{(1)} \equiv \mathbf{k}_{12} = 0$ and the other three belong to the stars of the vectors $\mathbf{k}_\lambda$, $\mathbf{k}_\Sigma$, and $\mathbf{k}_\Delta$. We denote the linear combinations of P_i forming 27 basis functions of the irreducible representations of O_h^1 as γ, λ_i ($i = 1$–8), σ_j ($j = 1$–12), and Δ_s ($s = 1$–6). For brevity, we write out only one function from each set, namely, a basis function for the first line of each of the four irreducible representations:

$$\gamma = \left(\sum_{i=1}^{27} P_i\right)\Big/\sqrt{27}, \qquad (7)$$

$$\lambda_1^{(1)} = \Bigg[P_1 + P_{11} + P_{12} + P_{13} + P_{17} + P_{18} + P_{19} + P_{23} + P_{27} - \frac{1}{2}\Bigg(\sum_{k=2}^{10} P_k + P_{14} + P_{15} + P_{16} + P_{20} + P_{21} + P_{22} + P_{24} + P_{25} + P_{26}\Bigg)\Bigg]/[\sqrt{6}/9], \quad (8)$$

$$\sigma_1^{(1)} = \Bigg[P_1 + P_2 + P_8 + P_{14} + P_{15} + P_{18} + P_{19} + P_{25} + P_{26} - \frac{1}{2}\Bigg(\sum_{i=1}^{8} P_i + \sum_{g=9}^{13} P_g + \sum_{l=20}^{24} P_l + P_{16} + P_{17} + P_{27}\Bigg)\Bigg]/[\sqrt{6}/9], \quad (9)$$

$$\Delta_1^{(1)} = \Bigg[P_1 + P_3 + P_4 + P_7 + P_9 + P_{10} + P_{18} + P_{19} + P_{22} - \frac{1}{2}\Bigg(P_2 + P_5 + P_6 + P_8 + \sum_{j=1}^{17} P_j + P_{20} + P_{21} + \sum_{l=23}^{27} P_l\Bigg)\Bigg]/[\sqrt{6}/9]. \quad (10)$$

The superscripts on $\lambda_1^{(1)}$, $\sigma_1^{(1)}$, and $\Delta_1^{(1)}$ indicate that the irreducible representations of O_h^1 appearing in the expansion of ρ are constructed from invariant representations of the space group of the vectors $\mathbf{k}_9^{(1)}$, $\mathbf{k}_4^{(1)}$, and $\mathbf{k}_8$. The subscript on $\lambda_1^{(1)}$, $\sigma_1^{(1)}$, and $\Delta_1^{(1)}$ is the number of the basis function. The other functions appearing in the expansion of ρ, namely, $\lambda_2^{(1)}$ to $\lambda_8^{(1)}$, $\sigma_2^{(1)}$ to $\sigma_{12}^{(1)}$, and $\Delta_2^{(1)}$ to $\Delta_6^{(1)}$, can be derived from Eqs. (7)–(10) using the symmetry operations of group O_h^1 [18] and the notation introduced in Eqs. (1). Using the linear symmetric combinations of P_i defined in Eqs. (7)–(10), we write the quadratic part of the nonequilibrium energy (5) in the form

$$E_2 = \frac{N}{27}\Bigg\{\mu_2(\Gamma^{(1)})^2 + b_{1\lambda}\sum_{i=1}^{8}(\lambda_i^{(1)})^2 + b_{1\Sigma}\sum_{i=1}^{12}(\sigma_i^{(1)})^2 + b_{1\Delta}\sum_{i=1}^{6}(\Delta_i^{(1)})^2\Bigg\}, \quad (11)$$

where

$$\mu_2 = (3W_1 + 6W_2 + 4W_3 + W_8) = 3u(1) + 6u(\sqrt{2}) + 4u(\sqrt{3}) + 3u(2) + 12u(\sqrt{5}) + 12u(\sqrt{6}) + \dots, \quad (12)$$

$$b_{1\lambda} = (-3W_1 + 3W_2 - W_3 + 2W_8)/2 = [-3u(1) + 3u(\sqrt{2}) - u(\sqrt{3}) - 3u(2) + 6u(\sqrt{5}) + \dots]/2, \quad (13)$$

$$b_{1\Sigma} = (-3W_2 + 2W_3 + 2W_8)/2 = [-3u(\sqrt{2}) + 2u(\sqrt{3}) - 6u(\sqrt{5}) + 6u(\sqrt{6}) + \dots]/2, \quad (14)$$

$$b_{1\Delta} = (3W_1 - 4W_3 + 2W_8)/2 = [3u(1) - 4u(\sqrt{3}) + 3u(2) + 12u(\sqrt{6}) + \dots]/2. \quad (15)$$

Expressions (12)–(15) account for the interactions in the first six coordination shells. This approximation is justified. Indeed, a model that includes only pair interactions (provided that the ordering energy sign is independent of the distance between the ions) can be used to describe a material having an additive free energy only if the interactions (i) propagate to a finite distance and (ii) sufficiently rapidly decrease with distance [15]. The term “sufficiently rapidly” implies that $u(R) \geq -\alpha/R^6$ for $\alpha > 0$ and $u(R) \geq -\omega(R)/R^3$ for $|\omega(R)| \leq \omega_0$ at large distances and that the interactions oscillate at finite distances.

We restrict ourselves to the model described by Eqs. (11)–(15) and construct the nonequilibrium Landau potential by considering only the configuration entropy associated with interchanges of B' and B'' ions, which is equivalent to the GBW approximation [2–5]. Let the pair interactions potentials $V_i(R)$ and ordering energies $u(R)$ decrease at large distances as $1/R^6$ (Lennard–Jones potential). For this model, it can be shown that, in the case of $u(1) \leq 0$, the ground state of a quasi-binary solid solution of B' and B'' ions below a certain temperature

$$T = T_B^{(1)} \approx 4.06u(1)c(1-c) \quad (16)$$

corresponds to a two-phase state, which arises as a result of decomposition of the solid solution [25]. The notation T_B is chosen since this temperature can be interpreted as the Burns temperature [26] of decomposition observed in $PbMg_{1/3}Nb_{2/3}O_3$ or $PbMg_{1/3}Ta_{2/3}O_3$ [26].

In the case of $u(1) > 0$, a stable ordered state arises in the quasi-binary solid solution of B' and B'' ions below the temperature

$$T = T_{\text{ord}}\left(\frac{1}{3}\lambda\right) \approx 1.33u(1)c(1-c). \quad (17)$$

This state is described by the equalities

$$\begin{gathered}\Gamma^{(1)} = \sqrt{27}; \quad \sigma_i = 0, \quad i = 1, 2, \ldots, 12; \\ \Delta_j = 0, \quad j = 1, 2, \ldots, 6; \\ \lambda_l = 0, \quad l = 2, \ldots, 8; \quad \lambda_1 \equiv \eta,\end{gathered} \tag{18}$$

where η is the order parameter dependent on T and c. The ordered state described by Eqs. (18) is shown in Fig. 1a.

Other possible types of ordered states predicted by the model that includes only pair interactions (in particular, the structures shown in Figs. 1b, 1c) do not occur, since they can arise from a disordered state only at lower temperatures [in comparison with the state described by Eqs. (18)]. Within the model used, the states that can be stable only at temperatures below $T_{\text{ord}}(1/3\lambda)$ given by Eq. (17) are metastable at all temperatures down to $T = 0$ [23]. Thus, the model considering only effectively pair interactions makes it possible to explain why the type of ordering shown in Fig. 1a arises in all ternary oxides $\mathrm{Ba}B'_{1/3}B''_{2/3}\mathrm{O}_3$ with perovskite structure. However, the existence of other possible states that cannot be predicted on the basis of Eq. (5) or (11) should also be examined.

5. PHENOMENOLOGICAL MODEL OF Pb-CONTAINING TERNARY OXIDES WITH PEROVSKITE STRUCTURE

Thus, we should examine the relative stability of 1 : 2 and 1 : 1 ordered states in the case of an arbitrary pair interaction potential. To this end, the calculations resulting in Eqs. (1)–(17) and carried out for 1 : 2 ordering should be repeated for 1 : 1 ordering. In the case of the latter ordering, we have a $2a \times 2a \times 2a$ extended unit cell. Let us enumerate the eight sites of the RSP $1(a)$ that compose extended cell 1:

$$\begin{gathered}1.\ (0, 0, 0);\ \ 2.\ (a, 0, 0);\ \ 3.\ (0, a, 0);\ \ 4.\ (a, a, 0); \\ 5.\ (0, 0, a);\ \ 6.\ (a, 0, a);\ \ 7.\ (0, a, a);\ \ 8.\ (a, a, a).\end{gathered} \tag{19}$$

Similarly to Eqs. (7)–(10), we construct, based on Eqs. (19), symmetric generalized coordinates that form bases for irreducible representations of group O_h^1 belonging to the stars of the vectors $\mathbf{k}_{12} = \mathbf{k}_\Gamma = 0$, $\mathbf{k}_{13} = \mathbf{k}_R = (\mathbf{b}_1 + \mathbf{b}_2 + \mathbf{b}_3)/2$, $\mathbf{k}_{11}^{(1)} = \mathbf{k}_M^{(1)} = (\mathbf{b}_2 + \mathbf{b}_3)/2$, and $\mathbf{k}_{10}^{(1)} = \mathbf{k}_X^{(1)} = \mathbf{b}_1/2$. Using notation similar to that in Eqs. (7)–(10), we write

$$\begin{aligned}\Gamma_1 &= (P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 + P_8)/\sqrt{8}, \\ R_4 &= (P_1 - P_2 - P_3 + P_4 - P_5 + P_6 + P_7 - P_8)/\sqrt{8}, \\ M_1 &= (P_1 - P_2 - P_3 + P_4 + P_5 - P_6 - P_7 + P_8)/\sqrt{8}, \\ X_1 &= (P_1 + P_2 + P_3 + P_4 - P_5 - P_6 - P_7 - P_8)/\sqrt{8}.\end{aligned} \tag{20}$$

In the case of eight perovskite structure sublattices, the Landau potential describing ordering in the RSP $1(a)$ to within the quadratic terms in the order parameter components is given by

$$\begin{aligned}F \approx \Big\{ & E_0 - \mu_1\Gamma + \mu_2\Gamma^2 + a_{1R}R^2 + a_{1M}(M_1^2 + M_2^2 + M_3^2) \\ & + a_{1X}(X_1^2 + X_2^2 + X_3^2) + T\frac{1}{2c(1-c)}(\Gamma^2 + R^2 \\ & + M_1^2 + M_2^2 + M_3^2 + M_4^2)\Big\} N/8.\end{aligned} \tag{21}$$

The phenomenological parameters μ_2, a_{1R}, a_{1M}, and a_{1X} of Landau potential (21) are related to the ordering energies $u(R)$ as

$$\begin{aligned}\mu_2 &= 3u(1) + 6u(\sqrt{2}) + 4u(\sqrt{3}) \\ &+ 3u(2) + 12u(\sqrt{5}) + 12u(\sqrt{6}) + \ldots,\end{aligned} \tag{22}$$

$$\begin{aligned}a_{1R} &= -3u(1) + 6u(\sqrt{2}) - 4u(\sqrt{3}) \\ &+ 3u(2) - 12u(\sqrt{5}) + 12u(\sqrt{6}) + \ldots,\end{aligned} \tag{23}$$

$$\begin{aligned}a_{1M} &= -u(1) - 2u(\sqrt{2}) + 4u(\sqrt{3}) \\ &+ 3u(2) - 4u(\sqrt{5}) - 4u(\sqrt{6}) + \ldots,\end{aligned} \tag{24}$$

$$\begin{aligned}a_{1X} &= u(1) - 2u(\sqrt{2}) - 4u(\sqrt{3}) \\ &+ 3u(2) + 4u(\sqrt{5}) - 4u(\sqrt{6}) + \ldots.\end{aligned} \tag{25}$$

From comparing Eqs. (22) and (12), it follows that, as it must be for any partition into sublattices and any $u(R) > 0$ dependence, theory predicts the same dependence of the temperature of possible decomposition of the solid solution on $u(R)$. If the Lennard–Jones model is used, then Eq. (16) for T_B remains valid. Using Eqs. (21) and (23), we can easily ascertain that, in the case of $u(1) > 0$, the disordered state of the solid solution loses stability with respect to the ordering shown in Fig. 2a at

$$T_{\text{ord}}(R) \approx 1.48u(1)c(1-c). \tag{26}$$

The loss of stability of the disordered state with respect to the ordering shown in Fig. 2b takes place at lower temperatures, where there already exists a stable state (shown in Fig. 2a) defined by the equalities

$$\begin{gathered}\Gamma = \sqrt{8}c, \quad R = \eta \neq 0, \\ M_1 = M_2 = M_3 = 0, \quad x_1 = x_2 = x_3 = 0.\end{gathered} \tag{27}$$

The loss of stability of the ordered state with respect to the ordering shown in Fig. 2c can occur only at temperatures below T_B given by Eq. (16), when the homogeneous state is absolutely unstable with respect to the formation of a two-phase state.

Thus, if the crystal features 1 : 1 ordering, there can occur only a transfer to the state shown in Fig. 2a or to an inhomogeneous state with the formation of solid solutions of two different compositions. A comparison of $T_{\mathrm{ord}}(1/3\lambda)$ in Eq. (17) and $T_{\mathrm{ord}}(R)$ in Eq. (26) shows that the loss of stability of the disordered state with respect to ordering allowed by the Lifshitz condition [1] occurs at a temperature that is higher than that at which instability arises with respect to ordering (18), characterized by tripling of the primitive cell period. This conclusion is independent of the solid solution composition for the model interaction potential used.

5. DISCUSSION

In the above, we used the approximation of effectively pairwise interactions and considered only the configuration entropy associated with possible interchanges of ordering ions. For numerical estimations of the effect of remote neighbors, we used a potential of the Lennard–Jones type. The theory developed in these approximations predicts the following.

(i) If ordering occurs into a homogeneous state allowed by the Lifshitz condition, then the order arising in the distribution of *B*' and *B*'' ions (or of *A*' and *A*'' ions, due to the features of the perovskite structure) is identical to the order that occurs in all lead-containing relaxors $\mathrm{Pb}Me^{2+}_{1/3}Me^{5+}_{2/3}\mathrm{O}_3$ (Fig. 2a) [3–14].

(ii) If ordering occurs into a homogeneous state with tripling of the perovskite lattice constants, then the stable ordered state will be that shown in Fig. 1a. This ordered state is characteristic of all compositions $\mathrm{Ba}Me^{2+}_{1/3}Me^{5+}_{2/3}\mathrm{O}_3$ [6–14].

(iii) The loss of stability of the disordered state with respect to the formation of the 1 : 2 ordered state (Fig. 1a) occurs at the temperature $T_{\mathrm{ord}}(\lambda)$ at which the disordered state is absolutely unstable. Near the temperature $T = T_{\mathrm{ord}}(\lambda)$, the 1 : 1 ordered state is stable (Fig. 2a), and this state is also stable in comparison with the 1 : 2 ordered state at all temperatures down to $T = 0$.

We note that all the results presented in Sections 2–4 were obtained under the assumption that the ordering energy depends on the distance as $u(R) \sim 1/R^6$ and with inclusion of the interaction in the first six coordination shells. Let us next discuss the possible occurrence of other ordered states in the case where the analytical $u(R)$ dependence varies and consider the influence of the number of coordination shells included in the calculations on the results. To illustrate the effect of the chosen model on the theoretical results, we restrict our consideration to 24 models in which (i) the sign of the ordering energy is independent of distance; (ii) the $u(R) \sim u^0/R^n$ dependence is valid, where $n = 1$–6; and (iii) the interaction extends to only the first three, six, eight, or eleven coordination shells.

First, we note that homogeneous states arise only at $u_0 > 0$. At $u_0 < 0$, only three phenomenological parameters of the Landau potentials in Eqs. (11) and (21) (μ_2, $b_{1\Delta}$, a_{1X}) can be negative, with $\mathrm{sgn}\,\mu_2 = \mathrm{sgn}\,u_0$ at any value of u_0. The sign of b_{1X} depends on both n and m (the numbers of coordination shells included in the calculations). For example, if $u_0 < 0$ and the first three coordination shells are included ($m = 3$), we have b_{1X} ($n = 1, 2, 3$) > 0. In other models (i.e., at $n = 1$–6 and $m = 3, 6, 8, 11$), $b_{1X} < 0$. The parameter $b_{1\Delta}$ likewise depends on n and m. Thus, generally speaking, it could be assumed that (taking into account the dependences of the entropy on the order parameter that are characteristic of the GBW model) there exist models in which $u_0 < 0$ and $u(R) = u_0/R^n$. Within these models, the ordered states shown in Figs. 1c and 2c are stable. However, this conclusion can only be preliminary without studying the dependence of the following two differences on n and m:

$$\begin{aligned}\mu_2 - a_{1X} = {} & 2u(1) + 8u(\sqrt{2}) + 8u(\sqrt{3}) + 8u(\sqrt{5}) \\ & + 16u(\sqrt{6}) + 10u(3) + 16u(\sqrt{10}) + 24u(\sqrt{11}),\end{aligned} \tag{28}$$

$$\begin{aligned}\mu_2 - b_{1\Delta} = {} & 3u(1)/2 + 6u(\sqrt{2}) + 6u(\sqrt{3}) + 3u(2)/2 \\ & + 12u(\sqrt{5}) + 18u(\sqrt{6}) + 6u(2\sqrt{2}) \\ & + 18u(3) + 6u(\sqrt{10}) + 12u(\sqrt{11}) + 16u(2\sqrt{3}).\end{aligned} \tag{29}$$

It is clear that the sign of differences (28) and (29) is always identical to that of u_0. Hence, at $u_0 < 0$, in all models considered above in which $u(R) = u_0/R^n$, the ordered states shown in Figs. 1c and 2c are not stable. On the contrary, these models predict that, in the case of $u_0 < 0$, the state in which the initial quasi-binary solid solution decomposed into two solid solutions with different contents of the components will be stable at low temperatures. At $u_0 > 0$, the phenomenological parameters $b_{1\lambda}$, $b_{1\Sigma}$, a_{1R}, and a_{1M} are negative. The stability of the corresponding ordered states is dictated by the three differences

$$\begin{aligned}b_{1\lambda} - a_{1R} = {} & [3u(1) - 9u(\sqrt{2}) + 7u(\sqrt{3}) - 9u(2) \\ & + 15u(\sqrt{5}) - 27u(\sqrt{6}) - 9u(2\sqrt{2}) + 33u(3) \\ & - 36u(\sqrt{10}) + 30u(\sqrt{11}) - 9u(2\sqrt{3})]/2,\end{aligned} \tag{30}$$

$$\begin{aligned}a_{1R} - a_{1M} = {} & -2u(1) + 8u(\sqrt{2}) - 8u(\sqrt{3}) - 8u(\sqrt{5}) \\ & + 16u(\sqrt{6}) - 10u(3) + 16u(\sqrt{10}) - 24u(\sqrt{2}),\end{aligned} \tag{31}$$

$$\begin{aligned}b_{1\lambda} - b_{1\Sigma} = {} & [-3u(1) + 6u(\sqrt{2}) - 3u(\sqrt{3}) - 3u(2) \\ & + 12u(\sqrt{5}) - 9u(\sqrt{6}) + 6u(2\sqrt{2}) - 9u(3) \\ & - 12u(\sqrt{10}) + 12u(\sqrt{11}) - 3u(2\sqrt{3})]/2.\end{aligned} \tag{32}$$

It follows from Eq. (30) that, for the assumed dependence of the ordering energy on the distance between ions, the stable ordered state is that shown in Fig. 1a rather than the state shown in Fig. 2a in 5 of the 24 models under discussion, namely, in the models with $n \leq 3$ and $m = 3$ and in the models with $n \leq 2$ and $m = 8$. At other values of m and n, the ordered state shown in Fig. 2a is stable.

However, the problem of the stable state can be solved only by comparing Eqs. (30)–(32). It turns out that, (I) at $m = 3$ and $n \leq 4$, the structure shown in Fig. 2b is most stable; (II) at $m = 8$ and $n = 1$, the ordered phase shown in Fig. 1b is most stable; and (III) the ordered phase that arises in Ba-based ternary oxides (Fig. 1a) is stable only in the model with $m = 8$ and $n = 2$. When interpreting the experimental data, only those theoretical conclusions have to be taken into account that do not change as the number of coordination shells for which the interatomic interactions are included tends to infinity. As is evident, refinement of the theory by including the interaction with more distant coordination shells shows that results (I)–(III) are nonphysical.

The main result of numerical calculations of the sums in Eqs. (28)–(32) with an accuracy of up to 1% in the case where the difference between the ordering temperatures is determined to within 10% can be formulated as follows. If ordering is due to only pair interactions caused by Van der Waals forces, $u(R) \sim 1/R^6$, then the ground state in oxides $AB'_xB''_{1-x}O_3$ with perovskite structure is an ordered state, the state described by the star of the vector $\mathbf{k}_R$, or a two-phase state.

Certainly, a simple inverse power-law dependence of the ordering energy on the distance between ions is not a realistic assumption. The analytical form of the $u(R)$ dependence is determined within more comprehensive microscopic theories than the theory developed in this paper. The objective of this study was to derive analytical relations defining the phase stability in terms of the dependence on distance of the pair interaction potentials $u(R)$ and to illustrate the influence of some model assumptions [such as the assumption concerning the number of coordination shells included in sums (28)–(32)] on theoretical predictions. Let us illustrate the utility of the results obtained using an example. In pursuing the goal to determine the pair interaction potential for which phase (1) (Fig. 1a) will be stable, we considered several two- and three-parameter dependences of the ordering energies on the distance between ions. Among other factors, we considered the most realistic form

$$u(R) = -1/R^6 + (\mathcal{A}\cos\mathbf{qr})/(\mathbf{qr})^3 + \mathcal{B}/R^9 \qquad (33)$$

at $q \sim 1$ [27, 28]. We succeeded in determining the ranges of $\mathcal{A}$, q, and $\mathcal{B}$ values over which the phases shown in Figs. 2b and 2c are stable. However, phase (18) (Fig. 1a) is not stable for any value of the parameters of ordering energy (33) in the phase diagram in the space of phenomenological parameters. This result agrees with the conclusions made in [7] and can be taken as a basis to explain the narrow region of external conditions (T and c) under which phase (18) is stable. However, final conclusions concerning the role of three-particle interactions in the stabilization of phase (18) (Fig. 1a) require a more detailed analysis than that performed in [7].

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 02-02-17921), the National Natural Science Foundation (China) (project no. 02-02-39016), and the Dynasty Noncommercial Program Foundation (grant of 2003–2004).

REFERENCES

1. L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 3rd ed. (Nauka, Moscow, 1976; Pergamon Press, Oxford, 1980), Parts 1 and 2.
2. M. A. Krivoglaz and A. A. Smirnov, *The Theory of Order–Disorder in Alloys* (Macdonald, London, 1964; Fizmatgiz, Moscow, 1958).
3. A. G. Khachaturyan, *The Theory of Phase Transformations and the Structure of Solids Solutions* (Nauka, Moscow, 1974).
4. N. M. Matveeva and É. V. Kozlov, *Ordered Phases in Metallic Systems* (Nauka, Moscow, 1989) [in Russian].
5. A. I. Gusev and A. A. Rampel’, *Nonstoichiometry, Disorder, and Order in Solids* (URS RAN, Yekaterinburg, 2001) [in Russian].
6. P. K. Davies, J. Tong, and T. Negas, J. Am. Ceram. Soc. **80** (7), 1727 (1977).
7. B. P. Burton, Phys. Rev. B **59** (9), 6087 (1999).
8. B. P. Burton and E. Cockayne, Phys. Rev. B **60** (18), R12542 (1999).
9. B. P. Burton, J. Phys. Chem. Solids **61**, 327 (2000).
10. Y. Yan, S. J. Pennycook, Z. Xu, and D. Viehland, Appl. Phys. Lett. **72** (24), 3145 (1998).
11. P. K. Davies and M. A. Akbas, J. Phys. Chem. Solids **61** (2), 159 (2000).
12. L. Bellaiche and D. Vanderbilt, Phys. Rev. Lett. **81** (6), 1318 (1998).
13. I. Molodetsky and P. K. Davies, J. Eur. Ceram. Soc. **21**, 2587 (2001).
14. V. I. Zinenko and S. N. Safonova, Fiz. Tverd. Tela (St. Petersburg) **46** (7), 1252 (2004) [Phys. Solid State **46** (7), 1291 (2004)].
15. Yu. M. Gufan, *Structural Phase Transitions* (Nauka, Moscow, 1982) [in Russian].
16. G. A. Smolenskiĭ and P. E. Pasynkov, Zh. Éksp. Teor. Fiz. **25** (1), 57 (1953).
17. L. D. Landau, *Complete Works* (Nauka, Moscow, 1969), Vol. 1, Chap. 16, p. 123 [in Russian].
18. O. V. Kovalev, *Irreducible Representations of the Space Groups* (Gordon and Breach, New York, 1965; Ukrai-

nian Academy of Sciences Press, Kiev, 1961), pp. 330–331.

19. L. Bouckaert, R. Smoluchowskiĭ, and E. Wigner, Phys. Rev. **50**, 58 (1936).
20. A. I. Grankina, A. M. Grudskiĭ, and Yu. M. Gufan, Fiz. Tverd. Tela (Leningrad) **29** (11), 3456 (1987) [Sov. Phys. Solid State **29** (11), 1980 (1987)].
21. *International Tables of X-ray Crystallography* (Kynoch, Birmingham, 1965), pp. 330–331.
22. Zh. Zhiao, A. N. Sadkov, Yu. V. Prus, and A. Yu. Gufan, Izv. Ross. Akad. Nauk, Ser. Fiz. **68** (5), 642 (2004).
23. J. Smart, *Effective Field Theories of Magnetism* (IBM Watson Research Center, New York, 1966; Mir, Moscow, 1968).
24. Ya. I. Frenkel, *Theory of Metals* (GIFML, Moscow, 1956) [in Russian].
25. A. Yu. Gufan, Izv. Ross. Akad. Nauk, Ser. Fiz. **5**, 651 (2004).
26. J. Chen, I.- Wei, Li Ping, and Wang Ying, J. Phys. Chem. Solids **57** (10), 1525 (1996).
27. V. Heine, M. L. Cohen, and D. Weaire, *Solid State Physics*, Ed. by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1970).
28. K. I. Portnoĭ, V. I. Bogdanov, and D. L. Fuks, *Calculation of Phase Interaction and Stability* (Metallurgiya, Moscow, 1981) [in Russian].

Translated by A. Kazantsev

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1143–1147. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1105–1109.
Original Russian Text Copyright © 2005 by Kamzina, Raevskiĭ.

MAGNETISM AND FERROELECTRICITY

Electric Field–Induced Birefringence in $Pb_{0.94}Ba_{0.06}Sc_{0.5}Nb_{0.5}O_3$ (PBSN–6) Solid-Solution Single Crystals

L. S. Kamzina* and I. P. Raevskiĭ**

*Ioffe Physicotechnical Institute, Russian Academy of Sciences, Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia
**Institute of Physics, Rostov State University, pr. Stachki 194, Rostov-on-Don, 344090 Russia

Received June 24, 2004

Abstract—The temperature dependence of birefringence induced by an electric field in PBSN–6 relaxors was studied. It was shown that the birefringence measured under different conditions of applying an electric field reveals a number of features not characteristic of the behavior of birefringence in conventional relaxors, more specifically, additional anomalies observed under sample cooling and heating in an electric field, a decrease in the thermal hysteresis width with an increase in the electric field strength, and the absence of birefringence saturation (impossibility to reach single-domain state of the sample during the experiment duration) at low temperatures in fields of up to 3 kV/cm. The results obtained are attributed to inhomogeneity of the cubic paraelectric relaxor phase caused by the coexistence of various polar regions having different dimensions and properties. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

In recent years, intense interest has been focused on compounds that exhibit, in addition to relaxor properties, characteristics typical of normal ferroelectrics. Among these materials are $PbSc_{1/2}Nb_{1/2}O_3$ (PSN), $PbSc_{1/2}Ta_{1/2}O_3$ (PST) with different extents of ion order [1, 2], and a number of PLZT-type ceramics [3, 4]. This interest stems from both purely scientific grounds and the application potential of these materials in electrostriction actuators, acoustic sensors, and pyroelectric transducers.

Stable relaxor behavior of the kind observed in the classical relaxor lead magnesium niobate (PMN) has yet to be successfully demonstrated in PSN and PST compounds even under conditions of complete disorder among the Sc and Nb (Ta) ions. Indeed, even in practically disordered compounds, a spontaneous phase transition (SPT) from the relaxor (microdomain) to a macroscopic-domain ferroelectric state occurs below the temperature at which the maximum in permittivity is observed [5]. To reach a stable relaxor state in these compounds, one usually creates additional disorder on the lead sublattice by substituting barium for the lead ions [6, 7].

It was shown in [6, 7] that an increase in the barium content to 6 mol % suppresses the SPT in these compounds and the relaxor state becomes frozen. At temperatures 40–50°C below the maximum in ε, the temperature dependences of ε and optical transmission in a zero electric field no longer have any anomalies [8] accompanying the SPT, but the macroscopic-domain ferroelectric phase in these compounds can be induced by comparatively weak electric fields. The dielectric and optical studies reported in [8, 9] lead to the conclusion that this Ba ion concentration range is actually a morphotropic region in which the cubic relaxor and rhombohedral macroscopic-domain ferroelectric phases coexist.

Phase coexistence in these compounds imparted to them a number of features that are characteristic neither of pure relaxor behavior nor of a normal ferroelectric. Indeed, in a zero electric field at temperatures below the maximum of permittivity, the cubic nonpolar matrix of the crystal revealed the presence of spontaneously polarized regions of the ferroelectric phase with dimensions comparable to a wavelength of light (greater than ~10^4 Å), with the size and number of these nanoregions growing with decreasing temperature [9]. This distinguishes the relaxor state of PBSN–6 compounds from the classical relaxor state, in which nanoregions do not grow noticeably in size and birefringence does not set in down to liquid-nitrogen temperature. Because the volume fraction of the rhombohedral phase in PBSN–6 is rather small (~30% at –20°C), no phase transition to the macroscopic-domain state occurred throughout the entire crystal and the birefringence (Δn) varied with decreasing temperature not in a jump but rather monotonically as the ferroelectric phase grew in volume [9].

Another feature of the PBSN–6 compounds is the fact that the paraelectric relaxor phase contains not only spontaneously polarized regions that can undergo SPT but also nonpolar regions in which the critical field needed to produce the ferroelectric phase is very low, and even weak internal electric fields are capable of driving the ferroelectric phase transition in these regions under crystal cooling. As a result, PBSN–6 compounds are characterized by several temperatures at which the physical properties exhibit anomalies, namely, the frequency-dependent temperature of the

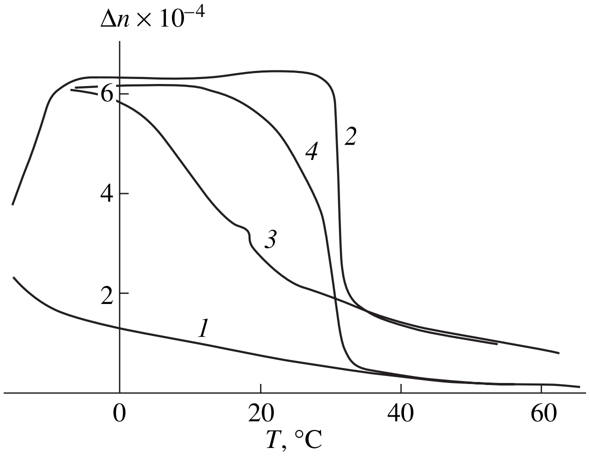


Fig. 1. Temperature dependences of the birefringence Δ*n* measured in a PBSN–6 crystal in different regimes of application of an electric field $E = 1$ kV/cm: (*1*) ZFC, (*2*) FHaZFC, (*3*) FC, and (*4*) ZFHaFC.

maximum in ε ($T_{\max} \cong 48$–50°C); the Vogel–Fulcher temperature $T_f \sim 20$–23°C, which is identified as the point of static freeze-out of electric dipoles and the transition to the dipole glass state; $T_{\mathrm{ds}} \sim 22$–25°C, corresponding to the depolarization temperature of regions that can undergo an SPT; and the depolarization temperature $T_{\mathrm{dr}} \sim 30$–32°C of the nanoregions in which cooling in an electric field induces the ferroelectric phase.

To better understand the nature of the processes occurring in the vicinity of these characteristic temperatures, we performed a study (reported in this paper) on the temperature dependence of the electric field–induced birefringence.

The few experimental studies that are available, with the exception of [8], have dealt primarily with the dielectric properties of PBSN. The observed anomalies were, however, not always clearly pronounced, which posed certain difficulties in their detection. Optical methods for studying crystals (optical transmission (OT), small-angle light scattering (SAS), birefringence Δ*n*) are more sensitive and suitable, particularly for investigating the processes connected with variations in the size of inhomogeneities. If a phase transition brings about changes in the size of optical inhomogeneities, these changes will affect not only light scattering and optical transmission but also the birefringence. Scattering depends on the ratio of the size of the scatterer (*a*) to the wavelength of light (λ). For small values of this ratio, the scattered intensity is small; therefore, the sample will be practically transparent and Δ*n* in the cubic phase will be zero. Optical methods have been successfully employed to study phase transitions in relaxors. The existence of polar nanoregions in a cubic nonpolar matrix is one of the reasons for the relaxor behavior. If the size of these nanoregions varies insignificantly at a phase transition and remains less than λ, Δ*n* should not undergo any changes.

2. GROWTH OF SINGLE CRYSTALS AND EXPERIMENTAL TECHNIQUES

PBSN–6 single crystals were grown by spontaneous crystallization. The growth technique, as well as the x-ray diffraction and dielectric studies of these crystals, is described in [6, 7]. The grown crystals were homogeneous in composition and were shaped as plates with {100} faces. To preclude the generation of additional stresses, all measurements were conducted on samples not subjected to mechanical processing. A dc electric field was applied in the [100] direction, and the light was propagated along [001]. Various electric-field application regimes were employed: zero field cooling (ZFC), zero field heating (ZFH), field heating after ZFC (FHaZFC), field cooling (FC), field heating after field cooling (FHaFC), and zero field heating after field cooling (ZFHaFC). Following each field application and prior to each new measurement, samples were heated at a temperature of 150°C for 0.5 h. To obtain reproducible results and to preclude the dielectric ageing effect, which is strongly manifest in PBSN–6 [7], dielectric and optical measurements were carried out directly after sample annealing. The rate of sample temperature variation could be varied from 2 to 10°C/min.

A He–Ne laser was used for optical measurements. The sample birefringence Δ*n* was derived from the relation $I = I_0 \sin^2 \pi \Delta n d / \lambda$, where I_0 is the intensity of incident light, *I* is the light intensity transmitted through a sample placed between two crossed polarizers (with the optical axis of the sample making an angle of 45° with the incident light polarization), λ = 632.8 nm is the wavelength of light, and *d* is the sample thickness.

3. EXPERIMENTAL RESULTS AND DISCUSSION

Figure 1 presents temperature dependences of birefringence for a PBSN–6 crystal measured in different regimes of application of an electric field of 1 kV/cm in the following sequence: ZFC ⟶ FHaZFC ⟶ FC ⟶ ZFHaFC. In contrast to PMN and 0.9PMN–0.1PT crystals, where Δ*n*, in the absence of an electric field, is practically zero down to the lowest temperatures [10], Δ*n* measured in PBSN–6 under ZFC (curve *1*) grows slightly with decreasing temperature due to the existence of SPT regions. When measured in the FHaZFC regime (curve *2*), Δ*n* increases noticeably and passes through a broad maximum (the width of the maximum is ~40°C). The variation in Δ*n* in the course of this measurement depends on the rate of sample temperature variation; namely, the higher the rate of temperature variation, the smaller the region of existence of the induced ferroelectric phase and, hence, the smaller the width of the Δ*n*(*T*) maximum. The difference in the temperature dependence of Δ*n* measured in the ZFC and FHaZFC regimes indicates that FHaZFC is a metastable-state process. However, the temperature dependence of Δ*n*(*T*) measured under ZFHaFC (curve *4*) reveals normal ferroelectric behavior; extrapolation of

Δn to zero yields the depolarization temperature $T_d \approx$ 30–32°C. In purely relaxor compounds, such as PMN and PLZT, the Vogel–Fulcher temperature (freeze-out point) coincides with the temperature T_d at which residual polarization disappears. This is also the temperature at which the Δn curves measured under FH (curve *2*) and FC (curve *3*) diverge. ($T_d \approx$ 30–32°C is the sample depolarization temperature observed in the FH regime.) The fact that the temperature T_d is uniform over the entire sample implies that the destruction of the macroscopic-domain state has become a collective process. Note that the depolarization temperature measured in PBSN–6 compounds whose cubic nonpolar matrix contains spontaneously polarized regions of the ferroelectric phase does not coincide with the Vogel–Fulcher temperature.

Figure 2 shows the behavior of birefringence in different electric fields. Below T_d, a pronounced thermal hysteresis induced by the freeze-out of polar clusters is observed. Note some features of this hysteresis. First, the existence of two types of regions in the cubic nonpolar crystal matrix below the temperature of the maximum in ε gives rise to the appearance of additional anomalies in the birefringence curves obtained under cooling and heating. The anomaly exhibited by the sample under field cooling in the vicinity of 15–19°C (which is dependent on electric field) originates from the phase transition from the relaxor to the ferroelectric state in the regions where the electric field strength is high enough to induce this phase. A further increase in Δn is associated with the transition to the ferroelectric phase in a small part of the SPT regions (curves *1*, *2*). The temperature of the field-induced phase transition under field cooling increases with the electric field strength.

Second, the hysteresis width decreases with increasing electric field (curves *1*–*3*), thus distinguishing these samples from typical relaxors, in which hysteresis increases with increasing electric field. Indeed, the temperature width of the hysteresis in PBSN–6 in a field of ~0.7 kV/cm is ~16°C (curve *1*) and decreases to ~10°C in a field of 1.5 kV/cm (curve *3*). The reason for this behavior of Δn remains unclear. It may be associated either with the coexistence of regions of various kinds in the cubic matrix of the crystal or with the possibility that the first-order phase transition occurring in weak fields transforms into a second-order transition in high fields.

Third, at low temperatures, Δn does not saturate in fields of up to 3 kV/cm (Figs. 2, 3). Saturation is characteristic of single-domain samples. It is conceivable that saturation of Δn could be achieved if the electric field were to be applied in the direction of spontaneous polarization, i.e., along ⟨111⟩. The maximum in the Δn curve appearing at fields in excess of 1.5 kV/cm (curve *3* in Figs. 2, 3) is due to the electrooptical effect induced in part of the sample volume and to the fact that the induced phase difference between the ordinary and

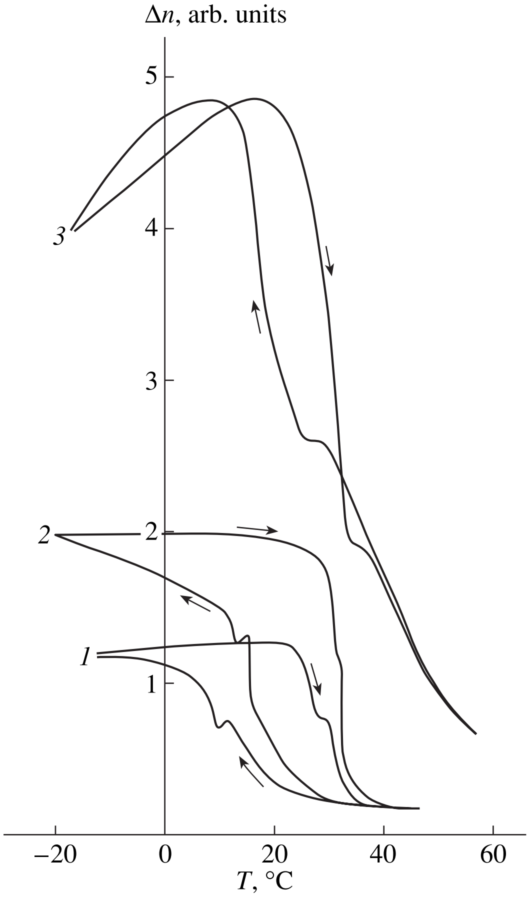


Fig. 2. Temperature dependences of birefringence Δn for a PBSN–6 single crystal measured in heating and cooling regimes in various electric fields E: (*1*) 0.7, (*2*) 1, and (*3*) 1.5 kV/cm.

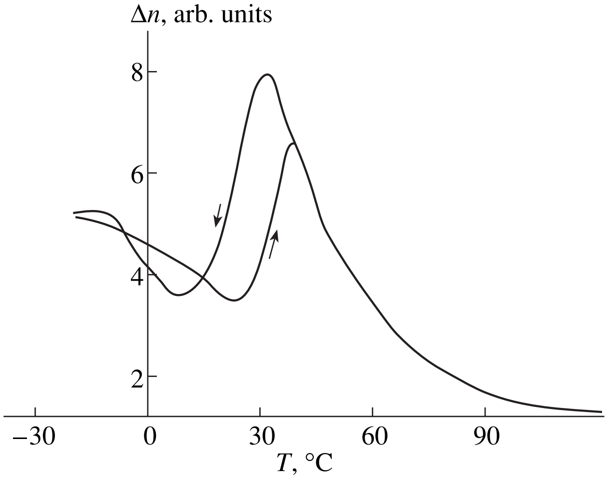


Fig. 3. Temperature dependences of birefringence Δn for a PBSN–6 single crystal measured under heating and cooling in an electric field of 3 kV/cm.

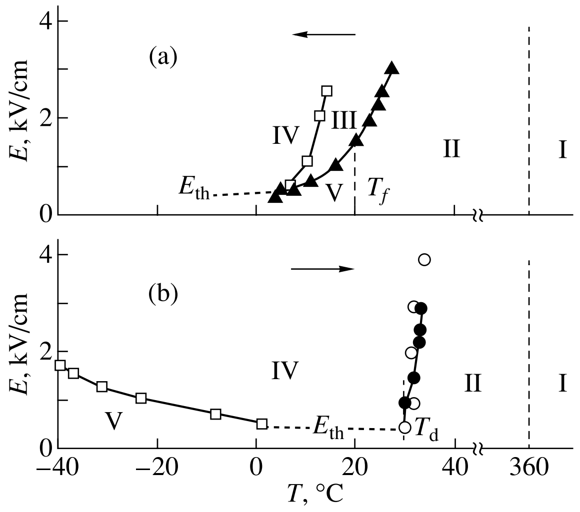


Fig. 4. E–T phase diagrams derived from optical (open symbols) and dielectric (filled symbols) studies under (a) cooling and (b) heating of a PBSN–6 crystal. The rate of sample temperature variation is 2°C/min. I is the paraelectric phase, II is the ergodic relaxor phase, III is the inhomogeneous (macroscopic-domain) ferroelectric phase, IV is the homogeneous macroscopic-domain ferroelectric phase, and V is the nonergodic relaxor phase. T_f is the Vogel–Fulcher temperature taken from [7]. The dashed line at T_f separates the ergodic from the nonergodic relaxor phase.

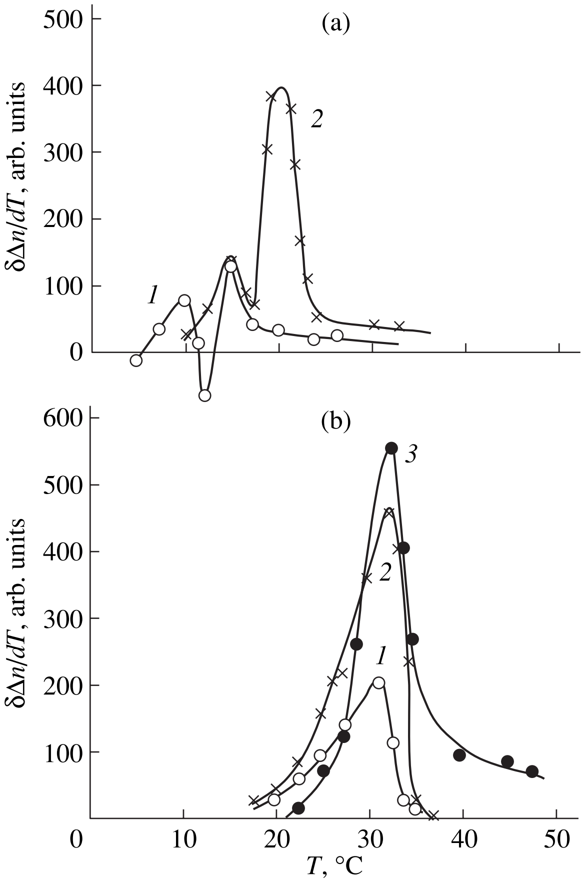


Fig. 5. Temperature dependences of the derivative of the birefringence obtained under (a) cooling and (b) heating in various electric fields E: (*1*) 0.7, (*2*) 1, and (*3*) 1.5 kV/cm.

extraordinary rays has reached π/2. The decrease in Δ*n* with a further decrease in temperature demonstrates inhomogeneity of the nonpolar sample matrix caused by the presence of regions with different properties and by the impossibility of producing a homogeneous single-domain state in a field-cooled sample. Furthermore, long-range order in PBSN–6 compounds sets in fairly slowly in the presence of an electric field, with a long incubation period. By cooling a crystal to lower temperatures and holding it at a low temperature for an extended period of time (more than 20–30 min), a single-domain state and Δ*n* saturation were attained in [9]. The duration of our experiment is considerably shorter than the incubation time, and this is probably why we did not reach Δ*n* saturation. However, in 0.9PMN–0.1PT compounds, in which a spontaneous phase transition and the relaxor state coexist in a zero field, a single-domain homogeneous state has been successfully induced under cooling during the experiment [10]. The evolution of Δ*n* in PBSN–6 compounds in time will be dealt with in a separate publication.

The fact that the cubic nonpolar matrix of the crystal is inhomogeneous is also argued for by the phase diagram obtained by us under crystal cooling and heating and published in [9] (Fig. 4). As is evident from Fig. 4a obtained under sample cooling, an intermediate inhomogeneous macroscopic-domain phase III exists within a certain temperature interval between ergodic relaxor phase II and homogeneous macroscopic-domain ferroelectric phase IV and the region of existence of this phase does not shrink even as the electric field increases. This phase is practically absent from PBSN–6 samples in the field heating regime (Fig. 4b).

The phase diagram constructed in [4] is supported by more comprehensive Δ*n* measurements performed in the present study. The temperature dependences of Δ*n* (Fig. 2) obtained under cooling and heating in various electric fields were used to calculate the derivative $(\delta\Delta n/dT)_{T=T_h} = (\Delta n_{h+1} - \Delta n_{h-1})/(T_{h+1} - T_{h-1})$, where Δn_h and T_h are the *h*th measurements of the birefringence and temperature, respectively. These relations are plotted in Fig. 5. The temperature dependence measured under cooling (Fig. 5a) clearly exhibits two maxima. The maximum seen at a higher temperature can be assigned to the induced phase transition occurring within a larger part of the polar nanoregions and to a transfer of the crystal to the inhomogeneous macroscopic-domain phase. The lower temperature maximum is associated with the transition to the ferroelectric state in regions supporting SPT. As the electric field increases (curve *2* in Fig. 5a), the maxima shift to higher temperatures, but the distance between them remains practically unchanged, which is consistent with the phase diagram in Fig. 4a. The presence of two maxima indicates that the homogeneous macroscopic-domain state cannot be obtained directly from the ergodic relaxor phase in a PBSN–6 sample under field cooling. Figure 5b presents analogous graphs derived by heating a sample in an electric field. Note that, in this case, the sample was held in an electric field at a temperature of ~–20°C for about 10 min, until the birefringence practically reached saturation, after which the sample was field-heated. It is seen that there is one maximum at the sample depolarization temperature T_d. The fact that the depolarization temperature obtained under heating is the same for the entire sample implies that the destruction of the macroscopic-domain state obtained by holding a sample for a long time in an electric field has become a collective process.

4. CONCLUSIONS

Thus, our comprehensive measurements of the birefringence in a PBSN–6 crystal conducted in various electric fields and in different field application regimes have once more demonstrated inhomogeneity of the cubic paraelectric relaxor phase of this crystal. This phase contains not only spontaneously polarized macroscopic-domain SPT regions but also a set of nonpolar nanoregions characterized by different critical fields needed to induce the ferroelectric phase. We have also verified the phase diagram obtained by us earlier for this crystal and refined the conditions under which the macroscopic-domain state can be attained.

ACKNOWLEDGMENTS

The authors are indebted to N.N. Kraĭnik and O.E. Kvyatkovskiĭ for helpful discussions.

This study was supported by a program of the Department of Physical Sciences of the RAS.

REFERENCES

1. N. Setter and L. E. Cross, J. Appl. Phys. **51** (8), 4356 (1980).
2. V. G. Smotrakov, I. P. Raevskiĭ, M. A. Malitskaya, S. M. Zaĭtsev, Yu. M. Popov, and N. A. Strekneva, Izv. Akad. Nauk SSSR, Neorg. Mater. **19** (1), 123 (1983).
3. Q. Tan and D. Viehland, Phys. Rev. B **53** (21), 14103 (1996).
4. O. Bidault, N. Menguy, and B. Dkhil, Ferroelectrics **240** (1–4), 191 (2000).
5. F. Chu, I. M. Reaney, and N. Setter, Ferroelectrics **151** (1–4), 343 (1994).
6. I. P. Raevskiĭ, V. G. Smotrakov, V. V. Eremkin, E. G. Gagarina, and M. A. Malitskaya, Ferroelectrics **247** (1–3), 27 (2000).
7. I. P. Raevskiĭ, V. V. Eremkin, V. G. Smotrakov, E. S. Gagarina, and M. A. Malitskaya, Fiz. Tverd. Tela (St. Petersburg) **42** (1), 154 (2000) [Phys. Solid State **42** (1), 161 (2000)].
8. L. S. Kamzina, I. P. Raevskiĭ, V. V. Eremkin, and V. G. Smotrakov, Fiz. Tverd. Tela (St. Petersburg) **44** (9), 1676 (2002) [Phys. Solid State **44** (9), 1754 (2002)].
9. L. S. Kamzina, I. P. Raevskiĭ, V. V. Eremkin, V. G. Smotrakov, and E. V. Sakhkar, Fiz. Tverd. Tela (St. Petersburg) **45** (6), 1061 (2003) [Phys. Solid State **45** (6), 1112 (2003)].
10. K. Fujishiro, T. Iwase, Y. Uesu, Y. Yamada, B. Dkhil, J.-M. Kiat, S. Mori, and Y. Yamamoto, J. Phys. Soc. Jpn. **69** (7), 2331 (2000).

Translated by G. Skrebtsov

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1148–1155. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1110–1117.
Original Russian Text Copyright © 2005 by Ankudinov, Titkov.

Atomic-Force Microscopy of Polarization Domains in Ferroelectric Films

A. V. Ankudinov and A. N. Titkov

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

Received July 14, 2004

Abstract—Single-crystal (001)-oriented $PbZr_{0.47}Ti_{0.53}O_3$ and polycrystalline (111)-oriented $PbZr_{0.47}Ti_{0.53}O_3$ thin ferroelectric films were studied using contact electrostatic force microscopy. Local electromechanical response measurements permitted study of the polarization vector distribution in natural and intentionally created polarization nanodomains in films. The principal components of an electromechanical response signal encountered in studies of ferroelectric films (the piezoelectric response and the additional capacitive contribution) were isolated and analyzed. The effect of tip–surface contact stiffness on the capacitive contribution to the electromechanical response signal was demonstrated experimentally and in terms of a model. It was shown that more accurate information on the distribution of the polarization vector in ferroelectric films can be gained by monitoring local variations in the tip–surface contact stiffness. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Studies of ferroelectric films by atomic force microscopy (AFM) play an essential role in a number of applications of ferroelectrics [1–3]. Since the radius of curvature of the AFM probe tip is very small (~10–100 nm), one can attain a high electric field in the gap between the tip and the ferroelectric sample ($E \sim 10^7$–10^8 V/m) by applying a rather small potential to the tip (~1 V). Such high fields can be employed to advantage both in probing the local polarization state and in polarization switching of nanosized regions in a film. This unique potential inherent in AFM makes it possible, for instance, to develop ferroelectric storage systems with a superhigh recording density of hundreds of gigabits per square centimeter [4]. The most informative method of diagnostics and domain structure modification in ferroelectric films is contact electrostatic force microscopy (EFM) [5, 6], which allows one to measure the local electromechanical response (EMR) of a sample. All ferroelectric materials are actually piezoelectrics, and the EMR signal turns out to be sensitive to the local inverse piezoelectric effect. This accounts for the method being frequently called piezoresponse force microscopy (PFM) [7]. When used in studies of thin ferroelectric films, the EMR signal is sensitive (with a resolution of better than 10 nm) to variations in the local coercive field [8, 9] and the piezoelectric coefficient [5, 10]. Measurement of these important parameters by EFM meets, however, with difficulties because the EMR signal has an additional capacitive contribution that is not directly connected with the ferroelectric properties of a sample.

It was shown in [11] that the capacitive contribution is proportional to the potential difference between the AFM tip and the bottom electrode supporting the ferroelectric film and is inversely proportional to the stiffness of the tip contacting the sample. Local relief curvature variations encountered by the tip displaced over the surface strongly affect the magnitude of this stiffness. As a result, the capacitive component of the EMR signal from a sample with a rough surface relief should be particularly pronounced. In this case, parts of the relief that differ in curvature should also differ strongly in terms of the EMR response; in other words, the EMR signal within a polarization domain may be spatially nonuniform. The capacitive contribution can be revealed by comparing the dynamic and pulsed voltage dependences of the EMR [11]. This approach, while being useful, is, however, fairly complex and time consuming, which hinders its application for the analysis of spatial variations in the capacitive contribution.

We propose a new, simpler procedure for revealing the capacitive contribution, basically consisting in recording, in parallel to the EMR signal, the variation in the local stiffness of tip contact with the sample surface. In this procedure, separation of the EMR signal components reduces to searching for correlations between the EMR and local stiffness images. The information that can be gained with this approach is illustrated in the particular example of a study of purposefully created and natural polarization domain structures in a (001) $PbZr_{0.47}Ti_{0.53}O_3$ single-crystal film and a polycrystalline $PbZr_{0.47}Ti_{0.53}O_3$ film with (111)-oriented grains.

2. EXPERIMENTAL TECHNIQUE AND SAMPLES

Let us consider a metal-coated AFM tip in contact with a ferroelectric film deposited on a conductive electrode (Fig. 1). If a voltage is applied between the tip and the conductive electrode, the thickness of the film in the contact region should change because of the inverse piezoelectric effect. The response ΔH of a piezoelectric plate to the applied bias U can be written as [12]

$$\Delta H = \mp d_{zz} U, \qquad (1)$$

where d_{zz} is the piezoelectric coefficient. The minus (plus) sign in this expression corresponds to the polarization vector in the film coinciding with (being opposite to) the direction of external bias.

To calculate ΔH under the AFM tip (Fig. 1), we have to solve the problem of contact between the piezoelectric plate and the indenter tip. An analysis of this problem performed in [10, 13, 14] reveals that Eq. (1) is also adequately satisfied in an AFM experiment and that the coefficient d_{zz} is proportional to the local value of the normal polarization vector projection. When the bias applied between the electrodes exceeds the polarization-switching threshold $U(\uparrow/\downarrow)$ or $U(\downarrow/\uparrow)$ in the film region under the tip, ΔH reverses sign in a jump. Polarization switching in a film depends on the original state of polarization and the prehistory of bias variation. For instance, a domain is switched from the original ↓ stable polarization state to another stable state ↑ as the voltage increases from zero to $U(\downarrow/\uparrow)$. Thereafter, as the voltage is reduced, the domain remains in the ↓ state and then is switched to the ↑ state at a voltage $U(\uparrow/\downarrow)$ different from $U(\downarrow/\uparrow)$, a process accounting for the hysteretic behavior in the dependence of ΔH on the dc bias.

In an AFM experiment, an ac (U_{ac}) and a dc (U_{dc}) bias are applied between the tip and the electrode carrying a ferroelectric film. The amplitude of the ac voltage is chosen to be considerably lower than the polarization-switching voltages. Local film thickness variations at the ac bias frequency (piezoresponse) give rise to cantilever vibrations, which are measured optically by recording cantilever deflections (Fig. 1). The distributed capacitive (C_{ps} in Fig. 1) interaction of the tip with the sample also causes cantilever vibrations at the ac bias frequency [15]. Note that the softer the cantilever, the stronger the additional vibrations. Because the term "piezoresponse" does not reflect the AFM tip vibrations adequately, one should preferably use the "electromechanical response" (EMR) instead. The above contributions to the EMR (in practice, the amplitude H_ω of the first harmonic of the EFM signal is usually measured) can be judiciously analyzed using a relation similar to the one proposed in [11]:

$$H_\omega = \left[\frac{1}{k^*}\frac{dC_{ps}}{dz}\left(U_{\mathrm{dc}} + \frac{V^{\uparrow} + V^{\downarrow}}{2}\right) \mp d_{zz}\right] U_{\mathrm{ac}}, \qquad (2)$$

where k^* is the stiffness coefficient of the AFM cantilever with both its ends kept fixed (one is clamped in the holder, and the other is in contact with the film surface) and $(V^{\uparrow} + V^{\downarrow})/2$ is the average value of the contact potential difference between the tip and the surface of the multidomain film. The film polarization directions ↓ and ↑ in the film correspond to the minus and plus signs, respectively, given to the parameter d_{zz} in Eq. (2). The sign becomes reversed at certain polarization-switching potentials $U(\uparrow/\downarrow)$ or $U(\downarrow/\uparrow)$.

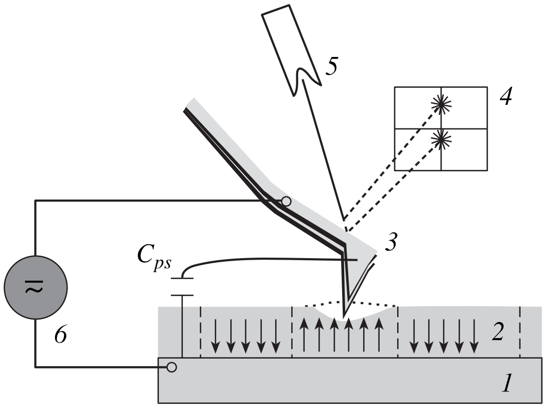


Fig. 1. Schematic diagram of an AFM experiment performed to study the polarization domain structure in ferroelectric films. (*1*) Conductive electrode, (*2*) ferroelectric film, (*3*) metal-coated AFM probe, (*4*) position-sensitive photosensor, (*5*) laser, and (*6*) ac and dc power supply.

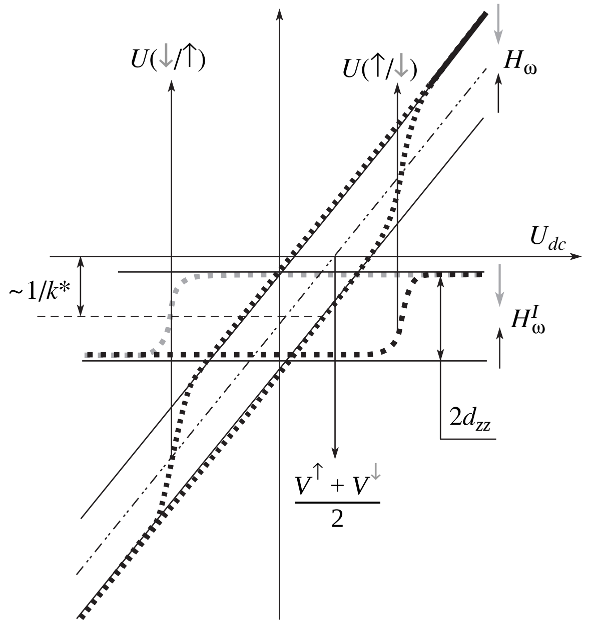


Fig. 2. Model H_ω and H^I_ω dependences given by Eq. (2).

Let us consider the $H_\omega(U_{\mathrm{dc}})$ graph in Fig. 2, which is a plot of Eq. (2) for $U_{\mathrm{ac}} = 1$. Because the polarization in the film switches at certain values of U_{dc} (equal to $U(\uparrow/\downarrow)$ and $U(\downarrow/\uparrow)$), this dependence has a hysteretic pattern. The hysteresis may have other than square shape and be markedly deformed because of the capacitive contribution, which is linear in U_{dc}. This distorted shape is indeed observed in AFM [11] (see also below). In experiment, however, there is always noise, which, combined with the shape distortion, reduces the accuracy of measurement of the piezoelectric coefficient d_{zz} and introduces uncertainty in the measurement of the switching potentials $U(\uparrow/\downarrow)$ and $U(\downarrow/\uparrow)$. This situation can be alleviated by using a short, rigid tip with a stiffness coefficient as high as 50 N/m (large values of k^* in Eq. (2) reduce the contribution of capacitive interaction). To preclude degradation of the resolution through abrasion of the tip, the rigid tip should be wear resistant. If conventional, moderately stiff tips (1–10 N/m) are to be employed, the capacitive contribution will have to be monitored.

This monitoring can be based on the shape of the pulse response $H_\omega^I(U_{dc})$ measured in the following manner [11]. Voltage pulses of different height are applied to the film, and H_ω^I is read off in a zero external field directly after termination of a pulse. The pulses probe the state of the film: in the case where their amplitude exceeds the potential $U(\uparrow/\downarrow)$ or $U(\downarrow/\uparrow)$, the polarization direction switches and H_ω^I jumps to a new value, whereas weaker pulses leave the polarization state and the value of H_ω^I unchanged. As shown in Fig. 2, the H_ω^I signal exhibits a clearly pronounced square hysteresis loop, thus making more accurate determination of the parameters $U(\uparrow/\downarrow)$, $U(\downarrow/\uparrow)$, and d_{zz} possible. Furthermore, the vertical translation of the H_ω^I curve as a whole is a measure of the distributed capacitive contribution (the actual direction of the translation, up or down, is determined by the relationship between the work functions of the tip and sample materials; see below). Thus, measurement of the pulse responses helps one to estimate the capacitive contribution, which can subsequently be taken into account in treating data obtained using standard tips with a moderate stiffness.

In actual practice, this procedure is difficult to apply to the determination of spatial variations in the parasitic capacitive contribution, because it is fairly time consuming to measure the pulse responses at each point of the scanning region. If, however, one fixes U_{dc}, then variations in the capacitive contribution will be dominated by the local value of the stiffness coefficient k^* of the AFM tip in contact with the surface, as can be seen from Eq. (2). This suggests a simpler approach to revealing the parasitic capacitive contribution from variations in the stiffness signal, which can be measured in the force modulation mode [16]. As will be shown below, the idea behind this approach is to analyze correlations and anticorrelations in the EMR and stiffness signals.

Two types of samples were chosen for the study. One type was a ≈37-nm thick (001)-oriented single-crystal $PbZr_{0.47}Ti_{0.53}O_3$ film, which was deposited by high-pressure plasma sputtering on a $SrRuO_3/SrTiO_3$ substrate. The other type was a ≈100-nm-thick polycrystalline $PbZr_{0.47}Ti_{0.53}O_3$ film sol-gel-deposited on a $Pt/TiO_2/SiO_2/Si$ substrate; the grains in the film were crystallographically aligned with the [111] axis normal to the substrate plane. An electrode grid was deposited on the top surface of a film for use in standard capacitive measurements of polarization versus external-field hysteresis loops [12]. Single-crystal films featured a coercive field of $\approx 4 \times 10^7$ V/m and a spontaneous polarization of ≈0.4 C/m^2. For polycrystalline films, these parameters were ≈10^7 V/m and ≈0.45 C/m^2, respectively. The description of the technology of film preparation and the results obtained in a study of the structure and integrated ferroelectric parameters of the samples can be found in [17, 18].

AFM experiments were conducted with a Solver P47 instrument (NT-MDT, Moscow), which made contact EFM mode operation, local-stiffness measurements, and raster lithography possible [19]. Silicon NSG11 tips were used that were coated with a Pt or W_2C conductive layer and had a stiffness of ≈5 N/m and a resonance frequency of about 150 kHz. The pyramid-shaped tip was 15- to 20-μm high and had a typical tip curvature radius of less than 40 nm.

3. EXPERIMENT AND DISCUSSION

Figure 3 illustrates parallel measurements of the voltage dependences of pulsed and dynamic EMR. The measurements are performed in the following way. Voltage pulses are applied, following a template (Figs. 3a, 3b), to the AFM tip scanning a region of the surface ~2 nm^2 in area in the raster lithography mode [19]. The EMR signal is read continuously during the process (Fig. 3d). Figure 3e shows an averaged section of the EMR signal, which reflects both the dynamic (H_ω) and pulsed (H_ω^I) dependences on the applied bias (i.e., the amplitude of the signal during and after application of a pulse, respectively). Figure 3c separately plots the dynamic H_ω response, and Fig. 3f shows the pulsed H_ω^I graph. The experimental curves in Figs. 3c and 3f agree well with the model dependences shown in Fig. 2. The dynamic H_ω course plotted in Fig. 3c has a hysteretic shape distorted strongly by the linear-in-U_{dc} capacitive contribution, which makes determination of the piezoelectric coefficient and switching voltages difficult. The hysteresis in the pulsed response is more distinct, which unambiguously indicates the existence of two polarization states in the film.

The data in Fig. 3f can be used to estimate the coercive field (polarization-switching field) and piezoelectric coefficient of the film. As one state switches to

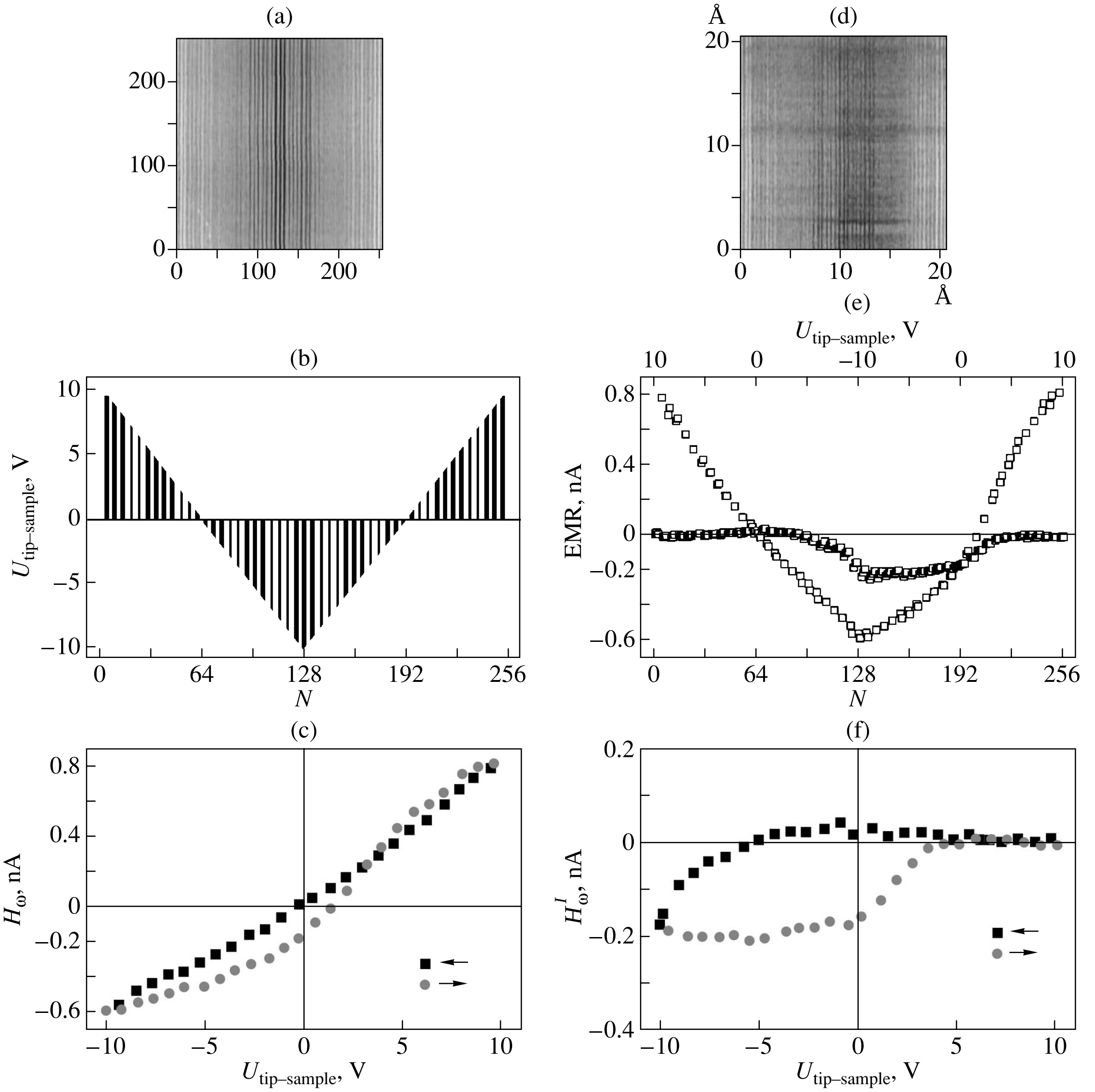


Fig. 3. Parallel measurement of local pulsed and dynamic responses of the EMR signal obtained on a (001)-oriented $PbZr_{0.47}Ti_{0.53}O_3$ single-crystal film. (a) Raster lithography template (256 × 256 pixels); (b) section of the image in panel (a); (c) voltage dependence of dynamic the EMR signal H_ω; (d) EMR signal image recorded by voltage lithography; (e) averaged section of the EMR image; and (f) voltage dependence of the pulsed EMR signal H^I_ω.

another, the pulsed H^I_ω response crosses the central line between the two states at external voltages of approximately +1.5 and –8 V. In a first approximation, the applied potential is distributed uniformly over the thickness (37 nm) of the single-crystal film. Therefore, the local coercive field is about $+4.2 \times 10^7$ and -21.6×10^7 V/m, which coincides with the average macroscopic coercive field of 4×10^7 V/m for the given sample for positive polarity and differs by a factor of more than 5 for negative polarity. The asymmetry between these values may be due to a built-in electric field at the interface between the film and the bottom electrode and to mechanical stresses in the ferroelectric film [20]. Note also that not all of the applied voltage must necessarily be the voltage drop across the ferroelectric film; indeed, this voltage may also include the voltage drop across the dielectric gap between the AFM tip and the film surface [10]. This implies that the above values of the local coercive fields may be considered an upper estimate. According to Eq. (2) and Fig. 2, the height of the H^I_ω hysteresis loop divided by the ac voltage amplitude yields twice the value of the piezoelectric coefficient d_{zz}. To calculate the piezoelectric coefficient, we substitute the amplitude of the ac voltage exciting the piezoresponse (1 V) into Eq. (2). The coefficient to reduce nanoamperes to nanometers as measured by the standard technique recommended for P47H [19] is 40 nm/nA, and the signal amplification is 400. As a result, we obtain $d_{zz} \approx (0.25 \text{ nA} \times 40 \text{ nm/nA})/(1 \text{ V} \times 400 \times 2) = 12.5$ pm/V. This value agrees well with the earlier data obtained by AFM on thin $PbZrTiO_3$ films and published in [9], where the value of d_{zz} was found to be a few picometers per volt.

As already mentioned, the capacitive contribution in a zero field ($U_{dc} = 0$) is determined by the vertical displacement of the pulsed response. Because of the asymmetry with which the top (+0.05 nA) and the lower (–0.2 nA) polarization states manifest themselves in H^I_ω (Fig. 3f), the capacitive contribution at the point where the hysteresis is measured is –0.075 nA, which is more than 30% of the hysteresis amplitude. The variation of the capacitive contribution at a fixed value of U_{dc} (in particular, at zero bias) is dominated by the local stiffness k^*. As follows from Eq. (2), an increase (decrease) in stiffness translates the H^I_ω hysteresis loop as a whole upward (downward); i.e., the EMR signal increases (decreases) irrespective of the polarization state. If one derives the k^* variation from the stiffness signal, then inclusion of the parasitic capacitive contribution can be reduced qualitatively to a search for and analysis of coincidences between the EMR and stiffness signal images.

Figure 4 combines the data illustrating the above approach. The study was conducted on a 37-nm thick (001)-oriented single-crystal $PbZr_{0.47}Ti_{0.53}O_3$ film with an average surface roughness on the level of a few nanometers (see the AFM image in Fig. 4a and the corresponding section in Fig. 4d). Figure 4e presents a hysteretic dependence of the pulsed EMR typical of the given film region, which reveals two polarization states in the film. The arrow-up polarization state corresponds to a signal level $H_\omega \approx +0.25$ nA, while the opposite, arrow-down polarization state is observed on a signal level of ≈–0.7 nA. Complete polarization switching occurs at a pulse amplitude above the threshold level, which is about 5 V, and a further increase in the pulse amplitude has almost no effect on the signal level.

Figure 4b displays an image of the EMR signal H_ω obtained for a surface region in which artificial polarization domains were formed by raster lithography [21]. The natural (single-domain) state of the film corresponds to the bright portions of the image. Against their background, we can see a periodic square (15 × 15) grid of darker regions produced in AFM lithography by applying 1-ms-long voltage pulses with an amplitude of –9 V. The inset to Fig. 4b shows a two-dimensional Fourier transform of the image with distinct reflections, which identify the 200-nm periodicity of the artificial domain structure. Figure 4f shows the cross section of the EMR H_ω image along the bright line crossing the lithographed region. In the range between 0.8 and 3.8 μm, we can see a periodic oscillation of the signal between the upper (≈+0.25 nA) natural state and the lower (≈–0.7 nA) artificial state.

Note that the topographic image in Fig. 4a obtained simultaneously with the data in Fig. 4b does not reveal any periodic structure and the Fourier transform of this image has a diffuse pattern. This argues for the artificial structure at the center of the image in Fig. 4b having predominantly a polarization contrast. A careful comparison of Figs. 4a and 4b brings to light, however, a large number of correlated changes in the EMR signal H_ω and topography. For instance, the arrow in Fig. 4a identifies a bright spot, a 5-nm-high hillock, which coincides with a dark area in Fig. 4b. Such dark areas in Fig. 4b are not related to the lower polarization state and reflect only a change in the additional capacitive component in the EMR signal H_ω.

Figure 4c shows a stiffness signal obtained simultaneously with the topography and the EMR signal. A comparison of the images in Figs. 4b and 4c, as well as of the corresponding sections in Figs. 4f and 4g, suggests that the periodic structure of polarization domains does not manifest itself in any way in the stiffness. The additional capacitive contribution to the EMR signal can be established by the coincidence of contrast in the stiffness and EMR signals. For samples with a relatively smooth surface, topography and stiffness data offer practically the same volume of information on the additional capacitive contribution to EMR. If, however, the surface is characterized by a strongly pronounced relief, the stiffness signal provides more information and makes it possible to gain a more accurate idea of the structure of the polarization domains.

The need for simultaneous measurement of three signals, namely, EMR, topography, and stiffness, is persuasively illustrated by Fig. 5, where the results of a study of a 100-nm-thick (111) $PbZr_{0.47}Ti_{0.53}O_3$ polycrystalline film are presented. The topography signal images (Figs. 5a, 5d) show the film to have a granular structure (with a characteristic height difference of 30 nm). Let us consider grains *1* and *2* surrounded by a contour in Fig. 5a, as well as grain *3* within a contour in Fig. 5d. The EMR contrast within grains *1*, *2*, and *3* varies (Figs. 5b, 5e). This suggests that each grain contains several differently polarized domains. The stiffness signal within grain *3* (Fig. 5f) is nonuniform in contrast and increases or decreases practically within the same nanosized regions as the EMR signal (Fig. 5e). Thus, grain *3* is not multidomain and is uniformly polarized. Within grains *1* and *2*, the stiffness signal has uniform contrast (Fig. 5c). Hence, each of these grains does indeed contain two differently polarized domains.

To clarify the case, we turn back to Eq. (2), according to which the capacitive contribution to the EMR signal (or, in other words, the vertical displacement of the H^I_ω hysteresis loop in Fig. 2) is proportional to $1/k^*$.

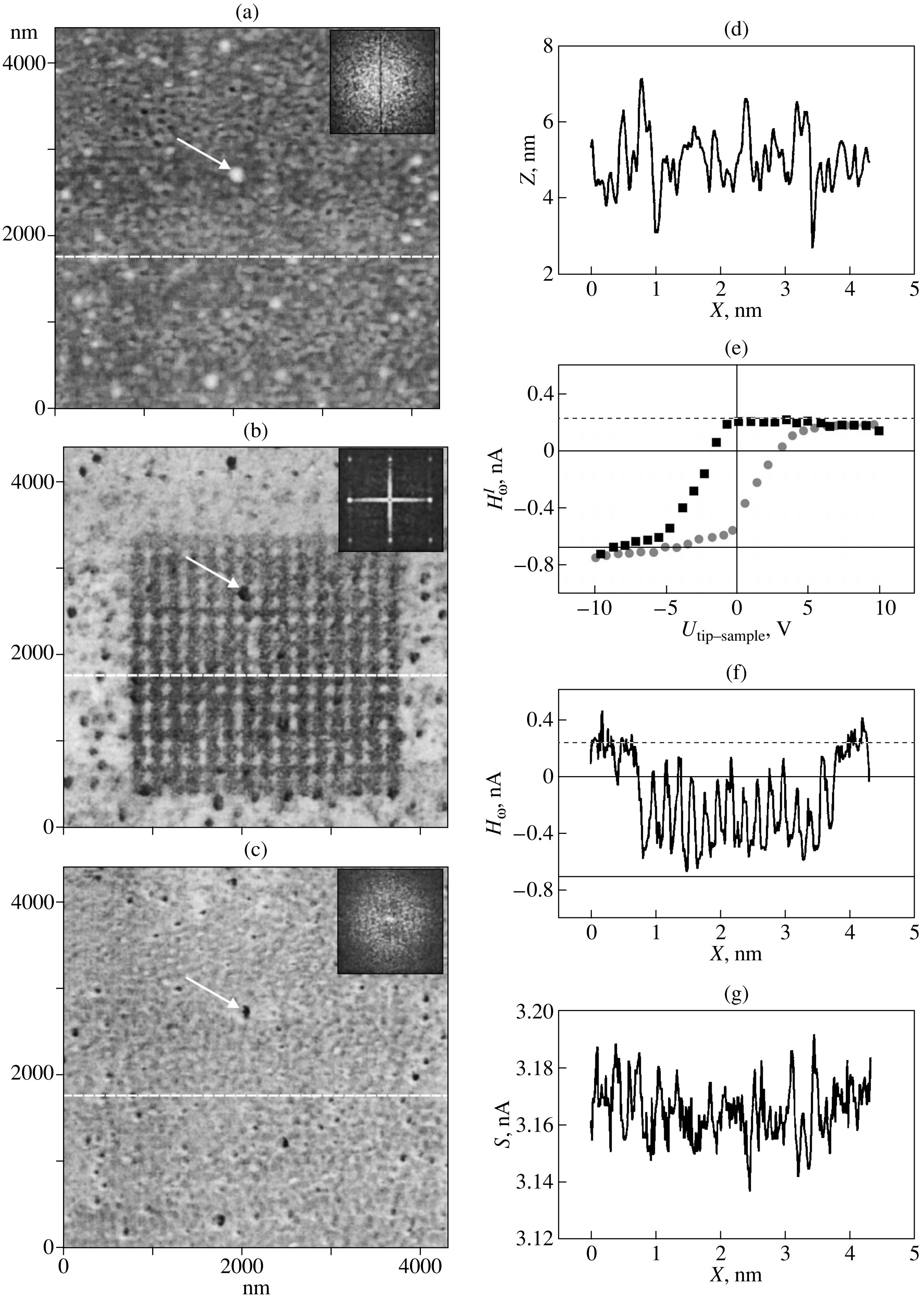


Fig. 4. Contact EFM study of a sample of a (001)-oriented $PbZr_{0.47}Ti_{0.53}O_3$ single-crystal film. (a) Topographic image; (b) EMR image; and (c) stiffness signal image. The EMR signal (H_ω, MagSin signal in Solver P47) was excited at a frequency of 30.175 kHz at an amplitude U_{ac} = 3 V and U_{dc} = 0; to excite the stiffness signal (S, Mag signal in Solver P47), an ac voltage 0.1 V in amplitude was applied at the same frequency to the z plates of the scanner carrying a sample. Each image specifies the pertinent Fourier data in the top right-hand corner. The sections of (d) topography, (f) EMR, and (g) stiffness signals were obtained along the bright lines in the corresponding images. (e) Hysteresis of the pulsed EMR signal response.

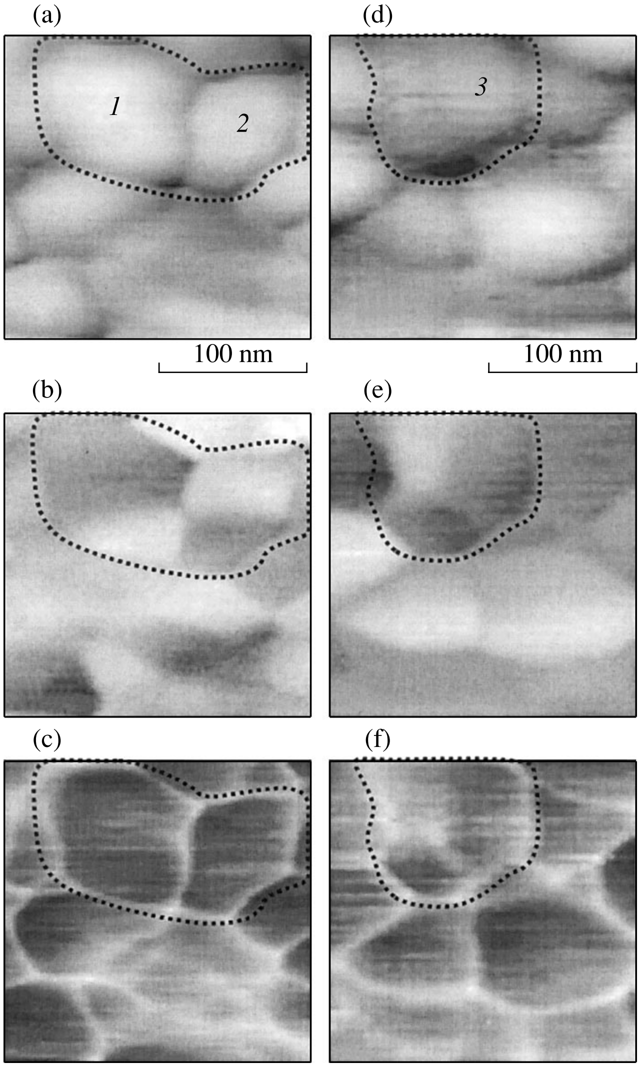


Fig. 5. Contact EFM study of a (111)-oriented, 100-nm-thick sample of a $PbZr_{0.47}Ti_{0.53}O_3$ polycrystalline film. (a, d) Topographic images, (b, e) EMR images, and (c, f) stiffness signal images. The dimensions of the images in panels (a–c) and (d–f) are specified by the scale bars under images in panels (a) and (d), respectively. The image regions with grains *1–3* discussed in the text are surrounded by contours.

It should be stressed that, as k^* increases, the loop may be displaced either upward or downward depending on the slope of the central line in the dynamic hysteresis loop and the position of the point at which this line crosses the abscissa axis (Fig. 2). Note that experimental data show the average contact potential difference to be positive (Fig. 3c); it is likewise positive for the sample regions shown in Figs. 4 and 5 [21]. It is also essential that the average slope of the dynamic response in these three cases was chosen to be positive (the slope of the EMR graph (Fig. 3c) may also be negative, because it is determined by the phase difference between the EMR and the lock-in detector reference signal). It is because of these two conditions that the EMR rises as the stiffness signal increases and the additional capacitive contribution is revealed by the coincidence of the contrast in the EMR and stiffness signals. If, however, the slope of the dynamic response is negative (positive) and the average contact potential difference is positive (negative), then the additional capacitive contribution will be revealed by the EMR and stiffness signals being in anticorrelation.

Finally, two points have to be clarified. First, the stiffness signal measured in the force modulation mode depends directly on variations in the stiffness of the tip–surface contact k_c [16, 19] rather than in the stiffness k^*. Second, the stiffness k^* of a probe in contact with the surface differs from that of a suspended probe k. Let us estimate first the range within which the k^*/k ratio can vary. To do this, it is expedient to treat the probe as a rod with one end clamped and the other free, touching a surface, or clamped. The lower the number of degrees of freedom of the second end of the rod, the higher the stiffness of the rod as a whole. The k^*/k ratio in these situations can be estimated from well-known analytical relations (see [22, p. 121]) for weak deflection of a rod under its own weight. Assuming for simplicity that the stiffness coefficient is proportional to the displacement of the rod center, we arrive at the conclusion that, as we pass over from the free end to the end touching a surface, the rod becomes stiffer by one order of magnitude and about twice as stiff again as the end becomes clamped. A similar behavior of the k^*/k ratio under variation of the constraints imposed on the rod end was revealed by numerical modeling of the response of an AFM probe in contact with a surface, which was carried out in [23]. The state of the end of the probe was specified in [23] by setting the contact stiffness k_c, which was calculated from an analytical solution to Hertz's contact problem (see [22, p. 46]). An essential point in substantiating our approach is that the coefficients k^* and k_c behave similarly with a variation in the local relief curvature, namely, that they both increase at depressions and decrease at protrusions. The same behavior is revealed by the stiffness signal measured in the force modulation mode, which is proportional to k_c/k [16, 19].

4. CONCLUSIONS

We have outlined a complex technique for studying thin ferroelectric films by AFM microscopy. It has been shown that the obtainment of topography and EMR images should be paralleled by measuring the stiffness signal. The stiffness signal provides a more accurate determination of the structure of polarization domains in a film and permits separation of the capacitive contribution from the piezoresponse in the EMR signal. This approach has been illustrated by a study of (001)-oriented $PbZr_{0.47}Ti_{0.53}O_3$ single-crystal and (111) polycrystalline $PbZr_{0.47}Ti_{0.53}O_3$ films with a grain size of about 100 nm.

ACKNOWLEDGMENTS

The authors are indebted to A. Tsvetkov for his assistance in the experiments and A. Emel'yanov, V. Kukhar', and N. Pertsev for helpful discussions.

This study was supported by the Russian Foundation for Basic Research (project no. 03-02-17635) and the Ministry of Education, Science, and Technology of the Russian Federation (contract "Diagnostics and Equipment for Metrology and Analysis of Components and Microstructures in Micro- and Nanoelectronics Technology").

REFERENCES

1. O. Kolosov, A. Gruverman, J. Hatano, K. Takahashi, and H. Tokumoto, Phys. Rev. Lett. **74**, 4309 (1995).
2. D. Damjanovic, Rep. Prog. Phys. **61**, 1267 (1998).
3. A. S. Sigov, Soros. Obraz. Zh. **10**, 83 (1996).
4. Y. Cho, K. Fujimoto, Y. Hiranaga, Y. Wagatsuma, A. Onoe, K. Terabe, and K. Kitamura, Appl. Phys. Lett. **81** (23), 4401 (2002).
5. K. Franke, H. Huelz, and M. Weihnacht, Surf. Sci. **415**, 178 (1998).
6. J. W. Hong, K. H. Noh, Sang-iL Park, S. I. Kwun, and Z. G. Khim, Phys. Rev. B **58**, 5078 (1998).
7. *Scanning Probe Microscopy and Spectroscopy: Theory, Techniques and Applications*, Ed. by D. A. Bonnell (Wiley-VCH, New York, 2001), Chap. 7.
8. A. Gruverman, B. J. Rodriquez, and R. J. Nemanich, J. Appl. Phys. **92**, 2734 (2002).
9. M. Alexe, C. Harnagea, D. Hesse, and U. Gösele, Appl. Phys. Lett. **79**, 242 (2001).
10. S. V. Kalinin and D. A. Bonnell, Phys. Rev. B **65**, 125408 (2002).
11. S. Honga, J. Woo, H. Shin, J. U. Jeon, Y. Eugene Pak, E. L. Colla, N. Setter, E. Kim, and K. No, J. Appl. Phys. **89**, 1377 (2001).
12. M. E. Lines and A. M. Glass, *Principles and Applications of Ferroelectrics and Related Materials* (Oxford Univ. Press, Oxford, 1977; Mir, Moscow, 1981).
13. S. Sridhar, A. E. Giannakopoulos, S. Suresh, and U. Ramamurty, J. Appl. Phys. **85**, 380 (1999).
14. A. E. Giannakopoulos and S. Suresh, Acta Mater. **47**, 2153 (1999).
15. S. V. Kalinin, Ph. D. Thesis, http://sergei2.kalininweb.com (2002).
16. P. Maivald, H. J. Butt, S. A. C. Gould, C. B. Prater, B. Drake, J. A. Gurley, V. B. Elings, and P. K. Hansma, Nanotechnology **2**, 103 (1991).
17. J. Rodriguez Contreras, J. Schubert, U. Poppe, O. Trithaveesak, K. Szot, Ch. Buchal, H. Kohlstedt, and R. Waser, Mater. Res. Soc. Symp. Proc. **688**, C8.10 (2002).
18. N. A. Pertsev, J. Rodriguez Contreras, V. G. Kurhar, B. Hermanns, H. Kohlstedt, and R. Waser, Appl. Phys. Lett. **83** (16), 3356 (2003).
19. http://www.ntmdt.ru.
20. A. Gruverman, A. Kholkin, A. Kingon, and H. Tokumoto, Appl. Phys. Lett. **78**, 2751 (2001).
21. A. V. Ankudinov and A. N. Titkov, in *Proceedings of the International Workshop on SPM-2004, Nizhni Novgorod, Russia, 2004*, p. 60.
22. L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, 3rd ed. (Fizmatlit, Moscow, 2001; Pergamon, Oxford, 1986).
23. U. Rabe, J. Janser, and W. Arnold, Rev. Sci. Instrum. **67** (9), 3281 (1996).

Translated by G. Skrebtsov

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1156–1160. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1118–1122.
Original Russian Text Copyright © 2005 by Belyaev, Gordeev, Konopleva, Nazarkin, Chekanov.

Martensitic Transformations in a Skeleton-Type TiC/TiNi Composite

S. P. Belyaev*, S. K. Gordeev, R. F. Konopleva***, I. V. Nazarkin***, and V. A. Chekanov*****

** Smirnov Research Institute of Mathematics and Mechanics, St. Petersburg State University, St. Petersburg, 198504 Russia*
e-mail: spb@smel.math.spbu.ru
*** Federal State Unitary Enterprise Central Research Institute of Materials, St. Petersburg, 191014 Russia*
**** St. Petersburg Institute of Nuclear Physics, Russian Academy of Sciences,*
Gatchina, Leningradskaya oblast, 188300 Russia

Received July 16, 2004

Abstract—The temperature kinetics of martensitic transformations in a skeleton-type TiC/TiNi composite with micron-sized structural constituents was studied. The composite is fabricated using a unique technique. Heat treatment and a change in the chemical composition of the titanium carbide are shown to affect the transformation temperatures. This effect is related to the redistribution of the chemical elements between the carbide and metallic components of the composite; as a result, the titanium nickelide becomes enriched in nickel. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The prospects for the creation of new materials with unique, novel properties are mainly related to the design of composites containing an alloy component that undergoes a thermoelastic martensitic transformation [1, 2]. It is assumed that optimum choice of the components of a composite and their proper spatial arrangement can provide structures belonging to smart materials in terms of their functional characteristics. The functional response of such composites to external factors is partly caused by a martensitic transformation (MT) occurring in one of their components. The sensory properties of materials with MTs are known to be unique, since, depending on the alloy grade, the phase transformation and structural changes can be stimulated by temperature changes, mechanical stresses, a magnetic field, hydrostatic pressure, or ultrasound. The responses of a material to these factors are diverse, from changes in its physical properties to mechanical work against applied forces [3].

Therefore, it seems promising to design a composite that has high strength and hardness and a sharp dependence of its physicomechanical properties on temperature and other physical factors. Such a composite is assumed to be based on carbide or nitride compounds in combination with a TiNi alloy, which undergoes a thermoelastic MT. The authors of [4–7] studied the properties of sintered composite materials in which the TiNi alloy was a binder and carbide or nitride particles served as solid strengthening inclusions to ensure high mechanical and operating properties of the composites.

In this work, we study the properties of a TiC/TiNi composite with a radically different structural arrangement, namely, with a skeleton-type structure. The material consists of two mutually penetrating frameworks, titanium carbide and titanium nickelide frameworks. The hard carbide phase provides high strength, hardness, wear resistance, and size stability of the material, and the metallic component improves the strength and fracture toughness of the material. If the metallic phase is assumed to undergo an MT, the physicomechanical properties of the composite will change with temperature, depending on the structural state (austenite, martensite, or austenite + martensite) of the alloy.

It should be noted that, in skeleton-type composites, a metallic phase occupies voids of a hard carbide skeleton that are several microns in size (specifically, 3–5 μm in the material under study). Therefore, one of the objectives of our study is to reveal possible MTs in TiNi fragments of these sizes.

2. MATERIALS

TiC/TiNi composite samples were prepared in two stages. In the first stage, a porous 5 × 5 × 25 mm billet was formed from titanium carbide. To this end, a titanium powder was compacted in a mold and the TiC compound was synthesized by heating the powder to 1120 K in a flow of natural gas followed by subjection to heat treatment. In a hydrocarbon (natural gas) atmosphere, carbon is synthesized on the surface of titanium particles. During further heat treatment, it reacts with titanium to form titanium carbide as a continuous skeleton throughout the entire billet. The composition of the formed stoichiometric (TiC) and nonstoichiometric TiC_y ($y < 1$) titanium carbides was controlled by varying the content of pyrocarbon deposited in the billet.

A study of voids in the porous titanium carbide skeleton showed that they were open and, thus, could be filled with a metallic melt. Therefore, in the second stage, the carbide billet was soaked with a melt of an equiatomic TiNi alloy in a vacuum furnace at 1570 K. After solidifying the metal, we fabricated a composite consisting of two phases, namely, a carbide and a metallic phase. The volume fractions of these phases and the characteristic size parameter of the metallic phase are specified by the initial void content and the void diameter in the TiC skeleton. Note that the properties of a composite that was produced by another procedure and had a similar structure were studied in [8].

It is obvious that an MT in a composite material can have characteristics other than those observed in conventional alloys, which can be caused by the following factors:

(1) constrained deformation of the lattice during the transformation without possible plastic shape accommodation in some cases;

(2) unusual grain sizes and other size parameters under conditions of confined geometry;

(3) a change in the chemical composition of an alloy due to diffusion through boundaries separating composite components; and

(4) thermal stresses induced by the difference in the thermal expansion coefficients of the components.

These and other factors can result in changes in the critical transformation temperatures and in the sequence of transformations and even in a complete suppression of martensitic phase transformations.

Therefore, the main purpose of this work, which is the first stage in the study of a skeleton-type composite material, was to investigate the development of MTs and the effect of heat treatment and composition on their temperature kinetics.

3. EXPERIMENTAL TECHNIQUE

We measured the electrical resistivity and the Young's modulus to study MTs in TiC/TiNi composite samples.

The Young's modulus was measured on a MUZA device under excitement and recording of forced mechanical vibrations of samples having a simple shape near their eigenfrequencies. Mechanical vibrations were excited and recorded with piezoelectric transducers attached to damped sound-conducting rods, and samples were fixed between the needle-shaped ends of these rods. Mechanical vibrations were excited with a sweep-frequency generator and were recorded with a unit intended for examining the amplitude–frequency characteristics. The apparatus allowed us to measure the parameters of the resonance peaks (the resonance frequency and the half-width and amplitude of peaks), which were used to calculate the elastic modulus and internal friction. The frequency range of the device was 30–300 kHz. Special-purpose equipment provided measurements in the temperature range 10–1300 K.

Sample compositions

Sample no.	y	Φ
1	0.76	0.5
2	0.83	0.5
3	0.89	0.5
4	1	0.59

The electrical resistivity was measured using the standard four-probe method. The temperature was changed at a rate of less than 2 K/min, and the temperature measurement accuracy was higher than 0.5 K. A computer was used to control the experiments and perform real-time data processing and visualization.

Samples contained different carbon contents y in the carbide phase of the TiC composite. The volume fraction of titanium nickelide Φ in the composite was close to 50%. The composition of the samples under study is given in the table.

4. RESULTS AND DISCUSSION

The measured properties of the as-prepared samples show that the skeleton-type composite material undergoes an MT as the temperature varies. Examples of the temperature dependences of the resistivity ρ and the Young's modulus (divided by its room-temperature value, E/E_{RT}) are shown in Fig. 1 for sample 4. It is seen that, during continuous cooling and heating, the Young's modulus varies with a minimum near 235 K and the temperature dependence of the resistivity is anomalous. Similar behavior of the physical characteristics is observed in titanium nickelide–based alloys in the MT temperature range [3]; hence, we may conclude that a martensitic transformation occurs in the compos-

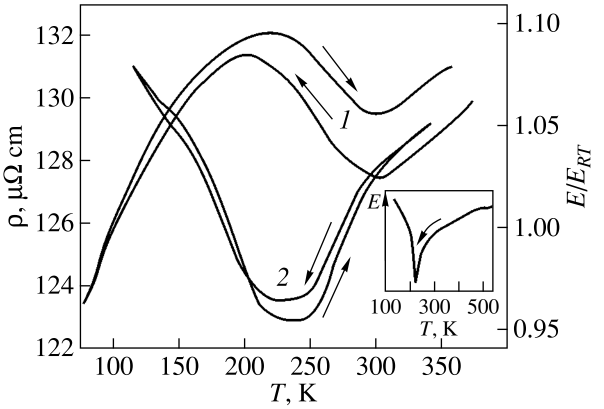


Fig. 1. Temperature dependences of (*1*) the electrical resistivity ρ and (*2*) reduced Young's modulus E/E_{RT} for composite sample 4 and (inset) for titanium nickelide.

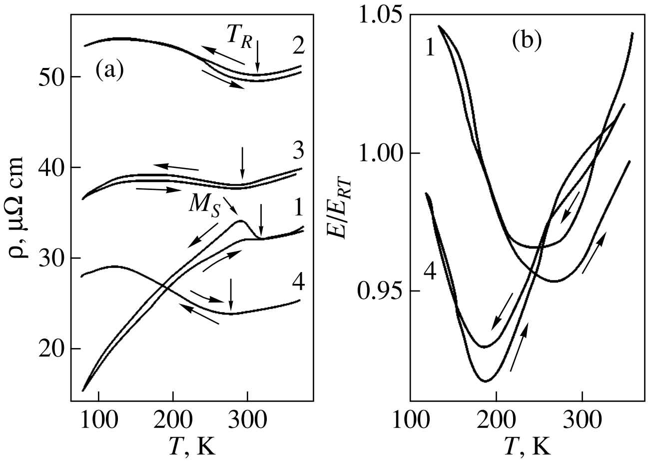


Fig. 2. Temperature dependences of (a) the electrical resistivity and (b) reduced Young's modulus of quenched samples. Numerals on the curves are the sample numbers.

ite material under study. Let us dwell on a few peculiar features. The minimum in the temperature dependences of the elastic modulus in the composite is substantially smeared, whereas it is sharp in TiNi (which is seen from comparing the curves in Fig. 1 and those in the inset to Fig. 1). Moreover, after a cooling–heating cycle is completed, the resistivity does not reach its initial value and the cycle turns out to be open. This behavior is likely to indicate a nonequilibrium structure of the composite and a substantial dispersion of its properties.

The TiNi solid solution in titanium nickelide–based alloys is usually homogenized by quenching from a temperature above 1070 K. Subsequent annealing of aging alloys at 500–800 K causes the formation of a number of phases in accordance with the phase diagram. As a rule, this treatment leads to a change in the temperature kinetics of the sequence of MTs in titanium nickelide. Therefore, it seems natural to use different heat treatments to find methods for controlling MTs in skeleton-type composites. To this end, samples were water-quenched from 1070 K and then annealed at 570, 670, 770, and 870 K for 2 h in an argon atmosphere.

During continuous cooling of the quenched samples, the temperature dependences of their resistivities (Fig. 2a) exhibit an anomaly, which consists in a deviation from a straight line at a point T_R with increasing resistivity. This deviation is known to be caused by the transformation from the high-temperature *B*2 cubic phase (CsCl type) to the martensite rhombohedral *R* phase [3, 9]. The $B2 \rightleftarrows R$ transformation occurs with a small temperature hysteresis and is close to a second-order phase transformation. Unlike other samples, sample 1 exhibits another phase transformation, from the rhombohedral to monoclinic *B*19' phase. The $R \longrightarrow B19'$ transformation begins at a temperature M_S corresponding to the beginning of a sharp drop in the resistivity upon cooling, and this transformation has a significant hysteresis. The temperature dependences of the Young's moduli of the quenched samples exhibit a minimum, as shown in Fig. 2b for samples 1 and 4. Note that the positions of both the minimum of the modulus on the temperature scale and the T_R point depend on the chemical composition of the TiC_y phase in the composite. As the titanium carbide composition approaches the

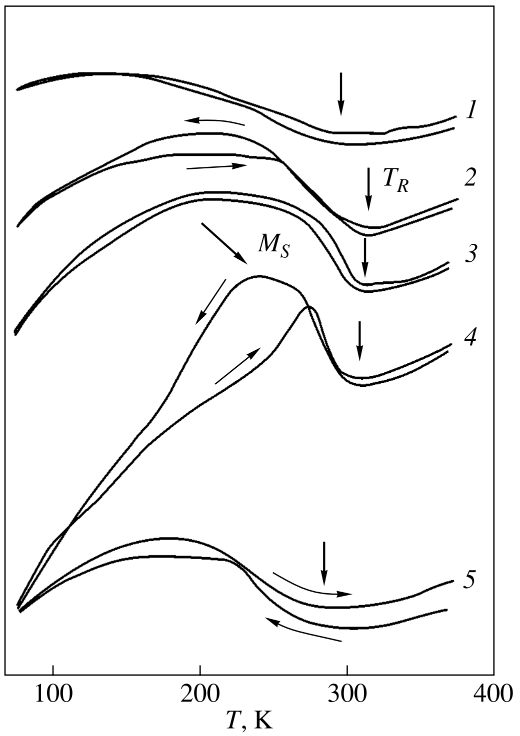


Fig. 3. Temperature dependences of the electrical resistivity of sample 2 (*1*) quenched from 1070 K and (*2*–*5*) then annealed at (*2*) 570, (*3*) 670, (*4*) 770, and (*5*) 870 K for 2 h.

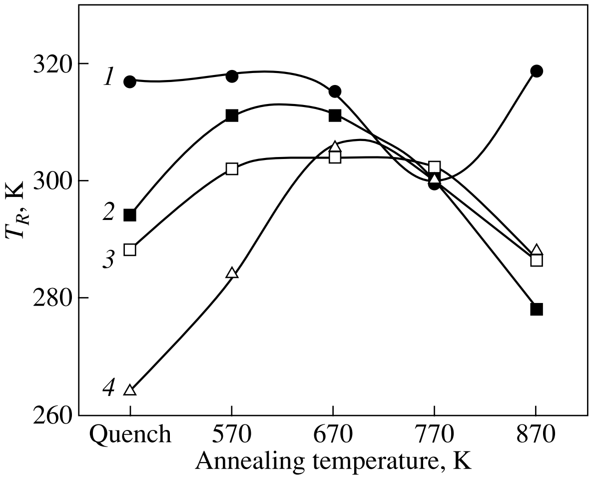


Fig. 4. Dependences of the transformation temperature T_R on the conditions of heat treatment. Numerals on the curves are the sample numbers.

stoichiometric composition, the T_R temperature and the minimum of the modulus shift toward low temperatures.

The shape of the temperature dependences of the physical characteristics shown in Fig. 2 for samples 2–4 is typical of nickel-rich TiNi alloys. In these alloys, the $R \longrightarrow B19'$ transformation temperature becomes lower than the boiling temperature of liquid nitrogen and, hence, only the transformation from the cubic to the rhombohedral phase occurs [10–12]. Therefore, we may suppose that, when the TiC/TiNi composite samples were prepared and quenched, titanium atoms diffused through the metal–carbide interface toward the carbide phase [4]. As a result, the titanium nickelide becomes enriched in nickel and the carbide is enriched in titanium. It is clear that the intensity of the titanium transport to the carbide phase is dictated by the initial titanium concentration in the carbide. That is why this process is less pronounced in sample 1, the TiNi composition remains close to stoichiometric, and two martensitic transformations ($B2 \longrightarrow R \longrightarrow B19'$) take place upon cooling.

The temperature kinetics of transformations in nickel-rich titanium nickelide is known to depend mainly on particles of the Ti_3Ni_4 phase forming upon annealing [13]. These particles are coherent with the surrounding TiNi matrix and form local stress fields, which favor heterogeneous nucleation of the rhombohedral martensite [9, 10, 13]. As the annealing temperature increases, the Ti_3Ni_4 particles coarsen and the T_R temperature increases. At the same time, the excess nickel is removed from the TiNi phase, its composition approaches the equiatomic composition, and two sequential transformations can occur in the alloy. The TiC/TiNi composite does behave like this. This behavior is seen in Fig. 3, which shows changes in the temperature dependences of the resistivity of sample 2 after quenching and subsequent annealing. High-temperature annealing at 870 K violates the coherence between the Ti_3Ni_4 and TiNi phases and qualitatively changes the temperature kinetics of the MT (Fig. 3) [14]. Figure 4 shows changes in the characteristic T_R temperature upon annealing of all samples under study. As follows from these data, the transformation temperature changes to form a maximum with increasing temperature.

The only exception is sample 1, in which the chemical compositions of the components are likely to be virtually unchanged during both fabrication and heat treatment. The concentration dependences of the transformation temperatures are most pronounced in the quenched samples (Fig. 5). As the annealing temperature increases, this dependence smoothes and the characteristic temperatures T_R of all samples become equal at 770 K.

In conclusion, we note that martensitic transformations in the TiNi phase of the designed TiC/TiNi composite materials with a skeleton-type structure develop according to laws identical to those occurring in the related homogeneous metallic alloys. The carbide component plays a significant role in the formation of the properties of the composite, since its chemical composition specifies the direction and intensity of redistribution of chemical elements between the phases. The carbide-phase composition and the conditions of heat treatment are the determining factors for controlling the sequence and temperatures of the martensitic transformations and, hence, the entire set of physicomechanical properties of the material.

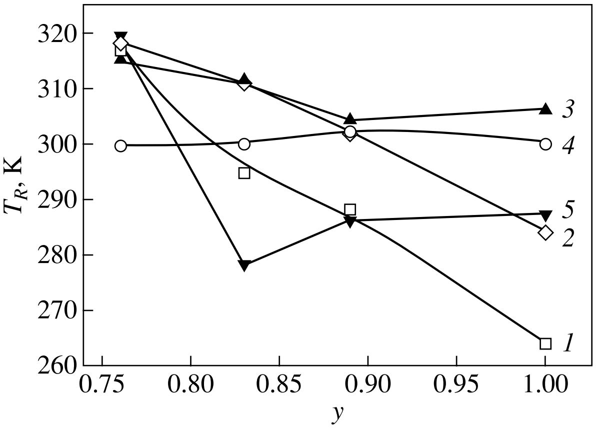


Fig. 5. Dependences of the transformation temperature T_R on y in TiC_y for samples (*1*) quenched from 1070 K and (*2*–*5*) then annealed at (*2*) 570, (*3*) 670, (*4*) 770, and (*5*) 870 K for 2 h.

The redistribution of chemical elements between the carbide and metallic components of the composite results in concentration gradients [8]. Since the martensitic-transformation temperatures in the TiNi alloy depend on the chemical composition, the transformations occur over a wide temperature range because of the concentration gradients. As a result, the transformation-induced features in the temperature dependences of the physical characteristics are strongly smeared.

ACKNOWLEDGMENTS

The authors are grateful to J. Cederstrom (Scandinavian Memory Metals AB) for his cooperation and useful discussion of the results and G.V. Kolobanov (St. Petersburg Institute of Nuclear Physics) for his assistance in the experimental work.

This work was supported by the Ministry of Science (project no. 40.012.1.1.1149) and the program for support of leading scientific schools (grant no. NSh-2180.2003.1).

REFERENCES

1. Z. G. Wei, R. Sandstrom, and S. Miyazaki, J. Mater. Sci. **33**, 3743 (1998).

2. Z. G. Wei, R. Sandstrom, and S. Miyazaki, J. Mater. Sci. **33**, 3763 (1998).
3. S. P. Belyaev, A. E. Volkov, V. A. Ermolaev, Z. P. Kamentseva, S. L. Kuz’min, V. A. Likhachev, V. F. Mozgunov, A. I. Razov, and R. Yu. Khaĭrov, *Shape Memory Materials* (Chemistry Research Inst., St. Petersburg Univ., St. Petersburg, 1998).
4. T. M. Poletika, S. N. Kul’kov, and V. E. Panin, Poroshkovaya Metallurgiya **7**, 54 (1983).
5. S. N. Kul’kov, T. M. Poletika, A. Yu. Chukhlomin, and V. E. Panin, Poroshkovaya Metallurgiya **8**, 88 (1984).
6. A. N. Ermakov, Yu. G. Zaĭnulin, V. G. Pushin, and E. V. Shchipachev, Fiz. Met. Metalloved. **92** (1), 43 (2001) [Phys. Met. Metallogr. **92** (1), 38 (2001)].
7. R. Vaidyanathan, M. A. M. Bourke, and D. C. Dunand, Acta Mater. **47** (12), 3353 (1999).
8. V. A. Sivokha, Yu. P. Mironov, V. V. Rudaĭ, and S. N. Kul’kov, Zh. Tekh. Fiz. **74** (1), 53 (2004) [Tech. Phys. **49** (1), 52 (2004)].
9. *Shape Memory Materials*, Ed. by K. Otsuka and C. M. Wayman (Cambridge Univ. Press, Cambridge, 1998).
10. V. N. Khanchin, V. G. Pushin, and V. V. Kondrat’ev, *Titanium Nickelide: Structure and Properties* (Nauka, Moscow, 1992) [in Russian].
11. T. Saburi, T. Tatsumi, and S. Nenno, J. Phys. (France) **42** (12), Suppl. ICOMAT-82, 261 (1982).
12. S. A. Shabalovskaya, A. I. Lotkov, and A. A. Baturin, Solid State Commun. **41** (1), 15 (1982).
13. G. A. Sobyanina and V. I. Zel’dovich, Fiz. Met. Metalloved. **86** (1), 145 (1998) [Phys. Met. Metallogr. **86** (1), 99 (1998)].
14. G. A. Sobyanina and V. I. Zel’dovich, Fiz. Met. Metalloved. **86** (1), 134 (1998) [Phys. Met. Metallogr. **86** (1), 91 (1998)].

Translated by K. Shakhlevich

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1161–1162. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1123–1124.
Original Russian Text Copyright © 2005 by Dekola, Sheleg, Tekhanovich.

LATTICE DYNAMICS
AND PHASE TRANSITIONS

Critical Behavior of Heat Capacity in the Vicinity of Phase Transitions in $[NH_2(CH_3)_2]_5Cd_3Cl_{11}$

T. I. Dekola, A. U. Sheleg, and N. P. Tekhanovich
Institute of Solid-State and Semiconductor Physics, National Academy of Sciences of Belarus,
ul. Brovki 17, Minsk, 220072 Belarus
e-mail: sheleg@ifttp.bas-net.by

Received May 11, 2004; in final form, July 29, 2004

Abstract—The heat capacity of a $[NH_2(CH_3)_2]_5Cd_3Cl_{11}$ crystal was studied calorimetrically in the temperature interval 100–300 K. The $C_p(T)$ dependence indicates that, as the temperature is lowered, phase transitions occur at temperatures $T_1 = 176.5$ K and $T_2 = 123.5$ K. The thermodynamic characteristics of this crystal were determined. It is shown that the transition at $T_2 = 123.5$ K is an incommensurate–commensurate phase transformation and that the transition at $T_1 = 176.5$ K is a normal–incommensurate phase transition. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The family of crystals containing a methyl ammonium cation exhibits various structural phase transitions (PTs). The comparatively new crystal $[NH_2(CH_3)_2]_5Cd_3Cl_{11}$, or $(DMA)_5Cd_3Cl_{11}$, which belongs to this family, reveals, according to different experimental studies, anomalies in its physical properties at 260, 180, 127 [1], 280 [2], and 127 and 178.5 K [3]. The nature of these transitions still remains unclear. It is known that the PT at 127 K is a first-order transformation and is associated with ordering of the dimethyl ammonium cation and that the PT at 178.5 K is considered to be second-order [3]. Dielectric studies have shown the $(DMA)_5Cd_3Cl_{11}$ crystal to be an incipient ferroelectric [1]. Investigations of the optical and elastic properties of $(DMA)_5Cd_3Cl_{11}$ suggested the possible existence of an incommensurate phase in this crystal within the temperature interval 127–178.5 K [3, 4]. It appeared of interest to perform high-precision measurements of the heat capacity of $(DMA)_5Cd_3Cl_{11}$ to refine the PT temperatures of this crystal, particularly in view of the fact that, as far as we know, no heat capacity studies of $(DMA)_5Cd_3Cl_{11}$ have thus far been attempted.

2. EXPERIMENTAL TECHNIQUE AND RESULTS

The heat capacity was measured in the temperature region 100–300 K using a vacuum adiabatic calorimeter with a discrete heat supply to the sample ($m = 7.32$ g). The sample heating rate was 0.03–0.08 K/min. The heat capacity measurements were performed in steps of 0.5–1.5 K with an accuracy of 0.3%. The sample temperature was monitored by a platinum resistance thermometer.

Figure 1 displays the temperature dependence of the specific heat of $(DMA)_5Cd_3Cl_{11}$. The $C_p(T)$ curve is seen to have anomalies at the PT temperatures $T_1 = 176.5$ K and $T_2 = 123.5$ K. The dashed line identifies the lattice part of the specific heat, which was obtained by interpolation with a polynomial of the form of $C = \sum_{i=0}^{3} A_i T^i$. Numerical integration yielded entropy and enthalpy changes of 9.6 J/K mol and 1185 J/mol for the former PT and 5.9 J/K mol and 931 J/mol for the latter PT, respectively. The smoothened values of the specific heat and the changes in the thermodynamic functions they correspond to, namely, the entropy S, the enthalpy H, and the Gibbs free energy Φ, are listed in the table.

Note that the peaked shape of the anomaly at $T_2 = 123.5$ K, as well as the increase in the experimental time needed for thermal equilibrium to be obtained in the region of this PT, suggests that the anomaly corre-

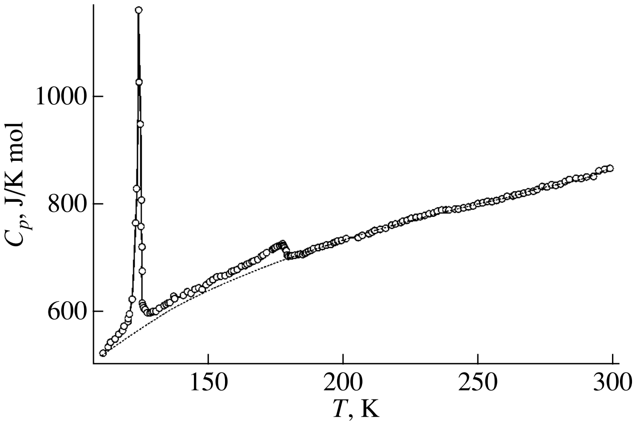


Fig. 1. Temperature dependence of the specific heat of $(DMA)_5Cd_3Cl_{11}$.

Smoothened values of the specific heat and the changes in the thermodynamic functions of $(DMA)_5Cd_3Cl_{11}$

T, K	$C_p(T)$	$S(T) - S(100\text{ K})$	$\Phi(T) - \Phi(100\text{ K})$	$H(T) - H(100\text{ K})$, J/mol
	J/K mol			
100	475.6	0.000	0.000	0.0
120	560.5	94.26	17.55	10360
140	614.1	184.8	44.18	22106
160	654.9	269.6	75.53	34796
180	699.1	349.4	109.3	48337
200	734.2	425.0	144.3	62670
220	762.7	496.3	179.8	77638
240	787.2	563.8	215.2	93138
260	810.3	627.8	250.3	109112
280	834.4	688.7	284.9	125559
300	862.1	747.3	318.9	142524

sponds to a first-order phase transition. Furthermore, the anomaly at T_2 has a clearly pronounced asymmetric shape (Fig. 1): on the high-temperature side, the specific heat is seen to fall off more slowly. According to [5], because of the soliton structure, the specific heat in the incommensurate phase near a PT should increase as $\Delta C_p \sim \{(T - T_c)[\ln(T - T_c)]^{-2}\}^{-1}$.

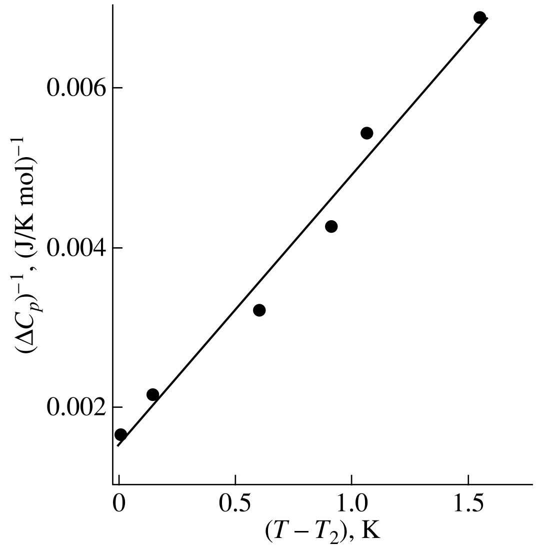


Fig. 2. $(\Delta C_p)^{-1}$ plotted as a function of $T - T_2$ for $(DMA)_5Cd_3Cl_{11}$ in the region of the PT at $T_2 = 123.5$ K.

Figure 2 plots a graph of $(\Delta C_p)^{-1}$ versus $T - T_2$ for $(DMA)_5Cd_3Cl_{11}$ in the vicinity of the PT at $T_2 = 123.5$ K. As seen from the figure, this plot is linear in the temperature interval from T_2 to $T_2 + 1.5$ K, which agrees with the theoretical prediction. Hence, one may suppose that the phase existing at temperatures $T > T_2$ in the $(DMA)_5Cd_3Cl_{11}$ crystal is incommensurate and that a PT from the incommensurate to the commensurate phase occurs at $T_2 = 123.5$ K.

The entropy of the incommensurate–commensurate phase transition is known to be always substantially smaller than that of a displacive transition, which is of the order of $0.1R$. The small value of the ΔS of lock-in transitions should be assigned to the fact that a soliton structure forms in the incommensurate phase in most crystals close to the transition temperature. Entropy changes $\Delta S \approx 0.1$ R due to incommensurate–commensurate phase transitions are observed in A_2BX_4-type incommensurate-phase crystals with both an atomic and an organic cation, with the exception of copper-containing crystals. The fairly large PT entropy change at T_2 for the $(DMA)_5Cd_3Cl_{11}$ crystal is apparently a result of considerable structural changes that accompany reorientation of the DMA molecular cations.

3. CONCLUSIONS

Our studies suggest that $(DMA)_5Cd_3Cl_{11}$ undergoes a PT sequence typical of crystals with an incommensurate phase. The second-order PT at $T_1 = 176.5$ K is a normal-to-incommensurate phase transition, and the first-order PT at $T_2 = 123.5$ K occurs from the incommensurate to a commensurate phase.

REFERENCES

1. Z. Czapla, S. Dacko, U. Krzewska, and A. Waskowska, Solid State Commun. **71** (2), 139 (1989).
2. K. Gesi and Z. Czapla, Ferroelectrics **159**, 37 (1994).
3. Z. Czapla, J. Przeslawski, and H. Schlemmbach, Solid State Commun. **91** (12), 981 (1994).
4. J. Furtak, Z. Czapla, and A. V. Kityk, Ferroelectric Lett. **22** (5–6), 147 (1997).
5. V. A. Golovko, Zh. Éksp. Teor. Fiz. **87** (3), 1092 (1984) [Sov. Phys. JETP **60** (3), 624 (1984)].

Translated by G. Skrebtsov

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1163–1168. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1125–1129.
Original Russian Text Copyright © 2005 by Murtazaev, Kamilov, Ramazanov.

LATTICE DYNAMICS AND PHASE TRANSITIONS

Critical Properties of the Three-Dimensional Frustrated Ising Model on a Cubic Lattice

A. K. Murtazaev, I. K. Kamilov, and M. K. Ramazanov
Institute of Physics, Dagestan Scientific Center, Russian Academy of Sciences, ul. 26 Bakinskikh Komissarov 94, Makhachkala, 367003 Dagestan, Russia
e-mail: m_akai@iwt.ru

Received June 8, 2004; in final form, September 7, 2004

Abstract—The critical properties of the three-dimensional fully frustrated Ising model on a cubic lattice are investigated by the Monte Carlo method. The critical exponents α (heat capacity), γ (susceptibility), β (magnetization), and ν (correlation length), as well as the Fisher exponent η, are calculated in the framework of the finite-size scaling theory. It is demonstrated that the three-dimensional frustrated Ising model on a cubic lattice forms a new universality class of the critical behavior. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The ideas underlying the scaling and universality hypotheses and the renormalization-group theory appeared most fruitful in constructing a unified modern theory of phase transitions and critical phenomena [1].

Until recently, it seemed that the theory of static phase transitions and critical phenomena in essence had been constructed and that progress on it had almost stopped. However, data obtained, for example, in the study of frustrated systems and spin systems with quenched nonmagnetic disorder have demonstrated that many results go far beyond the scope of the modern theory of phase transitions and critical phenomena [2].

In the majority of traditional theoretical and experimental methods for investigating such systems, serious problems are encountered in attempting to calculate the critical parameters and to elucidate the specific features, the nature, and the mechanisms of the critical behavior of these systems [2, 3]. For these reasons and others, phase transitions and critical phenomena have been extensively studied by Monte Carlo methods [3–7].

In the present work, the critical properties of the fully frustrated Ising model on a three-dimensional cubic lattice was investigated by the Monte Carlo method. Interest in this model is caused by the following basic factors.

First, frustrated systems on triangular and hexagonal lattices have been studied in sufficient detail [4, 5–7]. The critical properties of frustrated systems on a cubic lattice have not been adequately investigated.

Second, many important physical properties of frustrated systems depend substantially on the lattice geometry (the degree of frustration). Such a dependence can lead to a decrease in the number of universality classes of the critical behavior. This problem is still not clearly understood.

Third, the first attempts to investigate the above model were made when the power of computers and the algorithms employed in the Monte Carlo method were not sufficient to calculate the critical parameters with the necessary accuracy.

Moreover, the results obtained in [8–12] are ambiguous and contradictory and, hence, make the elucidation of all the specific features of phase transitions and the critical behavior of the model under consideration impossible. In particular, Chui *et al.* [8], Blankschtein *et al.* [9], and Grest [10] interpreted the phase transition revealed in the vicinity of the temperature $T_C = 1.34$ (hereafter, the temperature will be given in terms of $|J|/k_B$) as a first-order phase transition. However, Diep *et al.* [11] and Bernardi *et al.* [12] made the inference that, at this temperature, the system undergoes a second-order phase transition. It should be noted that, in the majority of the aforementioned works, primary attention was focused on the study of thermodynamic and magnetic properties, whereas the critical parameters were calculated in passing.

2. THE THREE-DIMENSIONAL FULLY FRUSTRATED ISING MODEL

Villain [13] was the first to propose the three-dimensional fully frustrated Ising model on a cubic lattice as a tool for adequately describing spin glasses. This model is illustrated in Fig. 1a. Note that the system can be divided into eight sublattices; however, sublattices *1–5*, *2–4*, *3–6*, and *7*, *8* are equivalent.

Figure 1b shows eight elementary cubes that describe a spin configuration of the given model in the ground state. It can be seen from this figure that each elementary cube involves three frustrated bonds.

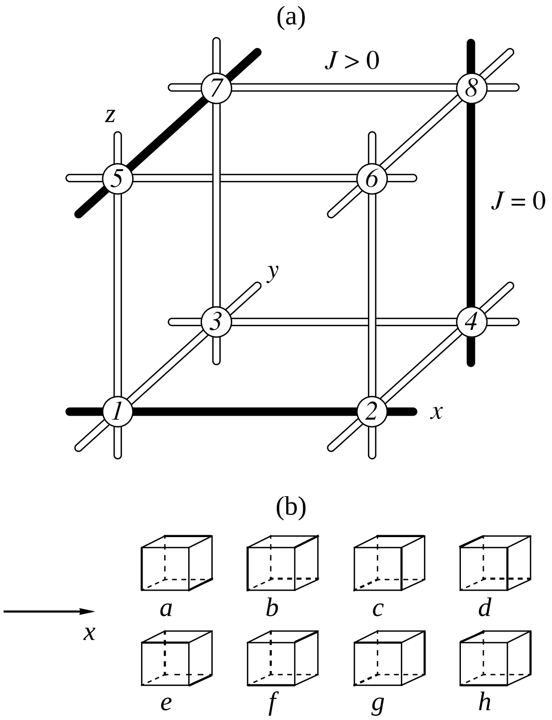


Fig. 1. (a) Three-dimensional fully frustrated Ising model on a simple cubic lattice and (b) eight elementary cubes.

The Hamiltonian of the system can be written in the following form:

$$H = -\sum_{\langle i,j\rangle} J_{ij}\sigma_i\sigma_j, \tag{1}$$

where σ is the Ising spin and J_{ij} is the exchange interaction ($J > 0$ and $J < 0$ for the ferromagnetic and antiferromagnetic bonds, respectively). Frustrations in this model are caused by competition between exchange interactions [11, 13].

The magnetic and thermodynamic properties of the above model were studied by the Monte Carlo method in [11, 12]. It was shown that the system is characterized by two phase transitions at $T_{C1} = 1.355$ and $T_{C2} \approx 0.7$ according to [11] and at $T_{C1} = 1.347$ and $T_{C2} \approx 0.7$ according to [12]. Bernardi *et al.* [12] rather conclusively demonstrated that the phase transitions at the temperatures T_{C1} and T_{C2} are second-order and first-order transitions, respectively. Most likely, Diep *et al.* [11] were the first to calculate a number of static critical exponents. The most important exponents ν and α were obtained by dividing the paraphase temperature range into two ranges $T_C \leq T \leq 1.45$ and $T > 1.45$. This procedure seems somewhat questionable, and the crossover temperature $T = 1.45$ determined in [11] is also in doubt. Furthermore, the origin of the crossover in the paraphase range remains unclear.

Moreover, the direct analysis of the Monte Carlo simulation data and the determination of the exponents from the slopes of the dependences of the thermodynamic parameters on a logarithmic scale are unconvincing, especially for the low Monte Carlo statistic obtained in [11].

Of special note is the work by Bernardi *et al.* [12], who determined the critical exponents α, β, ν, and η. However, the main objective of that work was to investigate the magnetic and thermodynamic properties of the model rather than to calculate the critical exponents. Moreover, the finite-size scaling method chosen by those authors for calculations, in our opinion, does not possess a high accuracy.

Nonetheless, the data obtained in the aforementioned works demonstrate that the critical parameters of the three-dimensional fully frustrated Ising model differ from those characterizing the universality class of the pure Ising model.

According to the modern theory of phase transitions and critical phenomena, the universality class of the critical behavior depends primarily on the following factors [1, 14]:

(i) the space dimension *d*,

(ii) the number of degrees of freedom of the order parameter *n*,

(iii) Hamiltonian symmetry, and

(iv) the length of the characteristic interaction.

However, a considerable amount of available data suggest that the universality class of frustrated systems can depend not only on the aforementioned factors. This is also confirmed by the Monte Carlo results obtained for lattices with different geometries [4, 5–7]. Note also that asymptotic values of the critical parameters for such systems have not been obtained with sufficient accuracy.

In this respect, the purpose of the present work was to attempt to determine the critical parameters of the three-dimensional fully frustrated Ising model by using reliable well-tried schemes within a unified approach and with as high an accuracy as possible.

3. COMPUTATIONAL TECHNIQUE

Frustrated spin systems are rather complex objects to be investigated even with the use of the Monte Carlo method. As is known, the Monte Carlo method in the vicinity of critical points is hampered by the problem of "critical slowing down," which, for frustrated systems becomes even more serious. Consequently, many new algorithms of the Monte Carlo method have been developed in recent years. Among them, the cluster algorithms of the Monte Carlo method have proved the most powerful and efficient for investigating critical phenomena in different spin systems and models [15–18]. These algorithms have provided a way to calculate the critical parameters of many model systems with a high accuracy [3]. Unfortunately, the use of the cluster algorithms for studying critical phenomena in frustrated systems has turned out to be inefficient. This is explained by the fact that frustrated systems undergo a phase transition at low temperatures and (or) a constructed cluster includes a very large region of the system. In a number of cases, the cluster algorithms have been used with a fitting parameter that makes it possible to control the size of the constructed cluster. Hence, frustrated systems have been investigated with special cluster algorithms, which are efficient only for low-dimensional models of frustrated systems [19, 20]. In the study of three-dimensional models, the efficiency of these algorithms has been shown to be even lower than that of the standard Metropolis algorithm (see references in [21]).

For this reason, the three-dimensional fully frustrated Ising model in our work was investigated using the classical Metropolis algorithm [22]. The calculations were performed for systems of size $L \times L \times L = N$ (L = 8–30) with periodic boundary conditions. The number N of spins in the simulated systems was equal to 512, 1000, 1728, 2744, 4096, 5832, 8000, 10648, 13824, 17576, 21952, and 27000. The initial configurations were specified so that all spins were aligned with the z axis. In order to bring the system into a state of thermodynamic equilibrium, a nonequilibrium segment of 5.0×10^5 MCS/spin in length was cut off. This segment was longer than the nonequilibrium segment itself by several factors. The thermodynamic quantities were averaged over a Markovian chain of a length of up to 3.0×10^6 MCS/spin.

4. RESULTS OF SIMULATION

The temperature dependences of the heat capacity and the susceptibility were examined using the relationships [23–25]

$$C = (NK^2)(\langle U^2\rangle - \langle U\rangle^2), \tag{2}$$

$$\chi = \begin{cases} NK(\langle m^2\rangle - \langle |m|^2\rangle), & T < T_C \\ NK\langle m^2\rangle, & T \geq T_C, \end{cases} \tag{3}$$

where $K = |J|/k_BT$, N is the number of particles, U is the internal energy, and m is the sublattice magnetization.

Figure 2 depicts the characteristic temperature dependences of the heat capacity C for the systems with linear sizes L = 8, 12, 16, and 20 (hereinafter, the error does not exceed the symbol sizes).

Note that the dependences of the heat capacity C for all the systems in the vicinity of the critical temperature exhibit clear maxima, which become more pronounced with an increase in the number of spins in the system. Moreover, these maxima, within the limits of experimental error, correspond to the same temperature even for the systems with the minimum size L. This suggests that, first, the periodic boundary conditions used in the calculation are highly efficient and, second, many parameters under investigation reach saturation with respect to the number N of spins.

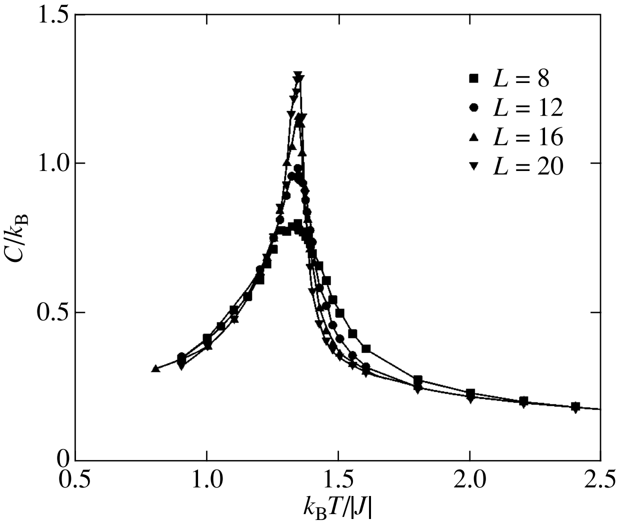


Fig. 2. Dependences of the heat capacity C/k_B on the temperature $k_BT/|J|$.

In order to determine the critical temperature T_C more precisely, we used the method of fourth-order Binder cumulants U_L. The fourth-order Binder cumulant has the form [26]

$$U_L = 1 - \frac{\langle m^4\rangle_L}{3\langle m^2\rangle_L^2}. \tag{4}$$

According to the finite-size scaling theory, the temperature dependences of the cumulant U_L intersect at the critical point [25].

Instead of the order parameter m in formula (4), we used the quantity $m = \frac{1}{4}\left(\sum_{\alpha=1}^{4} q_\alpha^2\right)^{1/2}$. Here, the Edwards–Anderson order parameter q_α is written in the form [11]

$$q_\alpha = \frac{8}{N}\sum_{i \in \alpha} \left|\langle \sigma_i^\alpha \rangle\right|, \tag{5}$$

where α indicates the sublattice number.

The characteristic temperature dependences of the Binder cumulant U_L for sublattices *3–6* are plotted in Fig. 3. The intersection point of these dependences corresponds to the critical temperature T_C = 1.344(2). The critical temperatures for the other sublattices were determined in similar fashion.

It can be seen from Figs. 2 and 3 that the maxima in the temperature dependences of the heat capacity for the systems with different numbers N of spins correspond to the critical temperature T_C determined by the Binder cumulant method. This demonstrates a high reliability of determining the critical temperature.

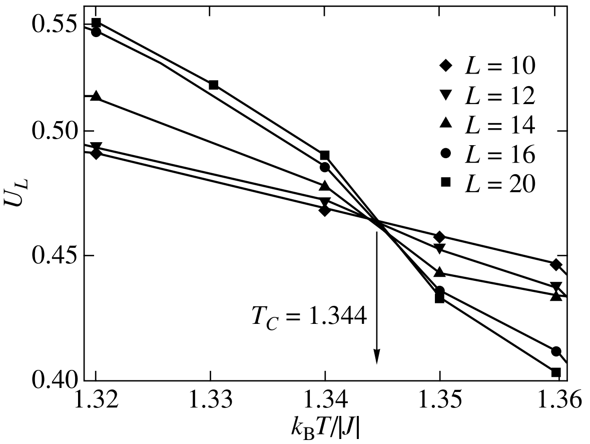


Fig. 3. Dependences of the Binder cumulant U_L on the temperature $k_BT/|J|$.

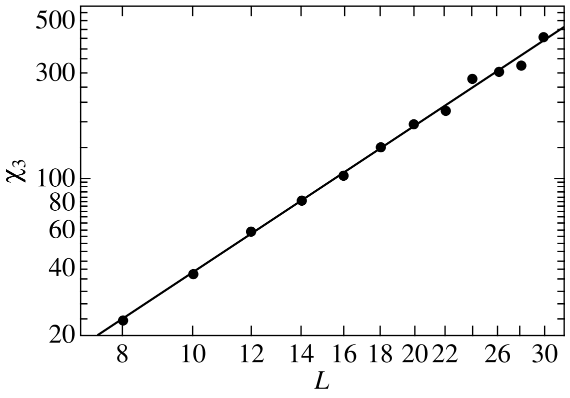


Fig. 4. Dependence of the susceptibility χ_3 on the linear size L of the system at $T = T_{C1}$.

The static critical exponents of the heat capacity α, the susceptibility γ, the magnetization β, and the correlation length ν were calculated from the relationships of the finite-size scaling theory.

According to this theory, the free energy of a sufficiently large system with periodic boundary conditions at a temperature T close to the critical temperature T_C of an infinite system can be represented in the following form [27, 28]:

$$F(T, L) \propto L^{-d}F_0(tL^{1/\nu}), \tag{6}$$

where $t = |T - T_C|/T_C$, $T_C = T_C(L = \infty)$, and ν is the static critical exponent of the correlation length for the infinite system ($L = \infty$).

Relationship (6) leads to similar dependences of the heat capacity, the susceptibility, and the spontaneous magnetization per spin [24, 27, 28]:

$$C(T, L) \propto L^{\alpha/\nu}C_0(tL^{1/\nu}), \tag{7}$$

$$\chi(T, L) \propto L^{\gamma/\nu}\chi_0(tL^{1/\nu}), \tag{8}$$

$$m(T, L) \propto L^{-\beta/\nu}m_0(tL^{1/\nu}), \tag{9}$$

where α, γ, and β are the static critical exponents of the system with $L = \infty$. These exponents are related by the hyperscaling expression $2 - \alpha = d\nu = 2\beta + \gamma$ [1].

Furthermore, a number of methods for determining the critical exponent of the correlation length ν on the basis of the finite-size scaling theory have been proposed to date [4, 29, 30].

As follows from this theory, at the phase transition point, we have

$$V_n = L^{1/\nu}g_{V_n}, \tag{10}$$

where g_{V_n} is a constant and the quantity V_n is defined by the expression

$$V_i = \frac{\langle m'E\rangle}{\langle m'\rangle} - \langle E\rangle, \quad (i = 1, 2, 3, 4). \tag{11}$$

It follows from formulas (8) and (9) that the magnetization and susceptibility in the system of size $L \times L \times L$ at $T = T_C$ with sufficiently large sizes L satisfy the following relationships [3, 24]:

$$m \propto L^{-\beta/\nu}, \tag{12}$$

$$\chi \propto L^{\gamma/\nu}. \tag{13}$$

These relationships were used to determine the critical exponents β and γ.

In our earlier work [31], it was shown that a similar relationship for the heat capacity does not hold. In practice, the dependence of the heat capacity on the size L, as a rule, can be approximated by the expression (see references in [31])

$$C_{\max}(L) = C_{\max}(L = \infty) - AL^{\alpha/\nu}, \tag{14}$$

where A is a coefficient.

The characteristic dependence of the susceptibility χ_3 on the linear size L of the lattice for sublattice *3* on the log–log scale is depicted in Fig. 4. It can be seen from Fig. 4 that the data fit a straight line whose slope determines the ratio γ/ν. A similar technique was applied to determine the quantities α/ν, β/ν, and 1/ν. Then, the exponents ν obtained were used to calculate the exponents α, β, and γ. All the exponents thus determined are listed in the table. For comparison, the data taken from [11, 12] are also presented in the table.

Of special note is the procedure used for determining the Fisher exponent η. By using the relationship

Critical exponents α, β, γ, ν, and η

Critical parameter	Data of this work	[11]	[12]	Unfrustrated Ising model (see references in [3])
α	0.46(2)	0.33(5)	0.32(2)	0.108
β	0.21(2)	–	0.25(2)	0.326
γ	1.18(3)	–	–	1.239
ν	0.55(2)	0.55(2)	0.56(2)	0.631
η	–0.15(5)	–0.28(6)	–0.10(2)	0.038
T_C	1.344(2)	1.335(2)	1.347(1)	4.5108

between the susceptibility χ and the correlation length ξ [32]

$$\chi \propto \xi^{\gamma/\nu} \tag{15}$$

and the expression $\eta = 2 - \gamma/\nu$ relating the exponents η and ν, we obtain

$$\ln(\chi/\xi^2) = c - \eta \ln \xi, \tag{16}$$

where c is a constant. For systems with finite sizes, the equality $\xi = L$ is satisfied. Then, at $T = T_C$, we have

$$\ln(\chi/L^2) = c - \eta \ln L. \tag{17}$$

Relationship (17) was used to determine the exponents η for all eight sublattices. These data are also presented in the table.

It should be noted that, to within the limits of error, the critical exponents ν calculated in our work are in agreement with those obtained in [12]. Our critical exponents α and β somewhat differ from the data of other authors [11, 12]. The critical exponent γ was most likely determined in our work for the first time (to the best of our knowledge, similar data are not available in the literature). The Fisher exponent η agrees with that calculated in [12] to within the limits of error. The critical temperature $T_C = 1.344(2)$ determined in our work nearly coincides with the critical temperature T_C = 1.355 obtained in [12].

5. CONCLUSIONS

Thus, the critical properties of the three-dimensional fully frustrated Ising model on a simple cubic lattice were investigated using the classical (Metropolis) algorithm of the Monte Carlo method. The results obtained made it possible to calculate all the static critical exponents. The critical exponents α (heat capacity), γ (susceptibility), β (magnetization), and ν (correlation length), as well as the Fisher exponent η, are calculated from the relationships of the finite-size scaling theory within a unified approach. The results of the calculations showed that the three-dimensional fully frustrated Ising model on the cubic lattice belongs to a new universality class. It seems likely that the critical exponent of the susceptibility for this model was calculated for the first time.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 04-02-16487), the State Program of Support for Leading Scientific Schools of the Russian Federation (project no. NSh-2253.2003.2), and the federal program "Integration." A.K. Murtazaev acknowledges the support of the Russian Science Foundation.

REFERENCES

1. A. Z. Patashinskiĭ and V. L. Pokrovskiĭ, *Fluctuation Theory of Phase Transitions*, 2nd ed. (Pergamon, Oxford, 1979; Nauka, Moscow, 1982).
2. Vik. S. Dotsenko, Usp. Fiz. Nauk **165** (5), 481 (1995) [Phys. Usp. **38** (5), 481 (1995)].
3. I. K. Kamilov, A. K. Murtazaev, and Kh. K. Aliev, Usp. Fiz. Nauk **169** (7), 773 (1999) [Phys. Usp. **42** (7), 689 (1999)].
4. D. Loison, A. I. Sokolov, B. Delamotte, S. A. Antonenko, K. D. Shotte, and H. T. Diep, Pis'ma Zh. Éksp. Teor. Fiz. **72** (6), 447 (2000) [JETP Lett. **72** (6), 337 (2000)].
5. H. J. Kavamura, J. Phys. Soc. Jpn. **61** (4), 1299 (1992).
6. H. J. Kavamura, J. Phys. Soc. Jpn. **56** (2), 474 (1986).
7. A. Mailhot, M. L. Plumer, and A. Caille, Phys. Rev. B **50** (10), 6854 (1994).
8. S. F. Chui, G. Forgacs, and D. M. Hatch, Phys. Rev. B **25** (11), 6952 (1982).
9. D. Blankschtein, M. Ma, and A. Nihat Berker, Phys. Rev. B **30** (3), 1362 (1984).
10. G. S. Grest, J. Phys. C: Solid State Phys. **18** (33), 6239 (1985).
11. H. T. Diep, P. Lallemand, and O. Nagai, J. Phys. C: Solid State Phys. **18** (5), 1067 (1985).
12. L. W. Bernardi, K. Hukushima, and H. Takayama, J. Phys. A: Math. Gen. **32** (10), 1787 (1999).
13. J. Villain, J. Phys. C: Solid State Phys. **10** (10), 1717 (1977).
14. I. K. Kamilov and Kh. A. Aliev, *Statical Critical Phenomena in Magnetically Ordered Crystals* (Dagest. Nauchn. Tsentr Ross. Akad. Nauk, Makhachkala, 1993) [in Russian].
15. U. Wolf, Phys. Rev. Lett. **62** (4), 361 (1989).
16. U. Wolf, Nucl. Phys. B **322** (3), 759 (1989).
17. A. M. Ferrenberg and R. N. Swendsen, Phys. Rev. Lett. **61** (23), 2635 (1988).
18. A. M. Ferrenberg and R. N. Swendsen, Phys. Rev. Lett. **63** (12), 1195 (1989).
19. D. Kandel, R. Ben-Av, and E. Domany, Phys. Rev. Lett. **65** (8), 941 (1990).
20. D. Kandel, R. Ben-Av, and E. Domany, Phys. Rev. B **45** (9), 4700 (1992).
21. P. D. Coddington and L. Hang, cond-mat/9402030 (1994).

22. *Monte Carlo Methods in Statistical Physics*, Ed. by K. Binder (Springer-Verlag, Berlin, 1979; Mir, Moscow, 1982).
23. K. Binder and J.-Sh. Wang, J. Stat. Phys. **55** (1), 87 (1989).
24. P. Peczak, A. M. Ferrenberg, and D. P. Landau, Phys. Rev. B **43** (7), 6087 (1991).
25. K. Binder and H. W. Heermann, *Monte Carlo Simulation in Statistical Physics: An Introduction* (Springer-Verlag, Berlin, 1988; Nauka, Moscow, 1995).
26. K. Binder, Phys. Rev. Lett. **47** (9), 693 (1981).
27. A. E. Ferdinand and M. E. Fisher, Phys. Rev. **185** (2), 832 (1969).
28. M. E. Fisher and M. N. Barber, Phys. Rev. Lett. **28** (23), 1516 (1972).
29. D. P. Landau, Physica A **205** (1), 41 (1994).
30. D. P. Loison, Phys. Lett. A **257** (1), 83 (1999).
31. A. K. Murtazaev, I. K. Kamilov, and M. A. Magomedov, Zh. Éksp. Teor. Fiz. **120** (6), 1535 (2001) [JETP **120** (6), 1330 (2001)].
32. Ch. Holm and W. Janke, Phys. Rev. B **48** (2), 936 (1993).

Translated by O. Borovik-Romanova

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1169–1171. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1130–1132.
Original Russian Text Copyright © 2005 by Zav'yalov, Kryuchkov, Meshcheryakova.

LOW-DIMENSIONAL SYSTEMS AND SURFACE PHYSICS

Effect of High Magnetic Fields on the Conductivity of a Quantum Cylinder under Stark Ladder Conditions

D. V. Zav'yalov, S. V. Kryuchkov, and N. E. Meshcheryakova
Volgograd State Pedagogical University, Volgograd, 400131 Russia
e-mail: sed@fizmat.vspu.ru

Received April 5, 2004; in final form, September 13, 2004

Abstract—The conductivity of a quantum cylinder with a parabolic lateral confinement potential and a superstructure is studied under conditions where uniform static quantizing electric and magnetic fields are applied along the cylinder axis. The charge carriers are assumed to be scattered by optical phonons. The dependence of the current density along the superlattice axis on the dc magnetic field is obtained. It is shown that, under certain conditions, the so-called Stark-hybrid-phonon resonance appears due to the hybridization of the electronic energy spectrum. In turn, this gives rise to a sharply nonmonotonic magnetic-field dependence of the current density. *© 2005 Pleiades Publishing, Inc.*

The modern development of nanotechnology has made it possible to fabricate surfaces of different curvature using strained GaAs/GaAlAs layers [1], including cylindrical surfaces with unusual physical properties [2]. Special attention has been paid to quantum nanostructures, where confinement can be modeled by a parabolic potential [3]. The optical properties of nanostructures with a parabolic confinement potential were studied in [3–5], where, in particular, it was shown that the application of a magnetic field results in resonance singularities at hybrid frequencies rather than at the cyclotron frequency.

The aim of this study is to calculate the longitudinal magnetoconductivity of a quantum cylinder with a superstructure. We assume that an additional periodic potential is created along the cylinder axis (thus, the z axis is both the axis of the superlattice and the axis of the quantum cylinder). Static quantizing magnetic and electric fields are applied along the z axis and produce hybridization in the electronic energy spectrum. Accordingly, observation of some interesting effects becomes possible. In particular, it is shown that the so-called Stark-hybrid-phonon resonance should be observed.

We describe a quantum cylinder with thin walls using the approach proposed in [3]. We assume that the dc electric field is sufficiently strong for the quantization effect to be noticeable, i.e., that the following inequalities are satisfied:

$$\tau\Omega_{\mathrm{st}} \gg 1, \tag{1}$$

$$\hbar\Omega_{\mathrm{st}} \ll E_g. \tag{2}$$

Here, $\Omega_{\mathrm{st}} = eEd/\hbar$ is the Stark frequency, e is the elementary charge, E is the dc electric field ($\mathbf{E} \parallel OZ$), d is the superlattice period, τ is the relaxation time, and E_g is the miniband gap. Condition (1) means that the discrete levels are not broadened by electron collisions with lattice irregularities, and condition (2) implies the possibility of using the single-miniband approximation.

The entire energy spectrum of charge carriers in a quantum cylinder in the situation described above is discrete and can be written as

$$\varepsilon_{nm\nu} = \hbar\Omega\left(n + \frac{1}{2}\right) + \lambda m^2 + \nu\hbar\Omega_{\mathrm{st}}, \tag{3}$$

where $\Omega = \sqrt{\omega_c^2 + \omega^2}$ is the hybrid frequency; $\omega_c = |e|B/m^*c$ is the cyclotron frequency; ω is the eigenfrequency of the confinement potential; $\lambda = 2\pi\hbar^2\omega^2/m^*L^2\Omega^2$ is the confinement energy; m^* is the carrier effective mass; $L = 2\pi R$ is the length of the channel; R is the radius of the cylinder; $n = 0, 1, 2, \ldots$; $m = 0, \pm 1, \pm 2, \ldots$; and $\nu = 0, \pm 1, \pm 2, \ldots$.

The electron wave function corresponding to spectrum (1) can be written as

$$\begin{aligned}\Psi_{nm\nu} = {} & \frac{1}{\sqrt{d}} J_{\nu - z/d}\left(\frac{\Delta}{\hbar\Omega_{\mathrm{st}}}\right)\frac{1}{L}\exp\left(\frac{imx}{R}\right)\frac{1}{\sqrt{2^n n!}} \\ & \times \exp\left(-\frac{1}{2l}\left(y + \frac{\omega_c l^2}{\Omega R}m\right)\right)H_n\left(\frac{1}{l}\left(y + \frac{\omega_c l^2}{\Omega R}m\right)\right),\end{aligned} \tag{4}$$

where $l = \sqrt{\hbar/m^*\Omega}$, $H_n(x)$ is the Hermite polynomial. Expressions (3) and (4) can be obtained from the corresponding formulas in [3] by replacing the continuous part of the spectrum $p_z^2/2m$ by the discrete component $\nu\hbar\Omega_{\mathrm{st}}$ and by replacing $\exp(ip_z z/\hbar)/\sqrt{2\pi\hbar}$ by the component of the wave function $J_{\nu - z/d}(\Delta/\hbar\Omega_{\mathrm{st}})/\sqrt{d}$ that

1063-7834/05/4706-1169$26.00 © 2005 Pleiades Publishing, Inc.

takes Stark quantization into account [6]. Such replacements are a consequence of the fact that **E** || **H** || *OZ*.

To calculate the current density, we use the general theory of electrical conductivity of semiconductors in quantizing fields, developed in [7]. Then, with regard to [8], we may write the current density along the superlattice axis in the form

$$j = j_1 + j_2, \tag{5}$$

where

$$\begin{aligned} j_{1,2} = {} & \frac{2\pi e d}{\hbar}\sum_{n,n'}\sum_{m m'}\sum_{\nu-\nu'}\sum_{\mathbf{q}}(\nu-\nu')f(n,m)H_{\mathbf{q}}^2 \\ & \times|M_{1,2}|^2\left(N_{\mathbf{q}}+\frac{1}{2}\pm\frac{1}{2}\right)\delta(\hbar\Omega(n-n') \\ & +\lambda(m'^2-m^2)+\hbar\Omega_{\mathrm{st}}(\nu'-\nu)\mp\hbar\omega_{\mathbf{q}}), \end{aligned} \tag{6}$$

$H_{\mathbf{q}} = C_{\mathbf{q}}\sqrt{\hbar/2\omega_{\mathbf{q}}\rho V}$, ρ is the density of the crystal, $C_{\mathbf{q}}$ is the electron–phonon interaction constant, $\omega_{\mathbf{q}}$ is the phonon frequency, **q** is the phonon wave vector, N_q is the Planck phonon distribution function, V is the normalizing volume ($V = L_xL_yL_z$), $f(n, m)$ is the nonequilibrium electron distribution function (to be calculated separately), $|M_{1,2}|^2$ is the square of the modulus of the matrix element of the operator exp($i\mathbf{qr}$) with wave functions (4) given by the expression

$$\begin{aligned} |M_{1,2}|^2 = {} & \frac{l^2 2^n n'!}{L^2 2^{n'} n!}\exp(-\Phi(q_x,q_y))[\Phi(q_x,q_y)]^{n'-n} \\ & \times[L_{n'}^{n-n'}(2\Phi(q_x,q_y))]^2 J_{\nu-\nu'}^2 \\ & \times\left(\frac{2\Delta}{\hbar\Omega_{\mathrm{st}}}\sin\frac{q_z d}{2}\right)\delta(m-m'\mp q_x R) \end{aligned} \tag{7}$$

and

$$\Phi(q_x, q_y) = \frac{\omega_c^2}{\Omega^2}\frac{q_x^2 l^2}{4}+\frac{q_y^2 l^2}{4}.$$

The current is due to electron hopping between the levels of the Stark ladder with a simultaneous change in the numbers $n \longrightarrow n'$, $m \longrightarrow m'$ and phonon emission (j_1) or absorption (j_2). In what follows, we consider in detail the case of electron interaction with optical phonons, whose dispersion is disregarded for simplicity (we assume that the quantities $H_{\mathbf{q}}$, $C_{\mathbf{q}}$, and $\omega_{\mathbf{q}}$ are independent of **q** and denote them by H_0, C_0, and ω_0, respectively). It follows from (6) that the dependence of the current density on the magnetic and electric fields resembles a series δ-function maxima, whose positions are determined by the condition of the Stark-hybrid-phonon resonance

$$\hbar\Omega(n-n')+\lambda(m'^2-m^2)+\hbar\Omega_{\mathrm{st}}(\nu'-\nu)\mp\hbar\omega_0 = 0. \tag{8}$$

This condition differs from that of [8, 9], where the magnetoconductivity of a one-dimensional superlattice was studied and the resonance was observed at the cyclotron rather than at the hybrid frequency.

At low temperatures, we can assume that the electrons are concentrated on the lowest energy level; hence, in the sums over n and m, we may retain only the terms with $n = m = 0$. We introduce the notation $f(0, 0) = n_0$, where n_0 is the concentration of electrons in the miniband. At low temperatures ($kT \ll \omega_0$), we have $N_q \ll 1$ and, therefore, $j_2 \ll j_1$. Thus, the expression for the current density assumes the form

$$\begin{aligned} j = {} & \frac{2\pi e d H_0^2 l^2 n_0}{\hbar^2\Omega L^2}\sum_n\sum_m\sum_{\nu-\nu'}\sum_{\mathbf{q}}(\nu-\nu')J_{\nu-\nu'}^2(z_q)\frac{2^n}{n!} \\ & \times\exp(-2\Phi(q_x,q_y))(\Phi(q_x,q_y))^n[L_0^n(2\Phi(q_x,q_y))]^2 \\ & \times\delta\left(-n-\beta m^2+\frac{\Omega_{\mathrm{st}}}{\Omega}(\nu'-\nu)-\frac{\omega_0}{\Omega}\right)\delta(m+q_xR), \end{aligned} \tag{9}$$

where $\beta = \lambda/\hbar\Omega$.

In the quasi-classical approximation, where $z_q \gg 1$, we can sum over $\nu - \nu'$ [8, 10, 11]:

$$\begin{aligned} & \sum_{\nu=-\infty}^{\infty}\nu J_\nu^2(z_q)\delta\left(n+\beta m^2-\nu\frac{\Omega_{\mathrm{st}}}{\Omega}+\frac{\omega_0}{\Omega}\right) \\ & = \frac{\Omega}{\pi\Omega_{\mathrm{st}}}\frac{\gamma_{n,m}\Theta(z_q-\gamma_{n,m})}{\sqrt{z_q^2-\gamma_{n,m}^2}}. \end{aligned} \tag{10}$$

Here, $\Theta(x)$ is the step function and $\gamma_{n,m} = n + \beta m^2 + \omega_0/\Omega$.

With regard to (10), the expression for the current density becomes

$$\begin{aligned} j = {} & \frac{2\pi e d H_0^2 l^2 n_0}{\hbar^2\Omega L^2}\sum_n\sum_m\sum_{\mathbf{q}}\frac{2^n}{n!}\frac{\Omega}{\pi\Omega_{\mathrm{st}}}\exp(-2\Phi(q_x,q_y)) \\ & \times\frac{\gamma_{n,m}\Theta(z_q-\gamma_{n,m})}{\sqrt{z_q^2-\gamma_{n,m}^2}}(\Phi(q_x,q_y))^n\delta(m+q_xR). \end{aligned} \tag{11}$$

After summing over **q** in (11), we obtain

$$\begin{aligned} j = {} & j_0\frac{l}{L}\sum_n\sum_m\frac{2^n}{n!}Q(n,m)\gamma_{n,m}\frac{\Theta(\alpha-\gamma_{n,m})}{\sqrt{\alpha^2-\gamma_{n,m}^2}} \\ & \times F\left(\arcsin\sqrt{\frac{\alpha^2-\gamma_{n,m}^2}{\alpha^2}},\sqrt{\frac{\alpha^2}{\alpha^2-\gamma_{n,m}^2}}\right), \end{aligned} \tag{12}$$

where we introduced the notation $j_0 = 16eH_0^2Vn_0/\hbar^2L^2\Omega_{st}\pi^2$, $\alpha = 2\Delta/\hbar\Omega$, $F(\varphi, k)$ is an elliptic integral of the first kind,

$$\begin{aligned} Q(n, m) = {} & \frac{1}{2\Gamma(-n)}e^{-2\chi^2}(\chi^{1+2n}\sqrt{\pi}\Gamma(-1/2 - n) \\ & \times {}_1F_1(1/2, 3/2 + n, 2\chi^2) + 2^{-1/2-n}\Gamma(-n) \\ & \times \Gamma(1/2 + n)\,{}_1F_1(-n, 1/2 - n, 2\chi^2)), \end{aligned} \tag{13}$$

$\Gamma(x)$ is the gamma function, ${}_1F_1(a, c, x)$ is the degenerate hypergeometric function, and $\chi = \omega ml/2R\Omega$.

The results of the simulation for the current density as a function of the magnetic field using expression (12) are plotted in the figure. The following typical superlattice and quantum cylinder parameters were used in the simulation: $\Delta = 0.01$ eV, $n_0 = 10^{14}$ cm^{-3}, $m^* = 10^{-29}$ g, superlattice period $d = 10^{-6}$ cm, phonon frequency $\omega_0 = 1.6 \times 10^{12}$ s^{-1} [12], $C_q = 5 \times 10^8$ eV/cm [13], frequency of the confinement potential $\omega = 1.6 \times 10^{13}$ s^{-1}, cylinder radius $R = 10^{-6}$ cm [3], and $\Omega_{st} \approx 2 \times 10^{12}$ s^{-1}. At these parameter values, we have $j_0 \approx 0.3$ A/mm^2, which is entirely attainable for experimental observation.

It is seen in the figure that the jumps in the current density occur as the magnetic field increases. These jumps can be explained in the following way. From the conservation of energy, we obtain the inequality

$$\frac{2\Delta}{\hbar\Omega} - n - \frac{2\pi\hbar\omega^2}{m^*L^2\Omega^3}m^2 - \frac{\omega_0}{\Omega} > 0, \tag{14}$$

which determines the selection rules for m and n. Thus, e.g., three terms correspond to the segment ab in the plot ($n = 1$, $m = 0$; $n = 0$, $m = -1$; $n = 0$, $m = 1$). The current surge $bcde$, caused by the addition of terms $n = |m| = 1$, occurs if

$$\begin{aligned} & \frac{(2\Delta - \omega_0\hbar)m^*c}{3\hbar e}\left(2\cos\frac{\varphi + 4\pi}{3} + 1\right) < B \\ & < \frac{(2\Delta - \omega_0\hbar)m^*c}{3\hbar e}\left(2\cos\frac{\varphi}{3} + 1\right), \end{aligned} \tag{15}$$

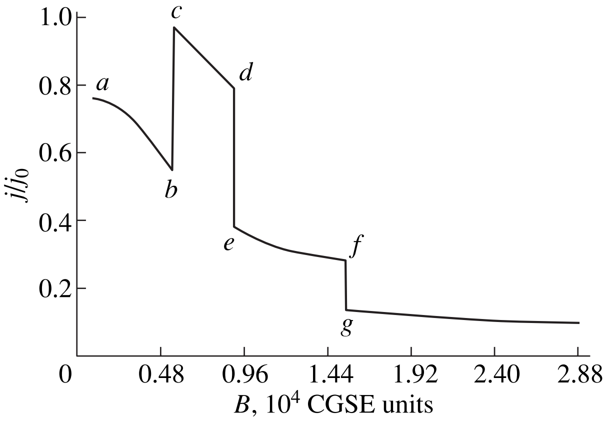
where B is the magnitude of the magnetic field vector and

$$\cos\varphi = 1 - \frac{27\pi\hbar^4\omega^2}{m^*L^2(2\Delta - \omega_0\hbar)^3}. \tag{16}$$

The jump fg is related to the disappearance of the sum over n ($n = 0$); this happens for

$$B > \frac{m^*c}{e}\sqrt{\left(\frac{2\Delta}{\hbar} - \omega_0\right)^2 - \omega^2}. \tag{17}$$

In the limit of high magnetic fields ($B \gg 10^5$ CGSE), the ratio j/j_0 tends to zero, since the energy gaps between the hybrid quantum energy levels increase with field, thus decreasing the carrier transition probability and, therefore, decreasing the current.



Dimensionless current density as a function of magnetic field. The parameter values are cited in the text.

ACKNOWLEDGMENTS

This study was supported by the Regional Competition AVO–RFBR "Povolzhye-2004," grant no. 04-02-96505.

REFERENCES

1. V. Ya. Prinz, V. A. Seleznev, V. A. Samoylov, and A. K. Gutakovsky, Microelectron. Eng. **30**, 439 (1996).
2. V. A. Geĭler, V. A. Margulis, and A. V. Shorokhov, Zh. Éksp. Teor. Fiz. **115**, 1478 (1999) [JETP **88**, 800 (1999)].
3. N. G. Galkin, V. A. Margulis, and A. V. Shorokhov, Fiz. Tverd. Tela (St. Petersburg) **43** (3), 511 (2001) [Phys. Solid State **43** (3), 530 (2001)].
4. É. P. Sinyavskiĭ and S. M. Sokovnich, Fiz. Tverd. Tela (St. Petersburg) **42** (9), 1685 (2000) [Phys. Solid State **42** (9), 1734 (2000)].
5. V. A. Margulis, Zh. Éksp. Teor. Fiz. **111**, 1092 (1997) [JETP **84** (3), 603 (1997)].
6. J. Callaway, *Energy Band Theory* (Academic, New York, 1964; Mir, Moscow, 1969).
7. V. V. Bryskin and Yu. A. Firsov, Zh. Éksp. Teor. Fiz. **61** (6), 2373 (1971) [Sov. Phys. JETP **61**, 1272 (1971)].
8. S. V. Kryuchkov and N. P. Mikheev, Fiz. Tekh. Poluprovodn. (Leningrad) **15** (11), 2277 (1981) [Sov. Phys. Semicond. **15**, 1323 (1981)].
9. V. M. Polyanovskiĭ, Fiz. Tekh. Poluprovodn. (Leningrad) **15** (10), 2051 (1981) [Sov. Phys. Semicond. **15** (10), 1190 (1981)].
10. S. V. Kryuchkov, Fiz. Tverd. Tela (Leningrad) **21** (5), 1595 (1979) [Sov. Phys. Solid State **21** (5), 921 (1979)].
11. L. Solymar, *Superconductive Tunnelling and Applications* (Chapman and Hall, London, 1972; Mir, Moscow, 1974).
12. F. G. Bass, A. A. Bulgakov, and A. P. Tetervov, *High-Frequency Properties of Semiconductors with Superlattices* (Nauka, Moscow, 1989) [in Russian].
13. K. Seeger, *Semiconductor Physics* (Springer, Berlin, 1974; Mir, Moscow, 1977).

Translated by I. Zvyagin

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1172–1175. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1133–1136.
Original Russian Text Copyright © 2005 by Wolf, Fedorov.

Evolution of Electronic States in Thin Ca(001) Films in Strong Electrostatic Fields

G. V. Wolf and D. V. Fedorov

Physicotechnical Institute, Ural Division, Russian Academy of Sciences, ul. Kirova 132, Izhevsk, 426001 Russia
e-mail: wolf@otf.fti.udmurtia.su

Received June 21, 2004; in final form, September 24, 2004

Abstract—The transformation of electronic states in Ca(001) films in strong electrostatic fields is studied using electron density functional theory. It is shown that an excess film charge of either sign pins the Fermi level (with respect to the conduction band edge) in a wide range of fields. For positively charged films, the change in the density of states at the Fermi level is small but the energy derivative of the density of states changes sign with increasing excess charge of the film. For negatively charged Ca(001) films, the change in the density of states at the Fermi level plays the main role in stabilizing the width of the occupied part of the conduction band; this should be manifested in the electronic thermodynamic and transport properties of negatively charged Ca(001) films with quantum confinement. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Variation in the electronic structure of a metal surface can play an important role in the formation of its physical properties in the presence of a strong external electrostatic field [1, 2]. The transformation of electron states at the Fermi level E_F caused by the charging of metal films is of great interest. Calcium can serve as a convenient object for studying these phenomena. In electrically neutral metallic calcium crystals, the Fermi level lies near the extrema in the local densities of states of *s*, *p*, and *d* types. In thin Ca(001) films, the density of surface states also has a peak near E_F [3, 4]. An applied field changes the potential within the limits of the penetration depth and thereby can cause a charge carrier redistribution and change the transport and thermodynamic characteristics [5, 6].

We are aware of very few calculations of the electronic structure of real metals in external electrostatic fields [7–12].[1]

This fact could be due to two main reasons. The first reason is mostly technical and is related to the long-range character of Coulomb interaction. When applying methods for calculating the electronic structure of electrically neutral crystals to charged crystals, the standard self-consistency procedure (which is an integral part of finding the electronic response of metals) does not converge [13–15]. Therefore, for charged crystals, it is necessary to develop special models and methods for calculating the electronic states. To this aim, the method of embedding potential has been developed for bulk crystals [16, 17]. We have developed a method of self-consistent calculation for electrically neutral and charged films that explicitly takes into account strong electron screening of the electric field at the metal surface [11].

The second reason is related to vanishing of the external electrostatic field in the bulk of the metal. Due to this, in bulk metallic crystals placed in an external field, the charge-dependent part of the total density of electronic states $n(E)$ is determined only by the surface contribution, is small as compared to the bulk contribution, and thus only slightly affects the physical properties of the sample.[2] A different situation occurs in films with quantum confinement consisting of 5–20 atomic layers. In this case, the bulk and surface contributions to $n(E)$ are comparable and a strong change in the electronic states at E_F and in the related physical properties is possible.

To the best of our knowledge, this study is the first to deal with the evolution of the density of electronic states in a real thin metal film in the presence of a strong electrostatic field.

2. STATEMENT OF THE PROBLEM AND DETAILS OF THE CALCULATIONS

The electronic structure of charged Ca(001) films was calculated using a new method [11] within the framework of electron density functional theory [18]. This method takes into account electronic screening at the metal surface. It should be noted that, today, electron density functional theory is probably the only

---

[1] Studies of the electrostatic response of metals by using the jelly model are more numerous. However, this model is very crude and is mainly used for finding the electronic density distribution near the metal surface.

[2] Surface-sensitive effects are an exception. For example, the intensity of the second harmonic generation of electromagnetic radiation at its reflection from a cubic crystal is entirely determined by the surface transformation of electronic states.

method that allows one to take account of exchange-correlation effects when calculating the electronic response of inhomogeneous systems with a complicated band structure [19].

The system of self-consistent equations to be solved is[3]

$$\begin{gathered}\{-\Delta + V[\varrho(\mathbf{r}, q)]\}\Psi_{n\mathbf{k}}(\mathbf{r}, q) = E_n(\mathbf{k}, q)\Psi_{n\mathbf{k}}(\mathbf{r}, q),\\ \Psi_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}_n, q) = \exp(i\mathbf{k}\mathbf{R}_n)\Psi_{n\mathbf{k}}(\mathbf{r}, q),\\ \int_{\Omega}|\Psi_{n\mathbf{k}}(\mathbf{r}, q)|^2 d\mathbf{r} = 1, \quad \varrho(\mathbf{r}, q) = \sum_{n,\mathbf{k}}|\Psi_{n\mathbf{k}}(\mathbf{r}, q)|^2.\end{gathered} \tag{1}$$

Here, $\mathbf{k}$ is the reduced two-dimensional quasi-momentum, $\mathbf{R}_n$ is the translation vector of the Bravais lattice of the film considered, and q is the number of “uncompensated” electrons in the unit cell Ω of the film. In calculating the electronic density $\varrho(\mathbf{r}, q)$, we sum over all states with energies $E_n(\mathbf{k}, q)$ that lie below the Fermi level $E_F(q)$.

The effective potential $V[\varrho(\mathbf{r}, q)]$ is the sum of the Coulomb and exchange-correlation contributions. The Coulomb contribution was calculated using the method developed in [20], in which the electrical neutrality of the bulk of the metal film was explicitly taken into account. This circumstance ensures the convergence of the self-consistency procedure used when calculating the electronic states of charged films. The exchange-correlation contribution to $V[\varrho(\mathbf{r}, q)]$ was calculated using the local density approximation and the interpolation Hedin–Lundqvist formula with the parameters listed in [21].

For an electrically neutral system ($q = 0$), the boundary conditions for $\Psi_{n\mathbf{k}}(\mathbf{r}, q)$ are naturally related to zero asymptotic behavior of the effective potential. The corresponding asymptotic expression for the wave functions of the discrete spectrum is a decaying exponential function. In the case of a charged film, the electrostatic potential in free space is a linear function of the distance from the metal surface, as is the potential of a charged plane [20]. An infinitely high potential barrier can be introduced in free space at a distance z_B from the film surface, where the electronic density is negligibly small; thus one may consider uncompensated charges of both signs using the stationary state description. As shown in [14, 22], if the barrier lies at a sufficient distance and the condition $V[\varrho(x, y, z_B, q)] > E_F(q)$ is satisfied, its effect on the energies of occupied states is negligible. A detailed description of the method used can be found in [11].

The lattice constant in the calcium film was set equal to its bulk value $A_{Ca} = 10.5296$ au and independent of the film charge; to some extent, this can be justified by the fact that the effect of surface relaxation on the surface density of electronic states in electrically neutral Ca(001) films can be disregarded [3].

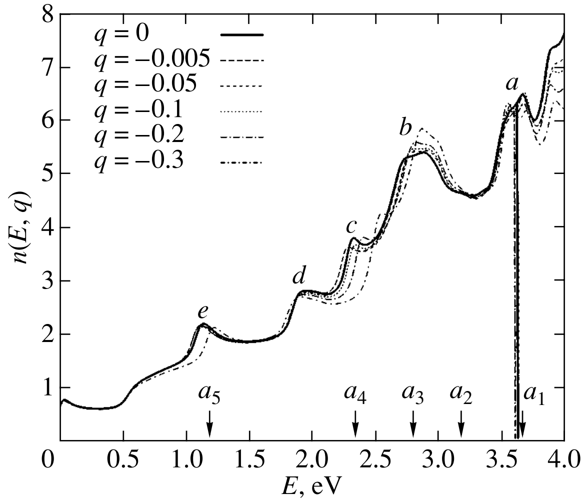


Fig. 1. Density of states in the occupied part of the conduction band of electrically neutral and positively charged ($Q > 0$) Ca(001) films. Vertical lines show the positions of the Fermi level $E_F(q)$ with respect to the bottom of the band.

The strongest electric fields $\mathbf{E}$ considered in this study ($|q| \le 0.3$, $|\mathbf{E}| \le 1.75 \times 10^{10}$ V/m) are large but quite accessible, e.g., at the electrode surface in an electrochemical cell [14].

3. RESULTS AND DISCUSSION

Figure 1 shows the evolution of the density of occupied electronic states in the conduction band of positively charged Ca(001) films. The changes in $n(E, q)$ with an increase in charge $Q = qe^-$ (here, e^- is the elementary charge) completely disagree with the simple picture based on the empirical rigid-band model [23]. Our calculations show that, in a five-layer Ca(001) film throughout the entire range of charge variation (down to 3 excess electrons per 10 film cells), the width of the filled part of the conduction band $\Delta(q)$ varies by no more than 0.048 eV, which is about 1% of $\Delta(q = 0)$. Thus, in the entire range of the variation of positive charge, a sort of pinning of the Fermi level near peak a occurs (Fig. 1). According to [4], this peak lies in the region of the energy localization of one of the surface states (peak a_1 in Fig. 1). The stability of $\Delta(q)$ is related to the transformation of $n(E, q)$ in the regions of energies close to the energies of surface states a_5, a_4, and a_3 (Fig. 1) obtained in [4]. The increase of electrons in number in the vicinity of state a_3 (peak b in Fig. 1) is almost fully compensated by the decrease in the concentration of electrons whose energies lie near surface states a_4 (peak c) and a_5 (peak e). A detailed evolution pattern of $E_F(q)$ and $n(E, q)$ near the Fermi levels for positively charged films is shown in Fig. 2. We see that the changes in $n(E_F, q)$ do not exceed 0.26 states/(cell eV) (number

---

[3] We use the atomic system of units with energy measured in Ry.

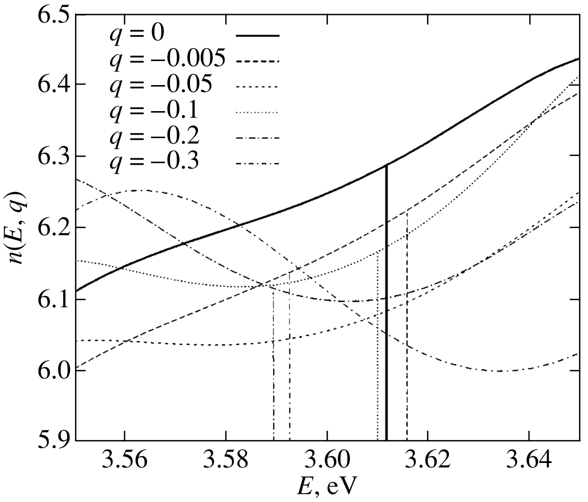


Fig. 2. Density of electronic states at the Fermi level for positively charged Ca(001) films.

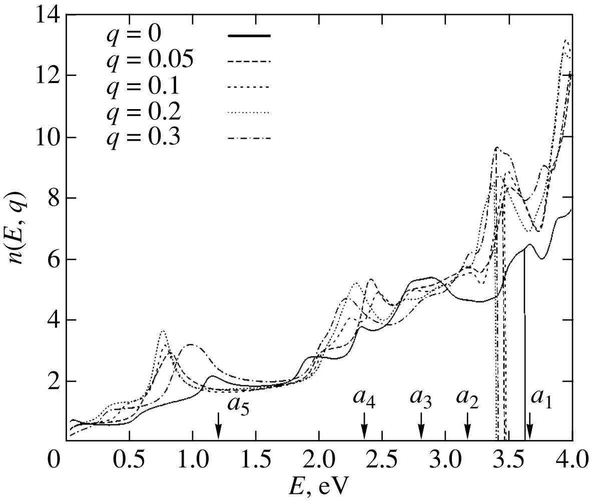


Fig. 3. Occupied part of the conduction band of an electrically neutral and negatively charged ($Q > 0$) Ca(001) films.

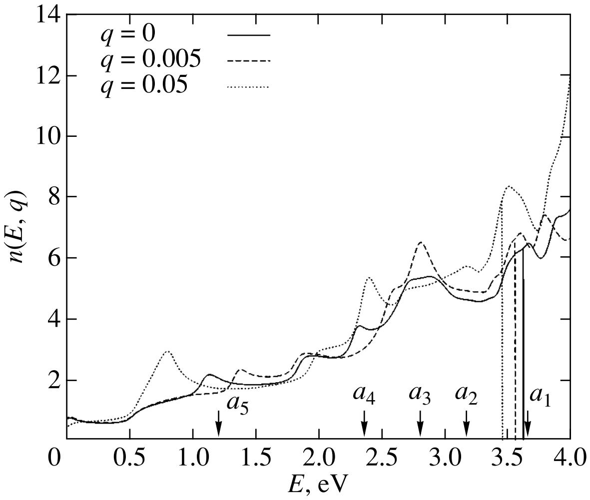


Fig. 4. Energy redistribution of negative screening charge in Ca(001) films at small q.

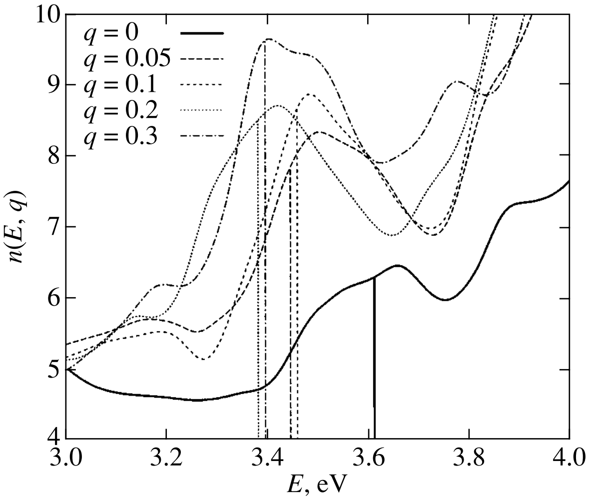


Fig. 5. Density of electronic states near the Fermi level for negatively charged Ca(001) films.

of states per unit cell per unit energy interval in electronvolts), or 4.1% of $n(E_F, q = 0)$. Therefore, the thermodynamics of the conduction electrons (electronic contributions to the entropy, heat capacity, temperature expansion coefficient, etc.) remains virtually unchanged for positive charging of the film. We see in Fig. 2 that, with increasing the positive charge of a film, the energy derivative of the density of electronic states at the Fermi level changes sign. For $|q| \leq 0.1$, we have $\partial n(E_F, q)/\partial E > 0$, whereas for $0.2 \leq |q| \leq 0.3$ we have $\partial n(E_F, q)/\partial E < 0$. This could prove important in studying the temperature dependence of the chemical potential and thermoelectric phenomena in charged metal films [6].

If the film charge is negative ($q > 0$), the evolution of the density of states has a somewhat different character. The Fermi level $E_F(q)$ is still pinned near the energy of surface state a_1 but $\Delta(q)$ is changed by 0.23 eV; this value is an order of magnitude greater than in the case of a positive charge and is 6.4% of the value $\Delta(q)$ for an electrically neutral film (Fig. 3). The main changes in $n(E, q)$ appear already at moderate negative charges. The presence of one excess electron per 20 film cells ($q = 0.05$) results in a significant (about 27%) increase in the density of states at the Fermi level. The value of $n(E, q)$ also appreciably increases near surface resonances a_4 and a_5 (Fig. 3). It is interesting that, at very small negative charges of a film cell ($q = 0.005$), the electronic states in the vicinity of surface state a_3 in an electrically neutral film are strongly perturbed. As q increases, overscreening appears and the increase in the number of electrons near peaks a_4 and a_5 in the surface density of states of an electrically neutral Ca(001) film becomes dominant even at $q = 0.05$ [4] (Fig. 4). At $q \geq 0.05$, the general structure in the $n(E, q)$ curve for the filled part of the conduction band of Ca(001) film is stabilized (Fig. 3). The changes in $\Delta(q)$ do not exceed 0.08 eV. This is due not only to the transformation of $n(E, q)$ near the energies of surface states a_4 and a_5 but also to a substantial increase in it at the Fermi level. From the data shown in Fig. 5, we see that, in the investigated range of negative charges, the quantity $n(E_F, q)$ changes by 3.362 states/(cell eV), i.e., by 53.4% of $n(E_F, q = 0)$. Accordingly, the electronic contribution to the heat capacity of the film $C_V^{(e)}(q) \sim n(E_F, q)$ increases by a factor of more than 1.5 [6].[4] Likewise, other physical quantities that depend on the density of electronic states at the Fermi level also change. In contrast to the case of a positively charged film, in the investigated region of negative charge, the energy derivative of $n(E, q)$ does not change sign at the Fermi level.

Thus, for both positively and negatively charged Ca(001) films, the Fermi level $E_F(q)$ is pinned near the surface state lying at E_F in the electrically neutral film. However, if in the case of a positive charge the stability of $\Delta(q)$ is ensured by the distribution of surface electrons with energies lying deep inside the conduction band of the film, in the case of a negative charge the increase in the number of electrons at the Fermi level plays a major role. This fact is expected to affect the thermodynamic and transport characteristics of electrons in negatively charged Ca(001) films.

ACKNOWLEDGMENTS

This study was supported by the Presidium of the Russian Academy of Sciences, program "Low-Dimensional Quantum Nanostructures."

REFERENCES

1. D. M. Kolb, W. Boek, Kai-Ming Ho, and S. H. Liu, Phys. Lett. **47** (26), 1921 (1981).
2. P. P. Boriskov, A. A. Velichko, A. L. Pergament, G. B. Stefanivich, and D. G. Stefanovich, Pis'ma Zh. Tekh. Fiz. **28** (10), 13 (2002) [Tech. Phys. Lett. **28**, 406 (2002)].
3. L. Ley, G. P. Kerker, and N. Martensson, Phys. Rev. B **23** (6), 2710 (1981).
4. G. V. Wolf and D. V. Fedorov, Fiz. Tverd. Tela (St. Petersburg) **45** (10), 1913 (2003) [Phys. Solid State **45** (10), 2012 (2003)].
5. Yu. F. Komnik, *Physics of Metal Films* (Atomizdat, Moscow, 1979) [in Russian].
6. I. M. Lifshitz, M. Ya. Azbel', and M. I. Kaganov, *Electron Theory of Metals* (Consultants Bureau, New York, 1973; Nauka, Moscow, 1971).
7. J. E. Inglesfield, Surf. Sci. **188** (3), L701 (1987).
8. G. C. Aers and J. E. Inglesfield, Surf. Sci. **217** (1–2), 367 (1989).
9. V. M. Silkin, L. Yurchishin, E. V. Chulkov, and M. Stenslitskaya, Poverkhnost **7**, 36 (1994).
10. S. Clarke, J. E. Inglesfield, M. Nekovee, and P. K. de Boer, Phys. Rev. Lett. **80** (16), 3571 (1998).
11. G. V. Wolf and D. V. Fedorov, Fiz. Tverd. Tela (St. Petersburg) **43** (3), 385 (2001) [Phys. Solid State **43** (3), 401 (2001)].
12. A. Y. Lozovoi, A. Alavi, J. Kohanoff, and R. M. Lynden–Bell, J. Chem. Phys. **115** (4), 1661 (2001).
13. N. D. Lang and W. Kohn, Phys. Rev. B **1** (12), 4555 (1970).
14. P. Gies and R. R. Gerhardts, Phys. Rev. B **33** (2), 982 (1986).
15. F. Schreier and F. Rebentrost, J. Phys. C: Solid State Phys. **20** (17), 2609 (1987).
16. J. E. Inglesfield, J. Phys. C **14** (26), 3795 (1981).
17. J. E. Inglesfield and G. A. Benesh, Phys. Rev. B **37** (12), 6682 (1988).
18. W. Kohn and P. Vashishta, *Theory of Heterogeneous Electron Gas* (Mir, Moscow, 1987), p. 86 [in Russian].
19. A. Williams and U. Bart, *Theory of Electron Heterogeneous Gas* (Mir, Moscow, 1987), p. 191 [in Russian].
20. G. V. Wolf and D. V. Fedorov, Poverkhnost **5**, 105 (1998).
21. O. Gunnarson and B. I. Lundqvist, Phys. Rev. B **13** (10), 4274 (1976).
22. P. Gies and R. R. Gerhardts, Phys. Rev. B **31** (10), 6843 (1985).
23. J. Friedel, Nuovo Cimento Suppl. **7** (2), 287 (1958).

Translated by I. Zvyagin

---

[4] We note that, for a positive charge of $|q| \leq 0.3$, the electronic heat capacity remains virtually unchanged ($C_V^{(e)}(q)/C_V^{(e)}(0) \leq 1.04$).

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1176–1185. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1137–1146.
Original Russian Text Copyright © 2005 by Zvezdin, Khval'kovskiĭ.

POLYMERS AND LIQUID CRYSTALS

Phase Transformations of the Magnetic Structure in Film Nanobridges

K. A. Zvezdin and A. V. Khval'kovskiĭ

Institute of General Physics, Russian Academy of Sciences, ul. Vavilova 38, Moscow, 119991 Russia

e-mail: khvalkov@ran.gpi.ru

Received March 18, 2004; in final form, July 26, 2004

Abstract—The magnetic structure of a plane nanobridge consisting of two ferromagnetic film electrodes connected by a nanosized crossbar of the same material is studied. Due to their magnetoresistive properties, such bridges are of considerable interest for microelectronics. Using a numerical micromagnetics method, it is shown that a domain wall is displaced from the center of the bridge crossbar as the anisotropy constant of the system decreases and reaches a critical value. A phase diagram is constructed, which makes it possible to determine the possible magnetic states of real nanobridges. The mechanism of the phase transformation is described in terms of an analytical model. This model explains the shape of the phase diagram of the nanobridge. Formally, the transformations of the magnetic structure of the nanocontact can be described in terms of the Landau theory of phase transitions in a certain range of parameters of the system. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Ferromagnetic nanocontacts and nanowires have recently been found to exhibit novel nontrivial effects, which opens up vast opportunities for the application of these structures in microelectronics. The first effect of note is giant magnetoresistance, which can be as high as several hundred percent at room temperature [1]. The mechanisms responsible for giant magnetoresistance have been studied in several publications. It has been found that the behavior and properties of domain walls (DWs) whose movement is restricted within a finite, nanosized volume (nanoconfinement) dictate the resistive properties of the magnetic nanocontact [2].

The behavior of DWs confined to a nanovolume is currently a subject of intensive study. In [3–5], the structure and properties of a DW confined to a nanocontact separating two large ferromagnetic regions were studied theoretically. It was found that, under these conditions, DWs differ from Néel and Bloch walls in terms of their structure and properties. One of the main distinctions of a confined DW is its extremely small size if the region to which the DW is confined is small (e.g., in the case of a point nanocontact). The effect of modulated anisotropy on the magnetization distribution in a nanosized region was studied experimentally in [6]. A theoretical model in which an external magnetic field causes a DW to escape from the confining region was considered in [4, 7]; it was shown that the increase in the DW surface energy is compensated for by a decrease in the magnetostatic energy.

Until recently, experiments were performed on nanocontacts with almost uncontrolled geometry. The study carried out in [1] was actually statistical in nature, because it was performed on contacts that formed randomly as two oppositely magnetized rods were joined or disjoined. In [8], it was shown that the magnetic structure of nanocontacts like those studied in [1] is extremely sensitive to even insignificant variations in their geometry. Obviously, nanocontacts with well-defined geometric parameters are required for their application in practice and for obtaining reliable experimental data.

In [9–11], such contacts are proposed to be fabricated in the form of a film nanobridge, i.e., two plane electrodes (banks) connected by a nanosized crossbar (Fig. 1). The magnetic structure of a nanobridge with oppositely magnetized banks was studied in [10] using computer simulation. It was found that, as the parameters of the system are numerically varied, there can occur a spontaneous displacement of the DW from the center of the nanobridge crossbar (spontaneous loss of stability). This DW displacement can be hysteretic or reversible, as is the case in first- and second-order phase transitions, respectively. However, the physical mechanism of this displacement is not understood. In analyzing these magnetic transformations, the main problem

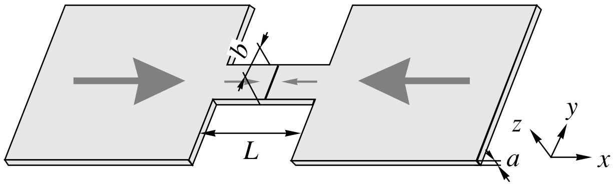


Fig. 1. Plane magnetic nanobridge with oppositely magnetized banks (schematic).

1063-7834/05/4706-1176$26.00 © 2005 Pleiades Publishing, Inc.

is correct inclusion of the long-range magnetostatic interaction, which is nonlocal and leads to the necessity of solving complicated integrodifferential equations using numerical methods.

In this paper, we perform a detailed micromagnetic study of the magnetic structure of a nanobridge. Based on the results of this study, we construct a phase diagram, which determines the possible magnetic states of the nanobridge, and propose a semiphenomenological model describing the physical mechanism of the phase transformations in the system and their dependence on the geometry and material parameters. Within this model, it is shown that, in a certain range of parameters, the transformations of the magnetic structure of the nanocontact can be formally described in terms of the Landau theory of phase transitions.

2. MICROMAGNETIC MODELING OF NANOSTRUCTURES

Experimentally, the study of the magnetic structure of a nanosystem is an extremely difficult problem, which sometimes cannot be solved unambiguously (see, e.g., [12]). Therefore, numerical calculations of the magnetization distribution in a small structure are of great importance. These calculations are based on micromagnetics theory, which was first formulated by Brown [13]. Modern micromagnetic methods combine a phenomenological theory developed for calculating the magnetization distribution in ferromagnetic structures subjected to given boundary conditions and numerical computing methods. These methods have been used to calculate the structure of many magnetic nanosystems, such as magnetic particles [14], nanocontacts [15], and three-layer spin switches [16].

In this study, we apply a much used numerical micromagnetic technique in which a plane ferromagnetic film is covered with a two-dimensional net and the magnetization vectors are determined at the cell centers. The stationary states and the dynamics of the magnetic system are studied by solving the three-dimensional Landau–Lifshitz magnetodynamics equations [13, 17]. To analyze the results obtained, we use the fact that the solution to the Landau–Lifshitz equations corresponds to a minimum of the total energy of the system.

Neglecting the magnetostriction and surface anisotropy, the total energy of the system E in the absence of external magnetic fields can be written in the form [13]

$$E = E_m + E_a + E_{\text{ex}}, \qquad (1)$$

where E_m is the magnetostatic energy, E_a is the anisotropy energy, and E_{ex} is the exchange energy. Let us consider each contribution in more detail [13].

2.1. Magnetostatic Interaction

The magnetostatic energy can be written as

$$E_m = -\frac{1}{2}\int_G (\mathbf{M}(\mathbf{r})\mathbf{H}_m(\mathbf{r}))d\mathbf{r}, \qquad (2)$$

where G is the region under consideration. The magnetostatic field $\mathbf{H}_m$ is given by

$$\mathbf{H}_m(\mathbf{r}) = \int_G \operatorname{div}\mathbf{M}(\mathbf{r}_1)\frac{\mathbf{r}-\mathbf{r}_1}{|\mathbf{r}-\mathbf{r}_1|^3}d\mathbf{r}_1. \qquad (3)$$

The magnetostatic energy is simple to understand physically but is difficult to calculate. The fundamental distinctive feature of this energy term as compared to the other contributions to the total energy is its nonlocality: the magnetostatic field at a point depends on the magnetization at all points of the solid. For this reason, modeling of the magnetic structure of various systems is a fairly difficult problem.

Equation (3) is an analog of the Coulomb law for magnetic charges, whose volume density is equal to $\operatorname{div}\mathbf{M}$:

$$\rho_m = \operatorname{div}\mathbf{M}. \qquad (4)$$

Therefore, we can determine $\operatorname{div}\mathbf{M}$ using micromagnetic modeling of the nanobridge or another model and calculate the integral in Eq. (3) by analogy with electrostatic problems. According to Eqs. (2) and (3), if there are magnetic charges on a boundary, then the magnetic field produced by these charges tends to reverse the magnetic dipoles located near this boundary. Therefore, $\mathbf{H}_m$ is often referred to as a demagnetizing field.

2.2. Exchange Interaction

The exchange energy E_{ex} is given by

$$E_{\text{ex}} = \int_G A(\nabla\mathbf{m})^2 d\mathbf{r}. \qquad (5)$$

The exchange interaction favors a parallel alignment of neighbor magnetic moments and thereby smoothes the magnetization distribution.

2.3. Anisotropy

The anisotropy energy E_a in an easy-axis ferromagnet is

$$E_a = -\int_G K(\mathbf{m}, \mathbf{n})^2 d\mathbf{r}, \qquad (6)$$

where K is the anisotropy constant, $\mathbf{m}$ is a unit vector directed along the vector $\mathbf{M}$, and $\mathbf{n}$ is a unit vector along the easy axis. In terms of the anisotropy energy, it is

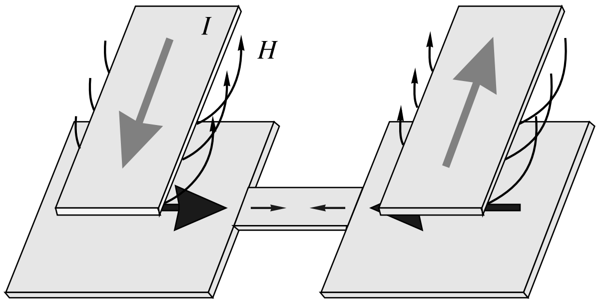


Fig. 2. Numerically simulated magnetization of the nanobridge with the use of two current lines.

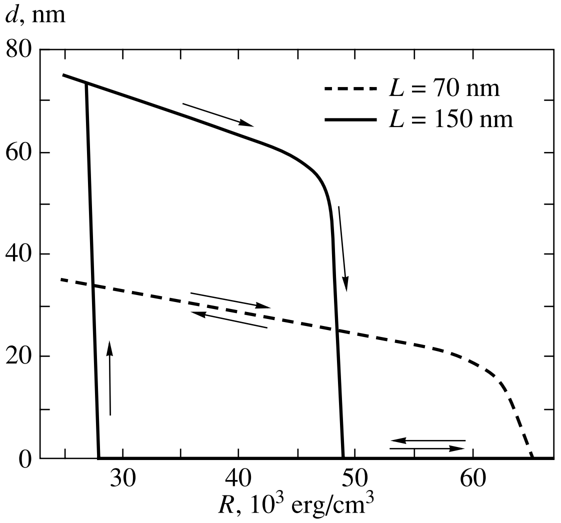


Fig. 3. DW displacement from the center of the crossbar as a function of the anisotropy constant for two values of the crossbar length.

favorable for the magnetization vector at each point to be parallel to the easy axis.

2.4. The Total Energy

The total energy can be expressed as an integral of the energy density ε:

$$E = \int_G \varepsilon(\mathbf{r})d\mathbf{r}, \tag{7}$$

$$\varepsilon = -\frac{1}{2}(\mathbf{H}_m\mathbf{M}) - K\cos^2\varphi + A(\nabla\mathbf{m})^2. \tag{8}$$

Here, φ is the angle between the magnetization **M** at a given point and the easy axis.

3. THE MAGNETIC STRUCTURE OF A NANOBRIDGE

We performed a numerical experiment to study a nanobridge consisting of two wide films connected by a narrow crossbar (Fig. 1). The magnetization and exchange energy are taken to be close to their values in a permalloy: M_s = 800 emu/cm^3 and $A = 1.25 \times 10^{-6}$ erg/cm. The x axis is taken to be parallel to the easy axis in the film. The crossbar width b is 40 nm, the leads (banks) are 500 × 500 nm in size, and the thickness of the structure is a = 1 nm (the notation coincides with that in Fig. 1). The uniaxial-anisotropy constant K and the crossbar length L are variable quantities.

The numerical experiment is performed as follows (Fig. 2). Through two current lines parallel to the y axis (Fig. 1), current pulses are passed in opposite directions perpendicular to the nanobridge having specific values of the crossbar length and anisotropy constant. The current pulses produce a magnetic field, which magnetizes the banks in opposite directions, so a DW forms at the center of the bridge crossbar. After current pulses are terminated, the magnetization relaxes toward an equilibrium distribution, in which the DW either remains at the crossbar center or shifts to one of the banks. Further, the reverse process is studied, in which the magnetization relaxes starting from the initial distribution with a DW displaced to one of the banks.

It is found that, at relatively large values of the uniaxial anisotropy constant, the symmetric magnetic configuration is stable. Therefore, in the equilibrium state, the head-to-head DW is located at the center of the nanobridge. As the anisotropy constant decreases and reaches a critical value K_c, the position of the DW at the center becomes unstable and the DW shifts to one of the banks.

This displacement of the DW from the center of the nanocontact crossbar is analogous to a phase transition in which the symmetry of the system is changed. For this reason, we will refer to this DW displacement as a phase transformation. The order of this transformation depends on the crossbar length, as seen from Fig. 3, where the DW displacement d from the center is plotted as a function of the anisotropy constant K. In the case of a short crossbar, the DW displacement varies continuously as the anisotropy constant decreases below the critical value (dashed curve), as is the case for a second-order phase transition. For a longer crossbar, hysteresis takes place, as in the case of a first-order phase transition (solid curves in Fig. 3). As the crossbar length increases further, only one branch of the hysteresis curve remains. In this case, once a DW appears at the crossbar center, it remains there at any positive value of the anisotropy constant K.

Figure 4 shows a phase diagram in which the regions of existence of possible magnetic states of the nanobridge are indicated. In region I, the DW is located near a bank; in region II, the DW is at the crossbar center; and in region III, there are two stable equilibrium positions of the DW (at the center and near a bank). The boundaries between regions I, II, and III correspond to phase transformations, which occur as the anisotropy constant K is varied in value. The A_1A_{tcr} line corresponds to a continuous transition, and the $A_{tcr}K'$ and $A_{tcr}K''$ lines correspond to loss of stability of the symmetric and asymmetric metastable phases (lability curves). On the phase diagram, there is a characteristic (tricritical) point A_{tcr} that separates the parameter range where the DW displacement from the center of the nanocontact occurs smoothly and the parameter range where the DW displacement exhibits hysteresis. In the vicinity of the tricritical point, all three magnetic states of the nanobridge can exist.

Although there are materials for which the uniaxial anisotropy constant depends on external factors (see, e.g., [18]), the possibility of varying the K value continuously is a distinctive feature of computer simulation and cannot be realized in practice. Nevertheless, the phase diagram obtained using computer simulation is of considerable practical importance, because this diagram makes it possible to determine the possible magnetic states and the magnetization-reversal parameters of real nanobridges.

Figure 5 shows magnetization-reversal curves for nanobridges of types I–III. A nanobridge of type I can be switched by an external magnetic field (parallel to the device axis) from one asymmetric state to another. Nanobridges of types II and III are switched by a magnetic field from the symmetric to an asymmetric state. At a zero field, an asymmetric state is unstable for type-II nanobridges and the DW returns to the crossbar center; for type-III nanobridges, this state is stable and the DW, once switched to this state, remains in it. Therefore, if the DW displacement is accompanied by a change in the resistance of the system (see, e.g., [9]), then a type-II nanobridge can be used as a memory cell and a type-III nanobridge as a magnetic-field sensor.

Micromagnetic modeling of the magnetic structure of a plane nanobridge was first performed in [10].

4. THE ENERGY OF THE NANOBRIDGE AS A FUNCTION OF ITS MAGNETIC STRUCTURE

As a result of computer simulation, the three components of the magnetization vector are found at each cell of the computational grid. In order to determine the mechanism of the transformations that occur in the nanobridge, the calculated magnetization distributions are analyzed qualitatively and quantitatively. As a result, the energy of the nanobridge is found as a function of the physical parameters and the magnetic structure of the system.

Let us consider the contributions from the magnetostatic interaction, anisotropy, and exchange interaction separately.

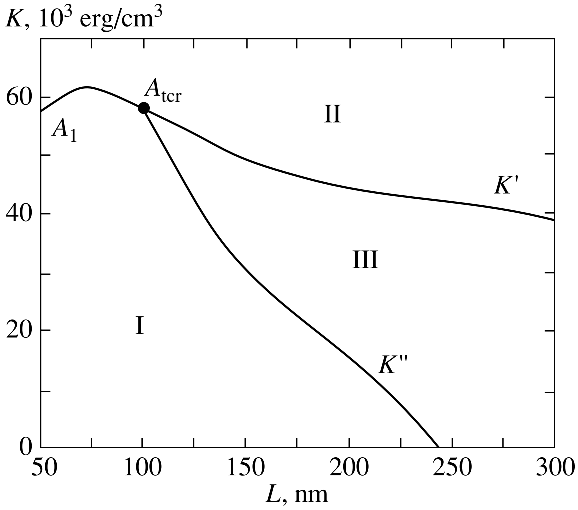


Fig. 4. Phase diagram of the magnetic nanobridge. Regions I–III correspond to different sets of magnetic states of real nanobridges. The boundaries between these regions correspond to phase transformations that occur as the anisotropy constant is varied. The A_1A_{tcr} curve corresponds to a continuous transition, and the $A_{tcr}K'$ and $A_{tcr}K''$ curves correspond to two hysteresis branches. The parameters of the nanobridge and explanation are given in the text.

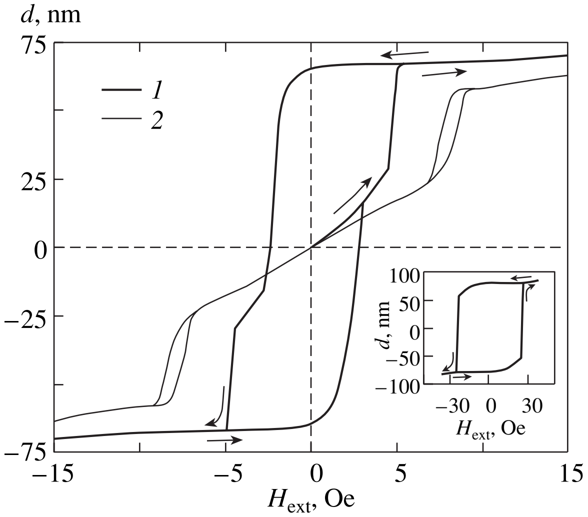


Fig. 5. Magnetic hysteresis loops for nanobridges with L = 150 nm and various values of K corresponding to regions III, II, and I in Fig. 4, respectively: (*1*) 45, (*2*) 60, and (inset) 20 kerg/cm^3.

4.1. Magnetostatic Energy

In order to estimate the magnetostatic energy, we need to find the distribution of magnetic charges in the system. First, we determine the charge density near the outer vertical sides (end faces) of the nanobridge, where the magnetization is virtually uniform and perpendicular to the end faces of the film (Fig. 6a). At a distance much larger than the film thickness a, the mag-

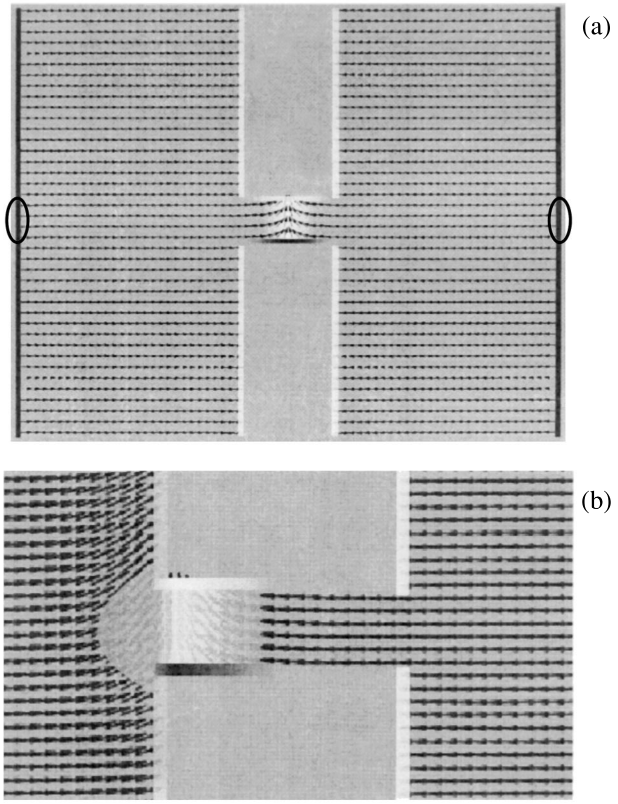


Fig. 6. Magnetization and magnetic-charge distributions (a) in the nanobridge in the case considered in the text and (b) near the crossbar in the case where the DW is near a bank. Positive and negative magnetic charges are marked in black and white, respectively. Ovals indicate the section of the face of a bank on which the charges are not compensated for by the unlike charges on the opposite face of the bank.

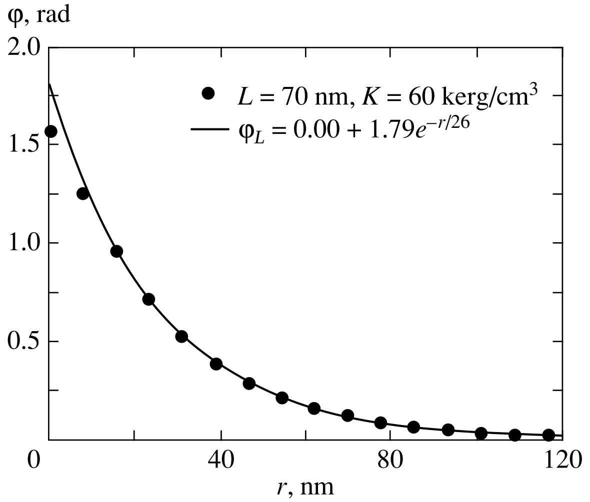


Fig. 7. Dependence of the angle φ on the distance to the DW r. Points are data from micromagnetic calculations, and the curve is a model dependence of the type $\varphi(r) = A + B\exp(-r/c)$.

netic field produced by these charges coincides with the field of a uniformly charged straight line with a linear magnetic-charge density λ. This density can be found from Eq. (4) to be

$$\lambda = Ma. \quad (9)$$

The distribution of magnetic charges in the nanobridge in the simplest case is schematically shown in Fig. 6a. In the banks, near the outer vertical sides, the magnetization is virtually uniform, the sides are uniformly charged with a positive linear charge density λ, and their field is identical to the field of a charged straight line. Calculations show that the influence of this field on the magnetic structure near the crossbar can be neglected because the distance to these sides is large (500 nm).

Near the inner vertical sides of a bank at a large distance along the y axis from the crossbar, the magnetization is uniform, and these sides are uniformly charged with a negative linear charge density –λ. Near the DW, the magnetic momenta are tilted from the x axis. As the DW approaches a bank (Fig. 6b), the magnetic moments in the bank tilt away from the horizontal plane and the magnetic-charge density on the vertical sides decreases in magnitude. According to Eqs. (2) and (3), the contribution from the magnetic moments of the bank to the magnetostatic energy decreases in this case. Therefore, the DW is attracted to a bank.

The force of attraction between the DW and a bank depends on the spacing between them. An analysis of the magnetization distribution (Fig. 7) shows that the dependence of the angle φ between the magnetization vector and the x axis on the distance from the DW center can be fitted with good accuracy by an exponential function,

$$\varphi(r) = \varphi_0 e^{-r/c}, \quad (10)$$

where c is the DW thickness. In the case where the anisotropy constant is larger than ~20 kerg/cm^3, the values of the parameter c of the DW in the crossbar and in a bank are of the same order of magnitude. However, these values are different, because the DW thickness in the crossbar is influenced (via the magnetostatic field) by the nearby faces of the crossbar. More specifically, the DW thickness in the crossbar is smaller (this effect is similar to that observed in [2–5]). In the case under study, the value of c is on the order of the crossbar thickness b, $c = (0.5–0.8)b$, depending on the value of K.

Thus, the force of attraction between the DW and a bank likewise decreases exponentially with distance, and the range of this force is on the order of b for the nanobridges under study.

One should separately treat the magnetic charges on the faces of the crossbar and in its bulk. Computer simulation shows that the magnetization distribution near the upper and lower faces of the crossbar is almost the same; therefore, the magnetic charges on these faces are also similar in magnitude and opposite in sign. These charges form a magnetic dipole, and its field tends to align the magnetic moments of a bank along the y axis in the negative direction. Since the y components of the magnetic moments of a bank and the DW are positive, the field of the magnetic dipole causes repulsion between a bank and the DW. As in the analogous case in electrostatics, the force of repulsion decreases as (or more slowly than) the inverse cube of the distance to the DW. Therefore, in contrast to the attractive force associated with a decrease in the magnetostatic energy of the banks, the repulsion between the DW and the magnetic charges of a bank is long-ranged. The charges of the magnetic dipole can be estimated as

$$q = \lambda \int_{hS} \sin\varphi dl, \tag{11}$$

where the domain of integration is the upper or lower side of the crossbar (hS). For the calculated magnetization distributions, numerical integration gives values of q lying in the range $(0.7–1.2)\lambda b$, depending on the nanobridge.

Since the magnetic charges on the faces of the crossbar are equal in magnitude and opposite in sign, the total charge of the crossbar is equal to its space charge. This charge can be estimated from the requirement that the total magnetic charge of the system is zero. Since the charge on the horizontal faces of the banks is virtually zero, the charge on the outer vertical faces of the banks is canceled by the total charge on the inner vertical faces of the banks and in the crossbar. The sections of the outer faces of the banks on which the charge is not compensated for by the charges on the inner vertical faces are indicated by ovals in Fig. 6a. The magnetization lines that begin at these sections terminate at the faces of the crossbar or at its interior points. Therefore, the total charge of the crossbar is $2\lambda b$ in the simplest approximation. Computer simulation shows that the space charge density in the crossbar decays exponentially with distance from the DW.

The field produced by the magnetic charges on the vertical faces of the banks is directed along the x axis. Since the magnetic moments in the DW are tilted to the x axis, the DW is repelled from the banks. The repulsive force is long-ranged because the field produced by the charges of a bank decreases as (or more slowly than) the inverse square of the distance from the bank. The magnetic moments of the banks are likewise repelled by the crossbar with its demagnetizing field.

4.2. Exchange Energy

Computer simulation shows that, if the DW is in one of the banks, its profile is changed only slightly.[1] However, since the DW volume is significantly larger in a bank, the energy of the system increases as the DW approaches a bank. Therefore, the DW is repelled from a bank. The force of repulsion is determined by the range over which the DW perturbs the magnetization distribution. This force is short-ranged, because the angle φ between the magnetization vector and the x axis decreases exponentially with distance from the DW, according to Eq. (10).

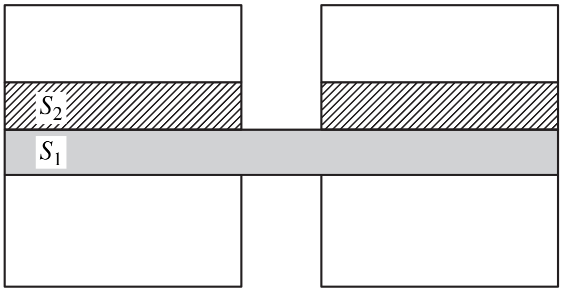


Fig. 8. Regions S_1 and S_2 used in calculating the total energy of the nanobridge.

4.3. Anisotropy Energy

The anisotropy energy also reaches a minimum when the DW is in the crossbar, because the DW volume is minimum in this case. Therefore, uniaxial anisotropy causes repulsion of the DW from the banks. This repulsion is likewise determined by the range over which the DW perturbs the magnetization distribution and is short-ranged.

5. MODEL PHASE DIAGRAM

5.1. Model Total Energy

Based on the results of micromagnetic modeling, we develop a semiphenomenological model. An expression for the energy density of the nanobridge is constructed, from which the total energy of the system is found. The total energy depends on the magnetic structure of the nanobridge. The minima of this energy determine the position of the DW for a given set of parameters of the system.

The model energy density is integrated over two regions, S_1 and S_2 (Fig. 8). Over each of these regions, the magnetization distribution is considered uniform. The energy of the entire nanobridge is found from the combined energy of these regions, which is calculated in a special way. Namely, normalization factors n_i are introduced into the expression for the energy density in the region S_2 in order to relate the energy of the banks (excluding the strip S_1) to the energy of this region:

$$\varepsilon = -\frac{1}{2}\left(\sum n_k \mathbf{H}_m^k \cdot \mathbf{M}\right) - n_{\mathrm{an}} K \cos^2\varphi + n_{\mathrm{ex}} A (\nabla \mathbf{m})^2. \tag{12}$$

---

[1] For values of K lying in the range 30–70 kerg/cm^3.

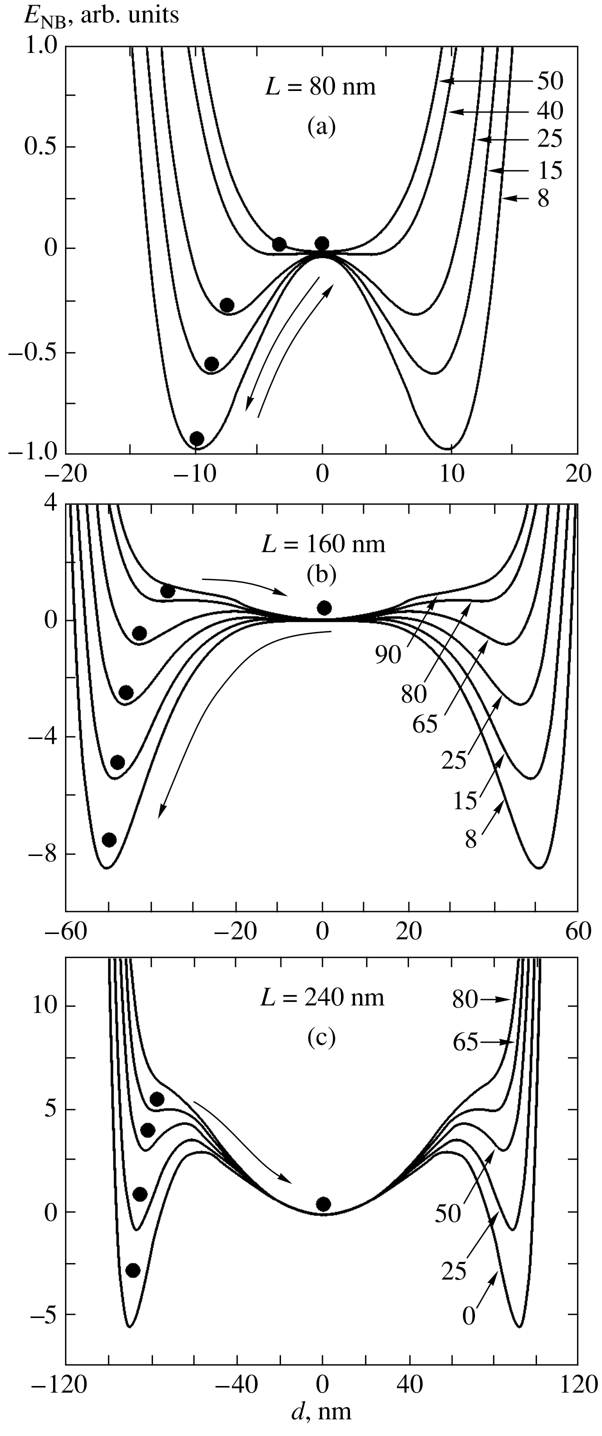


Fig. 9. Nanobridge energy as a function of the DW displacement d from the center of the crossbar calculated in a semiphenomenological model for the crossbar length L equal to (a) 80, (b) 160, and (c) 240 nm and for various values of the anisotropy constant K (indicated in kerg/cm^3 near the curves).

The normalization factors are introduced for the anisotropy, exchange interaction, and demagnetizing field $\mathbf{H}_m^k$ of each magnetic object separately, because the one-dimensional model (region S_2) used to estimate the energy of the banks is not equally effective for the contributions from the different interactions to this energy. For example, the contributions from the field of the magnetic dipole that forms in the crossbar and from the field produced by the space charge of the crossbar are different, because these fields decay differently with distance and also differ in terms of their direction, as follows from Eq. (2).

The fields of the magnetic dipole and the space charge of the crossbar and the field produced by the charges on the vertical faces of the banks are calculated using the formulas for the fields of a point magnetic dipole, a point charge, and a charged line, respectively. The divergence at the lower limit of integration of the form $H_m \propto r^{-n}$ is eliminated using the substitution

$$H_m \propto \frac{1}{(r^2 + \delta^2)^{n/2}}. \qquad (13)$$

Depending on its physical meaning, the parameter δ is chosen to be equal either to the nanobridge thickness a, for the fields of the magnetic dipole and the vertical faces of the banks, or to $b/2$, for the field of the space charge of the crossbar. Replacing δ by a quantity lying in the range (0.5–1.5)δ has almost no effect on the results of calculations.

It is assumed that the magnetization distribution in the crossbar and in the banks is described by Eq. (10). Inside the crossbar, the parameter c is taken to be $0.65b$ (which is the exact value of c at K = 40–60 kerg/cm^3). The value of c in the banks is adjusted to describe the pattern of phase transformations most adequately. The continuity of the magnetization distribution is achieved by fitting the parameter φ_0.

As the DW approaches a bank, the linear charge density on the inner vertical faces of the bank decreases. For this reason, this density is taken to be equal to $\lambda\cos\tilde{\varphi}$, where $\tilde{\varphi}$ is the angle between the magnetization vector and the x axis at the boundary between the bank and the crossbar. It should be noted that, in a nonzero approximation, one should take into account that the magnetization lines beginning at the encircled regions of the outer vertical faces of a bank do not all terminate at the crossbar; part of them terminate at the inner vertical faces of the bank. Therefore, the space charge of the crossbar is taken to be $(1 + \cos\tilde{\varphi})\lambda b$.

5.2. *Energy Profiles*

Figure 9 shows the energy of the nanobridge plotted as a function of the DW displacement from the center of the crossbar for various values of the crossbar length L and anisotropy constant K. The DW position is indicated by a circle and corresponds to a minimum energy. There are three types of curves.

The curves in Fig. 9a have two symmetric minima, which merge into one (central) minimum as K increases. Therefore, for any value of K, there is a single (to within the symmetry) equilibrium position of the DW relative to the center of the crossbar. This situation corresponds to second-order phase transitions.

For the curves shown in Fig. 9b, there is one (central) minimum at large values of K and two symmetric minima at values of K close to zero. In the intermediate range of K values, there are three equilibrium positions of the DW. In this case, the actual position of the DW depends on its history (as shown in Fig. 9b). This situation corresponds to first-order phase transitions.

In Fig. 9c, the curves have three minima even for a zero anisotropy constant. If the DW is in the central equilibrium position, it will remain in this position for any positive value of the anisotropy constant.

Note that, for energy profiles similar to those shown in Fig. 9a, the transformations of the magnetic structure of the nanobridge can be formally described in terms of the Landau theory of phase transitions. In this case, the order parameter is the DW displacement d and the coefficients of the expansion of the free energy in powers of d are determined by the geometric and physical parameters of the system, including the anisotropy constant, which is varied in our computer simulation. Near the tricritical point, the Landau theory can also be used to describe first-order phase transitions, as was done in [19]. This approach is inappropriate, however, in the case where the critical values of the anisotropy constant in the lability branches are significantly different.

5.3. Phase Diagram

The energy profiles depend on the crossbar length L, anisotropy constant K, and normalization factors n_i. For fixed values of the factors n_i, we can find the state of the nanobridge for any values of the crossbar length L and anisotropy constant K and thus construct the phase diagram of the system. Even this rough model (with values of n_i kept fixed over the entire range of L and K values) can qualitatively describe the pattern of phase transformations in the nanobridge. Figure 10 shows the phase diagram obtained within this model for certain values of the normalization factors.

6. THE MECHANISM OF PHASE TRANSFORMATIONS IN THE MAGNETIC NANOBRIDGE

We can separate the following competing interactions that determine the DW position:

(1) attraction between the DW and a bank that arises because the magnetostatic energy of the bank decreases as the DW approaches it and

(2) repulsion between the DW and a bank associated with (a) the magnetostatic interaction between the magnetic moments of the bank and DW and (b) an increase in the exchange energy of the system or (c) an increase in the magnetic anisotropy energy caused by the DW approaching the bank.

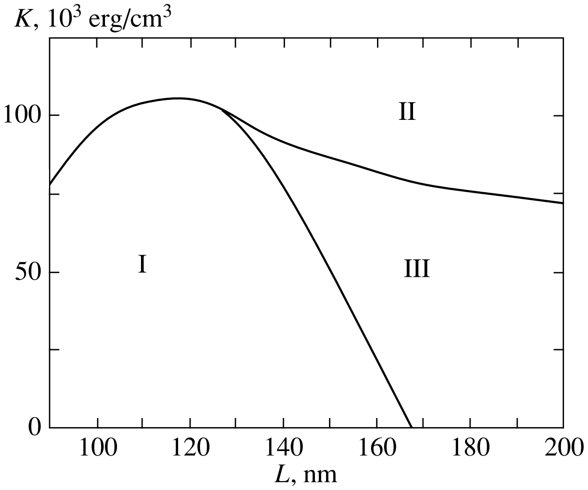


Fig. 10. Phase diagram calculated using a semiphenomenological model. Regions I–III are the same as in Fig. 4.

Interactions 1, 2b, and 2c are short-ranged, with their range being determined by the DW size. The forces exerted on the DW due to these interactions decay exponentially with the distance from a bank and vanish at a certain distance. In contrast, the magnetostatic repulsion (interaction 2a) is long-ranged, because the total demagnetizing field decreases as (or more slowly than) the inverse square of the distance.

Thus, if the crossbar is sufficiently long, the only force the banks exert on the DW located at the center of the crossbar is magnetostatic repulsion. Therefore, at the center, the DW has a minimum potential energy and is in a stable equilibrium position for any value of the anisotropy constant (Fig. 9c).

If the DW is near a bank at the same crossbar length, then the force of magnetostatic attraction (interaction 1) will be stronger than the force of repulsion (interactions 2a–2c) for a sufficiently small anisotropy constant. In this case, there will be a second stable equilibrium position. As the anisotropy constant increases, the side minimum disappears gradually and the DW jumps into the center (Fig. 9c). This explains the existence of a region in the phase diagram where there is only one hysteresis branch.

At smaller values of the crossbar length and a sufficiently small anisotropy constant, the attractive force (interaction 1) causes the DW to shift from the center. As the anisotropy constant increases, an energy minimum appears at the center of the crossbar; therefore, there are three equilibrium positions in this case: one is at the center and two are closer to a bank (Fig. 9b). The range of intermediate values of the anisotropy constant in which there are three energy minima corresponds to first-order phase transitions. In this case, the dependence of the DW position on K exhibits hysteresis.

As the crossbar length decreases still further, the relative contribution from interactions 2a and 2b to the

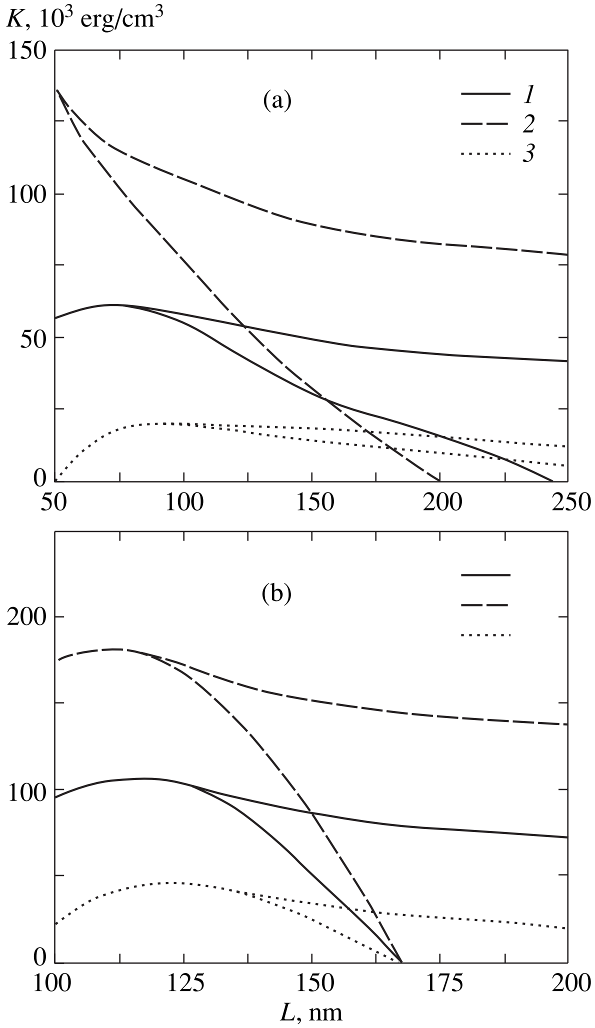


Fig. 11. Phase diagrams of (*1*) the initial nanobridge, (*2*) a nanobridge with $a = 1.5$ nm, and (*3*) a nanobridge with $M_S = 600$ emu/cm^3. The diagrams are calculated using (a) the micromagnetic method and (b) a semiphenomenological model.

potential energy of the DW at the center increases. As a result, the values of the anisotropy constant in the two lability branches approach each other. Under certain conditions,[2] both values of the anisotropy constant coincide at a certain crossbar length (corresponding to the tricritical point) and the energy profile of the system has two symmetric minima (Fig. 9a). At these values of the crossbar length, second-order phase transitions occur in the system.

7. PREDICTION OF THE FEATURES OF THE PHASE DIAGRAM OF THE NANOBRIDGE

Based on the mechanism of phase transformations in the magnetic nanobridge described above, we can predict the behavior of the system as its parameters are varied. For example, if the thickness *a* of the nanobridge increases slightly, then, according to Eq. (4), all magnetic charges will increase and, as a result, the magnetization distribution in the nanobridge will change. However, this change is a second-order effect with respect to the variation in the charge. Taking into account only first-order effects, we obtain the following pattern: (i) interactions 2b and 2c change only weakly, because they depend only on the magnetization distribution, and (ii) interactions 1 and 2a are enhanced with increasing magnetic charges.

Thus, the relative contribution from the attractive forces increases in comparison with that in the initial configuration of the system. Therefore, we may expect that the phase diagram will shift upward along the *K* axis, because an increase in the anisotropy constant is required to compensate for the increased repulsive forces.

Similarly, if the saturation magnetization M_S decreases slightly, then the magnetic charges will decrease and the phase diagram will shift downward along the *K* axis.

Figure 11a shows diagrams calculated using the micromagnetic technique for a nanobridge with thickness $a = 1.5$ nm and a nanobridge with saturation magnetization $M_S = 600$ emu/cm^3 and, for comparison, the diagram for the initial nanobridge with $a = 1$ nm and $M_S = 800$ emu/cm^3. Figure 11b shows the corresponding phase diagrams calculated using the semiphenomenological model with the initial values of the factors n_i. It can be seen that the phenomenological model makes it possible to predict the variations in the phase diagram of the nanobridge.

8. CONCLUSIONS

We have performed a comprehensive theoretical study into the magnetic structure of the plane magnetic nanobridge.

Using computer simulation, the dependence of the domain wall position on the anisotropy constant *K* has been investigated. It was shown that, as the anisotropy constant varies, the domain wall is displaced from the center of the nanobridge crossbar and approaches a bank. This process can be either irreversible (with hysteresis) or reversible, similar to first- and second-order phase transitions. The phase diagram of the nanobridge was calculated in the crossbar length versus anisotropy constant coordinates. The curves along which phase transformations occur with varying *K* (i.e., the symmetry of the system is changed) are the boundaries of closed regions of the phase diagram that correspond to a set of stable magnetic states of real nanobridges.

Based on the results of the micromagnetic calculations, we have constructed a semiphenomenological analytical model describing the formation of the magnetic structure of the nanobridge. Using this model, the

---

[2] In some cases, only first-order phase transitions occur regardless of the value of the crossbar length.

mechanism of first- and second-order phase transformations was investigated and the phase diagram of the nanobridge was obtained. For second-order phase transformations and for first-order phase transformations near the tricritical point, the dependence of the magnetic structure of the nanobridge on its parameters can be formally described in terms of the Landau theory of phase transitions.

This model makes it possible to predict the behavior of the system with variations in its parameters. As an example, we have considered the case where the thickness and saturation magnetization of the nanobridge are varied. A comparison of the results of micromagnetic calculations and the predictions of the phenomenological model showed that this model is adequate.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project nos. 04-02-17600, 05-02-16997) and INTAS (grant no. 03-51-4943).

REFERENCES

1. N. Garcia, M. Munoz, and Y.-W. Zhao, Phys. Rev. Lett. **82**, 2923 (1999); G. Tatara, Y.-W. Zhao, M. Munoz, and N. Garcia, Phys. Rev. Lett. **83**, 2030 (1999); S. H. Chung, M. Munoz, N. Garcia, W. F. Egelhoff, and R. D. Gomez, Phys. Rev. Lett. **89** (28), 287203 (2002).
2. K. Miyake, K. Shigeto, K. Mibu, T. Shinjo, and T. Ono, J. Appl. Phys. **91** (5), 3468 (2002).
3. P. Bruno, Phys. Rev. Lett. **83**, 2425 (1999).
4. V. A. Molyneux, V. V. Osipov, and E. V. Ponizovskaya, Phys. Rev. B **65**, 184425 (2002).
5. J. M. D. Coey, L. Berger, and Y. Labaye, Phys. Rev. B **64**, 020407 (2001).
6. S. P. Li, W. S. Lew, J. A. C. Bland, L. Lopez-Diaz, C. A. F. Vaz, M. Natali, and Y. Chen, Phys. Rev. Lett. **88**, 087202 (2002).
7. N. Garsia, V. V. Osipov, and E. V. Ponizovskaya, Phys. Rev. B **64**, 184412 (2001).
8. A. K. Zvezdin, A. F. Popkov, K. A. Zvezdin, and L. L. Savchenko, Fiz. Met. Metalloved. **9**, 165 (2001).
9. K. A. Zvezdin and A. V. Khval'kovskiĭ, Zh. Tekh. Fiz. **74** (3), 37 (2004) [Tech. Phys. **49** (3), 318 (2004)].
10. A. A. Zvezdin and K. A. Zvezdin, Pis'ma Zh. Éksp. Teor. Fiz. **75** (10), 613 (2002) [JETP Lett. **75** (10), 517 (2002)].
11. K. A. Zvezdin, Abstract of Candidate's Dissertation (Inst. of General Phys. RAN, Moscow, 2001).
12. L. Belliard, J. Miltat, A. Thiaville, S. Dubois, J. L. Duvail, and L. Piraux, J. Magn. Magn. Mater. **190**, 1 (1998).
13. W. F. Brown, Jr., *Micromagnetics* (Interscience, New York, 1963; Nauka, Moscow, 1979).
14. M. E. Schabes, J. Magn. Magn. Mater. **95**, 249 (1991).
15. R. P. van Garkom, S. J. C. H. Theeuwen, K. P. Wellock, N. N. Gribov, J. Caro, and S. Radelaar, Appl. Phys. Lett. **74** (3), 422 (1999).
16. K. A. Zvezdin, Fiz. Tverd. Tela (St. Petersburg) **42** (1), 116 (2000) [Phys. Solid State **42** (1), 120 (2000)].
17. Y. Nakatani, Y. Uesaka, and N. Haiashi, Jpn. J. Appl. Phys. **28**, 2485 (1989).
18. A. K. Zvezdin, V. M. Matveev, A. A. Mukhin, and A. I. Popov, *Rare-Earth Ions in Magnetooptic Crystals* (Nauka, Moscow, 1985) [in Russian].
19. K. P. Belov, A. K. Zvezdin, and A. M. Kadomtseva, *Spin Reorientation Transitions in Rare-Earth Magnets* (Nauka, Moscow, 1979) [in Russian].

Translated by Yu. Epifanov

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 1186–1190. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 1147–1150.
Original Russian Text Copyright © 2005 by Askhabov.

FULLERENES AND ATOMIC CLUSTERS

The Kvataron Model of Fullerene Formation

A. M. Askhabov
Institute of Geology, Ural Division, Russian Academy of Sciences, Pervomaĭskaya ul. 54, Syktyvkar, 167982 Russia
e-mail: xmin@geo.komisc.ru

Received May 17, 2004; in final form, August 2, 2004

Abstract—A new mechanism is proposed for the formation of fullerenes from clusters of a “latent” phase, which are referred to as kvatarons. According to this mechanism, hollow kvatarons initially arise in a supersaturated carbon-containing medium and then transform into rigid clusters (fullerenes) with characteristic icosahedral symmetry due to the formation of bonds between carbon atoms. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

A considerable amount of time has elapsed since the discovery of the stable carbon cluster consisting of 60 atoms and having a characteristic icosahedral symmetry [1], according to which this cluster was given the name fullerene C_{60}. However, the mechanism of formation of the C_{60} fullerene and other carbon clusters belonging to the fullerene family is still not clearly understood. In particular, the well-known scheme of assembling fullerenes from planar graphite fragments was not confirmed in elegant experiments performed by Ebbesen *et al.* [2], who attempted to answer the question as to whether fullerenes are formed by individual carbon atoms or fragments of graphite layers. In those experiments, the graphite layers were composed of ^{12}C carbon isotopes, whereas ^{13}C isotopes played the role of individual atoms. Hence, it could be expected that the formation of fullerenes from ^{12}C fragments should not be accompanied by the mixing of carbon isotopes; however, this mixing was observed experimentally. A more nonrealistic model was proposed by Kroto [3], who assumed that graphite layers should be initially curved. This model implies that the fullerene is formed as a result of accidental closure of curved carbon clusters that progressively increase in size. A number of models describing the fullerene formation [4–6] are based on the assumption that the fullerene-forming medium contains cluster precursors of fullerenes. In particular, Alekseev and Dyuzhev [6] assumed that the most probable precursor of fullerenes is a two-ring cluster with a single bond between the rings. According to the experimental data obtained by Tomilin *et al.* [7], the microclusters C_2 and C_{10} are synthesized first and then either an intermediate cluster nucleus or a final lower fullerene is assembled from them. Von Helden *et al.* [8] and Lozovik and Popov [9] suggested that a carbon cluster grows through the following sequence of transformations: linear chain ⟶ monocyclic ring ⟶ three-dimensional polycyclic cluster ⟶ fullerene. The kinetic model proposed by Krestinin *et al.* [10] for condensation of carbon vapors implies that a fullerene cage grows only due to the incorporation of C_2 fragments.

A detailed critical analysis of the existing models of carbon cluster formation suggests that the most probable way to transform carbon clusters into fullerenes is to crystallize a “liquid” cluster [9]. This hypothesis has been confirmed by the nuclear magnetic resonance studies of C_{60} fullerenes enriched in ^{13}C isotopes, according to which atoms that are neighbors in amorphous carbon clusters are not neighbors in fullerenes [10, 11]. However, up to now, the mechanism of formation of such a liquid-cluster precursor containing carbon atoms in amounts necessary for assembling fullerenes has defied explanation.

In this work, we proposed a radically new mechanism of fullerene formation from so-called “latent”-phase clusters, i.e., kvatarons, whose existence was recently substantiated in [12, 13].

2. THE KVATARON MODEL OF FULLERENE FORMATION

The energy of formation of a cluster involving n particles can be written as

$$\Delta G = S\sigma - n\varepsilon, \tag{1}$$

where S is the surface area of the cluster, σ is the interfacial energy per unit area, and ε is the energy per particle (bonding energy). Taking into account that the interfacial energy per unit area σ depends on the radius of curvature r, expression (1) for a spherical cluster formed by atoms of diameter δ takes the following form:

$$\Delta G = 4\pi r^2\sigma_0\left(1 - \frac{2\delta}{r}\right) - 8\left(\frac{r}{\delta}\right)^3\varepsilon. \tag{2}$$

Here, the dependence of the interfacial energy σ on the radius of curvature r is given by the relationship

$$\sigma = \sigma_0\left(1 - \frac{2\delta}{r}\right), \tag{3}$$

where σ_0 is the interfacial energy per unit area for a planar interface. Relationship (3) is a good approximation of the dependence $\sigma(r)$ for radii of curvature r lying in the range from molecular sizes to infinity. This relationship differs from the classical Tolman formula [14] (which is valid for $r \gg \delta$) and has a form similar to the expression obtained by Reiss *et al.* [15] for molecular particles in the framework of the scale-particle theory. It is this relationship between σ and r that most adequately describes the size dependence of the surface energy for small-sized particles [16, 17]. It should be noted that, within the quasi-thermodynamic approach, the parameter δ (the Tolman length), as a rule, is interpreted as the thickness of the interface region. In relationship (3), this parameter has the meaning of the minimum distance within which a cluster atom and an atom of the environment can come close to each other without forming bonds between them. Therefore, the parameter δ is approximately equal to the diameter of cluster-forming atoms.

Taking into account this estimate of the parameter δ, from expression (2), we can determine the number of atoms in the cluster; that is,

$$n = \frac{\frac{4}{3}\pi r^3}{\frac{4}{3}\pi(\delta/2)^3} = 8\left(\frac{r}{\delta}\right)^3. \tag{4}$$

It is interesting to note that the quantity $2\delta/r$ in relationship (3) has the meaning of the ratio between the number of surface atoms and the total number of atoms in the cluster. Actually, the number of surface atoms is determined by the equality

$$n_s = \frac{4\pi r^2}{\pi(\delta/2)^2} = 16\left(\frac{r}{\delta}\right)^2. \tag{5}$$

Hence, from formulas (4) and (5), we have $n_s/n = 2\delta/r$.

For the critical cluster size, which is determined by the condition $\partial\Delta G/\partial r = 0$, we obtain

$$\varepsilon = \frac{\pi\sigma_0\delta^3}{3r}\left(1 - \frac{\delta}{r}\right). \tag{6}$$

Substitution of this relationship into formula (2) gives the following expression for the energy of cluster formation:

$$\Delta G = \frac{4}{3}\pi r^2\sigma_0\left(1 - \frac{4\delta}{r}\right). \tag{7}$$

Expression (7) differs from the standard Gibbs formula for the energy of formation of crystal nuclei in that it allows for the spontaneous formation of clusters (at $\Delta G < 0$) whose radii satisfy the inequality $r < 4\delta$. These clusters are not nuclei of a new phase in the ordinary sense; most likely, they are noncovalently bonded intermediate aggregates. In [12], these specific clusters were referred to as latent-phase clusters or kvatarons.

In the most general form, the geometric interpretation of kvatarons is reduced to their interpretation in terms of Delaunay systems, or (R, r) systems. The arrangement of atoms in kvatarons is not strictly fixed. However, atoms in kvatarons can neither be farther apart from each other than the distance R in the (R, r) system nor be closer together than the distance r. These two conditions, which determine the Delaunay system [18], are entirely correct for kvatarons. Therefore, kvatarons can be treated as finite fragments of the Delaunay system, much as crystal nuclei are considered finite fragments of infinite crystal structures. Note that the distances R and r in kvatarons are specified within certain limits: the distance R does not exceed the parameter δ, and the distance r is no shorter than a critical distance close to twice the ionic or covalent radius of the atoms involved in the kvatarons.

It is known that crystals are considered a special case of the (R, r) systems and adhere to the axiom of regularity. A necessary and sufficient condition for a crystal to be formed is the existence of a local regularity within a region identical to the sphere of radius $4R$ (the local theorem) [19]. Since the parameter δ is nearly identical to the distance R in the kvataron (where R is the maximum interatomic distance at which the kvataron remains intact), the quantity $4\delta \approx 4R$ determines the region in which the system should become locally regular. Once this situation occurs, the cluster transforms into a crystal nucleus [20]. Thus, the fundamental importance of the local theorem lies in the fact that it determines the minimum number of atoms required from the system to form a crystal. Therefore, only a kvataron of size $r = 4\delta$, which, according to formula (4), contains 512 atoms, can be treated as a potential crystal nucleus. Clusters of smaller radius have an amorphous structure and are quasi-liquid aggregates.

Similarly, the energy of formation of hollow kvatarons with the inner and outer surfaces can be represented in the form

$$\Delta G = \frac{8}{3}\pi r^2\sigma_0\left(1 - \frac{4\delta}{r}\right). \tag{8}$$

Comparison of relationships (7) and (8) shows that the formation of hollow kvatarons is energetically more favorable at $r < 4\delta$, whereas dense clusters are formed at $r > 4\delta$. In actual fact, the filling of the kvataron volume occurs even at $r > 2\delta$, when, in accordance with expression (3), the swelling of the kvataron gives way to the action of a compressive force (the Laplace pressure).

If the above hypothesis is true, it would appear reasonable that hollow kvatarons can serve as precursors

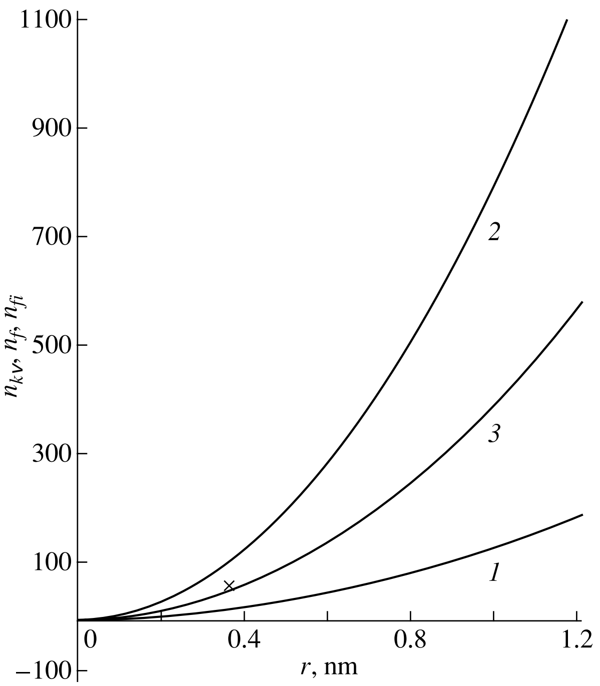


Fig. 1. Number of atoms located on (*1*) the surface of a kvataron n_{kv}, (*2*) the surface of a limiting cluster n_f, and (*3*) the surface of a fullerene n_{fi} as a function of the radius r of carbon clusters. The cross indicates the number of atoms on the surface of the C_{60} fullerene.

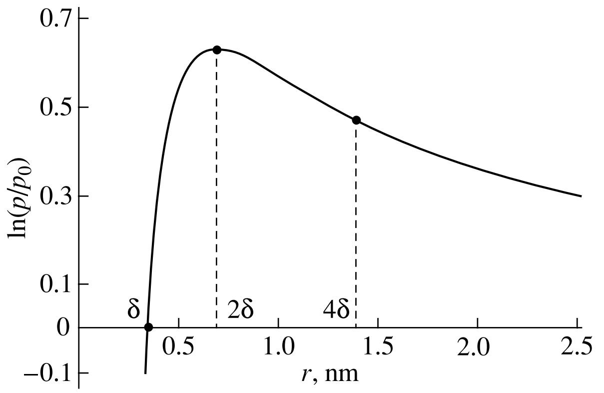


Fig. 2. Correlation between the radius of hollow carbon kvatarons and the supersaturation of the medium.

of fullerenes. In this case, the formation of fullerenes proceeds according to a very simple scenario: quasi-liquid kvatarons initially arise in a supersaturated medium and then, owing to the minimization of the kvataron energy (the formation of bonds between atoms), transform into rigid clusters (fullerenes) with fixed distances R and r in terms of the (R, r) system. This scheme excludes the sequential assembling of a fullerene cage from individual atoms and is seen to be an inevitable step in the evolution of the kvataron structure. The feasibility of the proposed scenario for the transformation of kvatarons into fullerenes was demonstrated by the molecular dynamics simulation.[1]

The maximum number n_{kv} of atoms occupying the surface of a kvataron of radius r can be determined from the condition

$$n_{kv} = \frac{4\pi r^2}{\pi(\delta/2)^2} = 16(r/\delta)^2. \quad (9)$$

In the case where the distance between atoms decreases to a value δ_1, the number n_f of atoms located on the surface of a rigid cluster formed in the process is given by

$$n_f = 16(r/\delta_1)^2. \quad (10)$$

According to the Euler theorem, a rigid icosahedral cluster can consist of 12 pentagons and $(n-20)/2$ hexagons. Hence, the number n_{fi} of atoms located on the surface of rigid icosahedral clusters is smaller than the number n_f and is determined by the formula

$$n_{fi} = n_f - \left(12 + \frac{n_f - 20}{2}\right) = \frac{n_f}{2} - 2. \quad (11)$$

The correlation of the numbers n_{kv}, n_f, and n_{fi} of atoms located on the surfaces of the corresponding clusters is shown in Fig. 1.

The radius of kvatarons depends on the supersaturation of the medium. In our model, the analog of the Kelvin equation that relates the supersaturation of a gas phase to the radius of hollow clusters has the following form:

$$\ln\frac{p}{p_0} = \frac{4\pi\delta^3\sigma_0}{3rkT}\left(1 - \frac{\delta}{r}\right), \quad (12)$$

where k is the Boltzmann constant, T is the temperature, and the supersaturation is expressed through the ratio between the vapor (gas) phase pressure p over the kvataron and the equilibrium vapor pressure p_0 over the flat surface.

Now, we substitute the following parameters into expression (12): $\delta \approx 0.345$ nm [this value, according to the physical meaning of the parameter δ, is somewhat smaller than the van der Waals diameter of the carbon atom (0.354 nm) but exceeds the length of the weakest π bond in graphite (0.335 nm)], $\sigma_0 = 350$ erg/cm^2 (calculated from the energy of the interaction between the networks of carbon atoms in graphite), and $T = 2500$ K. As a result, we obtain

$$\ln\frac{p}{p_0} = \frac{0.87}{r}\left(1 - \frac{0.345}{r}\right). \quad (13)$$

The curve calculated from relationship (13) is plotted in Fig. 2. As can be seen, this curve involves three charac-

---

[1] The results of the molecular dynamics simulation and the kinetic aspects of the fullerene formation will be published in a separate paper.

teristic points: (i) $r = \delta$ is the size of kvatarons under equilibrium conditions, (ii) $r = 2\delta$ is the radius of kvatarons at the maximum (limiting) supersaturation, and (iii) $r = 4\delta$ is the final size of kvatarons formed spontaneously.

Since the formation of fullerenes proceeds under nonequilibrium non-steady-state conditions, there exist several variants of the evolution of kvatarons with a change in their radius r and the number n of atoms located on the kvataron surface: (i) $r \uparrow n\uparrow$; (ii) $r \downarrow n\downarrow$; (iii) $r \downarrow n\uparrow$; (iv) $r \uparrow n$; (v) $r \uparrow n$; (vi) $r \downarrow n$; (vii) $r, n \uparrow$; (viii) $r, n \downarrow$; and (ix) r, n. It is clear that not all these variants lead to the formation of fullerenes. For example, an increase in the radius ($r \uparrow$) with a constant (n) or decreasing ($n\downarrow$) number of atoms [variants (iv), (v)] results in the decay of kvatarons, whereas a decrease in the radius at a constant number of atoms necessarily brings about the formation of fullerenes from kvatarons.

Under equilibrium conditions, the kvataron has a radius $r = \delta$. The maximum number of atoms located on the surface of this kvataron is determined to be $n = 16(r/\delta)^2 = 16$ atoms. Since the formation of covalent bonds between atoms occurs in a jumpwise manner, the interatomic distance decreases to $\delta_1 = 0.1415$ nm (this value is equal to the mean length of C–C bonds in fullerenes). As a result, the radius of a rigid limiting cluster formed in the process decreases to 0.138 nm. An increase in the number of atoms located on the surface of a kvataron with a constant radius can lead to the formation of a cluster containing 95 atoms ($r = 0.345$ nm). Therefore, the kvataron of radius δ can be a precursor to the formation of a large family of carbon clusters from C_{16} (at $n, r \downarrow$) to C_{95} (at $n \uparrow r$). This range of carbon clusters can be even wider provided both parameters n and r change simultaneously. Under conditions of a limiting (maximum) supersaturation, the kvataron has a radius $r = 2\delta = 0.69$ nm and can form rigid carbon clusters over a wide range from C_{64} (at $n, r \downarrow$) ($r = 0.276$ nm) to C_{380} (at $n \uparrow r$) ($r = 0.69$ nm). The kvataron of radius $r = 4\delta$ can provide the formation of carbon clusters in the range from C_{256} to C_{1522}. However, it is unlikely that the upper limit can be reached because kvatarons of such large sizes become double-walled, triple-walled, or dense aggregates and, consequently, form multilayer embedded structures, soot particles, or crystal nuclei (graphite, diamond).

If the medium is not supersaturated, the inequality $r < \delta$ is satisfied and the clusters contain a smaller number of atoms. For example, kvatarons whose radius falls in the range $\delta/2 < r < \delta$ form clusters containing from 3 to 23 atoms. It is interesting that, in some cases, the mass spectra of carbon clusters exhibit a double-hump distribution. One hump of this distribution in the range C_3–C_{23} can be associated with the clustering under conditions of incomplete saturation. This can be clearly seen in particular from the familiar figure given in [21].

Therefore, the carbon clusters, which are referred to as kvatarons, provide a basis for the formation of a wide variety of carbon clusters from C_{16} to C_{1522}. Among these clusters are the most stable aggregates with the highest symmetry and the smallest number of contacting pentagons.

It should be emphasized that the formation of the C_{60} fullerene is due to the evolution of both the kvataron of radius $r = \delta = 0.345$ nm (the initial number of atoms is 16) at $r \uparrow n \uparrow$ and the kvataron of radius $r = 2\delta = 0.69$ nm (the initial number of atoms is 64) at $r \downarrow n \downarrow$, with the second process being dominant. Consequently, the relatively widespread occurrence of the first magic fullerene C_{60} is not accidental. This is associated not only with the minimum energy of the C_{60} fullerene and its high symmetry responsible for the stability of the structure but also with the probable attainment of the limiting supersaturation under experimental conditions where clusters contain carbon atoms in amounts necessary for the formation of the architecture inherent in the C_{60} fullerene. Possibly, the necessity of preliminarily reaching the limiting supersaturation at which kvatarons have a radius $r = 2\delta$ is the main reason why other elements virtually do not form fullerene-like clusters of the C_{60} type. In particular, the limiting supersaturation for silicon, according to formula (12), is considerably higher than that for carbon because of the larger values of δ and σ_0.

In conclusion, one additional remark needs to be made. It is known that, according to x-ray diffraction data, the radius of the C_{60} fullerene is equal to 0.357 nm [22], which is several percent smaller than the radius predicted from formula (12). This discrepancy can be associated with the fact that the true shape of the C_{60} fullerene differs from the ideal icosahedron whose radius is determined from formula (12). Moreover, it is not improbable that this discrepancy results from the contraction of the fullerene due to electron density transfer into the fullerene cage.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 02-05-64883) and the Presidium of the Russian Academy of Sciences within the program “Fundamental Problems of the Physics and Chemistry of Nanosystems and Nanomaterials” [project “Cluster (Kvataron) Self-Organization of Matter on the Nanolevel and the Formation of Nanosystems of Crystalline and Noncrystalline Materials”].

REFERENCES

1. H. W. Kroto, J. R. Heath, S. C. O’Brian, R. E. Curl, and R. E. Smally, Nature **318**, 162 (1985).
2. T. W. Ebbesen, J. Tabuchi, and K. Tanigaki, Chem. Phys. Lett. **191**, 336 (1992).
3. H. W. Kroto, Science **242**, 1139 (1988).

4. T. Wakabayashi and J. Achiba, Chem. Phys. Lett. **190**, 465 (1992).
5. J. R. Dias, Chem. Phys. Lett. **209**, 439 (1993).
6. N. I. Alekseev and G. A. Dyuzhev, Zh. Tekh. Fiz. **71** (5), 67 (2001) [Tech. Phys. **46** (5), 577 (2001)].
7. F. N. Tomilin, P. V. Avramov, S. A. Vargonov, A. A. Kuzubov, and S. G. Ovchinnikov, Fiz. Tverd. Tela (St. Petersburg) **43** (5), 936 (2001) [Phys. Solid State **43** (5), 973 (2001)].
8. G. von Helden, M.-T. Hsu, N. G. Gotts, and M. T. Bowers, J. Phys. Chem. **97**, 8182 (1993).
9. Yu. E. Lozovik and A. M. Popov, Usp. Fiz. Nauk **167** (7), 751 (1997) [Phys. Usp. **40** (7), 717 (1997)].
10. A. V. Krestinin, A. P. Moravskiĭ, and P. A. Tesner, Khim. Fiz. **17** (9), 74 (1998).
11. J. M. Hawkins, A. Meyer, S. Loren, and R. Nunlist, J. Am. Chem. Soc. **113**, 9394 (1991).
12. A. M. Askhabov and M. A. Ryazanov, Dokl. Akad. Nauk **362** (5), 630 (1998).
13. A. M. Askhabov, *The Kvataron Concept of Cluster Self-Organization of Matter on the Nanolevel in Solving Problems of Crystallography, Mineralogy, and Related Sciences* (Geoprint, Syktyvkar, 2003) [in Russian].
14. R. C. Tolman, J. Chem. Phys. **16**, 758 (1948).
15. H. Reiss, H. L. Frisch, E. Helfand, and J. L. Lebowitz, J. Chem. Phys. **32** (1), 119 (1960).
16. A. D. Pomogaĭlo, A. S. Rozenberg, and I. E. Uflyand, *Metal Nanoparticles in Polymers* (Khimiya, Moscow, 2000) [in Russian].
17. A. I. Vavin and A. M. Askhabov, Zh. Fiz. Khim. **77** (11), 2103 (2003).
18. R. V. Galiulin, *Crystallographic Geometry* (Nauka, Moscow, 1984) [in Russian].
19. B. N. Delaynay, N. P. Dolbilin, M. I. Shtogrin, and R. V. Galiulin, Dokl. Akad. Nauk SSSR **227** (1), 19 (1976).
20. A. M. Askhabov and R. V. Galiulin, Dokl. Akad. Nauk **363** (4), 513 (1998).
21. E. A. Rohlfing, D. M. Cox, and A. Kaldor, J. Chem. Phys. **81**, 3322 (1984).
22. W. Kratschmer, K. Fostiropoulos, and D. R. Huffman, Chem. Phys. Lett. **170**, 167 (1990).

Translated by O. Borovik-Romanova

Physics of the Solid State, Vol. 47, No. 6, 2005, pp. 995–1007. Translated from Fizika Tverdogo Tela, Vol. 47, No. 6, 2005, pp. 961–973.
Original Russian Text Copyright © 2005 by Golovin, Dub, Ivolgin, Korenkov, Tyurin.

PROCEEDINGS OF THE CONFERENCE DEDICATED TO S. N. ZHURKOV (1905–1997)

Kinetic Features of the Deformation of Solids in Nano- and Microscopic Volumes

Yu. I. Golovin*, S. N. Dub, V. I. Ivolgin*, V. V. Korenkov*, and A. I. Tyurin***
** Tambov State University, Internatsional'naya ul. 33, Tambov, 392622 Russia*
e-mail: golovin@tsu.tmb.ru
*** Institute for Superhard Materials, National Academy of Sciences of Ukraine,*
Avtozavodskaya ul. 2, Kiev, 04074 Ukraine

Abstract—The deformation dynamics of nanovolumes in crystalline, quasicrystalline, and amorphous solids is studied experimentally using continuous nanoindentation with a resolution of 0.1 nm. The elastic limits of some materials are determined in a nanocontact region. A jumplike transition to a plastic flow (which is equivalent to a drop in yield in uniaxial macroscopic tests) is revealed and studied. The dynamics and statistics of an unstable plastic flow in strain-aging alloys are analyzed. The specific features of a local stress–strain curve associated with a phase transformation under an indenter and with microcrack nucleation are revealed. The load-carrying ability of a material upon nanocontact loading is shown to be many times its macroscopic yield strength and to approach the theoretical ultimate strength even in plastic materials. The relaxation processes occurring in submicrovolumes after unloading are found to induce an elastic aftereffect that is much larger than that in macroscopic tests. In Si and Ge single crystals, the effect of a jumplike increase in the cracking resistance with the strain rate is detected. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The kinetic approach to the strength and plasticity of solids developed by Zhurkov and his colleagues and followers over several decades as an alternative to various theories of the "limiting state" [1] has attracted considerable interest. Moreover, application of the powerful arsenal of modern physical methods for investigating the structural dynamics of solids on the mesoscopic, dislocation, atomic, and, currently, electron–spin scales has transformed the initial semi-intuitive concepts of the development of deformation and fracture in time into an experimentally verified system of knowledge of the mechanical behavior of solids as a multistage cascade of interrelated events at several hierarchical structural levels [2–4]. The difficulty with this approach consists in the fact that it is necessary to analyze objects and processes characterized by sizes L and relaxation times τ that differ by many orders of magnitude (up to 10^{10} times in L and up to 10^{20} times in τ). These large ranges are due to the wide variety of structural defects and their interactions responsible for damage accumulation upon loading, to differences in the dynamics and motion mechanisms of deformation carriers, to the different degrees of collectivization and self-organization of events at low structural levels, etc.

Macroscopic plastic deformation always results from a huge number of interrelated dynamic processes occurring in the defect structure of a material. The actual kinetics of a plastic flow at the microscopic level is masked due to the large number of microscopic deformation events per unit time and to the sluggishness of traditional mechanical means of testing materials and, therefore, is difficult to study.

To experimentally investigate the dynamics of nanostructures, *in situ* methods with high spatial and time resolution are needed. Atomic force microscopy and its related test methods and the dynamic nanoindentation techniques developed intensively over the past decade have made it possible to study mechanical properties in nanovolumes and, in the limiting case, at the level of individual molecules or small atomic groups. This opens up fresh opportunities for solving a number of unsolved fundamental problems in the physics of strength and plasticity. The designing of nanomechanical recorders, devices for ultrahigh-density data storage, and other nanoelectromechanical systems and the progress in nanotechnology on the whole have also raised new questions regarding the nature of the mechanical properties of nanoscale objects.

The available scarce experimental data indicate that the mechanical properties of most materials in submicron volumes are different, especially in regions with characteristic sizes $L \leq 100$ nm. Therefore, it is necessary to study them and to reveal the physical causes and mechanisms of the scale effects and kinetic processes under these conditions.

Apart from nanotechnology and nanostructures, there are many other common situations where nanocontact interaction plays a key role: dry friction of rough surfaces, abrasive and erosion wear, milling to form nanoparticles, mechanical alloying and activation in ball mills, etc. In all these processes, short-term

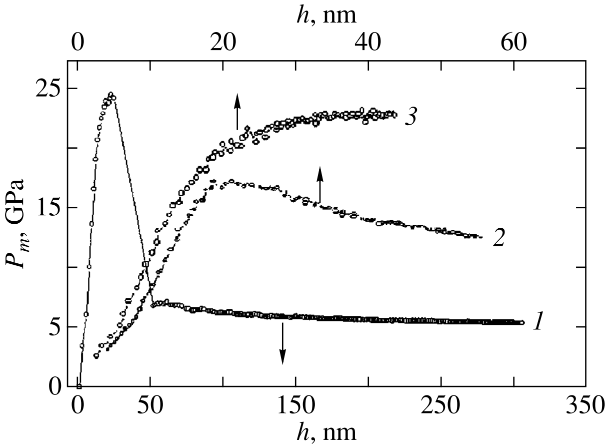


Fig. 1. Dependence of the mean contact pressure p_m on the indentation depth *h* in the initial stage of indentation for various materials: (*1*) W(001), (*2*) tetragonal metastable ZrO_2, and (*3*) cubic ZrO_2. In tungsten, a strain jump is detected at a pressure of 24.9 GPa.

intense nanocontact interactions take place and they can be modeled using dynamic nanoindentation.

In this work, we present data on the processes of transition from purely elastic to elastoplastic deformation and to collective behavior of dislocations, a viscoelastic aftereffect, phase transformations, and microcrack nucleation occurring in various crystalline, quasicrystalline, and amorphous materials under the conditions of local deformation of submicrovolumes.

2. EXPERIMENTAL

In this work, we use dynamic nanoindentation to study the mechanical properties and behavior of various solids at the submicron level. This method consists in precise local loading of a material surface with a certified probe (usually, a Berkovich diamond indenter) and continuous recording of the kinetics of its penetration with a resolution of a fraction of a nanometer. These data are used to construct a load–indentation depth (*P*–*h*) diagram, which is similar to the traditional σ–ε curve in its meaning. By analyzing the *P*–*h* diagram, one can derive about ten parameters characterizing the material at the nanoscopic level (including time-dependent parameters) and can even reconstruct the σ–ε curve using a special procedure.

A specific feature of local loading with a pyramidal indenter is a change in the characteristic indentation sizes $L \sim h$ and the material volume involved in deformation by several orders of magnitude during one test cycle. Therefore, the strain rate $\dot{\varepsilon} \approx (1/h)dh/dt$ also changes strongly with *h* during tests. At small values of *h* (~1–10 nm), indentation at even moderate linear velocities ($dh/dt \sim 1$ mm/s) produces deformation at rates $\dot{\varepsilon} \geq 10^5$ s^{-1}, which are characteristic of a surface explosion. Another specific feature of nanoindentation tests is the fact that the mean contact pressure in an indentation p_m (Meyer hardness) reaches a certain finite value in the elastic region. This value remains virtually constant upon subsequent loading, although the indentation depth increases (Fig. 1). In some materials, p_m reaches a maximum at $h \sim 10$–100 nm and then decreases smoothly or jumpwise (sometimes, by several times) and finally levels off. The nature of the scale effect at $h \leq 100$ nm is of particular interest. We also note that, except in this initial stage, deformation, to a first approximation, proceeds at a constant stress but involves a material volume ($V \sim h^3$) that increases by many orders of magnitude. All these factors make it possible to cover wide ranges of strain rates (from 10^5 to 10^{-3} s^{-1}) and of maximum indentation loads (from 0.01 mN to 2 N). Thus, unlike uniaxial tension or compression (where a sample with a constant volume is subjected to tests under increasing stresses and the strain increases at a constant rate), continuous nanoindentation can scan the deformation-zone sizes and the strain rates that vary by several orders of magnitude in one test (see table). Modern procedures of nanoindentation-data processing can separate the contributions

Comparison of two types of mechanical tests: continuous nanoindentation and uniaxial tension or compression

Characteristic	Test type	
	continuous indentation with a Berkovich indenter	uniaxial tension or compression
Machine type	“Soft”	“Stiff”
Character of macroscopic deformation	Nonuniform	Uniform
Maximum strain	$\varepsilon_{max} \approx \text{const} \approx 0.3$	$\varepsilon \longrightarrow \text{var}$
Maximum averaged stress	$\sigma = kP/h^2 \approx \text{const} \approx \text{H}$	$\sigma = P/d^2 \longrightarrow \text{var}$
Strain rate	$\dot{\varepsilon} \approx dh/hdt \approx V/h \longrightarrow \text{var}$	$\dot{\varepsilon} \approx \text{const}$
Deformed volume	$V \approx ch^3 \longrightarrow \text{var}$	$V = ld^2 \approx \text{const}$
Main features	Nonuniform deformation of a strongly increasing material volume under virtually static stresses	Uniform deformation of a constant volume under increasing stresses

from the scale and rate factors and give a large amount of information when one sample is subjected to nondestructive tests.

The problems that can be solved using nanoindentation methods are divided into two groups. One group includes methods for finding the boundaries of the scale invariance of mechanical properties and for determining the conditions under which destructive macroscopic tests can be replaced by nondestructive tests, which give the same information but are more productive and material-saving. The other group includes methods for revealing the character and causes of changes in the mechanical behavior in small-scale objects. In this work, we devote attention to both aspects of the application of nanoindentation for solving various problems.

To find the hardness H and the elastic modulus E from the indentation depth, we used the Oliver–Pharr technique [5–7]. This technique makes it possible to determine the actual indentation depth h_p from the measured indentation displacement h and calculate the indentation projection area A (which is needed to correctly determine $H = P/A$) with allowance for the actual shape of the indenter tip. With this technique, we can find the nanohardness only at one point on a loading curve, namely, at the maximum load, since only for this point can we determine the elastic bending flexure of the sample surface around an indentation. Researchers from the Institute for Superhard Materials designed a method for finding the mean contact pressure p_m for all points of a loading curve for a Berkovich indenter [8]; this method was also applied for processing the experimental data. These data were obtained using several models of nanohardness testers having different characteristics and possibilities: a NanoIndenter-II device (MTS Systems), a DUH-W201E ultramicrohardness tester (Shimadzu), and unique dynamic nanohardness testers designed at Tambov State University [9, 10]. These last testers have a high indenter-displacement resolution (~0.1 nm) and a time sampling of 0.3 μs (against 50 ms for DUH-W201E and 100 ms for NanoIndenter-II).

Structural studies and certification of the indenters were performed on a Solver-LS atomic force microscope operating in the contactless lateral mode. The sample surfaces were produced through quasi-brittle cleavage or mechanical polishing followed by chemical etching. These methods eliminate the dependence of results on the state of the surface layers that form during treatment or as a result of interaction with the environment. Check experiments showed that the parameters measured characterize the properties of a homogeneous material rather than of oxidized layers forming during sample preparation.

3. YIELD STRENGTH IN A NANOSCALE VOLUME

We studied the transition from elastic to elastoplastic deformation on a NanoIndenter-II nanohardness tester at very small limiting loads (~0.1–1 mN). Typical examples of P–h curves for materials that differ in terms of their type and hardness are shown in Fig. 2. The curves have a number of characteristic segments: a monotonic increase in h with increasing load, a jumplike increase in h, and a slow increase in h at P = const (creep). It should be noted that, in some materials, strain jumps occur not only in the stage of loading but also upon storage at a maximum load or even upon unloading (Fig. 2f). Although the Berkovich indenter is considered a cuspate trihedral pyramid (with an apex angle of 130.6°), the tip of a real indenter can be treated, to a first approximation, as a sphere with an effective radius R. From independent measurements, we determined the radius of our indenter to be R = 220 nm. Thus, to a depth of several tens of nanometers, our indentation tests are performed with a sphere rather than a trihedral pyramid. In this case, the load increases with depth according to Hertz's law $P \sim h^{3/2}$, which is characteristic of the interaction of a solid ball with a surface (Fig. 2a). In most cases, unloading to a strain jump results in completely reversible, i.e., purely elastic, behavior of a material (Fig. 3a). Some materials exhibit rather unusual irreversible deformation in the initial stage of indentation (Fig. 3b). Taking into account the specific features of this deformation, we may assume that it is caused by the generation and motion of nonequilibrium point defects.

Two types of deviations from Hertz's law are observed: either a loading curve goes below Hertz's curve beginning from a certain load or h increases jumpwise at P = const (a nanoindentometer is a "soft" testing machine maintaining a given force). In the latter case, we actually observe a local loss of stability that is analogous to a yield drop in stress–strain curves of certain materials subjected to uniaxial tension or compression in a "stiff" testing machine. The critical load P_{cr} and the jump amplitude Δh depend on many factors, such as the method of surface preparation, the degree of alloying, the dislocation density, the loading rate, the laboratory noise level, etc. For example, the surfaces prepared by cleavage or chemical etching are much more sensitive to jumps than are the mechanically treated (grinded, polished) surfaces and the former surfaces withstand a much higher load up to the onset of plastic flow (Fig. 2b). Jumplike deformation was detected even for the unpolished surfaces of KCl, Cu, Fe, and other single crystals, which are soft and smoothly deform under ordinary conditions (Fig. 2). An increase in the concentration of divalent Ca in KCl from 0.01 to 0.3 wt % leads to a fourfold increase in P_{cr} (Fig. 2a), whereas the hardness at $h \geq 100$ nm increases by less than 20% (from 0.24 to 0.28 GPa). As P_{cr} increases, the jump amplitude likewise increases. It

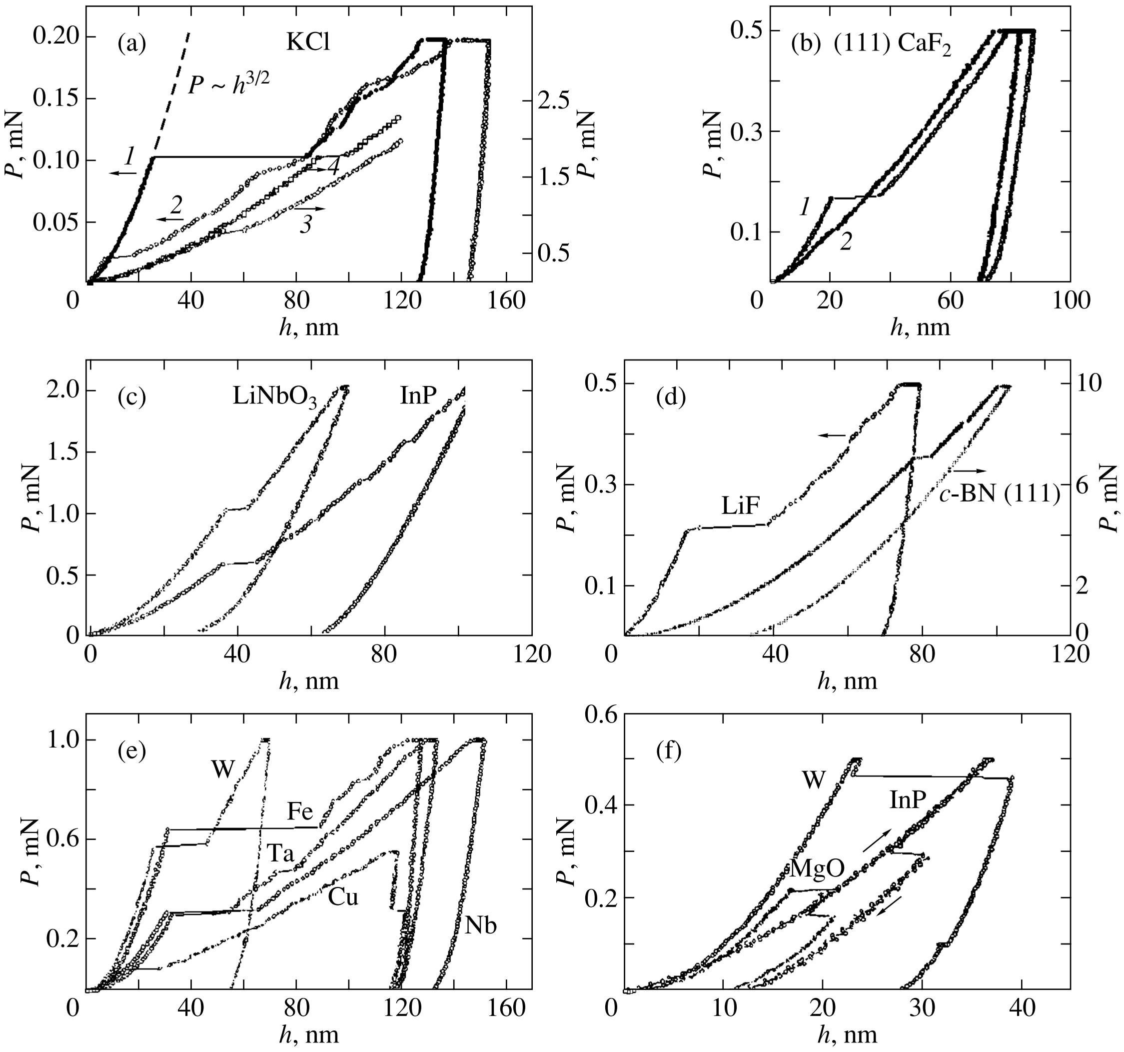


Fig. 2. Jumplike deformation under yield stress during nanoindentation of (a) KCl single crystals with various CaCl$_2$ impurity contents [(*1*) 0.01 and (*2*) 0.3 wt %] at strain rates $\dot{\varepsilon}$ equal to (*1*, *2*) 2, (*3*) 1.2×10^5, and (*4*) 1.7×10^5 s^{-1}; (b) CaF$_2$ single crystals with (*1*) cleavage surfaces and (*2*) mechanically polished surfaces; (c) InP and LiNbO$_3$ single crystals; (d) cubic boron nitride and lithium fluoride; and (e) metals with different hardnesses. (f) Strain jumps are possible not only during loading but also during unloading.

should be noted that, before a jump, the average stress in an indent (which is actually the yield strength in a nanoscopic volume) approaches $0.1E$ (i.e., the theoretical ultimate strength) even in these soft materials, whereas the macroscopic yield strengths of these materials are lower by one to two orders of magnitude! In other words, the load-carrying ability of a material under nanocontact conditions can exceed its macroscopic yield strength by many times. This factor increases the resistance of nanoscopic tools but creates difficulties upon treatment and local plastic deformation of substrates in various nanotechnologies [11].

Among all possible causes of the occurrence of large jumps at low loads ($P \leq 1$ mN), nucleation and motion of dislocations are assumed to be most probable. Let us consider the conditions of formation of dislocation loops under an indenter in the initial stage of indentation, when the indenter tip can be taken to be spherical. In various planes where dislocation loops can nucleate under an indenter, tangential stresses are operative [12]; their maximum is $\tau_{\max} = 0.31p_{\max}$, where $p_{\max} = (6E^{*2}P/\pi^3R^2)^{1/3}$ is the maximum contact pressure in the indentation $E^* = (1/E_m + 1/E_i)^{-1}$, and E_m and E_i are the Young's moduli of the material under study and the indenter, respectively. Note that the Meyer hardness is $H = (2/3)p_{\max}$. From energy considerations [13], it follows that shear stresses $\tau \geq Gb/r_{\mathrm{cr}}$ are necessary for the homogeneous nucleation of a dislocation loop. Therefore, we have

$$r_{\mathrm{cr}} \geq \frac{Gb}{0.31}\left(\frac{\pi^3R^2}{6E^{*2}P}\right)^{1/3}, \tag{1}$$

where G is the shear modulus, b is the Burgers vector, and r_{cr} is the critical radius of a nucleating dislocation loop. However, in the field of strongly inhomogeneous stresses, the following geometrical condition must be satisfied: the size r' of the region where stresses are sufficient for dislocation-loop nucleation must be greater than or equal to the radius r_{cr} of the critical loop. Assuming $r' = Ka$ (where $a = (\pi R/2E^*)(6E^{*2}P/\pi^3R^2)^{1/3}$ is the radius of the contact region and $K \approx 1$), we obtain

$$r_{cr} = \frac{K\pi R}{2E^*}\left(\frac{6E^{*2}}{\pi^3R^2}P\right)^{1/3}. \qquad (2)$$

By equating Eqs. (1) and (2), we find the critical value P_{cr} to be

$$P_{cr} = \left(\frac{2E^{*2}b}{K\pi R}\right)^{3/2}\frac{\pi^3R^2}{6E^{*2}} = 2.56E^*(b/K)^{3/2}R^{1/2}. \qquad (3)$$

The experimental dependence of P_{cr} on $E^*b^{3/2}R^{1/2}$ for a number of materials studied [14] agrees qualitatively with Eq. (3), i.e., with the model of homogeneous dislocation nucleation in nanoscopic volumes. From the jump Δh, the number of loops forming are estimated to be $N \approx \Delta h/b \approx 10$–$50$ for various materials, which seems reasonable. In actual practice, where the surface of a crystal is not perfectly planar and an indenter is not a perfect sphere, there may be stress concentrators that can decrease P_{cr}. However, the probability that such a concentrator exists in a nanoscopic volume subjected to loading is low. This conclusion is supported by the qualitative agreement between the experimental data and the results calculated using the model of homogeneous nucleation and also by the fact that, even after fine mechanical polishing, strain jumps disappear and the loading curve becomes smooth but lies below Hertz's curve almost from the very beginning (Fig. 2b).

Thus, the initial stage of nanoindentation has a purely elastic character even in plastic materials; as a result, before the jumplike transition to plastic flow, the contact stresses reach values $\tau_{max} \approx (0.05$–$0.10)G$, which are comparable to the theoretical ultimate strength. Since many modern elements of nanotechnology-based devices operate under conditions of dynamic nanocontact interaction (e.g., the probes in scanning atomic force microscopes), it is necessary to take into account a possible significant increase in the load-carrying ability of a material under these conditions.

The authors of [15, 16] described a technique for translating P–h diagram data into the traditional σ–ε coordinates for steels and even very brittle materials, which cannot be plastically deformed in uniaxial tests. If the scale factor is correctly taken into account, this technique allows partial or complete substitution for labor-consuming destructive tests of materials in standard macroscopic testing machines. Li [17] showed that a number of creep characteristics of metals could also be obtained using local loading under a constant load and continuous recording of indentation.

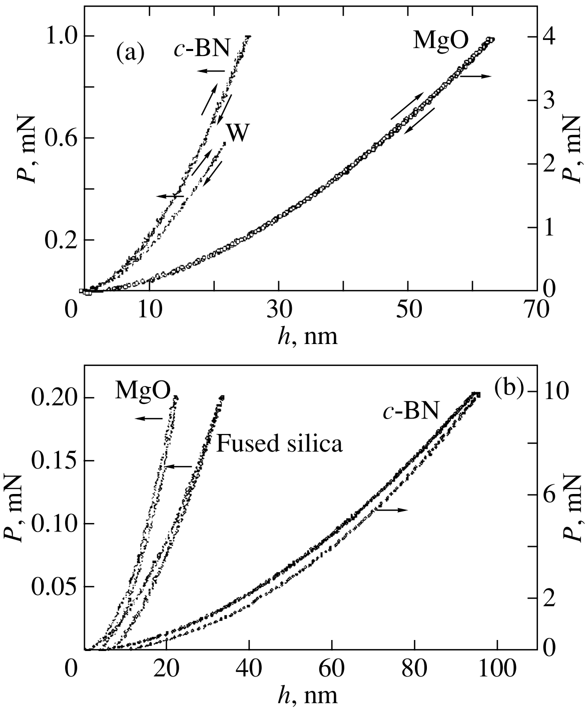


Fig. 3. (a) Completely elastic unloading and (b) almost completely reversible unloading in the same materials for different loads. The unusual plastic deformation shown in panel (b) is likely to be caused by point defects.

In [18], the activation parameters of plastic flow, namely, the activation energy U_0 and the activation volume γ, were determined by analyzing the smooth segments of σ–ε curves in the initial stage of elastoplastic deformation with the Zhurkov formula $\dot{\varepsilon} = \varepsilon_0\exp(-[(U_0 - \gamma\sigma)/kT])$. In [19], the values of γ characterizing the carrier type and the mass transfer mechanism were estimated. In the initial stages of rapid indentation, the values of γ for all materials under study were found to be comparable to the atomic volume; only in later stages and at greater depths did the values of γ attain several tens or hundreds of atomic volumes [20, 21]. This finding indicates unambiguously that point defects and their small clusters play a significant role in mass transfer at the initial stages of rapid indentation and in the initiating of dislocation mechanisms of plastic flow in subsequent stages.

4. UNSTABLE PLASTIC FLOW DURING LOCAL LOADING

Local deformation in the elastoplastic range is often accompanied by different types of instability. A typical example is the serrated flow observed in aluminum–magnesium alloys, which have been thoroughly studied from this standpoint under uniaxial deformation. This

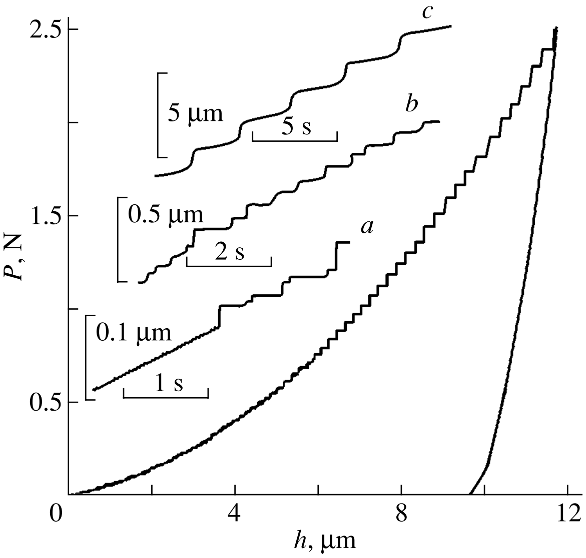


Fig. 4. Typical *P*–*h* diagram for nanoindentation of an Al–3 wt % Mg alloy under a linearly increasing load at a rate of 12.5 mN/s. The insets show typical *h*(*t*) kinetic indentation curves: (*a*) the initial stage of indentation (irregular jumps), (*b*) an intermediate stage, and (*c*) the stage of regular strain jumps.

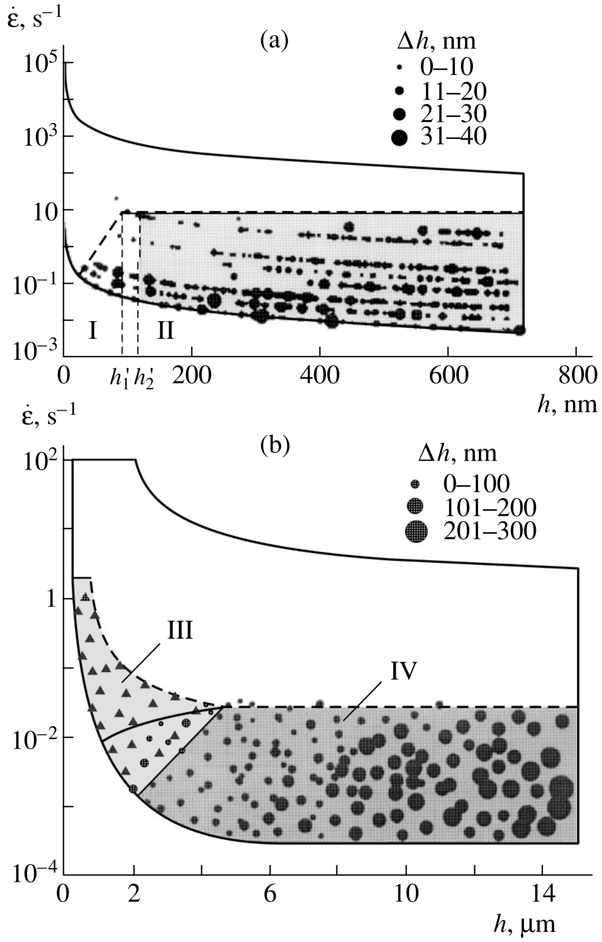


Fig. 5. Regions of the appearance of strain jumps in the *P*(*h*) diagrams of (a) bulk amorphous alloys and (b) Al–Mg alloys subjected to dynamic nanoindentation by a symmetric triangle force pulse (shown by points) at various rates. The tested region is limited by the capacity of the device and is shown by solid lines; *h* is the indentation depth; $\dot{\varepsilon}$ is the strain rate; and h'_1 and h'_2 are the depths at which the transition occurs from a monotonic to jumplike flow in bulk amorphous metallic alloys $Zr_{46.8}Ti_8Cu_{7.5}Ni_{10}Be_{27.5}$ and $Pd_{40}Cu_{30}Ni_{10}P_{20}$ during indentation. The region of the appearance of strain jumps in the $Zr_{46.8}Ti_8Cu_{7.5}Ni_{10}Be_{27.5}$ alloy is gray. I and II are the scale-dependent and scale-independent regions of the appearance of jumps, respectively; III and IV are the regions of irregular and regular strain jumps, respectively.

circumstance makes it possible to compare the behavior of the same material in macro- and microscopic volumes. Regular flow instabilities in strain-aging alloys (e.g., in Al–Mg alloys) are known as the Portevin–Le Chatelier (PLC) effect. This effect is explained by collective processes in a system of mobile dislocations that are periodically pinned by atmospheres of easily movable impurity atoms. Apart from the scientific interest that determining the role of point defects in the effects of self-organization of dislocation motion holds, the PLC effect is also important for many practical applications. For example, in many cases, the jumplike character of plastic flow makes it impossible to apply high-technology aluminum–magnesium alloys in the production of car parts, aircraft parts, disposable packaging, etc., since a strain-induced roughness of the surface makes the products unusable.

Berces *et al.* [22] showed that, under certain conditions, continuous indentation of Al–Mg alloys caused an unstable plastic flow with a characteristic serrated *P*–*h* curve, which can be interpreted as the PLC effect under the conditions of local plastic deformation of a microvolume. In the case of a linearly increasing load $P = \mu t$, two types of instabilities were observed; the typical *P*–*h* diagram obtained at a load rate $\mu = 12.5$ mN/s is shown in Fig. 4. In the first stage of loading, where the depth *h* varies from zero to a certain conventional critical depth h_{cr}, chaotic instabilities of plastic flow are observed. At moderate values of μ, this type of instability is damped with increasing *h* and is replaced by another (regular) type. The second type manifests itself as jumps in depth (Δh) and in force (ΔP), which increase with the indentation depth: $\Delta P = \gamma P$ and $\Delta h = 0.5\gamma h$, where $\gamma = \Delta H/H$ is a constant ($\gamma \approx 0.05$ for an Al–3 wt % Mg alloy). In the case of an exponentially increasing load, regular strain jumps of the second type become clearly defined [23], which makes it possible to perform statistic, autocorrelation, and other analyses of their formation. The conditions for the appearance of this type of instability are characterized by critical parameters that specify its lower boundary: $P_{cr} = \alpha\mu$ and $h_{cr} = \beta\mu^{1/2}$, where $\alpha \approx 10$ s and $\beta \approx 0.5$ μm $(s/mN)^{1/2}$

[24]. As is seen from Fig. 5, the generalizing critical parameter of a regular-type instability is the strain rate $\dot{\varepsilon}_{cr} = 0.05$ s^{-1}. The specific features of mechanical and heat treatments affect the numerical value of $\dot{\varepsilon}_{cr}$ only weakly. The data obtained for first-type instabilities agree with the mechanism of formation and evolution of slip bands, and the data obtained for second-type instabilities agree with the theory of strain hardening of random alloys.

The study of $h(t)$ loading curves by using wavelet analysis revealed another type of instability, which manifests itself in the form of jumps with an amplitude of about 1–2% of the depth h throughout the entire process of deformation. Because of its low amplitude, this type of instability is masked by instabilities of the regular type and cannot be revealed using other methods of analysis. By processing a large data array, we also revealed different types of jumps in the h–$\dot{\varepsilon}$ pseudophase diagram (Fig. 5b). As follows from this diagram, the scale effect begins to manifest itself at $h <$ 2–3 μm; at $h >$ 4–5 μm, the critical values $\dot{\varepsilon}_{cr}$ (corresponding to the onset of serrated flow) do not differ from their counterparts for macroscopic tests in spite of all the differences between local and uniform loading. This coincidence allows one to substitute nanoindentation for expensive macroscopic tests (at least in order to determine safe deformation conditions that do not cause a serrated flow).

Features of the P–h diagram that are qualitatively similar to those described above for Al–Mg alloys were also observed upon testing amorphous $Pd_{40}Cu_{30}Ni_{10}P_{20}$ and $Zr_{46.8}Ti_8Cu_{7.5}Ni_{10}Be_{27.5}$ alloys (Fig. 6a). These alloys do not exhibit the regular type of jumps [25–27]. We failed to measure the actual jump time despite the very high speed of response of our apparatus; in other words, the jump time is shorter than the time of sampling ($\tau_d = 50$ μs). By making allowance for the average jump amplitude ($\Delta h_m \approx 20$ nm for $Pd_{40}Cu_{30}Ni_{10}P_{20}$ and $\Delta h_m \approx 10$ nm for $Zr_{46.8}Ti_8Cu_{7.5}Ni_{10}Be_{27.5}$), we estimated the lower limit of the average velocity of the indentation surface during jumplike deformation to be $\langle v \rangle = \Delta h_m/\tau_d = 10^{-3}$ m/s. The actual velocity was likely much higher than this threshold value. However, at high values of $\dot{\varepsilon}$, both alloys demonstrate a clearly pronounced boundary between the stable and unstable plastic-flow modes (Fig. 5b). At $h > 100$ nm, we have $\dot{\varepsilon}_{cr} \approx 10$ s^{-1}, which is almost two orders of magnitude higher than in Al–Mg alloys. Obviously, this difference results from the different mechanisms of unstable plastic flow in these two cases.

The study of an indentation and the surrounding region using atomic force microscopy shows that nanoindentation jumps can be caused by strain localization in shear bands (Fig. 6b).

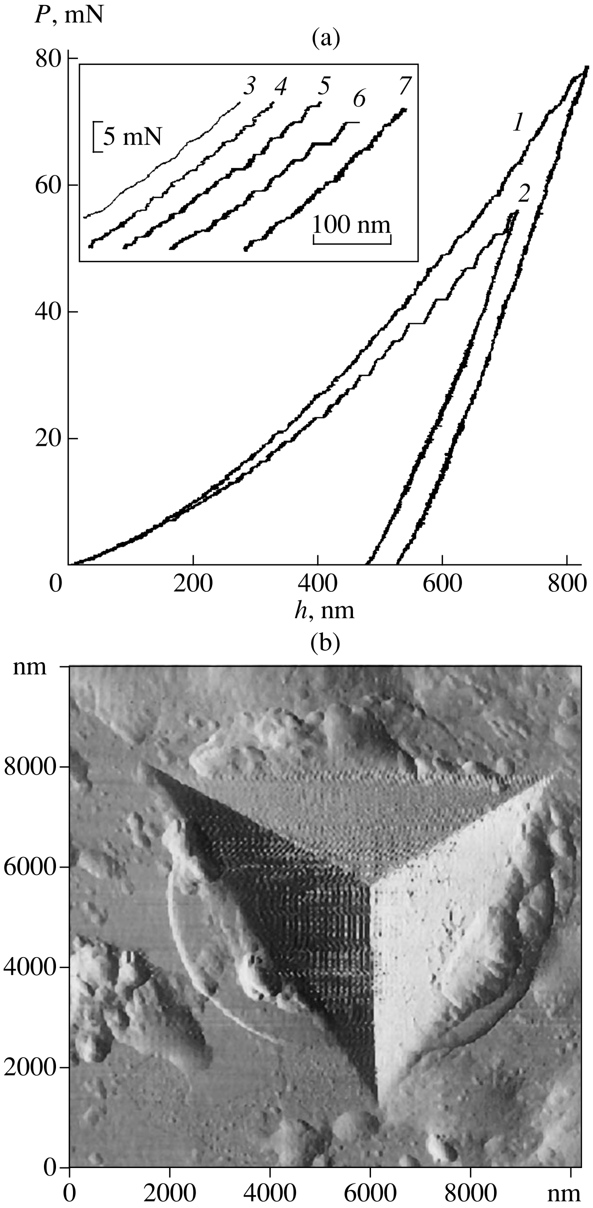


Fig. 6. (a) Unstable deformation during indentation of bulk amorphous alloys (*1*) $Zr_{46.8}Ti_8Cu_{7.5}Ni_{10}Be_{27.5}$ and (*2*) $Pd_{40}Cu_{30}Ni_{10}P_{20}$. The inset shows fragments of the loading curves for various loading rates: (*3*) 1120, (*4*) 56, (*5*) 1.12, and (*6*) 0.56 mN s^{-1} for $Pd_{40}Cu_{30}Ni_{10}P_{20}$ and (*7*) 0.80 mN s^{-1} for $Zr_{46.8}Ti_8Cu_{7.5}Ni_{10}Be_{27.5}$. (b) The atomic-force image of an indentation in $Zr_{46.8}Ti_8Cu_{7.5}Ni_{10}Be_{27.5}$ with sings of the material being carried toward the surface by localized shear bands.

5. PHASE TRANSFORMATIONS DURING LOCAL DEFORMATION

In the initial stages of indentation, a pressure of about 0.1 of the Young's modulus (see Section 3) appears in an indentation; this pressure can cause phase transformations (PTs) in certain materials. In contrast to a conventional high-pressure apparatus, which gen-

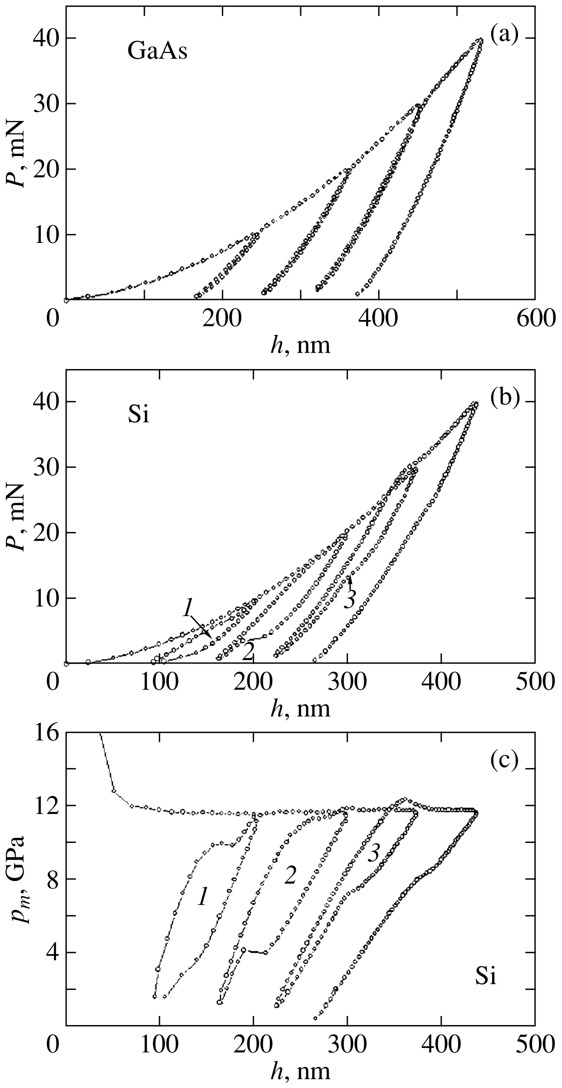


Fig. 7. Load dependences of the indenter displacement during multicycle loading of (a) gallium arsenide (which does not undergo phase transformations during indentation) and (b) silicon (which undergoes polymorphic transformations): (*1*) amorphous α-Si, (*2*) a Si-III and Si-XII mixture, and (*3*) an α-Si + Si-III + Si-XII mixture. (c) Dependences of the mean contact pressure on the indentation depth in silicon at a loading rate of (*1*) 2, (*2*) 2.5, and (*3*) 3 mN s^{-1}.

erates hydrostatic pressure, indentation generates a combined state of stress. Both high compressive and high shear stresses are generated in an indentation, and the relative shear stresses are higher than those generated upon uniaxial compression. This state of stress can substantially affect the PT parameters or can even cause PTs that do not occur with pure hydrostatic pressure. For example, the pressure required for the martensitic transformation of rhombohedral boron nitride into cubic boron nitride decreases from 55 GPa upon hydrostatic compression to 5.6 GPa upon uniaxial compression [28].

For most materials, the PT pressure is higher than the stress for the onset of plastic deformation; hence, no PT occurs in an indent during nanoindentation. However, in some solids, a structural collapse can take place before the onset of plastic flow. As the loading rate increases, the critical pressure for the transition from elastic to elastoplastic deformation can increase substantially, which increases the probability of formation of new phases and widens the range of materials that can undergo PTs during nanoindentation. PTs have been reliably detected to occur during indentation of some semiconductors and other materials with a high yield strength (e.g., silicon). The formation of high-pressure phases in Si is indicated by an increase in its electrical resistance in an indentation by a few orders of magnitude [29]; by the extrusion of a plastic high-pressure phase from an indentation [30]; and by electron diffraction, electron-microscopic, and Raman microspectroscopy data [31]. PTs in silicon under indentation are clearly manifested in *P*–*h* diagrams [32]. At $P_{\max} < 20$ mN, a specific “knee” is detected in the unloading curve, and steps are observed at $P_{\max} >$ 20 mN. In the former case, Raman microspectroscopy only reveals amorphous silicon, and in the latter case it reveals Si-III and Si-XII metastable phases [33]. Sometimes, the unloading curve contains a knee followed by a step. In this case, Raman microspectroscopy reveals both amorphous silicon and a mixture of metastable phases Si-III and Si-XII in an indentation.

It should be noted that the contact pressure in an indentation increases only in a short initial segment of the loading curve (Fig. 1) and then remains virtually unchanged during continuous indentation and depends only on the apex angle of the indenter. If an indentation is repeatedly loaded by a limiting load that is somewhat (15–25%) higher than that in the previous loading cycle, the deformation is mainly elastic and the contact pressure in the indentation increases from almost zero to the upper limit, which is bounded by the hardness.

If a load is repeatedly applied to a material that does not undergo a PT during indentation (Fig. 7a), then there appear narrow hysteresis loops, which are caused by inevitable losses of elastic energy. In this case, there are no specific features (knees or steps) in the unloading and repeated-loading curves. This behavior during nanoindentation is typical of the majority of crystalline materials. When an indentation in glasses or polymers is repeatedly loaded, large-area hysteresis loops form, which is caused by the viscoelasticity of these materials. For solid-phase polymorphic transformations, the formation of broad hysteresis loops and steps in the unloading and repeated-loading curves result from changes in the material density (Fig. 7b). This change can reach ~22–25% in silicon [33], whereas in most other materials it is substantially lower. For example, this change does not exceed 4–5% in zirconium dioxide–based ceramics. Accordingly, the strain-jump amplitude during a PT in silicon is much higher than that in ZrO_2. These ceramics do not exhibit a strain jump in the single-loading curve, but they exhibit a hysteresis with characteristic features of a PT in repeated loading cycles. Raman spectroscopy data support the fact that this hysteresis is related to the martensitic transformation of the initial metastable tetragonal structure of ZrO_2 into its stable monoclinic structure [34]. PT-induced strain jumps can be used to estimate the contact pressure that causes this PT. When studying the correlation between PTs in submicrovolumes and the shapes and sizes (the half-width at half-maximum and the area) of hysteresis loops appearing upon multicycle loading of an indentation, we detected PTs not only in covalent crystals (Si, Ge) but also in ionic–covalent crystals (CaF_2), quasicrystals (YMgZn), and intermetallic compounds (TiNi).

The competition between the developing plasticity and a contact pressure–induced PT can cause unstable deformation in submicrovolumes, which would substantially affect the operating properties of a material. This circumstance is particularly important for ultradense nanomechanical data recording, fine mechanical treatment, mechanochemical reactions, and tribological and other processes related to rapid contact interaction. However, only phase polymorphism in silicon during contact loading has been studied thoroughly and systematically [33]. The effects of the kinetic and scale factors on polymorphic transformations are poorly studied even for silicon.

Kailer *et al.* [31] showed that the phase composition in an indentation depends on the rate of decrease of the applied load (Fig. 7c). Upon rapid unloading, amorphous α-Si appears in an indentation. When the load decreases slowly, a mixture of two crystalline metastable phases (Si-XII and Si-III) forms in an indentation, with the amount of Si-III increasing as the unloading rate decreases [33]. This behavior is caused by the fact that, as the pressure in the indentation decreases, metallic silicon Si-II first transforms into the metastable phase Si-XII. When the pressure decreases further, the Si-XII phase transforms into Si-III. High residual stresses in the indentation hinder the completion of the transformation of Si-XII into Si-III after complete unloading of an indenter.

Since all available data [30–34] were obtained under quasi-stationary loading conditions with a time resolution of about 0.1 s, it was useful to study the behavior of different-type materials in contact spots with lifetimes ranging from 20 ms to 100 s with a higher time resolution ($\Delta t = 50$ μs). For a contact pressure–induced martensitic transformation in zirconium dioxide–based ceramics with a partly stabilized (PSZ) or totally stabilized (TZP) tetragonal structure, we found that the stress required for the formation of martensite in both types of ceramics increases with dP/dt. In our studies, the time it takes for the transformation from the tetragonal to monoclinic phase to be completed was found to be only several seconds. When a similar material (Y-TZP) was subjected to macroscopic tests, this time was found to be several hundreds of seconds (according to x-ray and neutron scattering data) [35]. This substantial discrepancy can be due to a number of factors, such as the effect of grain boundaries on the growth rate of the martensite phase in the initial tetragonal phase, the effect of shear stresses, and the presence of regions where contact pressures significantly exceed the average stress (e.g., near the edges and vertex of a pyramid) [36].

Apart from the loading rate, the scale factor substantially affects the PT parameters in submicrovolumes. For example, in zirconium ceramics with either partially or totally stabilized structures, the pressure inducing the martensitic transformation increases with decreasing indentation size.

In some cases, PTs in submicrovolumes can occur along with other manifestations of unstable deformation. Using atomic force microscopy and electron microscopy, Suresh and coworkers [37] obtained direct experimental evidence that the nanoindentation of bulk amorphous metallic alloys generates localized shear bands surrounded by a nanocrystalline structure that is identical to the structure forming upon 783-K annealing of these alloys without deformation. In our experiments on nanoindentation of a $Y_{10}Mg_{30}Zn_{60}$ quasicrystal, the loading curve contained ranges of both stable monotonic and unstable jumplike deformation. Since the internal structures of amorphous metallic alloys and quasicrystals are similar in many respects, we can assume that unstable deformation in this case results from the generation of localized shear bands. However, during repeated loading of an indentation, a hysteresis appears only after an indentation depth characteristic of the onset of unstable deformation is reached (Fig. 8a). Moreover, the rate sensitivity of the hardness and the activation volume γ change sharply at the indentation depth h_t corresponding to the first step in the P–h diagram. The obvious tendency toward an increase in γ with the indentation depth indicates that collective processes and the degree of ordering of the system become more important, which can be a direct consequence of the formation of a nanocrystalline phase. Finally, the hardness H of the crystalline phase should be lower, since the dislocation mechanism of plastic deformation becomes possible, and its Young's modulus E should be higher, since the crystalline phase is more ordered. These assumptions are confirmed experimentally (Fig. 8b). Thus, rearrangement of the structure can occur in shear bands with the formation of a crystalline phase in quasicrystals, as is the case in amorphous alloys [38].

In concluding this section, we note that PTs under nanocontact interaction can take place at pressures that are well below those typical of uniaxial tests (let alone those of hydrostatic compression) and at much higher rates. High residual stresses in the vicinity of a nanocontact zone help high-pressure phases to retain their metastable states. In other words, the role of PTs in the

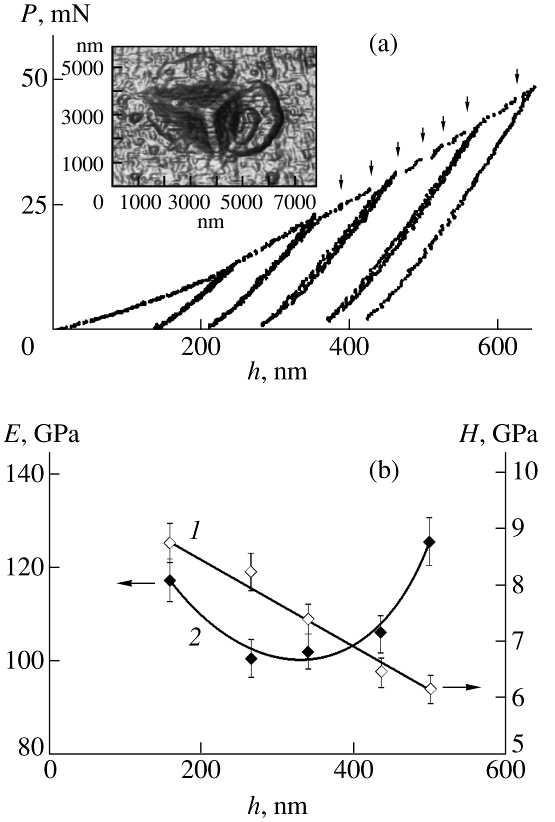


Fig. 8. (a) Unstable deformation in multicycle loading of a Y–Mg–Zn quasicrystal induced by deformation localization in shear bands (arrows show strain jumps) and by structural transformation inside the bands. (b) The variation in (*1*) the hardness *H* and (*2*) the reduced elastic modulus *E* with the formation of a crystalline structure in shear bands. The inset demonstrates a material flow from an indentation toward the surface in shear bands with simultaneous extrusion as a result of the phase transformation into a nanocrystalline state.

deformation behavior of materials in nanocontacts, especially dynamic nanocontacts, can be much more substantial than that in macroscopic tests.

6. VISCOELASTIC AFTEREFFECT

In order to design high-density storage matrices using dynamic local deformation of the surface of a recording medium, as well as to develop other nanomechanical devices, one needs to know detailed information on the mechanical behavior of materials not only under loading but also after unloading. In particular, the degree and kinetics of viscoelastic recovery of the indentation geometry upon unloading (which have not been studied under conditions of short-term local loading) can limit the speed of response and the reliability of such data recorders and readers. Moreover, dynamic reactions in nanocontacts determine the diverse characteristics of microscopically rough surfaces, including the parameters of dry friction and abrasive wear.

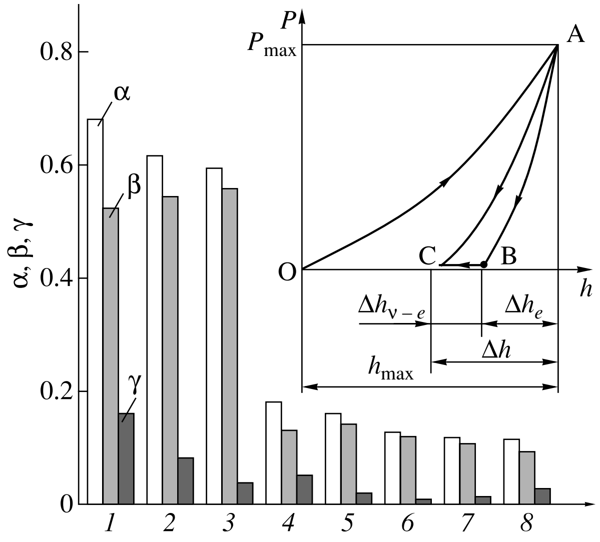


Fig. 9. Relation between the total ($\alpha = \Delta h/h_{\max}$), elastic ($\beta = \Delta h_e/h_{\max}$), and viscoelastic ($\gamma = \Delta h_{\nu-e}/h_{\max}$) recoveries of the nanocontact depth for (*1*) PMMA, (*2*) ZrO_2, (*3*) Ge, (*4*) Wood's alloy, (*5*) LiF, (*6*) Al, (*7*) Cu, and (*8*) Rose's alloy. The inset shows a schematic *P*–*h* diagram for slow (OAC curve) and rapid (OABC curve) unloading.

Instrumental nanoindentation has high spatial and time resolutions and allows one to model the situations described above under well-controlled conditions in both the stage of loading and the stage of elastic and viscoelastic recovery of an indentation after unloading [9, 10]. Using this method, we studied polymethyl methacrylate (PMMA), ZrO_2-based ceramics, single-crystal LiF and Ge, polycrystalline Al and Cu, Wood's alloy (50 wt % Bi–25 wt % Pb–12.5 wt % Sn–12.5 wt % Cd), and Rose's alloy (50 wt % Bi–28 wt % Pb–22 wt % Sn), which represent the main groups of solids in terms of hardness, the nature of bonding forces, and the microscopic mechanisms of plastic deformation. All measurements were performed at 296 K with a unique nanoindentation tester designed at Tambov State University. In the first series of experiments, a Berkovich indenter penetrated into a surface under the action of a triangular force pulse with various values of the amplitude $P_{\max}$ (from 4 to 235 mN), which were chosen so that the same maximum dynamic indentation depth $h_{\max}$ = 650 nm was reached in different materials for a force rise time τ_1 = 10 s and a force fall time τ_2 = 200 ms. To retain a reliable contact between the indenter and material after unloading, the load was maintained at a level of several percent of $P_{\max}$. Under these conditions, the nanotester operated as an indicator of the indentation depth varying with time under the action of relaxation processes.

Figure 9 shows the elastic recovery Δh_e (this quantity was actually taken to be the recovery within the unloading time τ_2 = 200 ms), the viscoelastic recovery $\Delta h_{\nu-e}$ (within τ_3 = 25 s after complete unloading), and the total recovery Δh (within the time $\tau = \tau_2 + \tau_3$) for various materials (normalized to $h_{\max}$). It was found that, at a fixed indentation depth ($h_{\max}$ = 650 nm), the quantity $\beta = \Delta h_e/h_{\max}$ was 50–55% in Ge, ZrO_2-based ceramics and PMMA and did no exceed 18% in the other materials. The relative magnitude of the viscoelastic recovery $\gamma = \Delta h_{v-e}/h_{\max}$ was rather high in all materials under study and varied from 5% in Al to 28% in Wood's alloy (which is even higher than that in PMMA). The quantities Δh, Δh_e, and Δh_{v-e} were found to be very sensitive to the scale factor, especially in the depth range 50 nm < $h_{\max}$ < 1 μm. In this range, a decrease in $h_{\max}$ is accompanied by a twofold or greater increase in the elastic recovery in metals and ionic and covalent crystals (whose recovery is mainly elastic). For materials with $\beta > \gamma$, a change in the indentation depth was found to have the largest effect on Δh_e, whereas in solids with $\beta \approx \gamma$ (PMMA, Rose's alloy) a decrease in h had a substantial effect on Δh_{v-e}. As h decreases, the relative magnitude of the elastic recovery increases and that of the viscoelastic recovery decreases in all solids under study. This effect is likely due to the fact that certain viscoelastic recovery mechanisms are blocked at small values of h (which provides an increase in the relative magnitude of Δh_e). The increase in γ at h < 200 nm in some materials can be explained by their surface properties and by specific structural-relaxation mechanisms becoming operative (which are dominated by traditional mechanisms at large indentation depths). In particular, in ionic and covalent crystals, the role of point defects increases and the role of dislocations decreases with decreasing $P_{\max}$ and $h_{\max}$ [19–21].

For all materials under study, the total recovery depends on the rate of load rise $\mu = dP/dt$ in the stage of loading. As μ increases from 2 mN/s to 1 N/s, the ratio $\alpha = \Delta h/h_{\max}$ increases by 15–40% in different materials. The rather high total recovery in Ge is explained by the significant fraction of the elastic component (up to 90%). In Wood's alloy, the elastic and viscoelastic recoveries are of the same order of magnitude, with Δh_{v-e} even exceeding Δh_e at small loading rates. As μ increases within the range indicated above, the ratio of Δh_{v-e} to the total recovery decreases by factors ranging from 1.5 in PMMA to 2 in Ge. It should be noted that the lower the unloading rate and the lower the time resolution of the apparatus, the greater the portion of the time-dependent viscoelastic recovery attributed to elastic recovery by an observer. Apparently, even our relatively high rates of unloading and data sampling (50 μs between measurements) were insufficient to completely exclude this error for some materials.

Thus, with dynamic nanoindentation, we found a relation between the elastic and viscoelastic components of the shape recovery for various solids after plastic deformation. The dependences of these quantities on the loading rate were determined, and their sensitivity to the scale factor was established.

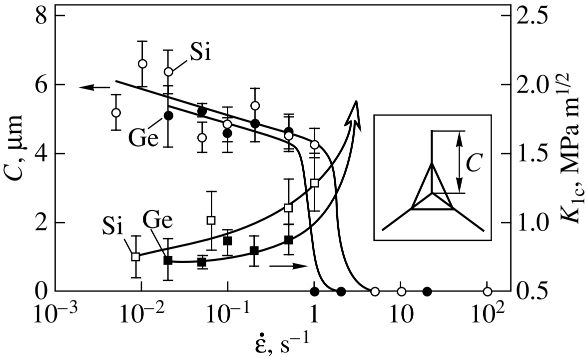


Fig. 10. Dependences of the crack length C and the critical stress-intensity factor of the first kind, K_{1c}, on the strain rate $\dot{\varepsilon}$ for Si and Ge. The inset shows a schematic representation of an indentation with cracks of length C at its corners.

7. MICROCRACK NUCLEATION IN A LOCAL-DEFORMATION ZONE

The method of local loading has been used for a few decades to determine the cracking resistance (or the critical stress-intensity factor of the first kind, K_{1c}) of high-strength and low-plasticity structural and tool materials, such as glasses, ceramics, hard alloys, nitrides, and carbides [39].

The nucleation and development of cracks under conditions of dynamic nanoindentation have been studied in various materials (fullerite C_{60}, silicon, germanium, lead tungstate $PbWO_4$) for various values of loads and strain rates. It was found that the nucleation and development of microcracks causes another mode of unstable deformation during nanoindentation; this mode manifests itself in separate jumplike changes that can occur under loading or unloading, depending on the sample material. By scanning residual indentations with an atomic force microscope, the formation of radial surface cracks extending from the corners of the indentations was observed and the rate dependences of K_{1c} in silicon and germanium were determined. Moreover, it was found that there occurs a jumplike increase in the dynamic viscosity and fracture in these materials with increasing strain rate (Fig. 10). Under ordinary conditions of uniaxial tension, most materials exhibit a decrease in K_{1c} with increasing loading rate; i.e., they become brittle. Therefore, at first glance, the termination of crack formation at a certain critical strain rate seems to be paradoxical. However, cracks can nucleate via conventional dislocation mechanisms if there is a sufficient number of dislocations interacting with each other at short distances. As follows from the results described in Section 3 (Figs. 2a, 3), dislocations are small in number and their formation at $h \leq 100$ nm becomes even more difficult with increasing $\dot{\varepsilon}$ and shifts toward higher loads. Obviously, dislocation-free, homogeneous microcrack nucleation requires still higher stresses. Nevertheless, a residual indentation can also form under these conditions, which indicates the presence of plasticity. All these facts taken together suggest that the sharp blocking of crack formation under conditions of high-rate local deformation is due to plastic strain occurring under these conditions via the generation and motion of point defects rather than dislocations.

8. CONCLUSIONS

(1) The properties of materials in nanovolumes change substantially as the nanovolume size decreases. The role and the relative proportion of jumplike deformation modes, as well as the contribution from elastic strain to the total deformation, increase as the load (the indentation size) decreases. However, in some cases, one can find normalizing factors and parameter ranges that provide invariance of the mechanical properties (such as a serrated flow in Al–Mg alloys; yield strength in metallic, semiconducting, and oxide single crystals; creep characteristics in metals and alloys). Therefore, nanoindentation can be used to solve two groups of problems: (i) to study the behavior of a material in nanovolumes and (ii) to replace (or to supplement) the methods of mechanical macroscopic tests (e.g., to determine the creep activation parameters, the characteristics of high pressure–induced phase transformations, the critical strain rates for various plastic-flow modes, etc.).

(2) Small characteristic sizes of a deformation region in the initial stages of indentation make it possible to achieve extremely high local strain rates ($\dot{\varepsilon} \sim 10^5$ s^{-1}) during indenter penetration at rather moderate linear velocities (~1 cm/s). This creates a unique opportunity for studying the role of the rate factors and the time-dependent characteristics of materials over a very wide range of $\dot{\varepsilon}$ values.

(3) The high spatial and time resolutions of modern nanoindenters allow one to study the kinetics and dynamics of elementary events important for the formation of mechanical properties, such as the nucleation and motion of dislocations, their damping by Cottrell clouds, and the formation and growth of new-phase nucleation centers under high pressure, in order to determine the instant of microcrack appearance, viscoelastic recovery after unloading, etc.

(4) The established regularities for the behavior of materials open up fresh opportunities for improving the operating properties of thin coatings, submicrovolumes of materials, nanomachines, etc.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 04-02-17198) and the Ministry of Education and Science of the Russian Federation (project no. E02-3.4-263).

REFERENCES

1. V. R. Regel’, A. I. Slutsker, and E. E. Tomashevskiĭ, *Kinetic Nature of the Strength of Solids* (Nauka, Moscow, 1974) [in Russian].
2. V. I. Vladimirov, *Physical Nature of Destruction of Metals* (Metallurgiya, Moscow, 1984) [in Russian].
3. V. E. Panin, V. A. Likhachev, and Yu. V. Grinyaev, *Structural Levels of Deformation of Solids* (Nauka, Novosibirsk, 1985) [in Russian].
4. N. A. Koneva and V. É. Kozlov, *Structural Levels of Plastic Deformations and Destruction* (Nauka, Novosibirsk, 1990), p. 123 [in Russian].
5. W. C. Oliver and G. M. Pharr, J. Mater. Res. **7** (6), 1564 (1992).
6. W. C. Oliver and G. M. Pharr, J. Mater. Res. **19** (1), 3 (2004).
7. J. B. Pethica, R. Hutchings, and W. C. Oliver, Philos. Mag. A **48** (4), 593 (1983).
8. N. V. Novikov, S. N. Dub, Yu. V. Mil’man, I. V. Gridneva, and S. I. Chugunova, Sverkhtverd. Mater. **18** (3), 36 (1996).
9. Yu. I. Golovin, A. I. Tyurin, V. I. Ivolgin, and V. V. Korenkov, Zh. Tekh. Fiz. **70** (5), 82 (2000) [Tech. Phys. **45** (5), 605 (2000)].
10. Yu. I. Golovin, V. I. Golovin, V. V. Korenkov, N. V. Korenkova, and B. Ya. Farber, Philos. Mag. A **82** (10), 2173 (2002).
11. P. Vettiger, G. Cross, M. Despont, U. Duerig, B. Gostmann, W. Haeberte, M. Lantz, H. Rothuizen, R. Stutz, and G. Binning, IEEE Trans. Nanotech. **1** (1), 39 (2002).
12. K. L. Johnson, *Contact Mechanics* (Cambridge Univ. Press, Cambridge, 1985; Mir, Moscow, 1989).
13. J. P. Hirth and J. Lothe, *Theory of Dislocations* (McGraw-Hill, New York, 1967; Atomizdat, Moscow, 1972).
14. Yu. I. Golovin and S. N. Dub, Dokl. Akad. Nauk **393** (2), 180 (2003) [Dokl. Phys. **48** (11), 612 (2003)].
15. Yu. V. Milman, B. A. Galanov, and S. I. Chugunova, Acta Metall. Mater. **41** (9), 2523 (1993).
16. S. Jayaraman, G. Hahm, W. Oliver, C. A. Rubin, and P. C. Bastias, J. Solid Struct. **35** (5–6), 365 (1998).
17. J. C. M. Li, Mater. Sci. Eng. A **322** (1–2), 23 (2002).
18. Yu. I. Golovin and A. I. Tyurin, Pis’ma Zh. Éksp. Teor. Fiz. **60** (3), 722 (1994) [JETP Lett. **60** (10), 742 (1994)].
19. Yu. I. Golovin and A. I. Tyurin, Fiz. Tverd. Tela (St. Petersburg) **42** (10), 1818 (2000) [Phys. Solid State **42** (10), 1865 (2000)].
20. Yu. I. Golovin, A. I. Tyurin, and B. Ya. Farber, J. Mater. Sci. **37** (4), 895 (2002).
21. Yu. I. Golovin, A. I. Tyurin, and B. Ya. Farber, Philos. Mag. A **82** (10), 1857 (2002).
22. G. Berces, N. Chinh, A. Juhasz, and J. Lendvai, J. Mater. Res. **13** (6), 1411 (1998).

23. V. I. Ivolgin and D. A. Sergunin, Kondensir. Sredy Mezhfaznye Granitsy **6** (1), 41 (2004).
24. Yu. I. Golovin, V. I. Ivolgin, and M. A. Lebedkin, Fiz. Tverd. Tela (St. Petersburg) **44** (7), 1254 (2002) [Phys. Solid State **44** (7), 1310 (2002)].
25. Yu. I. Golovin, V. I. Ivolgin, V. A. Khonik, K. Kotagawa, and A. I. Tyurin, Scr. Mater. **45** (8), 947 (2001).
26. Yu. I. Golovin, V. I. Ivolgin, A. I. Tyurin, and V. A. Khonik, Fiz. Tverd. Tela (St. Petersburg) **45** (7), 1209 (2003) [Phys. Solid State **45** (7), 1267 (2003)].
27. S. A. Schuh and T. G. Neih, Acta Mater. **51** (1), 87 (2003).
28. N. V. Novikov, I. A. Petrusha, L. K. Shvedov, S. B. Polotnyak, S. N. Dub, and S. A. Shevchenko, Diamond Relat. Mater. **8** (2–5), 361 (1999).
29. I. V. Gridneva, Yu. V. Milman, and V. I. Trefilov, Fiz. Tverd. Tela (Leningrad) **14** (1), 222 (1972) [Sov. Phys. Solid State **14** (1), 177 (1972)].
30. G. M. Pharr, W. C. Oliver, and D. S. Harding, J. Mater. Res. **6** (6), 1129 (1991).
31. A. Kailer, Y. G. Gogotsi, and K. G. Nickel, J. Appl. Phys. **81** (7), 3057 (1997).
32. Y. G. Gogotsi, V. Domnich, S. N. Dub, A. Kailer, and K. G. Nickel, J. Mater. Res. **15** (4), 871 (2000).
33. V. Domnich and Yu. Gogotsi, Rev. Adv. Mater. Sci. **3**, 1 (2002).
34. V. Domnich, Yu. Gogotsi, and M. Trenary, Mater. Res. Soc. Symp. Proc. **649**, 891 (2001).
35. N. Igawa, T. Nagasaki, Y. Ishii, K. Noda, H. Ohno, Y. Morii, and J. A. Fernandez-Baca, J. Mater. Sci. **33** (19), 4747 (1998).
36. Yu. I. Golovin, V. V. Korenkov, and B. Ya. Farber, Izv. Akad. Nauk, Ser. Fiz. **67** (6), 840 (2003).
37. J.-J. Kim, Y. Choi, S. Suresh, and A. S. Argon, Science **295** (5555), 654 (2002).
38. C. A. Schuh and T. G. Nieh, J. Mater. Res. **19** (1), 46 (2004).
39. B. R. Lawn and T. R. Wilshaw, J. Mater. Sci. **10**, 1049 (1975).

Translated by K. Shakhlevich