Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 1000–1010.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1153–1166.

SOLIDS
Structure

Atomic Structures of Two-Dimensional Strained InAs Epitaxial Layers on a GaAs(001) Surface: *in situ* Observation of Quantum Dot Growth

R. Z. Bakhtizin[a, *], Y. Hasegawa[b], Q.-K. Xue[b], and T. Sakurai[b]

[a]*Bashkortostan State University, Ufa, 450074 Bashkortostan, Russia*
[b]*Institute for Materials Research, Tohoku University, Sendai 980-8577, Japan*
** e-mail: raouf@bsu.bashedu.ru*

Received July 28, 1999

Abstract—Scanning tunneling microscopy and reflection high-energy electron diffraction under ultrahigh vacuum conditions were used to make an *in situ* study of atomic structures at the surface of an InAs/GaAs heterostructure grown by molecular-beam epitaxy. It was observed that the deposition of approximately 0.3 ML of indium on an arsenic-enriched GaAs(001)-2×4 surface leads to the formation of the 4×2 phase while the deposition of 0.6 ML indium leads to the appearance of a new 6×2 reconstruction. It is shown that layer-by-layer two-dimensional epitaxial growth of InAs on GaAs(001) as far as 13 monolayers can only be achieved if the growth front reproduces the 4×2 or 6×2 symmetry of the substrate and models of 4×2 and 6×2 reconstructions are proposed. Atomic-resolution images of faceted planes on the surface of three-dimensional islands in an InAs/GaAs(001) system were obtained for the first time and structural models of these were developed.

1. INTRODUCTION

Heteroepitaxial growth in lattice-mismatched systems is one of the most promising technological approaches to obtain low-dimensional quantum nanostructures (quantum dots, quantum wires) and to fabricate new optoelectronics devices using these [1, 2]. An effective method of producing these nanostructures is to use self-organization in heteroepitaxial systems which occurs as a result of the formation of elastically strained coherent three-dimensional islands [3]. In the best known system of this type, InAs/GaAs (with a 7.2% lattice mismatch) the transition from two-dimensional to three-dimensional growth, which can reduce the accumulated energy of the elastically strained layer grown by molecular-beam epitaxy, takes place when a critical thickness of approximately 2 monolayers is reached. Then in typical quantum-well nanostructures where the electron motion is bounded in one dimension, the morphological Stranski–Krastanow transformation [4] of the elastically strained layer into an array of three-dimensional coherent islands can be delayed or suppressed. This last factor can be used to achieve planar growth and to obtain abrupt interfaces, which is fundamentally important for optimizing the characteristics of optoelectronic devices fabricated using heterojunctions and superlattices.

The growth of any film by molecular-beam epitaxy is essentially a nonequilibrium process so that both the morphology and the evolution of the growing film are governed by the relationship between the kinetic and thermodynamic parameters which can be used to control the character of the growth. If the growth of the strained layer takes place by the Stranski–Krastanow mechanism [4], the formation of three-dimensional islands is accompanied by a reduction in the elastic strain energy although it leads to an increase in the total surface area and consequently a higher consumption of surface free energy. Thus, when the surface tension is low, relaxation of the elastic strain energy predominates, promoting a transition from two-dimensional to three-dimensional growth but as the surface tension increases, the situation changes. Thus, by varying the component ratio $[As_4]/[In]$ in the flux during the growth process and thereby varying the surface reconstruction from the arsenic-enriched 2×4 phase having low surface tension to the indium-enriched 4×2 phase, Schaffer, Lind, Kowalczyk, and Grant [5] showed that strained InAs epitaxial layers up to 2000 Å thick can be grown two-dimensionally on a GaAs substrate. Under conditions of strong indium enrichment, the formation of three-dimensional inhomogeneities is to a large extent suppressed because of the appreciable kinetic barrier for the formation of dislocations. This mechanism, being related to the higher surface tension of the indium-enriched reconstructed surface, requires a knowledge of the surface morphology and its atomic structure for its substantiation.

In the present paper we report results of a detailed *in situ* study of the surface structure of an InAs/GaAs system during its two-dimensional and three-dimensional growth as a function of the epitaxial layer thickness and the concentration ratio of the components in

the fluxes, made using scanning tunneling microscopy under ultrahigh vacuum conditions. The purpose of the study was to understand the growth mechanism of the films at the atomic level and also the process of formation of InAs quantum dots formed at the surface of GaAs as a result of self-organization [6].

2. METHOD

All the experiments were carried out using an ultrahigh-vacuum (base pressure 3×10^{-11} Torr) scanning tunneling microscope (STM) at Tohoku University which was combined with a molecular-beam epitaxy chamber [7]. Gallium arsenide substrates measuring 4×10 mm^2 were cut from wafers oriented in the [001] crystallographic direction and were etched in a standard mixture of H_2SO_4 and H_2O_2 before being placed in the growth chamber. The oxide layer on the surface was removed by annealing at 600°C in an As_4 stream. A 400 nm thick GaAs buffer layer was grown at 550°C at a rate of 200 nm/h with an $[As_4]/[Ga]$ concentration ratio in the flux of 40. Silicon (1×10^{18} cm^{-3}) was used as the dopant. The growth process was monitored by measuring the intensity oscillations of the reflection high-energy electron diffraction (RHEED) spots. After growth of the buffer layer the sample was annealed at 470°C and transferred to the microscope chamber. The STM images were usually observed in the filled state regime at bias voltages V_s between –1.6 and –3.5 V and at tunnel currents $I_t = (20–40) \times 10^{-12}$ A.

The most ordered β-phase was used as the substrate for the InAs growth [7]. The initial InAs wetting layer was prepared by depositing a submonolayer (0.5 ML) In coating on the arsenic-enriched surface at a substrate temperature of 450°C and a chamber pressure of 1×10^{-10} Torr. This layer had a 2×4 structure similar to the 2×4 reconstruction of the GaAs(001) surface although the dimer rows on the InAs surface were not as straight as those on the GaAs surface [7] and they have a higher density of kinks resembling the 2×4 structure on the surface of a solid InAs(001) crystal [8]. We observed that after depositing approximately 0.3 ML of indium and then annealing for 5 min at 450°C, the surface exhibited a very sharp RHEED pattern corresponding to 4×2 symmetry. The subsequent layer-by-layer growth of the InAs was achieved using migration-enhanced epitaxy [7, 9]. STM observations showed that in this case two parameters are critical: the $[As_4]/[In]$ concentration ratio in the flux (to obtain the 4×2 phase this was 3–5 at 450°C) and the switching time τ of the shutters of the Knudsen cells in the growth chamber.

3. EXPERIMENTAL RESULTS AND DISCUSSION

3.1. 4 × 2 Surface Reconstruction

Figure 1 shows typical filled-state STM images of a 4×2 surface which demonstrate its high degree of perfection. The images consist of straight lines separated by 16 Å gaps and equidistant humps along the [110] direction (the 4×2 unit cell is shown by the white rectangle). Also shown are scanning profiles along the $L1$, $L2$, and $L3$ lines. The fourfold (4×) periodicity of the surface in the $[\bar{1}10]$ direction and the twofold (2×) periodicity in the [110] direction are observed most clearly on the cross-section profiles of the $L1$ and $L2$ lines, respectively. As the bias voltage decreased to $V_s = –1.6$ V no substantial changes were observed in the contrast between the lines and the humps but an additional single (1×) periodicity appeared along the line formed by the smaller projections as can be seen on the scanning profile of the $L3$ line. A more thorough examination of this image revealed that the smaller hump forming the 1× periodicity are always situated on either side of larger humps in the [110] direction (some of the small humps near the unit cell are shown white) which indicates that the humps having 2× and 1× periodicity are attributable to tunneling from different species. The observed characteristics differ appreciably from those reported for homoepitaxial growth of an indium-enriched (001)InAs-4×2 surface [8, 10] and an arsenic-enriched GaAs(001)-2×4 surface [7] so that they should have different structures. Since the 2× direction coincides with the direction of indium dimerization, we can postulate that the large 2× humps are caused by tunneling from In dimers. The STM images of the empty states of this surface revealed the

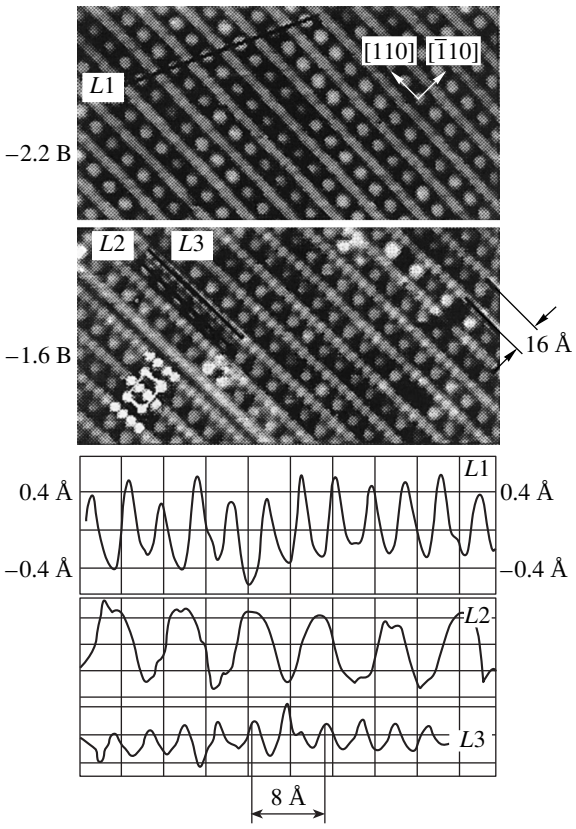

Fig. 1. High-resolution STM images of an indium-enriched GaAs(001) surface and corresponding cross-section profiles.

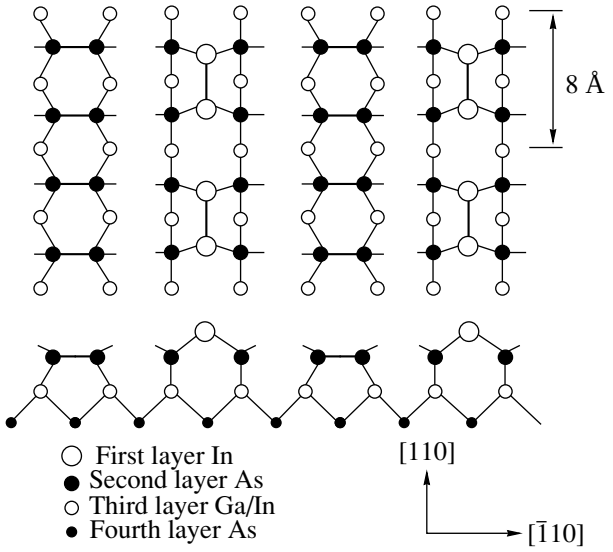

Fig. 2. Geometric model of 4×2 GaAs(001) surface.

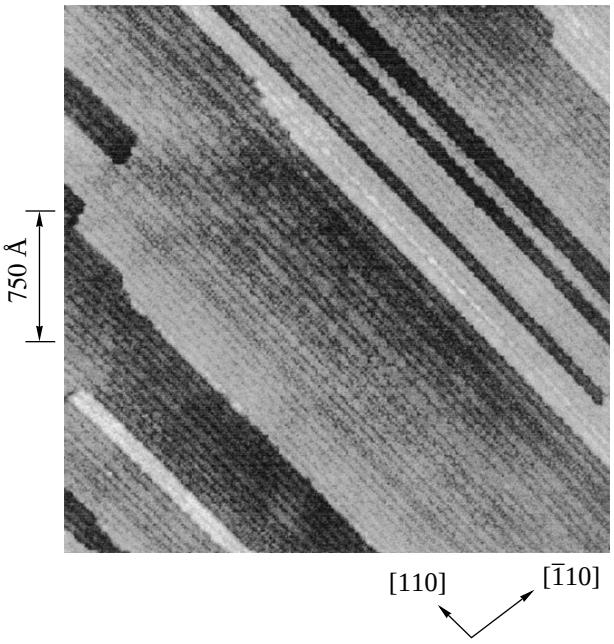

Fig. 3. Surface morphology of sample after 13 monolayers of InAs were deposited on a GaAs(001)–As-2×4 substrate, $V_s = -2.5$ V.

same hump-plus-line features but the 2× humps become brighter than the lines. This change in the contrast is consistent with the concept that the 2× humps are caused by tunneling from In dimers in the first layer and the lines are produced by tunneling from As atoms in the second layer, and this allowed us to propose a model of 4×2 reconstruction (Fig. 2). The 2× periodicity occurs as a result of dimerization of In adatoms in the [110] direction while the regular missing of In dimers is responsible for the formation of fourfold periodicity in the [$\bar{1}$10] direction. The exposed As adatoms in the missing rows dimerize to their neighbours and their uniform arrangement explains the single periodicity along the *L*3 line in Fig. 1. The unit cell consists of a single In dimer at the top of the layer and two As dimers in the second layer and assumes that the surface has an approximately 0.3 ML thick In coating, which agrees with the experiment. The proposed model does not contradict the Pashley electron counting rule [10] and implies that the dangling bonds of each As atom are completely filled with two electrons while the dangling bonds of the In atoms remain empty, without trapping charge, which corresponds to a stable semiconducting surface. The observed difference in contrast between the images of the filled and empty states can be attributed to a difference in the position of the energy levels of the dangling arsenic bonds and the antibonding orbitals of the In dimers. Note that this model gives a 0.5 ML thick In surface coverage which agrees with the experiment.

Ohkouchi and Tanaka [11] proposed a single-dimer structure which also agrees with the observed STM images and could be used as an alternative model of an indium-enriched 4×2 surface. However, this model predicts a difference in contrast of 2.95 Å (step height of a double InAs layer) between the In dimers of the first and third layers, which was not observed, and assumes an In surface coverage of approximately 0.75 ML.

After the 4×2 layer described above had been fabricated on the substrate, a two-dimensional multilayer InAs coating with a smooth surface could be grown systematically, layer by layer, by selecting and strictly maintaining the ratio of the [In] and [As_4] atomic concentrations in the flux, the substrate temperature, and the cooling rate such that the growth front reproduced the 4×2 symmetry of the substrate. This factor clearly demonstrates the specific function of the 4×2 surface which serves as a template for the two-dimensional growth. STM observations of the morphology of a surface coated with 13 InAs monolayers (which is considerably thicker than the critical thickness of two monolayers) confirmed its planar growth and suppression of the formation of three-dimensional islands (Fig. 3). It can be seen that within an area of 2300×3000 Å^2 we can identify four levels of planar terraces (each corresponds to approximately 3 Å which is slightly greater than the value of 2.8 Å which is the step height of a double layer of bulk GaAs). Generally, the growth of new-phase islands is determined by two processes: diffusion of adatoms toward the island and transitions of atoms across the interface with the island, i.e., the boundary kinetics. Using migration-enhanced epitaxy under conditions of indium enrichment ensured a fairly long diffusion path length $L = \sqrt{D\tau}$, where D is the diffusion coefficient of the adatoms, and in this case, an earlier transition to the formation of three-dimensional islands should be predicted. Thus the layer-by-layer growth observed by us was attributed to the increasing surface

tension: as we know, indium-enriched structures have a higher coefficient of surface tension γ than arsenic-enriched structures [5]. These results show good agreement with the criterion for the critical layer thickness t_{cr} (deposition time) obtained by Snyder, Mansfield, and Orr [12] based on a kinetic approach:

$$t_{cr} \approx \gamma^2/K^2\varepsilon^4 L, \quad (1)$$

where K is the bulk modulus and ε is the lattice mismatch (7.2%). For thicknesses $t < t_{cr}$ we observed two-dimensional growth of a metastable elastically strained film whereas for $t > t_{cr}$ we observed the formation of three-dimensional coherent islands where the expression for the minimum temperature T_{min} below which the formation of islands is suppressed [12]

$$T_{min} = \left(\frac{k_B}{E_A}\ln\frac{1.44D\tau K^2\varepsilon^4}{\gamma^2}\right)^{-1}, \quad (2)$$

gives a temperature close to that used by us: T_{min} = 450°. Here k_B is the Boltzmann constant and E_A is the activation energy of an adatom. We note that although the surface morphology is fairly smooth, the STM image in Fig. 3 does not exhibit such a high degree of ordering as in Fig. 1 although the RHEED patterns in both cases corresponded to 4×2 symmetry.

Another important topic is the mechanism for relaxation of the elastic energy accumulated in the two-dimensional epitaxial layer since no stacking faults nor the formation of mismatch dislocations were observed at this stage of the growth process. Returning to Fig. 3, we note that the surface consists mainly of 4×2 domains and at the same time is modulated by characteristic dark lines in the [110] direction from one edge of the step to the other, which form a unique structure with domain walls separated by the distance Na_0 (a_0 is the surface lattice constant). For all coatings in the range of 4–13 ML the value $N = 6$ predominated. In the STM images of empty states, domain walls were observed on the same positions which demonstrates their geometric origin. The observed characteristics are similar to the $2 \times N$ structures on the strained Ge/Si interface and the lines of vacancy defects caused by the presence of Ni on the Si(100) surface, and are the result of the relaxation of surface stress [13–15]. We postulate that the regularly distributed domain walls are a new mechanism for the relaxation of elastic strain and may be considered as potential sites for the nucleation of misfit dislocations.

3.2. 6 × 2 Surface Reconstruction

At the initial stage of growth of a strained InAs layer on a GaAs(001) surface we observed a new 6×2 reconstruction whose formation appreciably improved the morphology and structure of the substrate. The 6×2 phase was prepared by depositing 0.6 ML of indium on an arsenic-enriched GaAs(001)-2×4 surface at 500°C, i.e., under essentially the same conditions as those used to obtain the 4×2 phase. The InAs layers were then grown at 450°C at a rate of 0.2 ML/s and in order to achieve two-dimensional growth the $[As_4]/[In]$ concentration ratio in the flux was maintained at 2–3 in the molecular-beam epitaxy regime or 6 in the migration-enhanced epitaxy regime (with τ = 1 s for As and 2 s for In). In both cases the RHEED pattern corresponded to 6×2 symmetry and the layer-by-layer growth process typical of a superlattice structure could be maintained up to 13 (sometimes even fifty!) monolayers.

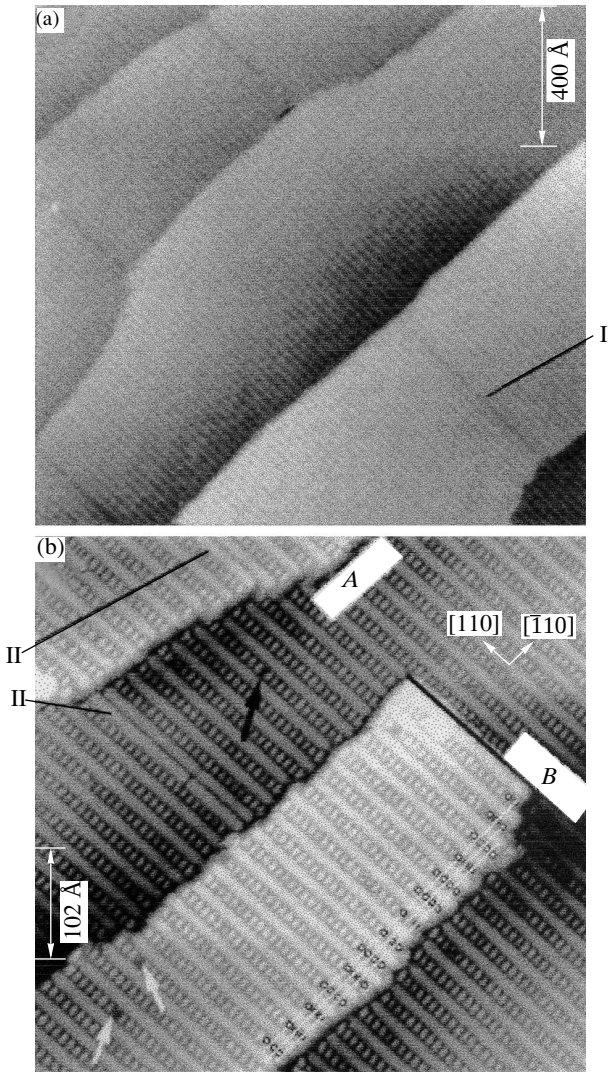

Fig. 4. STM images of a 6×2 surface: (a) large-scale scan image and (b) the zoom-in image of the same surface showing step structure and surface defects, V_s = –2.5 V.

Figure 4 shows STM images of a surface with the 6×2 phase where we can clearly discern large flat terraces approximately 500 Å wide in the [110] direction and two-layer steps 2.8 Å high (step height of a double GaAs layer) but we do not observe any adsorption-induced artifacts or step bunching. Note that we could not obtain images of structures at bias voltages $|V_s| \le 1$ V;

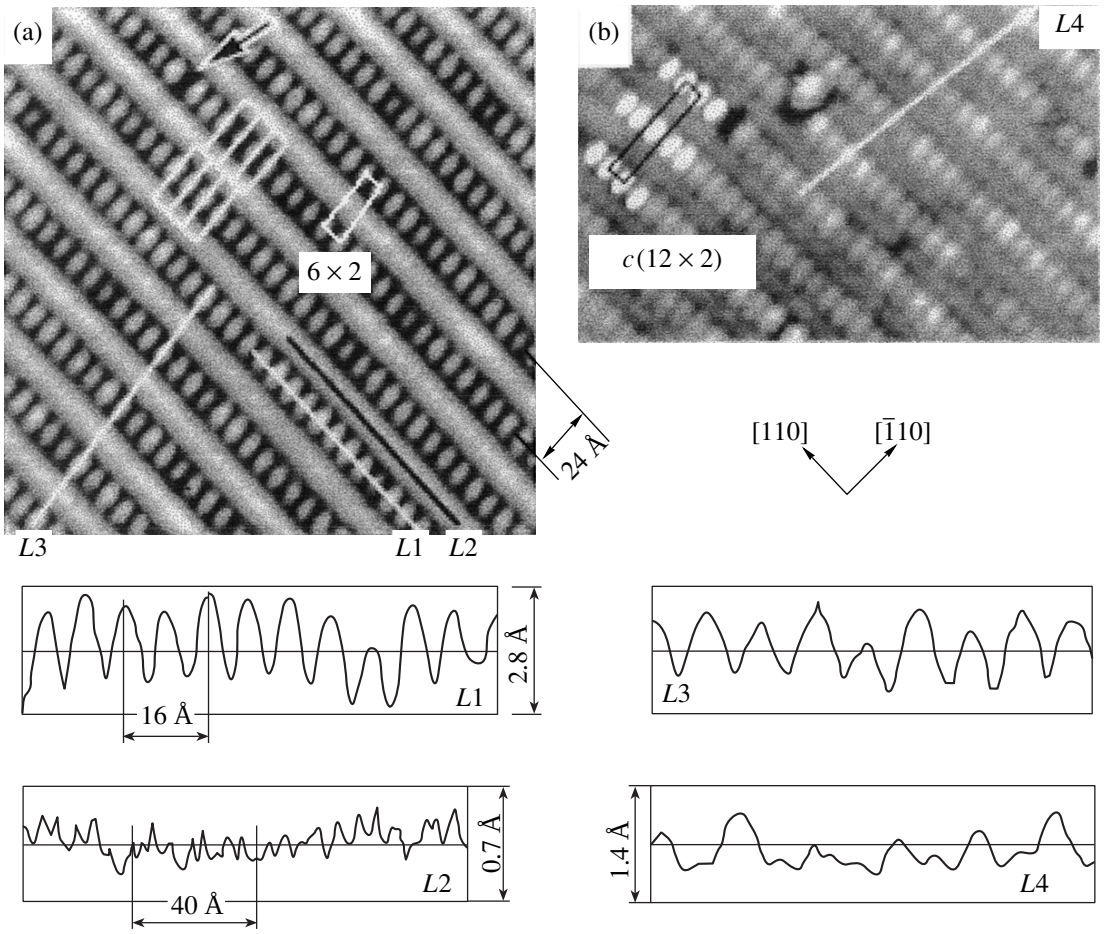

Fig. 5. High-resolution STM images and cross-section profiles obtained in the filled state regime (a), V_s = –2.5 V and in the empty state regime (b), V_s = +2.2 V demonstrating a kink type defect and a transition from a 6 × 2 structure to the $c(12 \times 2)$ antiphase induced by it.

this implies a semiconducting rather than a metal surface. The 6× periodicity on the STM images is characterized by uniformly distributed rows of bright spots passing across the entire terrace without any kinks and separated by 24 Å in the [$\bar{1}$10] direction which forms an appreciable contrast with the arsenic-enriched 2×4 phase [7]. The spacing between neighboring spots is 8 Å and corresponds to 2× periodicity. We shall assume that the high coherence in the [$\bar{1}$10] direction is caused by the stronger lateral interaction at the surface during deposition of the indium so that the kink formation energy should be considerably higher.

The STM image in Fig. 4 demonstrates the structure of both types of single-layer steps: *A* (running along the rows of dimers of the upper terrace in the [$\bar{1}$10] direction) and *B* (running perpendicular to the dimer rows in the [110] direction). We know that on an arsenic-enriched 2×4 surface, type *A* steps are straight while type *B* steps are highly kinked because of a difference in their formation energies [11, 16]. The situation changes after the formation of the 6×2 phase: first the rough edge of step *B* becomes straight while the smooth edge of step *A* becomes rougher so that in the 6×2 phase step *A* is characterized by a higher formation energy than step *B*. Since step *B* contains no kink and step *A* is smoother than *B* on an arsenic-enriched surface, the absolute value of the kink formation energy on the 6×2 surface will be higher than that on a 4×2 surface.

Figures 4 and 5 show several groups of surface defects. In the [1[bar]10] direction we observe two types of domain walls, I and II, and in the [110] direction we observe two different types of defects: the missing of bright protrusions (shown by the white arrows in Fig. 4) and kinks where the spacing between neighboring protrusions is shifted to 4 Å (shown by the black arrow in Fig. 4 and more clearly by the arrow on the magnified image of the surface in Fig. 5a). The sequence of the neighboring rows near the kink changes and leads to the antiphase $c(12 \times 2)$ (Fig. 5b). High-resolution STM images of this surface in the filled (Fig. 5a) and empty (Fig. 5b) state regime exhibit the same structural features. The 2× periodicity in the

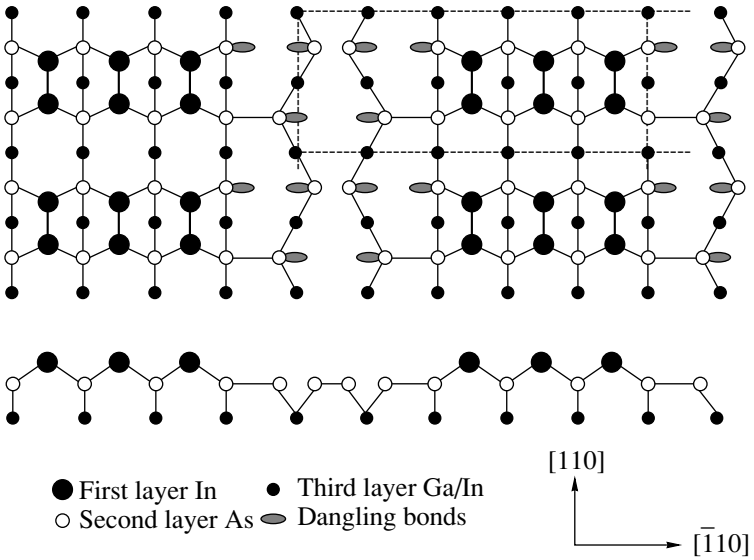

Fig. 6. Geometric model of 6×2 reconstruction.

[110] direction is illustrated by typical linear chains consisting of regularly spaced oblong protrusions ($L1$ scanning profile) and the $1\times$ periodicity is illustrated by the $L2$ scanning profile. In the image of the filled states, the rows look brighter than the protrusions at 0.30 Å ($L3$ profile) whereas in the empty state regime the protrusions are imaged more brightly than the rows at 0.45 Å ($L4$ profile), For the (001) polar face of a covalent InAs crystal where the filled states are localized predominantly at the anion sites and the empty states are localized at cation sites, the observed dependence of the contrast on the polarity of the bias voltage implies that the bright rows correspond to As atoms in the second layer and the oblong protrusions correspond to In dimers. The model of the 6×2 reconstruction developed on the basis of this reasoning (Fig. 6) consists of three In dimers in the first layer and two dimerized As atoms in the second layer (six dangling bonds per unit cell) and shows good agreement with the Pashley electron counting rule [11]. Essentially this model is a compromise between the surface density of the dangling bonds and the surface elastic strain. A high density of dangling bonds (a maximum of 8 per unit cell) is energetically unfavorable and thus, as they tend to equilibrium, the surface atoms will form additional bonds, and in particular they will form pairs or dimers in order to reduce the number of dangling bonds. In the limit we obtain four dangling bonds per unit cell which leads to an excessively high surface tension.

In order to check this model we carried out a series of experiments on the annealing of the 6×2 phase under ultrahigh vacuum conditions. As the temperature increased, the number of protrusions decreased monotonically which, bearing in mind the higher binding energy of the GaAs crystal compared with InAs, means that these can be ascribed to indium atoms. Another conclusion is the appearance of a locally ordered 6×6 phase (Fig. 7) which is characterized by large oval protrusions formed after the desorption of all the indium atoms in the first layer and also the arsenic atoms in the second layer. On comparing this phases with the gallium-enriched GaAs(001)-4 $\times$ 6 phase which was obtained under similar conditions [17], we reached the conclusion that the oval protrusions are gallium clusters.

No 6×2 phase was observed, even as a transition phase, on any static phase diagram for the homoepitaxial growth of a (001) GaAs or InAs surface so that it

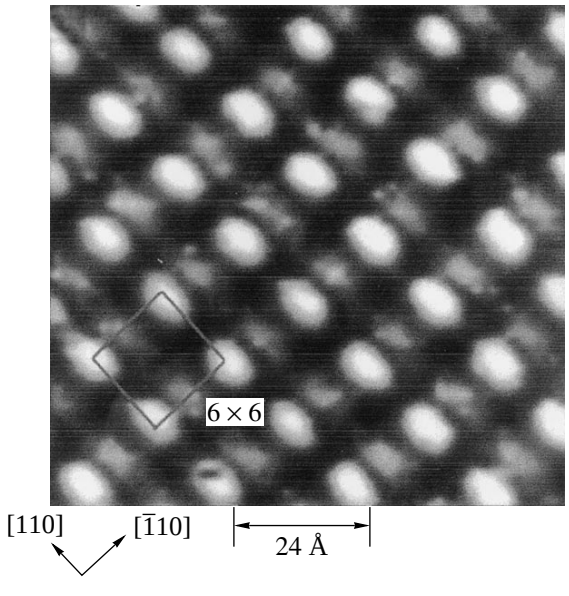

Fig. 7. STM images of an InAs/GaAs surface illustrating the formation of the 6×6 phase (unit cell indicated) as a result of the desorption of In dimers in the first layer and As atoms during annealing at 580°C for 6 min, $V_s = -2.2$ V.

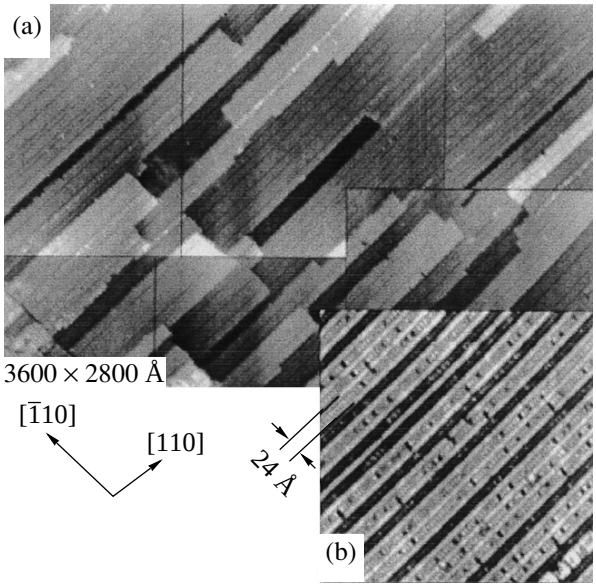

Fig. 8. (a) Evolution of the surface morphology of InAs/GaAs during deposition of 13 epitaxial InAs layers (six successive stages shown, V_S = –2.5 V); (b) Magnified image of the same surface demonstrating typical modulation of the 6×2 structure by domain walls and confirming the removal of elastic strain at the growth fronts, V_S = –2.3 V.

was important to determine how it is formed. The 6×2 phase is commensurate with the substrate structure and only accumulates elastic strain energy as a result of the 7.2% lattice mismatch although similar preparation conditions may lead to the appearance of a coherently strained (i.e., containing no dislocations) 4×2 reconstruction so that an additional mechanism must be included for the appearance of a specific 6×2 phase. In order to demonstrate the special role of the 6×2 structure in the two-dimensional growth of InAs, Fig. 8a shows the evolution of the surface morphology during the successive deposition of 13 InAs monolayers. The layer-by-layer growth is evident as a result of the presence of a terrace–step structure over the entire scanning area although the high rate of surface diffusion under our selected growth conditions should promote a Stranski–Krastanow morphological transition. The fact that this transition did not take place is evidence of an increase in the surface tension of the 6×2 phase. It can also be seen from Fig. 8a that the edges of both types of steps (*A* and *B*) are straight while the islands are more anisotropic than those on the GaAs(001) surface. Since subsequent annealing did not change the configuration, this behavior should be the result of anisotropic diffusion up the steps and/or different sticking coefficients at steps *A* and *B*. In addition, as is shown in Fig. 8b, the surface structure is no longer uniform: the former 6×2 terraces are modulated and contain a considerable number of type I domain walls, and no screw dislocations or stacking faults are observed at the surface. The formation of domain walls is a clear indication that the elastic strain energy is redistributed over the surface of thick strained layers [18]. We postulate that this strain could be partially relieved by the formation of edge dislocations at the interface [13] or by expansion of the lattice in the direction of growth [4, 11] as a result of the elastic longitudinal deformation of the surface, which cannot be observed in an STM.

3.3. Structure of Faceted Faces of Three-Dimensional Islands

We shall now consider the growth of coherently strained (i.e., containing no dislocations) islands in the Stranski–Krastanow mode in a heteroepitaxial InAs/GaAs system. In this case, the deposited materials initially form a two-dimensional pseudomorphic wetting layer on the substrate and after a critical thickness has been reached, they form a three-dimensional island structure. Undoubtedly the main reason for the change in the growth mechanism is that when the next layer is filled, the lattice parameter changes. The formation of three-dimensional islands is usually explained on the basis of the energy balance of the elastic strain and the surface free energy because as a result of the lattice mismatch, the elastic strain energy is accumulated as the wetting layer grows. The main reason for the formation of islands is a possible reduction in the strain as a result of elastic relaxation accompanied by bending of the atomic planes of the lattice which takes place far more efficiently in three-dimensional islands than in two-dimensional layers. During the formation of these islands some of the accumulated energy may be released but as a result of an increase in the total surface area, a higher surface energy is required. Thus, the formation of an array of three-dimensional islands occurs at thickness for which the increase in the surface free energy is compensated by a reduction in the elastic strain energy. Broadly the growth stage of coherently strained islands concludes with the formation of the first mismatch dislocation stimulated by an increase in the elastic strain energy proportional to the volume of the island. If the volume of the island is constant its shape, characterized by its height-to-length ratio, will play an important role in this process. The critical size of the island at which the first mismatch dislocation forms and the residual strain relaxes may be calculated as in the two-dimensional growth regime. In [19–21] the authors proposed various models using the concept of energy balance, and the distortion of the unit cells in an island is considered as a longitudinal uniform plate deformation (similar to that studied in [22]) which allowed them to obtain expressions for the strain/stress in the [001] direction of growth:

$$u_{001} = -2u\frac{\sigma}{1-\sigma}, \tag{3}$$

where σ is the Poisson coefficient of the epitaxial layer (the strains/stresses in the [110] and [$\bar{1}$10] directions

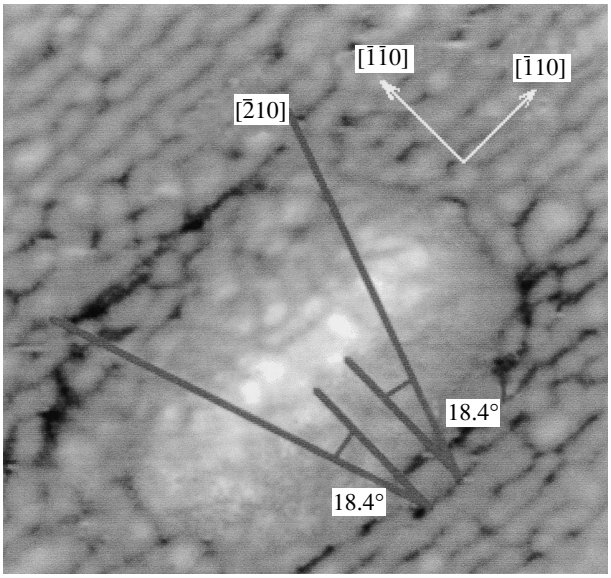

Fig. 9. A magnified STM image of three-dimensional island on a GaAs(001) surface after deposition of 1.6 ML InAs, V_s = –3.5 V.

$u_{110} = u_{\bar{1}10} = u$ are assumed to be isotropic and equal) and the elastic strain energy:

$$E_{\text{strain}} = \frac{2\mu(1+\sigma)}{1-\sigma} V u^2, \tag{4}$$

where V is the volume of the island and μ is the shear modulus of the epitaxial layer, which agree with the results obtained and accurately describe the creation of a new phase. Coalescence of the islands at the later stages of growth is responsible for the formation of new dislocations and smearing of the edges of the epitaxial layer.

At present in most theoretical studies devoted to the mechanism for the formation of three-dimensional islands, these have been assumed to be disk-shaped or hemispherical [23–25]. However, it has been found that the shape of the islands is usually quite intricate and facets having lower surface energy may appear at the surface during reconstruction. In order to study the mechanism for the formation of three-dimensional islands we need to study their atomic structure. Using an atomic force microscope proved to be ineffective [26, 27] and thus, it was advisable to use an STM in the same line as the growth chamber and to observe the for-

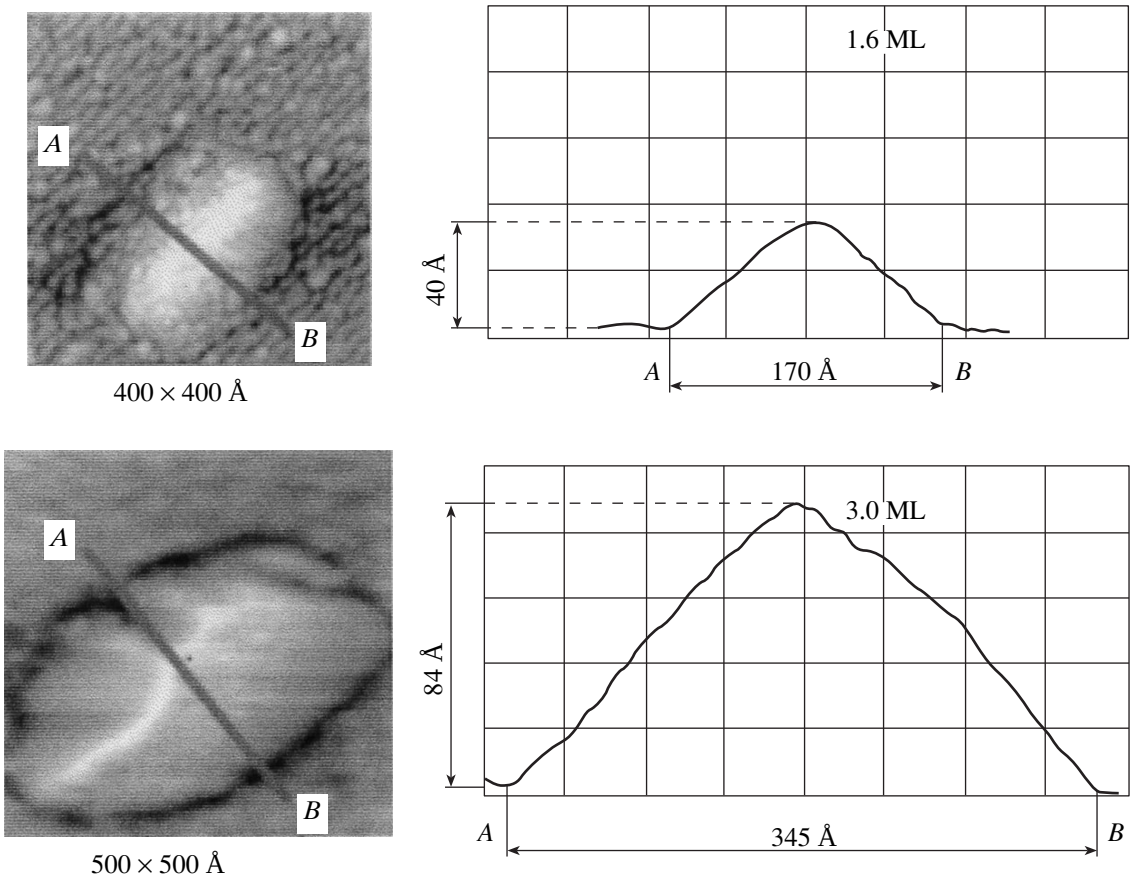

Fig. 10. STM images of isolated quantum dots obtained for various InAs coverages and their cross sections.

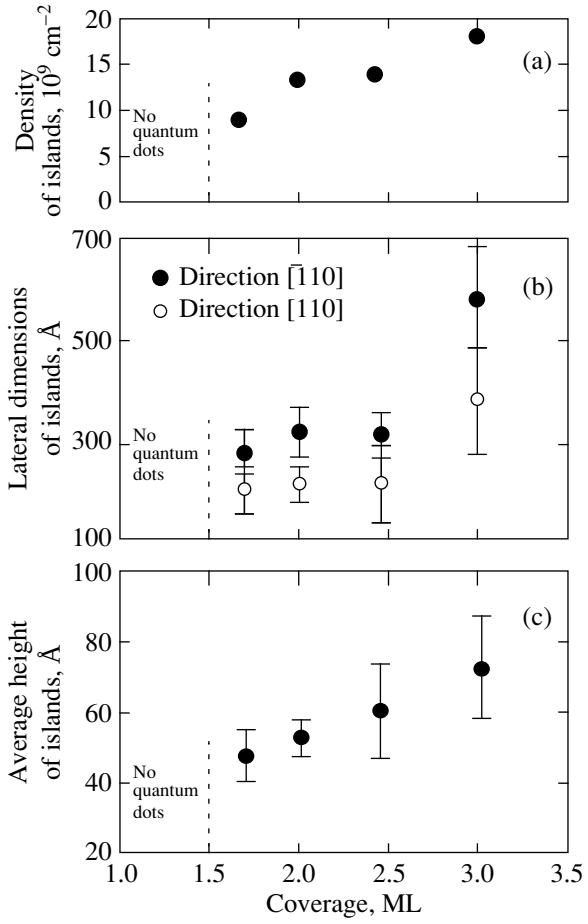

Fig. 11. Characteristics of three-dimensional island structure in an InAs/GaAs(001) heterosystem as a function of the coverage: (a) density of islands; (b) lateral dimensions; (c) average height.

mation of the islands *in situ* with atomic resolution under ultrahigh-vacuum conditions [7]. Figure 9 shows three-dimensional islands formed on a two-dimensional InAs wetting layer. An analysis of the STM images revealed that the preferred sites for the nucleation of three-dimensional islands are steps on the wetting layer. A typical island has an approximately rectangular shape elongated in the [$\bar{1}$10] direction along rows of dimers (Fig. 9). Since the indium atoms must diffuse over a large distance in this direction, the elongated rectangular shape is due to the higher growth rate of the islands in the [$\bar{1}$10] direction. Note that the wetting layer has a 2×4 structure [8–10] but the dimer rows are curved unlike the dimer rows on a GaAs(001) 2×4 surface [7]. Figure 10 shows STM images, cross section profiles (pyramidal shape), and characteristic dimensions of two islands measured along the lines *A–B*. Surface diffusion of the deposited atoms has an appreciable influence on the shape and position of the islands whose density N_s can be estimated using the simple relationship [28]:

$$N_s \sim \frac{1}{\pi L^2} \exp\left(\frac{E_A}{k_B T}\right). \quad (5)$$

In our case, the density of islands for a 1.6 ML InAs coating was approximately 9×10^9 cm^{-2}, the average height was 40 Å, the width in the [110] direction was 170 Å, and in the [$\bar{1}$10] direction 230 Å. The Miller indices of the faceted plane were determined from its angle of inclination relative to the substrate plane.

We used a sequence of STM images obtained for various InAs coatings to obtain information on the geometric parameters of the three-dimensional islands

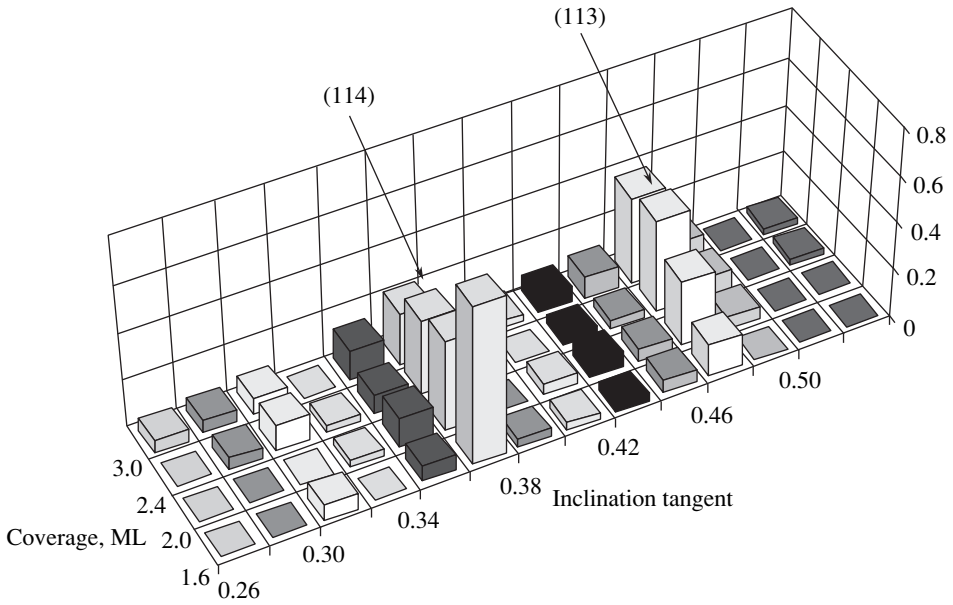

Fig. 12. Histogram showing static distribution of faceted planes over angles of inclination and its dependence on the coverage. The ordinate gives the relative contribution of the various planes with the (113) and (114) planes indicated.

(Fig. 11). It can be seen that in the 1.6–2.6 ML range the lateral dimensions do not change significantly and are approximately 200 Å ([110] direction) and 250 Å ([$\bar{1}$10] direction). For a 3.0 ML coating the dimensions of the islands in these directions increase to 350 and 550 Å, respectively, evidently as a result of the formation of dislocations at the interface of the InAs islands with the GaAs substrate. The average height of the islands increased continuously from 45 Å (1.6 ML) to 70 Å (3.0 ML). Since the lateral dimensions of the islands in the 1.6–2.4 ML range remained constant, the increase in height convincingly indicates that as the coating increases, the side plane becomes steeper with respect to the substrate. These data do not contradict the results obtained by other independent researchers [27–32].

The angles of inclination of the faceted planes in the [110] direction were measured for each island and Fig. 12 gives a complete three-dimensional histogram which clearly reveals maxima for the (113) and (114) planes. In a GaAs crystal these planes have a lower surface energy and thus are stable like the planes with low Miller indices such as (100) and (111). In fact, it was shown in [33, 34] that (113) and (114) oriented GaAs substrates are suitable for preparing flat interfaces even in systems with large lattice mismatch parameters. It can be postulated that these arguments are also applicable to the (113) and (114) planes on an InAs surface. As regards the dependence on the coverage, it was found that at 1.6 ML the (114) faceting [tilted by 19.5° from the (001) substrate] dominates whereas with increasing coverage, the peak corresponding to the (113) plane inclined at an angle of 25.2° increased. In the [$\bar{1}$10] direction the faceting of the islands was not so obvious and it was difficult to determine the angle of inclination of the facets as clearly as in the [110] direction.

On the faceted planes of the three-dimensional islands we observed atomic-scale features. In particular, on the (113) plane along the [210]-axis toward the tip of the island, we observed linear structures forming a pattern similar to a chevron (Fig. 9). As is indicated on the STM image, this direction forms an angle of 18.4° with the [110]-axis and the distance between the lines is approximately four times the lattice constant of the heterostructure so that this structure may be considered to be a 4×1 reconstruction of the InAs(113) surface. Since the surface of the sample was enriched in arsenic in accordance with the preparation conditions and the observed STM images were obtained in the filled state regime ($V_s < 0$), these linear structures may be considered to be rows of As dimers in the [$\bar{2}$11] direction. On the basis of these results and satisfying the electron counting rule [11] we developed a structural model (Fig. 13) according to which, in addition to the formation of As dimers, two neighboring In atoms in the unit cell also form a bond in order to saturate their dangling bonds. We note that the GaAs(113) surface has been studied by various methods including STM and atomic-resolution images revealed similar linear structures [35] although the direction along which these linear features run was [$\bar{3}$32] rather than [$\bar{2}$11] as on the faceted planes of the InAs islands. Similar structures were also present on the STM images of the (114) faceted plane and these also differ from those observed on a planar GaAs(114) surface [36]. This suggests that the faceted planes of three-dimensional islands have a different atomic structure compared with their planar counterparts.

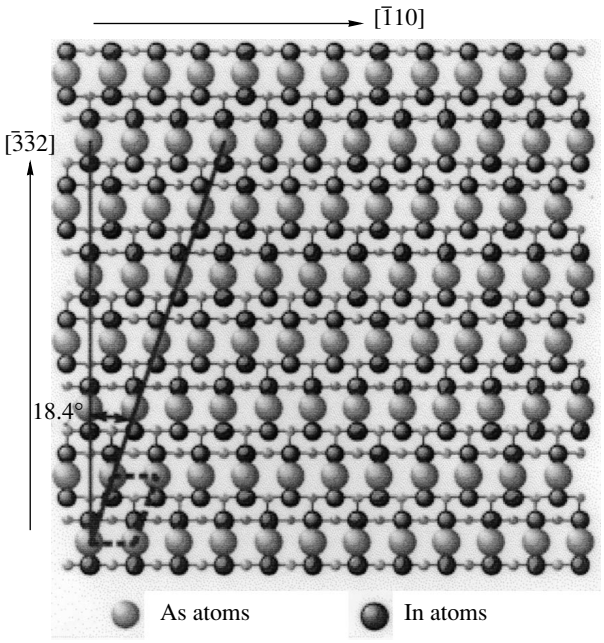

Fig. 13. Structural model of faceted (113) plane.

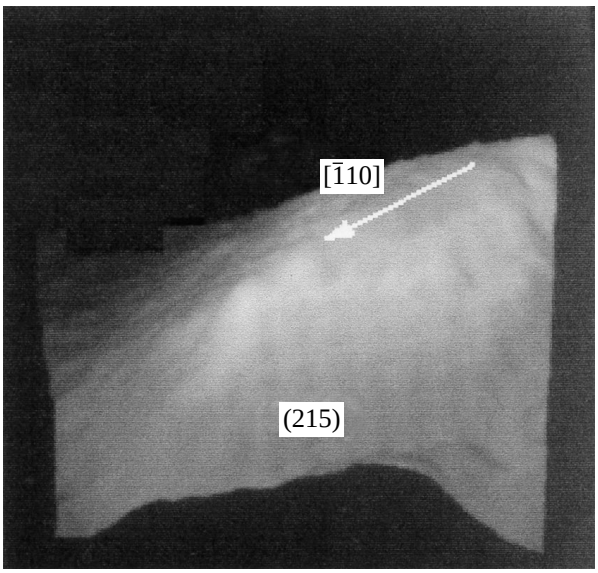

Fig. 14. Three-dimensional STM images of InAs quantum dot on GaAs(001) surface, scanning area 130×130 Å^2, V_s = –3.5 V; the faceted plane (215) and the [$\bar{1}$10] direction are indicated.

In addition, a careful study of the STM images revealed that on some islands the faceted planes are tilted from the (113) plane. These systems of facets were observed most frequently when other islands were localized close to the faceted plane. Figure 14 shows an example of a similar facet orientation whose analysis allowed us to ascribe the index (215) to the faceted plane. On this plane we can clearly identify regularly distributed protrusions mapping As dimers. We developed a structural model of this plane with a 1×1 unit cell containing a single As dimer and one In bond similar to the model of the (113) plane which is also consistent with the electron counting rule.

4. CONCLUSIONS

We have used atomic-resolution STM to study *in situ* the initial stages of InAs/GaAs(001) heterostructure growth according to the Stranski–Krastanow mechanism, and the ensuing surface reconstructions. We established the important role of the 4×2 phase as a template for the layer-by-layer growth of a high-quality interface. We observed the formation of domain walls which facilitated the two-dimensional growth of epitaxial InAs films. For the first time we observed a new 6×2 phase whose formation appreciably improved the morphological stability and structure of the substrate.

We investigated the nucleation of three-dimensional InAs islands on a GaAs substrate and studied their characteristics. For the first time we obtained STM images which reveal the atomic structure of the faceted planes of these islands and we proposed structural models for these. We showed that the atomic structures of the faceted planes differ from the structures of their flat counterpart and geometric models of flat surfaces cannot be applied to faceted surfaces.

ACKNOWLEDGMENTS

This work was supported by the Federal Program "Surface Atomic Structures" of the Russian Ministry of Science and Technology (project 3.4.99) and by the Russian Foundation for Basic Research (project no. 99-02-17382).

REFERENCES

1. *Proceedings of the JRDC International Symposium on Nanostructures and Quantum Effects, Tsukuba, Japan, 1993*, Ed. by H. Sakaki and H. Noge (Springer-Verlag, Berlin, 1994).
2. L. Jacak, P. Hawrylak, and A. Wojs, *Quantum Dots* (Springer-Verlag, Berlin, 1998).
3. *Mesoscopic Physics and Electronics*, Ed. by T. Ando, Y. Arakawa, K. Furuya, S. Komiyama, and H. Nakashima (Springer-Verlag, Berlin, 1998).
4. I. N. Stranski and L. von Krastanow, Akad. Wiss. Lit. Mainz, Abh. Math. Naturwiss. Kl. **146**, 797 (1939).
5. W. J. Schaffer, M. D. Lind, S. P. Kowalczyk, and R. W. Grant, J. Vac. Sci. Technol. B **1**, 688 (1983).
6. J. Tersoff, C. Teichert, and M. G. Lagally, Phys. Rev. Lett. **76**, 1675 (1996).
7. R. Z. Bakhtizin, T. Sakurai, T. Hashizume, and Qikun Xue, Usp. Fiz. Nauk **167**, 1227 (1997) [Phys. Usp. **40**, 1175 (1997)].
8. H. Yamaguchi and Y. Horikoshi, Jpn. J. Appl. Phys., Part 2 **33**, L1423 (1994).
9. Y. Horikoshi, M. Kawashima, and H. Yamaguchi, Jp. J. Appl. Phys. **25**, L868 (1986).
10. M. D. Pashley, Phys. Rev. B **40**, 10481 (1989).
11. S. Ohkouchi and I. Tanaka, Appl. Phys. Lett. **59**, 1588 (1991).
12. S. W. Snyder, J. F. Mansfield, B. G. Orr, *et al.*, Phys. Rev. B **46**, 9551 (1992).
13. A. Trampert, E. Tournie, and K. H. Ploog, Appl. Phys. Lett. **66**, 2265 (1995).
14. Y. W. Mo and M. G. Lagally, J. Cryst. Growth **111**, 876 (1991).
15. R. Butz and S. Kampers, Appl. Phys. Lett. **61**, 1307 (1992).
16. N. Ikoma and S. Ohkouchi, Jpn. J. Appl. Phys., Part 1 **34**, 5763 (1995).
17. R. Z. Bakhtizin, Qikun Xue, T. Sakurai, and T. Hashizume, Zh. Éksp. Teor. Fiz. **111**, 1858 (1997) [JETP **84**, 1016 (1997)].
18. Q.-K. Xue, Y. Hasegawa, T. Ogino, *et al.*, J. Vac. Sci. Technol. B **15** (4), 1270 (1997).
19. R. Vincent, Philos. Mag. **19**, 1127 (1969).
20. N. Cabrera, Surf. Sci. **2**, 320 (1994).
21. W. A. Jesser and J. H. van der Merwe, Surf. Sci. **31**, 229 (1972).
22. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 7: *Theory of Elasticity* (Nauka, Moscow, 1965; Pergamon, New York, 1986).
23. S. Priester and M. Lannoo, Phys. Rev. Lett. **75**, 93 (1995).
24. H. T. Dobbs, D. D. Vvedensky, and A. Zangwill, Phys. Rev. Lett. **79**, 897 (1997).
25. S. A. Kukushkin and A. V. Osipov, Usp. Fiz. Nauk **168**, 1083 (1998) [Phys. Usp. **41**, 983 (1998)].
26. D. Leonard, K. Pond, and P. M. Petroff, Phys. Rev. B **50**, 11687 (1994).
27. G. S. Solomon, J. A. Trezza, and J. S. Harris, Jr., Appl. Phys. Lett. **66**, 991 (1995).
28. K. Tillmann, D. Gerthsen, P. Pfundstein, *et al.*, J. Appl. Phys. **78** (6), 3824 (1995).
29. N. P. Kobayashi, T. R. Ramachandran, P. Chen, and A. Madhukar, Appl. Phys. Lett. **68**, 3299 (1996).
30. J. M. Moison, F. Houzay, F. Barthe, *et al.*, Appl. Phys. Lett. **64**, 196 (1994).
31. A. Madhukar, Q. Xie, P. Chen, and A. Konkar, Appl. Phys. Lett. **64**, 2727 (1994).
32. Y. Nabetani, T. Ishikawa, S. Noda, and S. Sasaki, J. Appl. Phys. **76**, 347 (1994).
33. S. Shimomura, A. Wakejima, A. Adachi, *et al.*, Jpn. J. Appl. Phys., Part 2 **32**, L1728 (1993).
34. Y. Hsu, W. I. Wang, and T. S. Kuan, Phys. Rev. B **50**, 4973 (1994).
35. M. Wassermeier, J. Sudijono, M. D. Johnson, *et al.*, Phys. Rev. B **51**, 14721 (1995).
36. T. Yamada, H. Yamaguchi, and Y. Horikoshi, J. Cryst. Growth **150**, 421 (1995).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 1011–1016.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1167–1173.

SOLIDS
Structure

Order–Disorder–Order Phase Transitions Between Metastable Modifications of Biperiodic Stripe Domain Structures

G. V. Arzamastseva, F. V. Lisovskiĭ*, and E. G. Mansvetova

Institute of Radio Engineering and Electronics, Russian Academy of Sciences, Fryazino, Moscow oblast, 141120 Russia
**e-mail: lisf@dataforce.ru*

Received December 28, 1999

Abstract—Magnetooptic methods were used to observe the existence of magnetic-field-induced order–disorder–order phase transitions between metastable modifications of biperiodic stripe domain structures in magneto-uniaxial iron garnet films having a low positive anisotropy constant. It is shown that the loss of long-range order in the system in a certain range of variation of the field is caused by the loss of correlation between the quasi-harmonic surface distortions of the profile of neighboring domain walls. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In the present paper we report results of an experimental investigation of the properties of biperiodic stripe domain structures in thin uniaxial magnetic films having a low positive anisotropy constant β_u and the axis of easy magnetization directed along the normal **n** to the surface (subsequently called the *z*-axis). These structures which were observed by the authors in [1] and exist in a specific range of film thicknesses $L_{cr}^{*} < L < L_{cr}^{(1)}$ are characterized by an ordered system of continuous stripe domains of period *d* oriented in some specific direction (subsequently called the *y*-axis), separated by walls whose profile undergoes quasi-harmonic surface modulation with the period Λ. In the absence of a magnetic field all the domain walls undergo "in-phase" modulation, i.e., the distance between neighboring domain walls along the *x* axis is a constant equal to *d*.

It was shown in [2] that as the film thickness increases, the amplitude of modulation of the domain wall profile a_Λ increases monotonically from zero for $L = L_{cr}^{*}$ (a second-order phase transition in terms of thickness from a monoperiodic to a biperiodic domain structure) to a value approximately equal to *d*, undergoing a jump of approximately $d/4$ for $a_\Lambda/d = 0.5$ (a first-order phase transition in terms of thickness between two modifications of biperiodic domain structures with in-phase modulation of the domain wall). When $L > L_{cr}^{(1)}$ the domains in the structure begin to branch and then cease to be stripe domains. The vast majority of published studies have been devoted to the properties of biperiodic domain structures in films of varying thickness and the construction of various theoretical models (using the approximation of structureless "geometric" domain walls) to explain the reasons for the formation of these structures (see the definitive studies [1–4] and also the bibliography in [5, 6]). In earlier experiments in the presence of a magnetic field the emphasis was on studying hysteresis loops and mechanisms for the magnetic reversal of films; no field-induced phase transitions were reported in the range of existence of biperiodic domain structures.

It was recently observed [5, 6] that in the presence of a magnetic field

$$\mathbf{H} = H_{\parallel}\mathbf{e}_z + H_{\perp}\mathbf{e}_{\perp}$$

several types of regular biperiodic domain structures (DS) may exist in a certain range of field strength and orientation in uniaxial magnetic films with $\beta_u \leq 1$. These structures include: DSI where the profile of all the domain walls undergoes quasi-in-phase modulation at each surface of the film (in each domain wall the modulation of the profile at different surfaces is in antiphase); DSII where the profile of neighboring domain walls undergoes quasi-antiphase modulation at each surface of the film (in each domain wall the modulation of the profile at different surfaces takes place independently); DSIII is a hybrid domain structure having twice the period of the blocks generating it, in which sections with in-phase and antiphase modulation of the neighboring domain wall profile alternate systematically in pairs. As the strength of the magnetizing field increases or decreases, first- or second-order phase transitions take place between various types of domain structure. For example, in the presence of a weak field $H_{\parallel}$ = const and a field $H_{\perp}$ whose strength increases continuously from zero, we observe the following chain of phase transitions: DSI ⟶ DSIII ⟶ DSII ⟶ simple (monoperiodic) DS ⟶ uniformly magnetized state. For $H_{\parallel} = 0$ the component corre-

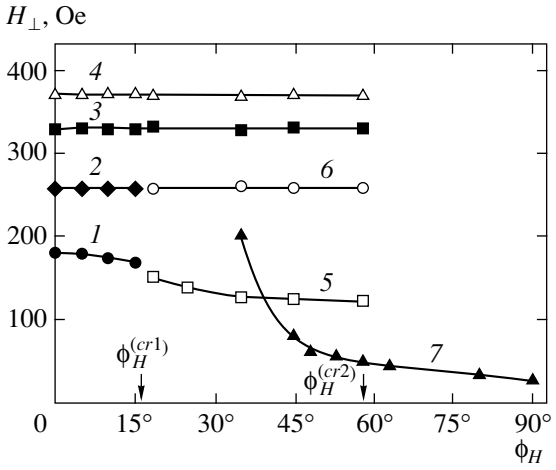

Fig. 1. Phase diagram of the 16 μm thick $Lu_{2.1}Bi_{0.9}Fe_5O_{12}$ film No. 1 on the plane (ϕ_H, $H_\perp$). Explanations to curves *1–7* are given in the text.

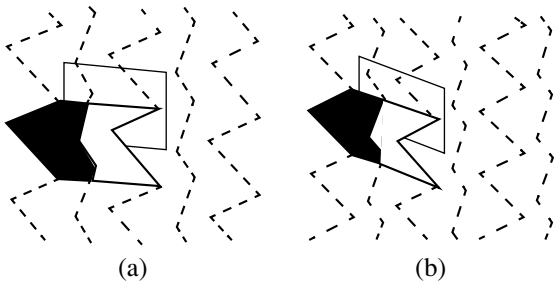

Fig. 2. Schematic diagram of two possible types of biperiodic domain structure where the unit cell profile differs (a) weakly and (b) strongly from rectangular.

sponding to the hybrid DSIII is omitted from this chain of phase transitions. The symmetry of the optical diffraction patterns observed for normally incident light on a film having any of these types of domain structure is characterized by the same *mm2* point group. The phase diagram of the films on the plane ($H_\perp$, $H_\parallel$) was determined for the case when the component $H_\perp$ is collinear to the direction in which the stripe domains are oriented (the *y* axis).

In the present study we report results of an experimental study of the stability of various types of biperiodic domain structures relative to a symmetry-perturbing external action, specifically the magnetic field component H_x, and we also analyze the ensuing metastable distributions of the magnetic moment.

2. EXPERIMENTAL RESULTS

Experiments were carried out using uniaxial films of magnetic garnets (for a detailed description of the properties of the films see [5, 6]) at $T = 293$ K and $H_\parallel = 0$ by the following method. First we determined the saturation field $H_\perp > H_\perp^*$ which was then decreased continuously to zero with the result that a regular biperiodic domain structure with domain walls along the *y*-axis formed in the film. The film was then turned about the normal to the surface by a certain angle ϕ_H and the response of the domain structure to a continuous increase in the field $H_\perp$ was studied using a polarizing microscope.[1] We also monitored the diffraction patterns (using 0.6328 μm laser radiation) which played a decisive role in identifying the types of domain structure observed. As a result of the existence of a rotational coercive force [7] the initial direction of the domain wall was conserved in a certain range of $H_\perp$ where the width of this range depended on the selected value of ϕ_H.

We shall illustrate the main results of a study of the phase transitions between different types of domain structure in the presence of the component H_x for one film 16 μm thick having the composition $Lu_{2.1}Bi_{0.9}Fe_5O_{12}$ (film No. 1) for which the period of the domain structure *d* in the absence of a magnetizing field was 5.6 μm and the period of the quasi-harmonic distortions was $\Lambda = 2.1$ μm.[2] The apparent value of the uniaxial anisotropy constant $\beta_u^* = H_\perp^*/4\pi M$ for this film was 0.21. For clarity the data obtained are given as a phase diagram (Fig. 1) on the plane (ϕ_H, $H_\perp$). Curves *1–7* give a set of points corresponding to the upper limit of the range of stability of a specific type of domain structure for a continuously increasing field $H_\perp$ for $H_\parallel = 0$.

(1) For small ($\phi_H \leq \phi_H^{(cr1)} \approx 17°$) deviations of the direction of the field $H_\perp$ from the *y* axis the evolution of the domain structure with increasing field was the same as that for $\phi_H = 0$, i.e., the following chain of transitions was observed: DSI ⟶ DSII ⟶ monoperiodic DS ⟶ uniformly magnetized state. The only characteristic feature was that for any nonzero value of H_x the structures of the neighboring domain walls (and neighboring domains) in the domain blocks differed and the shape of the unit cell also exhibited a small (a few degrees) deviation from rectangular, as is shown schematically in Fig. 2a. The motif-forming element is shown as a black and white figure having its perimeter outlined by a heavy solid line, the domain walls are shown by the dashed lines, and the unit cell is a parallelogram (thin solid lines) having a corner angle close to π/2; the black and white sections correspond to domains with $M_z > 0$ and $M_z < 0$. The field component H_x amplified the peri-

[1] For the case $\phi_H \neq 0$ we cannot use the more informative procedure of cyclic magnetic reversal of the films, as was used to study the domain structures observed for $\phi_H = 0$ [5, 6] because after saturation of the film followed by a reduction in the magnetic field the incipient stripe domains are oriented in the direction $\mathbf{H}_\perp$. This implies that all the nonuniform distributions of the magnetic moment described in this study are metastable.

[2] Results of an investigation of the phase transitions between different types of domain structure for $\phi_H = 0$ were described in detailed for this film in [5, 6] so that to avoid repetition we shall subsequently refer to the data in these studies as necessary.

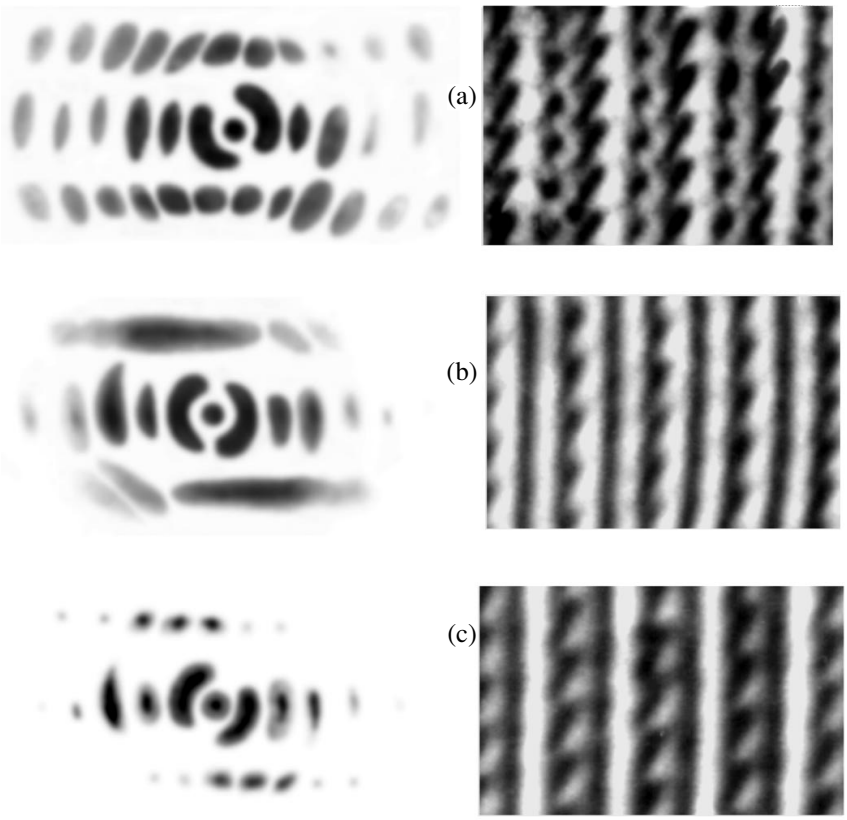

Fig. 3. Photographs of observed domain structures (right) and their corresponding diffraction patterns (left) for film No. 1 for $\phi_H = 35°$ and $H_\perp$ = (a) 100, (b) 120, and (c) 130 Oe.

odic surface modulation of the magnetization distribution profile at one of the walls of each domain and suppressed it at the other. The diffraction pattern corresponding to this domain structure lost the mirror reflection planes and its symmetry was based on point group *2*, not on the *mm2* group as for $\phi_H = 0$.

As the field increased continuously (from zero), at only slightly differing field values, a transition took place from an initial quasi-in-phase domain structure to a quasi-antiphase structure (curve *1* in Fig. 1) with different modulation periods of the domain wall profile at the free surface ($\Lambda_{1a} = 1.8$ μm) and at the film–substrate interface ($\Lambda_{2a} = 3.6$ μm). With increasing field, the modulation initially disappears at the first of these surfaces (curve *2*) and in stronger fields, at the second (curve *3*), i.e., a transition takes place to a monoperiodic domain structure.[3] For fields corresponding to curve *4* the film is converted to the single-domain state (compare with Fig. 2 in [6]).

(2) If the angle ϕ_H is larger than $\phi_H^{(cr1)}$, the processes observed with continuously increasing field $H_\perp$ were initially the same as those described above: a quasi-in-phase biperiodic domain structure with a slightly distorted rectangular unit cell was stable in weak fields (Fig. 2a). Photographs of the domain structure for this case and the corresponding diffraction pattern are shown in Fig. 3a. As the field $H_\perp$ increased further in a certain narrow range we observed some smearing of the diffraction peaks caused by modulation of the domain walls whose character indicated that the period of the modulation distortions of the magnetization distribution profile remained almost the same for all domain walls but the spatial phase shift between distortions in neighboring domain walls was not constant (see Fig. 3b), i.e., the corner angle of the unit cell fluctuated perceptibly, as shown in Fig. 2a. Above a certain critical value (curve *5* in Fig. 1) however, these fluctuations disappeared and a regular domain structure formed, characterized in that the quasi-harmonic distortions of the magnetization distribution profile in domain walls of the same type (distributed alternately) were shifted relative to each other by approximately half a period. In this case, all the diffraction peaks $J_{(p, q)}$ with $q \neq 0$ were shifted rela-

[3] In this range of fields the modulation period Λ increases monotonically for the in-phase domain structure. First it is the same for both developed surfaces ($\Lambda_{1s} = \Lambda_{2s} = \Lambda$) but as we approach the range of stability of the antiphase domain structure the values of $\Lambda_{1s} > \Lambda_{1a}$ and $\Lambda_{2s} < \Lambda_{2a}$ begin to differ, the difference between them increasing monotonically with increasing field, remaining less than $\Lambda_{2a} - \Lambda_{1a}$ (see Fig. 2 in [6]). As a result of the inevitable nonuniformity of the properties in the film in the narrow range of fields a transition takes place from a in-phase to an antiphase domain structure, both types of structure coexist and domain walls having four modulation periods of the profile are observed simultaneously. On the diffraction patterns this is observed as splitting of all the diffraction peaks $J_{(p, q)}$ with $q \neq 0$ into doublets.

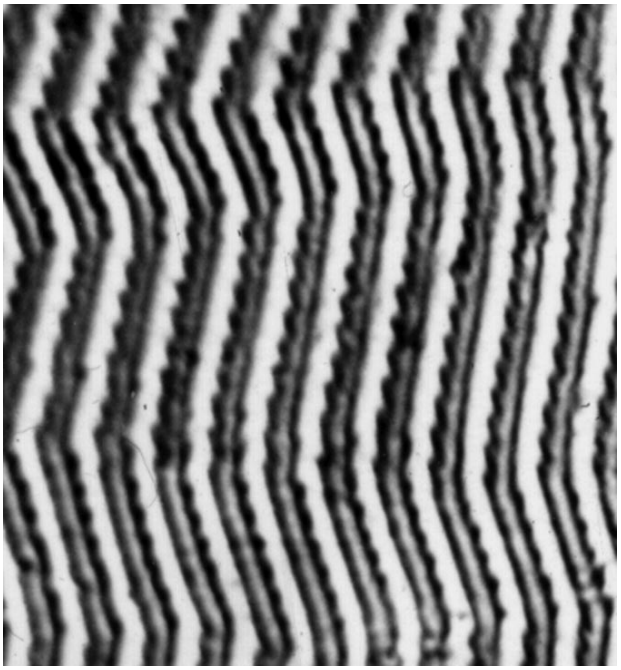

Fig. 4. Photograph of the domain structure of 10 μm thick $Lu_{2.1}Bi_{0.9}Fe_{4.83}Mg_{0.17}O_{12}$ film No. 2 at $H_\perp = 0$.

tive to the peaks $J_{(p, q-1)}$ parallel to the abscissa by an amount close to half the spacing between the neighboring initial reflexes (see Fig. 3c). This may be treated as a first-order phase transition between states having different unit cell profiles and specifically for values of $H_\perp$ below curve *5* in Fig. 1 the acute angle at the corner of the parallelogram is close to $\pi/2$ whereas above the critical value of $H_\perp$ this angle is approximately $\arctan(2d/\Lambda)$, see Fig. 2.

A characteristic feature of this phase transition is the exact agreement between the modulation periods of the domain wall profile Λ at the free surface of the film and at the film–substrate interface which is confirmed by the complete absence of any tendency to form typical doublets on the diffraction patterns which are observed for $\phi_H < \phi_H^{(cr1)}$.

(3) As the magnetic field increases, the domain structure formed having an oblique-angled unit cell remains stable over a fairly wide range of $H_\perp$ but after a certain critical value has been exceeded (curve *6* in Fig. 1) this domain structure becomes unstable with respect to strong fluctuations of the unit-cell corner angle which leads to such severe smearing of the diffraction peaks $J_{(p, q)}$ with $q \neq 0$ that they almost merge into continuous bands. The diffraction pattern then has a form similar to that shown in Fig. 3b. This smearing persists until a transition takes place to a monoperiodic domain structure (curve *3* in Fig. 1); conversion to the single-domain state takes place in a stronger field (curve *4* in Fig. 1).

(4) For any value of $\phi_H \neq 0$ the asymmetry of the distribution of the magnetization vector in neighboring domain walls becomes increasingly strong as the field increases; a visual observation under conditions when $\phi_H > \phi_H^{(cr1)}$ shows that the modulation of the profile for half the domain walls (every other one) at each surface of the film becomes almost indiscernible in fairly weak fields (see the photographs in Figs. 3b and 3c obtained by focusing on the outer surface of the film). Each of the domain walls has a modulated profile on one surface and an unmodulated one on the other and at neighboring walls the modulation is suppressed at different surfaces.

(5) There is a second critical value of the angle $\phi_H = \phi_H^{(cr2)}$ (for film No. 1 this value is around 30°) above which in a fairly strong field (above curve *7* in Fig. 1) biperiodic domain structures become unstable with respect to the formation of large-scale (relative to the periods d and Λ) kinks (breaks) in the domain walls with the result that these walls are transformed into zigzag structures with two types of subblocks. In the first type of subblock the orientation of the domain walls remains close to the initial orientation whereas in the second type of subblock all the domain walls are turned through a small angle relative to the y axis in the direction of the vector $\mathbf{H}_\perp$. With increasing $H_\perp$, as a result of the creation and migration of kinks subblocks of the first type become displaced and the domain walls in all the subblocks rotate in the direction of the magnetic field. If the kinks are small-angled (a few degrees), as is the case for $\phi_H \lesssim 60°$, they have almost no influence on the processes of rearrangement of the domain structure; for $\phi_H > 60°$ even in comparatively weak fields the domain wall kinks become so strong that the evolution of the domain blocks follows a completely different scenario.

(6) If the field intensity is reduced without allowing reorientation of the domain walls and (or) conversion to the single domain state, the chain of phase transitions described above takes place in the reverse order with appreciable hysteresis whose width in terms of the magnetic field depends on the angle ϕ_H.

This pattern of phase transitions can be made more complex if the magnetic anisotropy energy of the films includes a contribution attributable to the crystallographic (cubic) anisotropy. For the orientation of the films used in our experiments [the substrates were cut in the (111) plane] even for $\phi_H = 0$ the neighboring domain walls will only have the same magnetization distribution profile if the field $\mathbf{H}_\perp$ is assigned to one of the mirror reflection planes passing across the [111] axis. If this condition is not satisfied, only domain structures of the type shown in Fig. 2a having different distribution profiles of the magnetization vector in neighboring domain walls will exist. In addition, in this case we can also observe the formation of large-scale

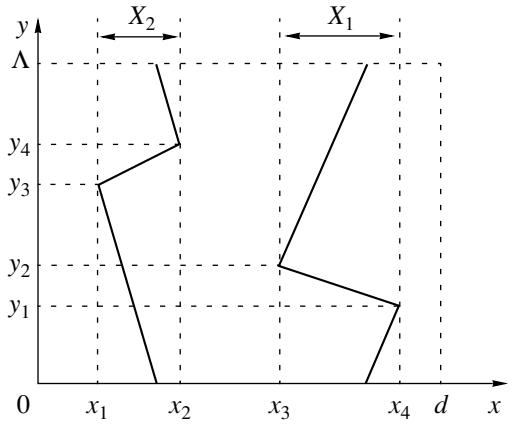

Fig. 5. Model of domain walls used to calculate the intensity of the diffraction peaks.

kinks in the domain walls similar to those formed under the action of the field H_x.

This is illustrated in Fig. 4 which shows a photograph of the domain structure of a 10 μm thick, (111)-oriented demagnetized ($H_\perp = 0$) film No. 2 whose composition only differed from that of film No. 1 in that it contained a small (0.17 per formula unit) quantity of magnesium ions (the influence of cubic anisotropy was manifest most strongly in these films although weaker effects were observed in films of any composition). Large-scale kinks can be observed and a difference between the images of neighboring domain walls is also clearly visible. The orientation of the stripe domains was specifically selected so that the reduction in the symmetry of the magnetization distribution was most clearly defined. Note that this effect should be observed not only in biperiodic but also in monoperiodic domain structures.

The presence of a cubic component of the magnetic anisotropy also has the result that when domains are created from the saturated state, the stripe domains are generally not strictly oriented parallel to the field $\mathbf{H}_\perp$ but are deflected from it by some angle (up to 10°) where the magnitude of the deflection depends on the sign of the field (with its orientation unchanged).

In order to check whether the various modifications of the biperiodic domain structures observed using the magnetooptic diffraction of light are correctly identified, we used a well-known scheme (see, e.g., [8–10]) to make theoretical calculations of the intensity of the diffraction peaks. The model shown in Fig. 5 was used to calculate the structure factor for various types of domain structures. The profile of the domain walls was approximated by broken lines (cf. Fig. 2), the amplitudes of the distortions of neighboring domain walls $X_1 = x_4 - x_3$ and $X_2 = x_2 - x_1$ were generally assumed to be different, and the positions of the projections and indentations on the domain walls were characterized by the values y_i defined by the following relationships:

$$y_1 = c\Lambda, \quad y_2 = \Lambda/2 - y_1,$$
$$y_3 = \Lambda/2 + y_1, \quad y_4 = \Lambda - y_1,$$

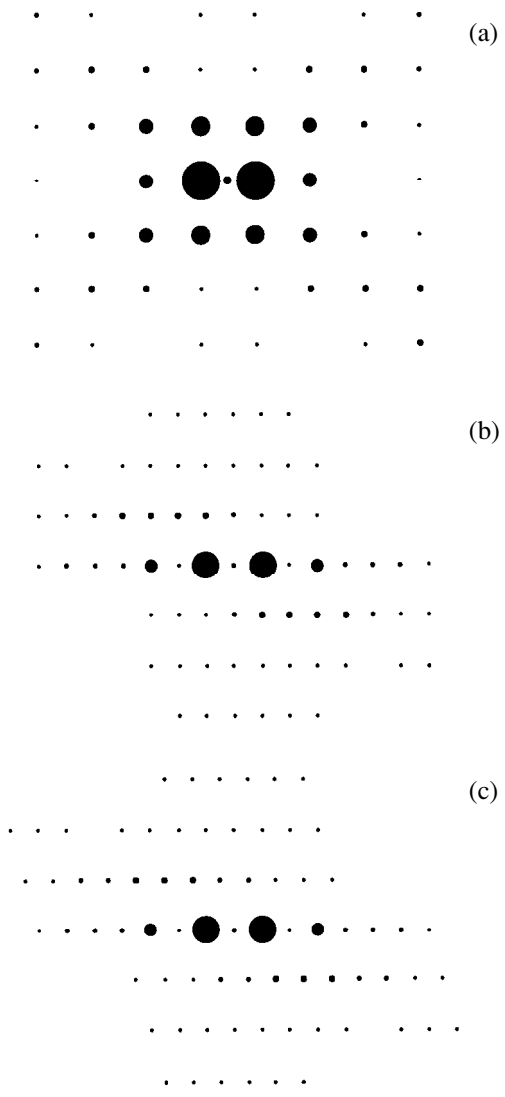

Fig. 6. Theoretical diffraction patterns for various types of biperiodic stripe domain structures: (a) biperiodic domain structure with in-phase modulation of the domain wall profile and a rectangular unit cell; (b) biperiodic domain structure with a negligibly small deviation of the unit cell shape from rectangular; (c) biperiodic domain structure with an oblique-angled unit cell.

where the numerical parameter c determines the degree of deviation of the domain structure from a symmetric ($H = 0$) in-phase structure for which $c = 0$. It was also assumed that

$$\frac{x_1 + x_2}{2} = \frac{d}{4}, \quad \frac{x_3 + x_4}{2} = \frac{3d}{4}.$$

We used a step approximation for the distribution of the component M_z and a model with a triangular distribution where M_z goes to zero at the domain wall. For the low-order diffraction peaks observed in real experiments both approaches yield almost the same results.

Although calculations were made for all possible types of domain structure, we shall merely confine our analysis to the three examples shown in Fig. 6 and relating to film No. 1. The first (a) describes diffraction at a symmetric in-phase domain structure ($y_1 = 0$, $y_2 = y_3 = \Lambda/2$, $y_4 = \Lambda$, $X_1 = X_2 = 1.1$ μm), while the second and third (b and c) model the situation shown in Figs. 3a and 3c, respectively. The values of the geometric parameters for the second and third examples which reflect the real experimental results were selected as follows: $y_1 = y_3 = \Lambda/4$, $y_3 = y_4 = 3\Lambda/4$, $X_1 = 1.1$ μm, for a domain structure with a negligible deviation of the cell profile from rectangular (Fig. 6b) $X_2 = 0.3$ μm, and for a domain structure with an oblique-angled cell (Fig. 6c) $X_2 = 0$. In this last case, in the calculations of the structure factor integration was performed over a doubled cell (compared with that shown in Fig. 5) comprising a combination of two initial cells shifted relative to each other by the distance d along the abscissa in one of which the domain walls were shifted by $\Lambda/2$ along the ordinate. The diffraction peaks are shown by the black circles whose area is proportional to the intensity and the centers are positioned at the nodes of a reciprocal lattice formed by translations of the basis vectors. A comparison with the experimental results (see also Figs. 4 and 7 from [6]) shows that good agreement is observed between the calculated and experimental diffraction patterns.

3. CONCLUSIONS

An analysis of the experimental results shows that in the presence of a magnetic field directed at an angle to the plane of the domain walls, only those biperiodic domain structures which possess certain unit cell corner angles, either close to 90° or differing negligibly from $\arctan(2d/\Lambda)$, are stable. The magnetic-field-induced first-order phase transition between these domain structures belongs to the order–disorder–order type, i.e., it takes place via an intermediate amorphized state characterized by a specific type of partial loss of long-range order caused by strong fluctuations of the unit-cell corner angle. The phase transition from a biperiodic domain structure (having a unit cell which differs substantially from rectangular) to a monoperiodic structure also takes place by the same scheme.

To conclude, we note that in [5, 6], in order to discriminate between various modifications of biperiodic domain structures, the present authors used a symmetry description of domain structure "transforms" (photographs) based on the framework of two-dimensional space groups which corresponded to a description of the symmetry of the distribution of the component $M_z(x, y)$ for z = const in any plane parallel to the surfaces of the film. A truncated symmetry description was used because in experiments to observe domain structures and their corresponding diffraction patterns the Faraday effect was used which only gives information on the z component of the magnetization vector; information on the $M_x(x, y, z)$ and $M_y(x, y, z)$ distribution is completely lost. A complete description of these objects in terms of magnetic symmetry can only be given (if we neglect the difference between the properties of the film at the interfaces with free space and the substrate) after obtaining the necessary information, for example, using the Kerr effect or magnetic force microscopy.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 99-02-17404).

REFERENCES

1. B. W. Roberts and C. P. Bean, Phys. Rev. **96**, 1494 (1954).
2. Ya. Katser, Zh. Éksp. Teor. Fiz. **46**, 1787 (1964) [Sov. Phys. JETP **19**, 1204 (1964)].
3. J. Goodenough, Phys. Rev. **102**, 356 (1956).
4. A. Hubert, Phys. Status Solidi **24**, 669 (1967).
5. G. V. Arzamastseva, F. V. Lisovskiĭ, and E. G. Mansvetova, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 701 (1998) [JETP Lett. **67**, 738 (1998)].
6. G. V. Arzamastseva, F. V. Lisovskiĭ, and E. G. Mansvetova, Zh. Éksp. Teor. Fiz. **114**, 2089 (1998) [JETP **87**, 1136 (1998)].
7. R. J. Spain, Appl. Phys. Lett. **3**, 208 (1963).
8. B. Kuhlov, Optik (Stuttgart) **53**, 115 (1979).
9. B. Kuhlov, Optik (Stuttgart) **53**, 149 (1979).
10. F. V. Lisovskiĭ, E. G. Mansvetova, and Ch. M. Pak, Zh. Éksp. Teor. Fiz. **108**, 1031 (1995) [JETP **81**, 567 (1995)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 1017–1028.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1174–1187.

SOLIDS
Structure

Mesoscopic Magnetic Inhomogeneities in the Low-Temperature Phase and Structure of $Sm_{1-x}Sr_xMnO_3$ ($x < 0.5$) Perovskite

V. V. Runov*, D. Yu. Chernyshov, A. I. Kurbakov, M. K. Runova, V. A. Trunov, and A. I. Okorokov

Petersburg Nuclear Physics Institute, Russian Academy of Sciences, Gatchina, Leningrad oblast, 188350 Russia
**e-mail: runov@hep486.pnpi.spb.ru*

Received May 12, 2000

Abstract—Results are presented of studies of the $^{154}Sm_{1-x}Sr_xMnO_3$ system using neutron powder diffraction and small-angle polarized neutron scattering. An analysis of the neutron diffraction spectra showed that at $T <$ 180 K these exhibit typical Jahn–Teller distortions of the manganese–oxygen octahedrons which persist under further cooling and on transition of the sample to a metallic magnetically ordered state. The magnetic contribution to the diffraction is satisfactorily described using the $(A_x(A_y)F_z)$ model and is interpreted as the coexistence of ferromagnetic and antiferromagnetic phases. The exaggerated widths of the diffraction lines indicate an appreciable contribution from microdeformations evidently associated with the inhomogeneity of the system. Small-angle polarized neutron scattering showed that the Sm system for $x = 0.4$ and 0.25 is magnetically inhomogeneous in the low-temperature phase. Ferromagnetic correlations occur on scales of around 200 Å and having dimensions greater than 1000 Å which, combined with the temperature hysteresis of the magnetic small-angle scattering intensity observed for an $x = 0.4$ sample in the low-temperature phase, suggests that the transition is of a percolation nature. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Manganites having the general formula $R^{3+}_{1-x}A^{2+}_xMnO_3$, where R is a rare earth and A is a divalent metal, have recently attracted increased interest following the observation of colossal negative magnetoresistance in these materials and also because they exhibit a broad spectrum of structural and magnetic transitions which evidently lead to the complex mesoscopic magnetic structure of these materials. Researchers have always noted the magnetic inhomogeneity of these systems (beginning with [1]) and have put forward assumptions that magnetic inhomogeneities exist, these being an integral and fundamental characteristic which is responsible for the colossal negative magnetoresistance (see, e.g., the reviews [2, 3] and [4]). This has recently led to an increasing number of studies on small-angle neutron scattering to identify magnetic inhomogeneities associated with phase separation using the electrical conductivity, principally in La perovskites, and analyses of magnetic scattering mainly in the paramagnetic phase [5–11].

The structural homogeneity of the manganese samples used for these studies has frequently been cast into doubt. Indeterminacy of the stoichiometric composition [12, 13], anisotropic diffraction line broadening associated with the presence of a microstructure [14], exaggerated Debye–Waller factors, and inhomogeneities visible by electron microscopy [15] have all been noted. For fundamentally inhomogeneous samples the structural characteristics extracted from diffraction data have the meaning of averaged values. In this context, a study of mesoscopic-scale inhomogeneities using small-angle neutron scattering should be highly informative for investigating substituted rare-earth manganites and may significantly complement and refine the averaged picture visible from Bragg diffraction.

The $Sm_{1-x}Sr_xMnO_3$ system is of particular interest for these studies since, as a result of a difference between the ionic radii of samarium and strontium (r_{Sm} = 1.132 Å, r_{Sr} = 1.31 Å), it is assumed that there will be appreciable local distortions of the lattice (for example, for $x = 0.4$ we have $\sigma^2 = 7.604 \times 10^{-3}$ Å^2, where $\sigma^2 = \sum x_i r_i^2 - \langle r \rangle^2$) which substantially influence the transport and magnetic properties of the manganites [16]. In addition, there are disparities between the results of investigations of thin films and powder samples. For example, data on the optical absorption and resistivity of $Sm_{0.6}Sr_{0.4}MnO_3$ thin films [17] indicate that this compound remains an insulator over the entire range studied, it does not undergo ferromagnetic ordering, and exhibits indications of charge ordering. An investigation of this compound in powder form using electron diffraction analysis and microscopy combined

with a study of the transport and magnetic properties revealed ferromagnetic ordering, the existence of a metal–insulator transition, and the coexistence of several structural types [18]. Assuming that the film and the powder have the same chemical composition, differences in the macroscopic characteristics should be sought in a difference between the characteristic sizes of the films (around 100 nm) and the powder grains (5–10 μm) which emphasizes the need for studies on the mesoscopic scale.

The present study is an attempt to make a more detailed investigation of samarium–strontium manganite $Sm_{0.6}Sr_{0.4}MnO_3$ which includes a study of the macroscopic, magnetic mesoscopic (small-angle polarized neutron scattering, SAPNS), and microscopic (high-resolution neutron powder diffraction, NPD) characteristics of the same powder sample. As far as we are aware, no detailed structural analyses have been made of a samarium–strontium system. Results of preliminary studies of the macroscopic properties and structure of this system with $x = 0.25$ and 0.4 were published in [19–21], and results for the magnetic mesoscopic characteristics were given in [22] where SAPNS was mainly used to study the $x = 0.25$ composition, and in [23]. It was established in [19] that the system has a distorted perovskite structure, where (1) a transition to a magnetically ordered phase (110–130 K) accompanied by a change in the type of conductivity to "metallic" was observed for a sample with $x = 0.4$ and (2) an increase in the magnetic susceptibility (90–100 K) was observed in a sample with $x = 0.25$, the magnetic structure was not determined, and no transition to the metallic state was observed. The magnetic and electrical properties of the samples studied in [19] show a good correlation with the properties of the Sm system described in [18]. With regard to the mesoscopic characteristics, it was shown in [22, 23] that in an Sm system with $x = 0.25$ and 0.4, ferromagnetic correlations and magneto-nuclear cross correlations on scales of 180–250 Å and ferromagnetic correlations on scales of thousands of angstrom exist in the low-temperature range.

In the first section we present the necessary information on the sample, we describe its magnetotransport properties and the characteristics of the neutron scattering experiments. We then analyze the magnetic and crystal structures using the diffraction data and present an interpretation of the small-angle scattering and depolarization. In Section 4 we search for a model for a joint treatment of the macroscopic, mesoscopic, and microscopic data. The main results are presented in the Conclusions.

2. SAMPLE AND EXPERIMENTS

The initial reagents were samarium and manganese oxides and strontium carbonate. The samples were prepared by stepwise solid-phase synthesis from stoichiometric mixtures of the initial components with intermediate grinding and pressing at 50–10 000 kg cm^2. The samples were synthesized in air using alundum crucibles at temperatures of 1000, 1100, and 1200°C, the duration of a single synthesis stage was 6–97 h, and the samples were quenched in air. The synthesized samples were tested to determine the content of the metal components using chemical methods of analysis: a semimicro complexometric titration method was used to determine Sm, Sr, and Mn with a random error not exceeding 1% for Sm and Mn and 2% for Sr relative to oxygen. The Sm content was also determined by a photometric technique. The accuracy of these methods of analysis was confirmed by analyzing artificial solutions and comparing the complexometric and photometric results. Finally we prepared a completely enriched ^{154}Sm powder having the composition $^{154}Sm_{0.590(6)}Sr_{0.410(8)}Mn_{1.00(1)}O_3$ and grain size 5–10 μm. Data on the magnetoresistance of this sample were published in [19] and results of an investigation of the magnetic susceptibility and magnetization were kindly supplied by A. Maignan [24].

The neutron powder diffraction measurements were made using the G4.2 Franco-Russian high-resolution neutron powder diffractometer at the Leon Brillouin Laboratory (Saclay, France). Monochromatic neutrons having the wavelength $\lambda = 2.3433$ Å were used. Most of the data were obtained in a heating regime at temperatures $T = 1.5$, 19.5, 50.5, 72, 87.1, 106, 124.9, and 300 K. In addition, diffraction patterns were also measured at $T = 250$, 180, 145, and 120 K in the sample cooling regime. For the NPD measurements the powder sample was placed in a vanadium cylindrical container 8 mm in diameter.

The small-angle scattering measurements were made using the VEKTOR small-angle polarized neutron scattering device [25] (WWR-M reactor, Gatchina). For the SAPNS measurements we used a sample of the same powder in a 2-mm thick aluminum container measuring 8×45 mm. VEKTOR was fitted with a twenty-counter (^{3}He) detector and a multichannel analyzer which could be used to make a comprehensive polarization analysis in the range of scattering vectors $0 < q < 3 \times 10^{-1}$ Å^{-1} in a slit geometry ($\mathbf{q} = \mathbf{k} - \mathbf{k}'$ where $\mathbf{k}$ and $\mathbf{k}'$ are the wave vectors of the incident and scattered neutrons, respectively), i.e., the four scattering cross sections $S^{\pm,\pm}$ could be measured where $\pm$ are the neutron spin states relative to the magnetic field before and after the sample, respectively. The polarization of the neutron beam was defined as

$$P = \frac{I^+ - I^-}{I^+ + I^-},$$

where $I^\pm$ is the intensity of neutrons having the corresponding spin state relative to the magnetic field. The initial polarization of the neutron beam incident on the sample was $P_0 \simeq 0.94$. The measurements were made in a magnetic field $0 < H < 4500$ Oe (the field lay in the scattering plane at an angle of around 55° to the **z**-axis

directed along **k**) at the wavelength $\lambda = 9.2$ Å ($\Delta\lambda/\lambda = 0.25$). The measurements were made in an RNK10-300 cryorefrigerator in the temperature range $15 \le T \le 300$ K with temperature stabilization of around 0.1 K. The samples were placed in a gaseous helium atmosphere which was used as the heat-exchange gas.

3. MAGNETOTRANSPORT PROPERTIES

The temperature dependence of the resistance of this $Sm_{0.6}Sr_{0.4}MnO_3$ sample given in [19] shows that the compound undergoes a transition to the metal state at $T = 110$ K. The application of a field substantially reduces the resistance at the transition point [$R_{110}(0)/R_{110}(24\ \text{kOe}) = 7$] which allows us to classify $Sm_{0.6}Sr_{0.4}MnO_3$ as a compound having "colossal magnetoresistance" (Fig. 1). At $T > 120$ K the conductivity is of the hopping (polaron) type and at $T \simeq 180$ K an activation energy jump occurs. In a field of 24 kOe this jump disappears. Measurements of the susceptibility and magnetization show that at $T = 110$ K a spontaneous magnetic moment appears in the sample. Characteristics of this compound should include some drop in the magnetization and magnetic susceptibility at $T < 40$ K and the presence of an elongated section on the susceptibility curve at $T > 110$ K [18]. The magnetotransport properties of this sample in the paramagnetic range were analyzed in greater detail in [26] where results of measurements of the second harmonic of the susceptibility were taken as the basis for assuming that antiferromagnetic correlations with a weak ferromagnetic component develop in the paramagnetic phase. In addition, it is also assumed in [26] that regions of charge ordering of the manganese ions form in the sample, where antiferromagnetic ordering occurs. The size of these regions is hundreds of angstrom and their bulk fraction should be extremely small (this was not estimated). This treatment correlates with the results of measurements of the magnetization, and the electron diffraction and microscopy data obtained for a manganite having the same composition [18].

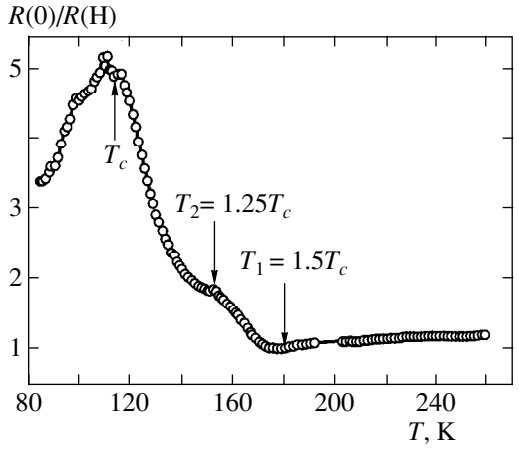

Fig. 1. Temperature dependence of the magnetoresistance $R(0, T)/R(H, T)$ of an $Sm_{0.6}Sr_{0.4}MnO_3$ sample for $H = 24$ kOe. The features on the temperature dependence indicated by the arrows are as follows: (*1*) T_1 is associated with an *O*–*O*' structural transition; (*2*) T_2 is associated with the appearance the second harmonic of the magnetization; (*3*) T_c is associated with a metal–insulator transition and the appearance of spontaneous magnetization.

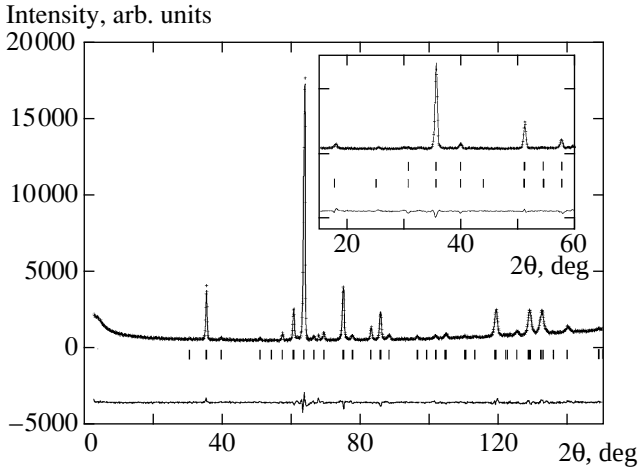

Fig. 2. Experimental and calculated (almost the same) diffraction patterns of $^{154}Sm_{0.6}Sr_{0.4}MnO_3$ sample at 300 K. The difference curve is shown by the lower line. The markers below the diffraction pattern indicate the position of the peaks. The inset shows the same pattern at 1.5 K, the upper row of markers corresponds to the crystal-structure peaks and the lower row corresponds to the magnetic structure peaks.

4. ANALYSIS OF NEUTRON DIFFRACTION DATA

Diffraction spectra of Sm manganite obtained at room temperature and at $T = 1.5$ K are shown in Fig. 2. An analysis of the width of the diffraction lines revealed that the lines are broadened appreciably compared with the resolution function of the diffractometer. This discrepancy is indicative of effects associated with the microstructure such as the influence of the size of coherent regions and microdeformations. By examining suitable corrections it was established that the line widths actually observed can be described in terms of an orthorhombic microdeformation in which the lattice constants are assumed to have a Gaussian distribution over the sample while conserving the orthorhombic symmetry of the unit cell. Parameters to be refined are the dispersions of the lattice constants of the model structure. A different degree of "smearing" of the lattice constants leads to different line broadening with different indices *h*, *k*, and *l*. This broadening of the diffraction peaks, described as anisotropic, was observed in manganites in an earlier study [14] and was attributed to the multiphase composition of the samples and to the existence of a continuous distribution of lattice constants caused by microdeformations. At the final stages of a Rietveld analysis, two parameters were used to

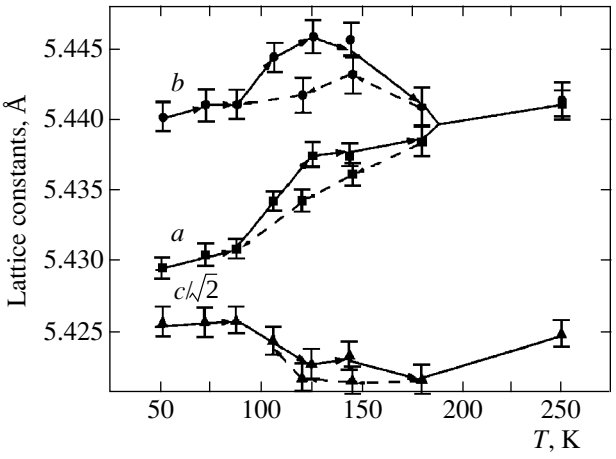

Fig. 3. Temperature dependences of the lattice constants of $^{154}Sm_{0.6}Sr_{0.4}MnO_3$ (*Pbmn*). The arrows indicate the direction of change in temperature.

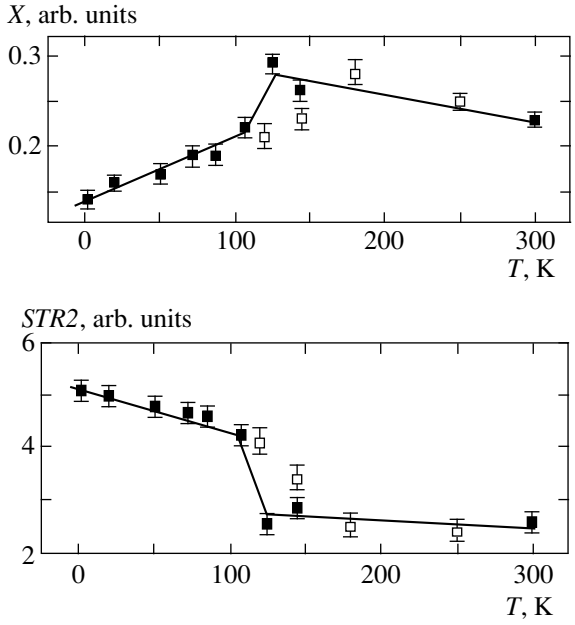

Fig. 4. Temperature dependences of the microdeformations parameters proportional to the dispersions of the lattice parameters: *X*—isotropic component; *STR2*—anisotropic component along the *b*-axis of the unit cell; ■—measurements in heating regime (top to bottom: 300, 250, 180, 144, 120, 106, 87, 72, 50.5, 19.5, and 1.5 K); □—measurements in cooling regime.

described the line broadening, an isotropic broadening component and an anisotropic component proportional to the dispersion of the distribution of the lattice constant *b* over the sample. The latitude parameters were fixed at the values of the diffractometer resolution function. The other parameters of the orthorhombic microdeformation model had zero averages over temperature, allowance for these did not improve the quality of the fit, and in order to reduce the number of parameters to be refined, these were eliminated from the refinements at the final stage of the treatment. Temperature dependences of the refined lattice parameters are plotted in Fig. 3.

The microdeformations which when taken into account provided a satisfactory description of the observed diffraction line broadening, are associated with the size and shape distribution of the unit cells. The refined microdeformation parameters are proportional to the second moments of these distributions. Thus, all the structural parameters given below have the meaning of averages. The existence of microdeformations is associated with the inhomogeneity of the sample and the characteristics of the degree of inhomogeneity may be taken to be the dispersions of the lattice constants which, as we can see from Fig. 4, vary with temperature. The temperature dependences of the microdeformation parameters exhibit a jump when magnetic ordering of the manganese atoms appears. This jump is also retained when only the anisotropic component is used to refine the microdeformation parameters [20].

5. CRYSTAL STRUCTURE OF $Sm_{0.6}Sr_{0.4}MnO_3$

The structural parameters were extracted by the Rietveld method using the FULLPROF program [27]. The parameters were determined in the context of an orthorhombic structure (space group 62, *Pbnm*). Temperature dependences of the refined lattice constants given in Fig. 3 and Table 1 indicate that at room temperature the structure is metrically close to tetragonal. The phase *O* ($a \geq b > c/\sqrt{2}$) denotes the orthorhombic (*Pbnm*) structure obtained from the cubic aristotype by rotating the regular (undistorted) MnO_6 octahedrons [28]. This structure has three crystallographically independent Mn–O spacings and in the *O*-phase these are the same. The *O*' phase denotes the orthorhombic structure where the axes are related as $b > a > c/\sqrt{2}$. This structure may be obtained from *O* if the spacings in the manganese–oxygen fragment cease to be the same. On the basis of the temperature dependences of the lattice constants, the transition temperature is $T_{O-O'} \simeq 180$ K. The transition to a magnetically ordered metallic state corresponds to a jump on the temperature dependences of the lattice constants. The lattice constant *b* in particular, has a maximum at $T = 120$ K. In addition, as for the microdeformation parameters, when $T < T_{O-O'}$, a slightly different temperature dependence of the lattice constants *a* and *b* is observed during heating and cooling of the sample (Figs. 3 and 4). The difference reaches a maximum and exceeds two standard deviations at $T \simeq 120$ K. For the lattice constant *c* the difference in the heating and cooling regimes does not exceed two standard deviations.

Table 1. Refined values of lattice constants and magnetic moments for manganese atoms

T, K	*a*, Å	*b*, Å	$c/\sqrt{2}$, Å	m_z, μ_B (*F*-type)	m_y, μ_B (*A*-type)
1.5	5.4297(7)	5.439(1)	5.4260(9)	2.50(6)	0.53(4)
19.5	5.4294(8)	5.439(1)	5.426(1)	2.55(6)	0.55(4)
50.5	5.4297(8)	5.440(1)	5.426(1)	2.40(6)	0.50(4)
72.0	5.4306(8)	5.441(1)	5.426(1)	2.14(6)	0.48(4)
87.1	5.4310(7)	5.441(1)	5.426(1)	1.99(6)	0.44(5)
106.0	5.4344(8)	5.445(1)	5.425(1)	1.28(8)	0.35(5)
124.9	5.4376(9)	5.446(1)	5.423(1)	–	–
300.0	5.4415(6)	5.4419(9)	5.4256(6)	–	–

Table 2. Bond lengths and angles in a manganese–oxygen fragment

T, K	Mn–O1, Å	Mn–O2(1), Å	Mn–O2(2), Å	Mn–O1–Mn	Mn–O2–Mn
1.5	2.06(2)	1.935(2)	1.86(2)	158.1(2)°	164.9(5)°
19.5	2.06(2)	1.935(2)	1.86(2)	157.7(2)°	164.8(5)°
50.5	2.07(2)	1.934(2)	1.84(2)	157.8(2)°	165.2(5)°
72.0	2.06(2)	1.934(2)	1.86(2)	157.6(2)°	165.4(6)°
87.1	2.08(2)	1.931(2)	1.84(2)	157.9(2)°	166.8(6)°
106.0	2.06(2)	1.937(2)	1.86(2)	158.2(2)°	163.8(5)°
124.9	2.12(2)	1.936(2)	1.78(2)	160.3(1)°	163.9(5)°
300.0	1.96(2)	1.952(3)	1.94(2)	160.0(2)°	158.7(4)°

The *O*–*O*' transition corresponds to a loss of equality between the lengths of the three manganese–oxygen bonds. The temperature dependence of these bond lengths is given in Table 2 and is typical of the cooperative Jahn–Teller effect. The splitting of the bond lengths in the MnO_6 octahedron corresponds to the Mn–O2 square lying approximately in the plane *ab* being distorted to form a rhomb at $T < 180$K where the maximum distortion occurs at the magnetic ordering temperature. It should be noted that the distortion of the manganese–oxygen fragment is unusually large for a relatively dilute system of Jahn–Teller ions (60% Mn^{3+}) and is comparable with that for undoped $LaMnO_3$ (100% Mn^{3+}) [29]. This well-defined distortion of the MnO_6 octahedron is also conserved in the region of metallic conductivity despite the fact that these distortions imply localization of the Mn^{3+} ions.

The overall structural data, specifically the appreciable line broadening, the temperature hysteresis of the lattice constants, and the high level of Jahn–Teller distortions which are conserved in the temperature region of magnetic ordering and metallic conductivity, indicate that the sample is structurally inhomogeneous. By this we understand that the sample has a multiphase composition which, because of the similarity between the crystal structures of the different phases, only appears within the limits of the diffractometer resolution for the effects listed.

6. MAGNETIC STRUCTURE OF $Sm_{0.6}Sr_{0.4}MnO_3$

A preliminary analysis of the contribution of magnetic scattering to the neutron diffraction spectra at $T < T_C$, where T_C is the Curie temperature, was made in [20]. Here we shall give refined data which can be used to provide a more accurate model of the magnetic ordering in $Sm_{0.6}Sr_{0.4}MnO_3$. The diffraction spectra at low temperatures contain ferro- and antiferromagnetic contributions. Magnetic scattering almost disappears at $T = 130$ K. The magnetic contribution was analyzed using the FULLPROF program. We only considered a single-phase homogeneous model for the Mn sublattice based on an orthorhombic (*Pbnm*) crystal lattice.

We found that the magnetic structure can be described as a mixture of ferromagnetic and antiferromagnetic components. The magnetic moment of the Mn ions has components m_y (or m_x) and m_z, respectively, along the *b*- (or *a*-) and *c*-axes in the orthorhombic (*Pbnm*) structure. The $m_z$components are ordered ferromagnetically in the *c* direction; the m_y (or m_x) components are ordered ferromagnetically in the *ab* plane and antiferromagnetically in the *c* direction, forming a so-called *A*-type antiferromagnetic structure. The total magnetic moment of the Mn ions at saturation at $T = 1.5$ K was defined as $m = 2.52(5)\mu_B$. The temperature dependences of the components of the magnetic moment are given in Table 1. The experimental data

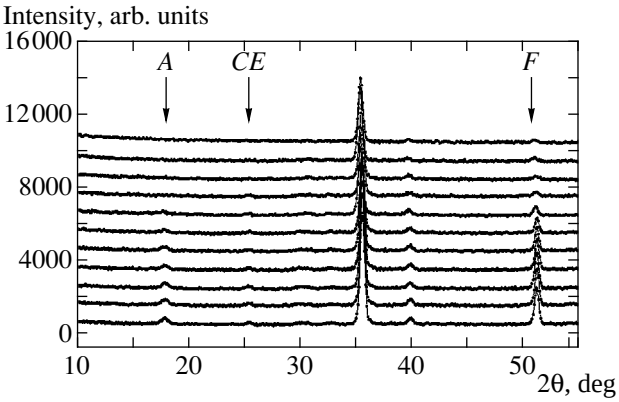

Fig. 5. Fragments of experimental diffraction patterns of $^{154}Sm_{0.6}Sr_{0.4}MnO_3$ at various temperatures (top to bottom: 300, 250, 180, 144, 120, 106, 87, 72, 50.5, 19.5, and 1.5 K). The arrows indicate the changes in intensity caused by the *A*, *F*, and *CE* types of magnetic ordering.

cannot be used to make a definitive choice between the *a* and *b* directions of the antiferromagnetic moment. Thus, we say that there is an equal probability of an A_xF_z or A_yF_z magnetic structure having the wave vector $\mathbf{k} = (0, 0, 0)$ being formed. Moreover, a choice between a noncollinear ferromagnetic or a multiphase system consisting of ferromagnetic and antiferromagnetic regions cannot be made merely from powder diffraction data. Using data from [30], the A_xF_z magnetic structure can be described as a superposition of nonrecurrent irreducible $\tau_1'' + \tau_3$ representations of the *Pbnm* group while the A_yF_z magnetic structure can be described as a superposition of recurrent irreducible $\tau_3' + \tau_3$ representations. In the first case, we have different irreducible representations and therefore two exchange multiplets which most likely corresponds to a two-phase (ferro- and antiferromagnetic) state of the compound. In the second case, however, we have a single irreducible representation τ_3 and thus a single exchange multiplet and there is good reason to consider the system to be a canted ferromagnet.

A subsequent more detailed analysis of the neutron diffraction data showed that the description of the crystal and magnetic structures of this $Sm_{0.6}Sr_{0.4}MnO_3$ compound given above may be slightly improved if we assume that there is another antiferromagnetic phase having the wave vector $\mathbf{k} = (1/2,1/2, 0)$. This corresponds to doubling the magnetic unit cell, compared with the crystal one, along the *a*- and *b*-axes. The orientation of the magnetic moments at the Mn ions in this phase corresponds to a *CE*-type antiferromagnet (the *a*- and *b*-axes are again not isolated as in the *A*-type antiferromagnetic structure considered above). Figure 5 shows parts of the experimental spectra measured for an $Sm_{0.6}Sr_{0.4}MnO_3$ sample in the temperature range $1.5 \le T \le 300$ K. The positions of the (001) peak from the *A*-type antiferromagnetic lattice and the (1/2 1/2 1) peak from the *CE*-type antiferromagnetic lattice are indicated. It can be seen that whereas the *A*-type magnetic ordering disappears above 120 K, the *CE*-type magnetic ordering is still observed at 150 K but not at 180 K. We also note that no *CE* structure is observed for an $Sm_{0.75}Sr_{0.25}MnO_3$ sample. The magnetic moment of the Mn ions forming the *CE* structure calculated from the experimental data reaches $m_{CE} = 0.31(3)\mu_B$ at low temperatures ($\simeq$1.5 K).

In many studies the formation of a *CE* antiferromagnetic structure is attributed to charge ordering of the Mn^{3+} and Mn^{4+} ions. Our neutron diffraction data broadly agree with the results of [18] where local regions of charge ordering were observed from electron diffraction and electron microscopy data for $Sm_{1-x}Sr_xMnO_3$ manganites in the concentration range $0.4 \le x \le 0.6$ and the magnetic, $T_C \simeq 125$ K, and charge ordering temperatures $T_{CO} \simeq 140$ K were determined from the temperature dependence of the magnetization for $Sm_{0.6}Sr_{0.4}MnO_3$.

7. RESULTS AND ANALYSIS OF SMALL-ANGLE POLARIZED NEUTRON SCATTERING MEASUREMENTS

Typical temperature dependences of the magnetic small-angle scattering intensity $I_m(T)$ and the polarization $P(T)$ are plotted in Figs. 6 and 7. The values of $I_m(T)$ were calculated as

$$I_m(T) = I(T) - I(300\ \text{K}), \tag{1}$$

where $I(T)$ is the experimentally measured intensity, and the polarization $P(T)$ is normalized to the polarization of the incident neutron beam. Figure 7b gives the temperature dependence of the integrated magnetic scattering cross section $\Sigma(T)$ in the range of small q (in practice in the central counter of a detector with $q < 0.003$ Å^{-1}) which is determined from the depolarization data. According to [31] (see also [32] and the literature cited there) the depolarization of neutrons after passing through the sample may be considered to be the result of integral scattering at magnetic inhomogeneities within the angular width of the transmitted beam:

$$P = P_0 \exp(-g\Sigma L), \tag{2}$$

where $g < 2$ is a coefficient which depends on the orientation of $\mathbf{P}_0$ relative to $\mathbf{k}$, and L is the sample thickness. Thus, by measuring the depolarization it is possible to study the total scattering cross section in the range $q < q_{\min}$ ($q_{\min}$ is the resolution of the counter) and to obtain integrated information on large-scale magnetic inhomogeneities having the characteristic dimension $R > 1/q_{\min}$.

On analyzing the dependences $P(T)$ and $I_m(T)$ we can confirm the following.

(a) Scattering increases with decreasing T in ranges of small $q < q_{\min}$ and $q > 0.003$ Å^{-1} i.e., $\Sigma(T)$ and $I_m(T)$,

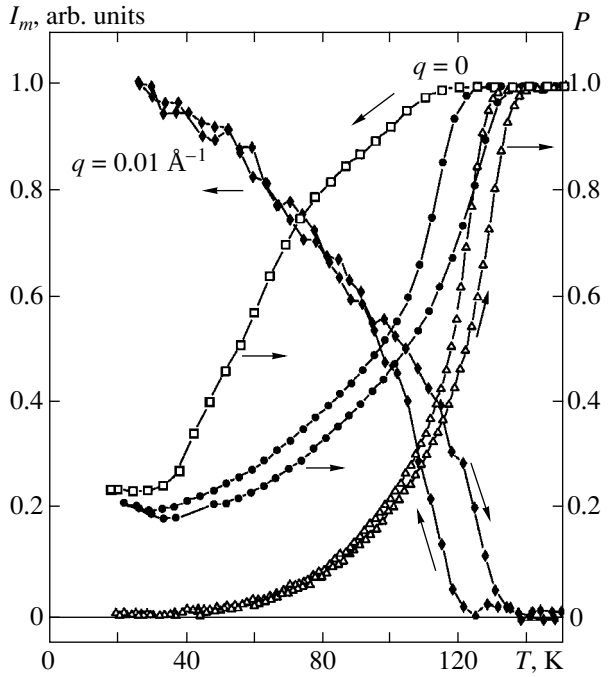

Fig. 6. Temperature dependences of the polarization P in the central counter ($q < 0.0003$ Å^{-1}) and the neutron magnetic scattering intensity I_m for an $Sm_{1-x}Sr_xMnO_3$ sample with $x = 0.4$ for $H = 0$ (●, ◆) and $H = 4.2$ kOe (△). The neutron depolarization data for the $x = 0.25$ sample (□) for $H = 0$ are taken from [22].

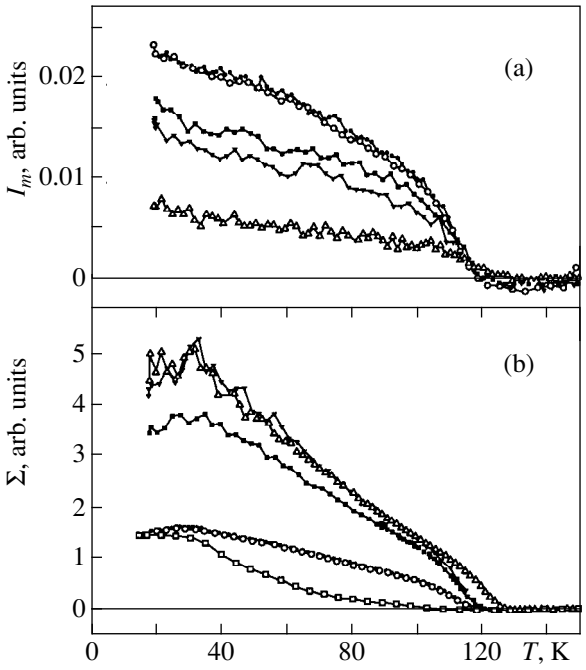

Fig. 7. Temperature dependences of (a) the magnetic scattering intensity I_m with $q = 0.01$ Å^{-1} and (b) the integrated magnetic scattering in the range $q < 0.003$ Å^{-1} for $Sm_{1-x}Sr_xMnO_3$ samples with $x = 0.4$ and $H = 0$ (●), 130 (○), 800 (■), 1240 (▼), 4200 Oe (△). The data for the $x = 0.25$ sample (□) for $H = 0$ were taken from [22].

respectively. The transition takes place over a larger range of T and the temperature dependences of the scattering have typical sections which are very different for $x = 0.4$ and 0.25 (data for $x = 0.25$ were taken from [22]).

(b) Temperature hysteresis is observed on the dependences $I_m(T)$ and $P(T)$.

(c) The depolarization $\Delta P(T) = 1 - P(T)$ in weak fields increases with decreasing T (not exceeding 80%) and decreases at $T < 40$ K.

(d) The scattering exhibits a strong dependence on the field H where the scattering cross section increases with H in the range of small $q < q_{\min}$ [depolarization or $\Sigma(T)$] and decreases in the range $q > 0.003$ Å^{-1} [i.e., $I_m(T)$]. An exception is the range $T = 130$–110 K where $I_m(T)$ depends weakly on H while $\Sigma(T)$, conversely, increases relatively rapidly in the field $H = 4200$ Oe. In addition, in this field we observe that the transition temperature is shifted to 130 K.

The SAPNS data show that in the low-temperature phase the Sm system is magnetically highly inhomogeneous. The authors will not attempt to interpret all the nuances of the observed temperature and field dependences of small-angle scattering. Our task is to attempt to give a qualitative explanation of the observed effects taking into account research by other methods. The first question is the extent to which the observed magnetic scattering is associated with the aggregate inhomogeneity of the system, since the samples being studied are powder samples in which magnetic scattering can take place at grain/pore boundaries. Using a polarized neutron method, we can give an estimate of the contribution of this scattering to the total magnetic scattering. The intensity of the scattering of polarized neutrons at a magnetized sample $I(T, q)$ may be expressed in the form

$$I(T, q) \propto F_n^2(T, q) + 2F_n(T, q)F_m(T, q) + F_m^2(T, q),$$
$$I_n(T, q) = F_n^2(T, q), \quad I_m(T, q) = F_m^2(T, q), \quad (3)$$
$$I_{mn}(T, q) = F_n(T, q)F_m(T, q),$$

where I_n and I_m are the nuclear and magnetic scattering at magnetic density fluctuations having amplitudes F_n and F_m, respectively, and I_{mn} is the magneto-nuclear interference term. For simplicity we shall assume that F_m includes a geometric factor for the relative orientation of the vectors **q** and **m**, where **m** is the magnetic moment of the sample. All three terms in (3) can be measured independently and as a result we can find the ratio of the magnetic scattering intensities obtained from measurements of the interference term and caused by scattering at grains/pores (see below) to the total magnetic scattering intensity (1).

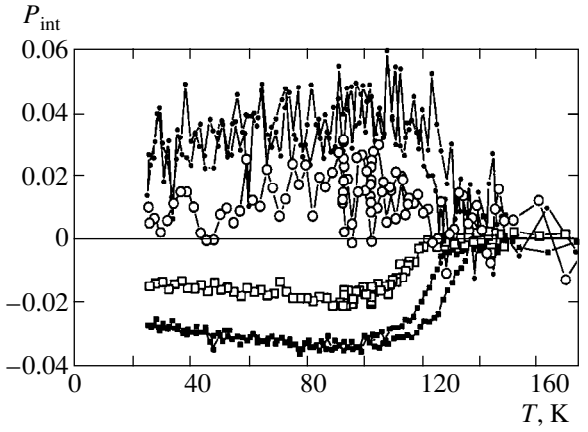

Fig. 8. Temperature dependence of the normalized magneto-nuclear interference effect $P_{int} = (I^+ - I^-)/I_m$ in fields H = 800 Oe (□, ○) and 4200 Oe (■, ●) for an $Sm_{0.6}Sr_{0.4}MnO_3$ sample with $q = 0$(■, □) and $q = 0.01$ Å^{-1} (●, ○).

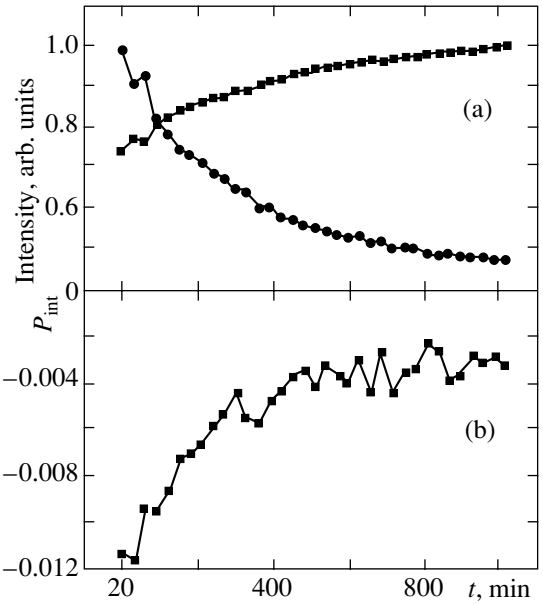

Fig. 9. Time dependences of: (a) the neutron intensity in the central counter (■) and the scattering for $q = 0.01$ Å^{-1} (●) and (b) the normalized magneto-nuclear interference effect $P_{int} = (I^+ - I^-)/I_m$ during condensation of N_2 in sample pores: $Sm_{0.6}Sr_{0.4}MnO_3$ sample, T = 70 K, H = 130 Oe.

The interference term I_{mn} can be measured as the difference Δ between the scattering intensities of neutrons polarized parallel (I^-) and antiparallel (I^+) to the applied magnetic field H:

$$\Delta = I^+ - I^- = 4F_n(T, q)F_m(T, q). \quad (4)$$

It was shown in [22] that for Sm samples with $x = 0.25$ and 0.4 an interference effect is observed whose magnitude normalized to the total magnetic scattering $P_{int} = \Delta/I_m$ does not exceed 2% in fields up to 1 kOe. In this study and references cited in it the authors determined the conditions for observation of interference scattering at magnetic-nuclear cross correlations which, in accordance with the optical theorem, changes sign for $q \longrightarrow 0$.

Figure 8 gives the temperature dependence $P_{int}(T)$ for a sample with $x = 0.4$ in fields up to 4.2 kOe. The interference effect may be caused by scattering at cross-correlated intragranular magnetic and structural fluctuations on scales of hundreds of angstrom and at grain/pore boundaries (when the grains are magnetized) or at both. The fundamental possibility of studying the cross-correlation magnetic and lattice subsystems will be discussed below. The existence of the second scattering channel is confirmed by the dependence of the interference effect with varying nuclear contrast using a method of gas condensation [33] (in this case nitrogen) in the sample pores. Figure 9 gives time dependences of the change in the interference effect (b), the intensity in the central counter ($q \simeq 0$) and the scattering for $q = 0.01$ Å^{-1} (a) during the nitrogen condensation process.

From independent measurements of I_n and I_m we found that, for example at T = 100 K and H = 130 Oe the ratio is $\alpha = F_n^2/F_m^2 \lesssim 3$. [In these order-of-magnitude estimates we neglect the difference in the momentum dependences $I_m(q)$ and $I_n(q)$. The dependence $I_m(q)$ will be considered subsequently and the nuclear scattering which is mainly determined by scattering at grain/pore boundaries is satisfactorily described by the asymptotic form $I_n(q) \propto q^{-4}$ for $q \geq 4 \times 10^{-3}$ Å^{-1}.] Bearing in mind that $P_{int}(T)$ does not exceed 6–7% in all measurement regimes and that the magnetic scattering increases no more than fourfold relative to that at 100 K as the temperature decreases, the value of β, which is the ratio of the values of F_m^2 obtained from measurements of the interference term and from measurements of the magnetic scattering may have the upper estimate (i.e., maximum P_{int} and minimum α)

$$\beta = \frac{P_{int}^2}{16\alpha(T, H)} \lesssim 10^{-3}.$$

This ratio is an upper estimate of the fraction of the magnetic scattering at grain/pore boundaries in the total magnetic scattering since we assumed that all the interference scattering takes place merely in one channel.

Hence, to within 10^{-3} the magnetic scattering whose temperature dependences are plotted in Fig. 7 involves scattering at magnetic fluctuations in grains. An analysis of the dependences $I_m(q)$ showed that these are accurately described by the expression:

$$I_m(q) = \frac{A}{(q^2 + \kappa^2)^2}, \quad (5)$$

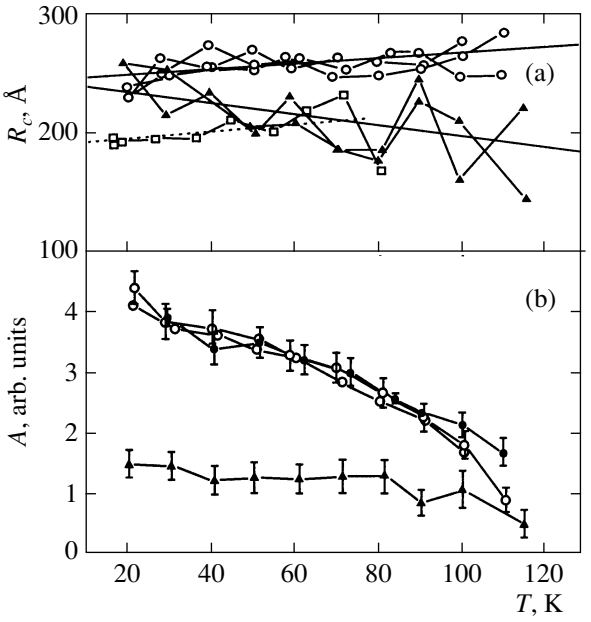

Fig. 10. Temperature dependences of (a) the correlation radius R_c and (b) the parameter A for $Sm_{1-x}Sr_xMnO_3$ samples under various measurement conditions: a—x = 0.4, cooling in fields H = 0 (○) and 4.2 kOe (▲); x = 0.25, cooling at H = 0 (□) [22]; b—x = 0.4, cooling at H = 0 (○) and 4.2 kOe (▲) and heating at H = 0 (●). The lines in the upper diagram were drawn by eye.

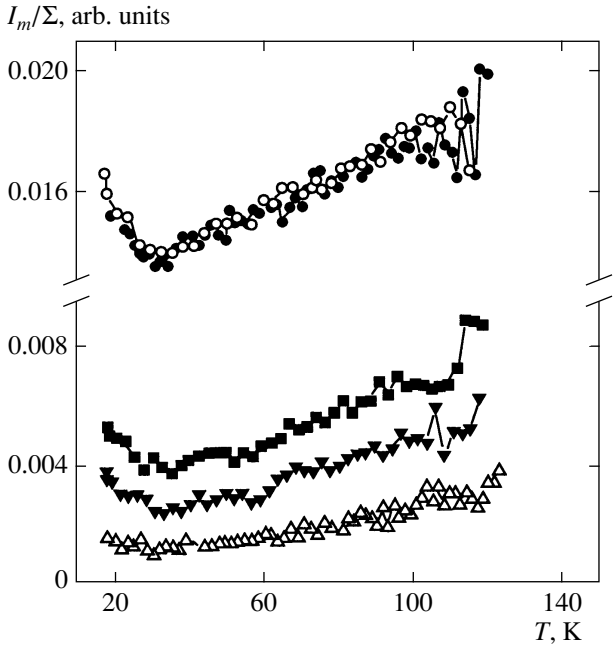

Fig. 11. Temperature dependence of the ratio of the intensity of neutron magnetic scattering I_m with q = 0.01 Å^{-1} to the magnetic scattering Σ in the range q < 0.003 Å^{-1} for an $Sm_{0.6}Sr_{0.4}MnO_3$ sample under cooling in various magnetic fields: H = 0 (●), 130 (○), 800 (■), 1200 (▼), and 4200 Oe (△).

where A and $\kappa = 1/R_c$ are free parameters and R_c has the meaning of the characteristic correlation radius. In the coordinate representation expression (5) corresponds to scattering at the spin correlation function $\langle S_i, S_j\rangle$ which decreases exponentially with distance r:

$$(S_i, S_j) \propto \exp(-r/R_c). \qquad (6)$$

Values of the parameters A and R_c obtained by convoluting the Lorentzian (5) with the resolution function of the system are plotted in Fig. 10. It can be seen that the characteristic correlation radius R_c depends weakly on temperature unlike the parameter A which, in accordance with the variation of $I_m(T)$, follows the temperature variations of the magnetic moment and the magnetic inhomogeneity density having the characteristic dimension R_c.

Figure 11 gives the temperature dependence of the ratio $I_m(q)/\Sigma$ which primarily characterizes the change in the topology of the magnetic inhomogeneities since in the magnetic cross section ratio the dependence on the induction should be reduced in a first approximation. This ratio is essentially the ratio of the magnetic scattering in the detector counter at the angle θ, which corresponds to q = 0.01 Å^{-1}, to the magnetic scattering in the central counter of the detector (θ = 0). It can be seen that I_m/Σ shows a stable tendency to decrease with decreasing temperature and increasing magnetic field which may be interpreted as an increase in the density of large-scale scattering inhomogeneities with decreasing temperature and increasing magnetic field. An exception is the temperature range T < 40 K. An increase in $P(T)$ with decreasing T is typical of a reentrant spin-glass transition accompanied by the formation of a cluster spin glass phase [34].

The change in the topology of the magnetic inhomogeneities is also confirmed by the strong depolarization with decreasing temperature and increasing field up to 4200 Oe. The increase in the depolarization in fields up to 4 kOe implies that the magnetic moments in the low-temperature ferromagnetic phase are not completely oriented parallel to the field in this range, i.e., the system exhibits fairly strong magnetic anisotropy. For instance, the maximum value of ΔP is approximately 80% for H = 0 and reaches 100% for H = 4200 Oe. We can show that even for the maximum magnetic moment of the Mn atom scattering at ferromagnetic fluctuations on the scale R_c cannot give depolarization greater than 20%. For estimates we can use the formula [32]

$$\frac{P}{P_0} = \exp\left[-\frac{1}{2}\left(\frac{\gamma B(T)}{v}\right)^2 R_c L\right], \qquad (7)$$

where γ is the neutron gyromagnetic ratio, $B(T) = 4\pi M(T)$ is the induction of the ferromagnetic region of dimension R_c, v is the neutron velocity, and L is the effective thickness of the sample. [Formula (7) agrees with (2) provided that the change in the phase of the

neutron wave at fluctuations R_c is small, i.e., $\gamma BR_c/v \ll 2\pi$, and apart from a numeric factor agrees with the classical depolarization formula obtained by Halpern and Holstein [35]. For example, if the magnetic moment of the Mn atom is $4\mu_B$ and the effective sample thickness is L = 1.5 mm the depolarization is $\lesssim$15%. However, if we assume homogeneous ferromagnetic ordering on the grain scale the depolarization for $H \approx 0$, the minimum magnetic moment of the Mn atom around $1\mu_B$, and the given sample thickness, should be 100%. The experimental data on the temperature dependences of the scattering $I_m(T)$ and the polarization $P(T)$ (see Figs. 6 and 7) may agree if we assume that in addition to fluctuations having the characteristic scale R_c, the system also has larger-scale ferromagnetic fluctuations from which scattering takes place almost in the central counter so that it is only recorded from the change in polarization. For instance, the presence of approximately 20% ferromagnetic fluctuations having dimensions of 5000–6000 Å can give a depolarization of around 80%. Data on the depolarization only give a qualitative estimate of the sizes of the magnetic regions since its value depends (apart from the size) on the magnitude and distribution of the induction in these regions and also on the concentration of these regions.

8. DISCUSSION OF RESULTS

An analysis of the neutron diffraction data and magnetic small-angle polarized neutron scattering data for this manganite unambiguously indicate that this contains magnetic and nuclear inhomogeneities which evolve with temperature. This pattern agrees with the results of electron diffraction and microscopy which have revealed the coexistence of structural *P*- and *I*-types for a compound having this composition. A reduction in temperature leads to the appearance of incommensurate superstructures in the *P*-phase which is treated in [18] as the appearance of a local charge-ordered state at T = 140 K. This conclusion is consistent with the (1/2, 1/2, 1) peak associated with the magnetic *CE* structure observed by us in the 1.5–150 K range. The magnetic nature of this peak is confirmed by the appearance of hysteresis of the second harmonic of the magnetization at $T \simeq$ 160 K and below [26], and its low intensity indicates that the charge-ordered regions have a small bulk fraction. At the same time, no *I*-type structure could be clearly identified from the neutron diffraction data. Possible reasons for this are the following.

(1) The similarity between the structural parameters of the *P*- and *I*-structures and the small fraction of the *I*-phase which make it impossible to resolve the structures at the level of instrumental resolution used. This point is supported by the line broadening and the exaggerated Debye–Waller factors. In addition, neutron diffraction gives a pattern averaged over the bulk of the sample whereas electron methods study regions near the surface.

(2) The absence of a magnetic field in the neutron diffraction experiments. For example, a magnetic field of the order of 1–2 T was applied in the electron microscope experiment [18]. In particular, the difference between the neutron diffraction and electron microscopy data in $Pr_{0.5}Sr_{0.5}MnO_3$ manganite was ascribed to this factor [36]. In our case, the sensitivity of the structure to a magnetic field follows from the magnetoresistance data: in a 24 kOe field the characteristic kink on the temperature dependence of the resistance disappears at $T \simeq$ 180 K which corresponds to a *OO′* structural transition. In addition, the results of SAPNS measurements reveal a strong dependence of the mesoscopic inhomogeneity topology on the magnetic field.

An analysis of the magnetic diffraction results revealed the coexistence of several types of magnetic ordering (*A, F, CE*). The reason for the appearance of different magnetic phases in manganites is usually assumed to be competition between ferromagnetic double exchange and antiferromagnetic superexchange (see, for example, [37]). Using the nominal magnetic moment for manganese at a given concentration, Mn^{3+}/Mn^{4+}, and the experimental magnetic moments, we can approximately estimate the fraction of manganese atoms in the different phases. At T = 1.5 K 60% of the manganese atoms form the ferromagnetic phase, 15% are antiferromagnetically ordered, and the remaining fraction of the manganese atoms (25%) makes no contribution to the magnetic diffraction. The reason for this may be the small sizes of the magnetically ordered regions containing 25% of the manganese atoms and forming at low temperatures an additional magnetic phase of the cluster spin glass type whose presence in these perovskites has been discussed in many studies [7, 38, 39]. Indications of a spin glass state may include a difference in the measurements of the magnetonuclear interference term as a function of the cooling regime [in zero field (ZFC) followed by application of the field H or in the field H (FC) [22]], a difference in the magnetization in the ZFC and FC regimes at $T <$ 40–50 K [24], and characteristic depolarization behavior at $T <$ 40 K in weak fields [34]. The formation of magnetic inhomogeneities having characteristic dimensions of around 200 Å which depends weakly on temperature and field is attributed to the evolution of ferromagnetic correlations with decreasing temperature against a background of competing antiferromagnetic regions. In order to complete the scenario we also need to include the existence of large-scale ferromagnetic correlations (thousands of angstrom) which ensure the observed depolarization. A contribution to the small-angle magnetic scattering may also be made by antiferromagnetic regions having scales of hundreds of angstrom formed inside large-scale ferromagnetic regions, i.e., “magnetic holes” in ferromagnetic regions. The transition to the ferromagnetic phase is evidently accomplished by a percolation manner where ferromagnetic correlations having dimensions of around 200 Å merge into clusters of around 1000 Å. This scenario can at least reconcile

the small-angle scattering and neutron depolarization data. Similar scales of ferro- and antiferromagnetic regions were observed by electron microscopy in an (La, Pr, Ca)MnO_3 system in [40]. In this system, however, the authors observed oriented ordering of ferromagnetic regions in fields up to 4 kOe which is not the case in an Sm system, as we have already noted because in this range of magnetic fields the depolarization is far stronger than in zero field. We can postulate that in an Sm system the anisotropy is stronger. The SAPNS data indicate that the magnetic mesostructure of an Sm system with $x = 0.4$ and 0.25 differs (see Figs. 6, 7, and 10). These differences possibly characterize the absence of a transition to the metal state at deficient strontium concentration.

An important characteristic of the transition to the magnetically ordered state is the strong hysteresis of the depolarization and small-angle scattering and the appreciable hysteresis of the lattice constants. This hysteresis is natural for a percolation transition. In this context it should be noted that the concept of the Curie temperature T_C used to describe the phase diagrams of the Sm system and similar systems is very arbitrary.

Using polarized neutrons in the small-angle experiments makes it possible to measure the magneto-nuclear interference on a suitable scale. Here these measurements were mainly used to give an upper estimate of the ratio of the magnetic scattering intensity at grains/pores to the total magnetic scattering assuming that all the interference scattering takes place only in one channel. Such an estimate is required for powder samples. However, the method in principle provides the unique experimental possibility of studying intercorrelations between the magnetic and lattice subsystems on a scale of 10–1000 Å. There is some basis for their existence. First, experiments show (Fig. 9) that the time dependence of the interference term in nuclear contrasting clearly does not tend to zero and second the diffraction peaks are only satisfactorily described using a microdeformation model. However polycrystalline or single-crystal samples are required for these measurements. Nevertheless, we note once again that the SAPNS method gives more adequate information for powder experiments. This in turn provides a unique possibility for directly comparing the SAPNS and NPD results obtained for the same sample.

9. CONCLUSIONS

A combined analysis of experimental data on the macroscopic properties (magnetoresistance, magnetization, magnetic susceptibility), microscopic structural parameters (neutron powder diffraction), and mesoscopic characteristics (magnetic small-angle polarized neutron scattering) suggests the following pattern for the temperature evolution of $Sm_{0.6}Sr_{0.4}MnO_3$.

(a) At $T \simeq 180$ K a structural transition takes place from the *O* to the *O*' phase associated with the evolution of well-defined Jahn–Teller distortions of the manganese–oxygen octahedrons. The temperature dependence of the resistance, having a polaron character, exhibits a singularity corresponding to an increase in activation energy.

(b) At $T \simeq 160$ K the antiferromagnetic correlations appearing in part of the sample lead to charge and related magnetic *CE* ordering. These effects lead to the formation of an elongated section on the magnetic susceptibility and magnetization curves [18], the appearance of the (1/2, 1/2, 1) magnetic peak, and hysteresis of the second harmonic of the magnetization [26].

(c) From $T \simeq 110$–120 K a spontaneous magnetic moment appears in the system and increases. In this case, ferromagnetic (*F*) and antiferromagnetic (*A* and *CE*) phases coexist. A magnetic diffraction analysis shows that even at low temperatures, around 25% of the manganese atoms make no contribution to the magnetic diffraction peaks. However, an analysis of the small-angle scattering in this temperature range indicates that magnetic inhomogeneities having characteristic dimensions of around 200 Å and of the order of a few thousand angstrom coexist in the system.

(d) At $T < 40$ K some of the manganese atoms which make no contribution to the magnetic diffraction form a magnetic phase, evidently of the cluster spin glass type, which leads to an increase in the polarization of the neutrons transmitted by the sample and explains why the temperature dependences of the magnetic characteristics differ from the magnetic prehistory of the sample.

Quite clearly, the main result of this study is that we have obtained experimental evidence that in a system with competing magnetic interactions the coexistence of magnetically different but structurally similar phases is energetically favorable. The mechanism for the phase separation remains unclear although the scales of the mesoscopic inhomogeneities do not suggest electron phase separation. The existence of magnetic inhomogeneities in the system and the dependence of their topology on the magnetic field indicates that the negative magnetoresistance effect may be attributed to characteristic features of the magnetic mesostructure.

ACKNOWLEDGMENTS

The authors are grateful to A. Maignan for carrying out the magnetic measurements, H. Glattli for cooperation and discussions, and S.M. Dunaevskii, G.P. Kopitse, S.V. Grigor'ev, and S.A. Klimko for discussions and assistance with the measurements.

This work was supported by the Russian Foundation for Basic Research (project nos. 98-02-17632, 96-15-96775, and 00-15-96814) and the State Scientific-Technical Program "Neutron Studies of Condensed Media."

REFERENCES

1. E. O. Wollan and W. C. Koehler, Phys. Rev. **100**, 545 (1955).
2. É. L. Nagaev, Usp. Fiz. Nauk **166**, 833 (1996) [Phys. Usp. **39**, 781 (1996)].
3. E. L. Nagaev, Aust. J. Phys. **52**, 305 (1999).
4. John B. Goodenough and J. S. Zhou, Nature **386**, 229 (1997).
5. J. M. de Teresa, M. R. Ibarra, J. Blasco, *et al.*, Phys. Rev. B **54**, 1187 (1996).
6. J. M. de Teresa, M. R. Ibarra, P. A. Algarabel, *et al.*, Nature **386**, 256 (1997).
7. J. M. de Teresa, C. Ritter, M. R. Ibarra, *et al.*, Phys. Rev. B **56**, 3317 (1997).
8. C. Ritter, M. R. Ibarra, J. M. de Teresa, *et al.*, Phys. Rev. B **56**, 8902 (1997).
9. M. Viret, H. Glattli, C. Fermon, *et al.*, Europhys. Lett. **42**, 301 (1998); Physica B (Amsterdam) **241-243**, 430 (1998).
10. S. Rosenkranz, R. Osborn, J. F. Mitchell, *et al.*, Physica B (Amsterdam) **241-243**, 448 (1997).
11. R. Caciuffo, J. Mira, M. A. Senaris-Rodríguez, *et al.*, Europhys. Lett. **45**, 399 (1999).
12. B. C. Tofield and W. R. Scott, J. Solid State Chem. **10**, 183 (1974).
13. Q. Huang, A. Santoro, J. W. Lynn, *et al.*, Phys. Rev. B **58**, 2684 (1998).
14. P. G. Radaelli, D. E. Cox, M. Marezio, *et al.*, Phys. Rev. Lett. **75**, 4488 (1995); P. G. Radaelli, D. E. Cox, M. Marezio, and S.-W. Cheong, Phys. Rev. B **55**, 3015 (1997).
15. M. Hervieu, G. van Tendeloo, V. Caignaert, *et al.*, Phys. Rev. B **53**, 14274 (1996).
16. L. M. Rodríguez-Martínez and J. Paul Attfield, Phys. Rev. B **54**, R15622 (1996).
17. Y. Moritomo, A. Machida, K. Matsuda, *et al.*, Phys. Rev. B **56**, 5088 (1997).
18. C. Martin, A. Maignan, M. Hervieu, and B. Raveau, Phys. Rev. B **60**, 12191 (1999).
19. S. M. Dunaevskiĭ, A. I. Kurbakov, V. A. Trunov, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **40**, 1271 (1998) [Phys. Solid State **40**, 1158 (1998)].
20. D. Yu. Chernyshov, V. A. Trounov, A. I. Kurbakov, *et al.*, Mater. Sci. Forum **321-324**, 812 (2000).
21. D. Yu. Chernyshov, A. I. Kurbakov, and V. A. Trounov, Physica B (Amsterdam) (in press).
22. V. V. Runov, H. Glattli, G. P. Kopitsa, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 323 (1999) [JETP Lett. **69**, 353 (1999)].
23. V. Runov, H. Glattli, G. Kopitsa, *et al.*, Physica B (Amsterdam) **276-278**, 795 (2000).
24. A. Maignan, private communication.
25. S. V. Grigor'ev, O. A. Gubin, G. P. Kopitsa, *et al.*, Preprint No. 2028, PNPI (Petersburg Nuclear Physics Institute, Russian Academy of Sciences, 1995).
26. I. D. Luzyanin, V. A. Ryzhov, D. Yu. Chernyshov, *et al.*, Preprint No. 2342, PNPI (Petersburg Nuclear Physics Institute, Russian Academy of Sciences, 1999).
27. J. Rodríguez-Carvajal, Physica B (Amsterdam) **192**, 55 (1993).
28. A. M. Glazer, Acta Crystallogr. B **B38**, 3384 (1972).
29. J. Rodríguez-Carvajal, M. Hennion, F. Moussa, *et al.*, Phys. Rev. B **57**, R1389 (1998).
30. A. N. Pirogov, A. E. Teplykh, V. I. Voronin, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **41**, 103 (1999) [Phys. Solid State **41**, 91 (1999)].
31. S. V. Maleev and V. A. Ruban, Zh. Éksp. Teor. Fiz. **62**, 415 (1972) [Sov. Phys. JETP **35**, 222 (1972)].
32. S. V. Grigor'ev, S. A. Klimko, S. V. Maleev, *et al.*, Zh. Éksp. Teor. Fiz. **112**, 2134 (1997) [JETP **85**, 1168 (1997)].
33. A. I. Okorokov, V. V. Runov, A. D. Tret'yakov, *et al.*, Zh. Éksp. Teor. Fiz. **100**, 257 (1991) [Sov. Phys. JETP **73**, 143 (1991)].
34. V. V. Runov, S. L. Ginzburg, B. P. Toperverg, *et al.*, Zh. Éksp. Teor. Fiz. **94**, 325 (1988) [Sov. Phys. JETP **67**, 181 (1988)].
35. O. Halpern and T. Holstein, Phys. Rev. **59**, 960 (1941).
36. F. Damay, C. Martin, M. Hervieu, *et al.*, J. Magn. Magn. Mater. **184**, 71 (1998).
37. Michel van Veenendaal and A. J. Fedro, Phys. Rev. B **59**, 1285 (1999).
38. A. Maignan, C. Martin, G. van Tendeloo, *et al.*, Phys. Rev. B **60**, 15214 (1999).
39. M. Muroi and R. Street, Aust. J. Phys. **52**, 205 (1999).
40. M. Uehara, S. Mori, C. H. Chen, and S.-W. Cheong, Nature **399**, 560 (1999).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 1029–1032.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1188–1192.

SOLIDS
Structure

Bifurcations of the Shape of a Magnetic Fluid Droplet in a Rotating Magnetic Field

K. I. Morozov* and A. V. Lebedev

Institute Mechanics of Continua, Ural Division, Russian Academy of Sciences, Perm, 614013 Russia
**e-mail: mrk@icmm.ru*

Received May 26, 2000

Abstract—An analysis is made of the behavior of a magnetic droplet suspended in a liquid in a high-frequency uniform, rotating magnetic field. In weak fields the droplet is spheroidal while in strong fields it is disk-shaped. The observed change in the shape of the droplet as the amplitude of the field increases depends on the magnetic permeability μ of the liquid and takes place according to three scenarios: (a) for small μ the spheroidal droplet is continuously converted into a disk; (b) for intermediate μ there is a range of fields in which the droplet becomes a triaxial ellipsoid with its major axis lying in the plane of the field, and spheroid–triaxial ellipsoid–disk transitions take place as a result of a soft bifurcation; (c) at high μ both transitions are hard. Theoretical calculations are made of the stability curve for the various droplet shapes. It is predicted that a change in the types of droplet shape bifurcations will occur in strong fields. A comparison is made with the experimental data.

Studies of equilibrium shapes of rotating volumes of liquid in order to describe the shape of the planets were started as early as the mid nineteenth century [1] when Plato also conducted the first experiments. Since it is clearly impossible to simulate gravitating volumes of liquid under laboratory conditions, in these experiments the role of the planet was played by an ordinary liquid droplet whose shape was determined as a result of the competition between surface and centrifugal forces. The main Plato methodology, that is to say the neutral buoyancy of a droplet and setting it in motion using a rotating drum, are still retained to some extent today. Alongside the traditional Plato technique, in modern experiments the droplet is suspended under conditions of weightlessness [2] or by the action of an ultrasonic wave [3]. Rotation of the droplet is also achieved by acoustic methods [3]. There is also a simpler method of rotating the droplet. If a magnetic fluid, comprising a colloidal suspension of a magnetic substance in an ordinary liquid [4], is used as the droplet material, the droplet will rotate when it is placed in a rotating magnetic field. This approach was implemented by us in [5] where we studied the behavior of a magnetic droplet in a low-frequency (~1 Hz) external field and observed that the droplet breaks up into two smaller volumes as its rotation speed increases. An increase in the frequency of the field does not generally lead to a significant increase in the rotation speed of the droplet. This observation is exactly the same as the so-called rotation effect, in which a layer of magnetic fluid is entrained by a rotating magnetic field [4]. In this effect the angular rotation speed of the liquid is several orders of magnitude slower than the rotation speed of the field. However, it is possible in principle to achieve rapid rotation of a droplet under the action of a rotating magnetic field exactly the same as the suspension of a droplet in air in a graded magnetic field. Before conducting such an experiment we must first determine how the magnitude of a rapidly rotating magnetic field and the magnetic properties of the magnetic fluid influence the equilibrium shape of the droplet.

In the present study we report an experimental investigation of the behavior of droplets of magnetic fluid in a uniform rotating magnetic field at 55 Hz. This field frequency is high in the sense that it is much higher than the rotation frequency of the droplet (≤1 Hz). For this frequency ratio the predicted droplet shape is an oblate ellipsoid of revolution whose maximum cross section coincides with the plane of rotation of the field. We know that in a static field the droplet is an ellipsoid of revolution elongated in the direction of the field [6]. Thus, when the field rotates rapidly and the droplet shape cannot keep up with its change, it is predicted that the prolate ellipsoid will become "smeared" in the plane of the field and will be converted into an oblate ellipsoid. We denote the semiaxes of the droplet ellipsoid by a, b, c where the first two lie in the plane of the field and $a \geq b$. In our experiment as the amplitude of the field increased, the droplet was successively converted from a sphere to a spheroid ($a = b > c$) and then to a disk, i.e., a highly oblate ellipsoid of revolution ($a = b \gg c$). However, this sequence of continuous flattening of the droplet, beginning with a spherical droplet in the absence of a field and ending with a disk-shaped droplet in strong fields, only occurs when the magnetic permeability of the magnetic fluid is low. For $\mu > 5$ we observed unexpected behavior of the droplet shape which underwent two successive bifurcations. First, above a certain field G_A the spheroidal droplet became

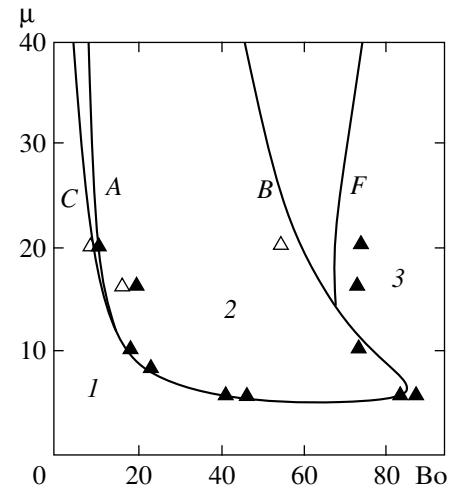

Fig. 1. Stability diagram of equilibrium droplet shapes. The numbers indicate the ranges of parameters for which the droplet is a spheroid (*1*), a triaxial ellipsoid (*2*), and a disk (*3*). The solid curve gives the results of calculations using formulas (2) and (3), the filled triangles give the experimental data with increasing field and the unfilled triangles give the data for decreasing field. The notation *A*, *B*, *C*, *F* is defined in the text.

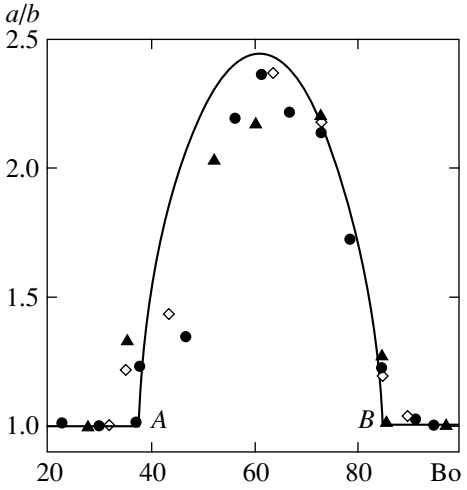

Fig. 2. Droplet semiaxis ratio a/b in the plane of the field as a function of the Bond number for $\mu = 5.94$. The symbols give the experimental data for three droplets having different initial radii, *A* and *B* are the bifurcation points. The solid curve gives the results of the calculations using formulas (2) and (3).

elongated in the plane of the field, being converted into a triaxial ellipsoid ($a > b > c$), then it began to be compressed, and finally it was converted into a disk in the field G_B.

In order to make a detailed study of the effect, we used six concentrated magnetic fluids having magnetic permeabilities between 5.8 and 20.4. For each magnetic fluid we prepared samples of between three and seven droplets, whose radius R in the absence of the field was 2–5 mm. We suspended each droplet in an aqueous solution of zinc chloride. We determined the coefficient of surface tension σ at the interface between the droplet and the solution by a standard method [6] using the elongation of the droplet in a static field. We then placed the droplets in a high-frequency rotating magnetic field and we determined the limits of the range of field amplitudes G_A and G_B for its existence in the form of a triaxial ellipsoid. Both shape-bifurcation fields for droplets of fixed magnetic permeability depended on their dimensions as $G_{A,B} \propto R^{-1/2}$. Thus, we calculated the corresponding values of the Bond magnetic numbers $\mathrm{Bo} = G^2R/\sigma$ which were averaged over a set of droplets having the same magnetic permeability μ. In this way we determined the range of μ–Bo parameters for which the droplet exists in the form of a spheroid, triaxial ellipsoid, or disk. The maximum amplitude of the magnetic field in the experiments was 50 Oe which corresponded to Bond numbers of ~150. The experimental data are plotted in Fig. 1. An important characteristic of the conversion of the droplet from a spheroid to a triaxial ellipsoid and back is that for $\mu < 11$ both transitions are soft. This property is reflected in Fig. 2 which gives the experimental results for the semiaxis ratio a/b as a function of the Bond number. It can be seen that as the field increases near the bifurcation points *A* and *B*, the ratio a/b varies monotonically. Bifurcations of the droplet shape at high μ take place completely differently. In our experiments when the field reached the value G_A a droplet with $\mu = 20.4$ was abruptly converted from a spheroid to a triaxial ellipsoid with the semiaxis ratio $a/b = 6$. The reverse transition in strong fields is also hard: a triaxial ellipsoid with the semiaxis ratio $a/b = 10$ is abruptly converted to a disk. The hard instability at high magnetic permeabilities is accompanied by the appearance of hysteresis of the droplet shape, i.e., the magnitudes of the transition fields and the geometric dimensions of the droplet become different for increasing and decreasing amplitude of the field.

We shall now make a theoretical analysis of the problem of droplet behavior in a rotating high-frequency magnetic field. In this problem there are several physical mechanisms determining the shape of the droplet, i.e., the magnetic field, the surface tension at the interface, and the flow created inside and outside the droplet. However, the role of this flow is insignificant. This is deduced from a simple estimate of the characteristic values of the viscous and hydrodynamic pressures and from our experimental data obtained as the field frequency increases. It was found that the frequency of the field merely (weakly) influences the rotation speed of the droplet but not its shape. Thus, we can predict that the droplet shape is determined as a result of competition between magnetic and surface stresses. We take the axes of a coordinate system rotating with the droplet as follows: the x axis is positioned along the

major axis of the ellipsoid and the z axis lies in the same direction as the vector of the angular rotation velocity of the field ω. Let us assume that G is the amplitude of the field and the rotation speed of the field is much higher than the rotation velocity Ω of the droplet, $\omega \gg \Omega$. The droplet energy is equal to the sum of the surface and magnetic energies:

$$E = \sigma S - \frac{VG^2}{4}(M_1 + M_2). \tag{1}$$

Here S and V are the surface area and volume of the droplet; $M_1 = (\mu - 1)/4\pi(1 + (\mu - 1)n_1)$, M_2 is obtained from M_1 by substituting $1 \longrightarrow 2$ in the index; n_1 and n_2 are the demagnetizing factors in the plane of the field, along the x and y axes, respectively. The magnetic contribution to the energy was formulated using the well-known result for the magnetization of a uniformly magnetized ellipsoid [7] assuming that μ is independent of the field, averaging over the period of the field, and neglecting the dispersion of the magnetic permeability of the magnetic fluid which is vanishingly small at these field frequencies (~100 Hz) [4]. By varying (1) according to the semiaxis ratios a/b and a/c at constant droplet volume, after simple but cumbersome calculations we obtain a system of equations to determine the droplet shape. After writing the ellipsoid semiaxes a, b, and c in units of the droplet radius R without the field, this system finally has the form

$$\begin{aligned}\tilde{n}_1(a^2 - b^2) + c^2(\tilde{n}_3 - \tilde{n}_2) \\ = \frac{2\pi}{3}\mathrm{Bo}(M_1^2(n_2'' + 2n_3'') - M_2^2(n_1'' + 2n_3'')),\end{aligned} \tag{2}$$

$$\begin{aligned}-\tilde{n}_1(a^2 + b^2) + \tilde{n}_2(2a^2 - c^2) + \tilde{n}_3(2b^2 - c^2) \\ = 2\pi\mathrm{Bo}(M_1^2 n_2'' + M_2^2 n_1''),\end{aligned} \tag{3}$$

where $n_1'' = (n_2 b^2 - n_3 c^2)/(b^2 - c^2)$, the values of n_2'' and n_3'' are obtained from n_1'' by means of a cyclic permutation of the indices 1, 2, and 3 and their corresponding ellipsoid semiaxes a, b, and c. The quantities with tildes $\tilde{n}_1$, $\tilde{n}_2$, and $\tilde{n}_3$ denote the demagnetizing factors of a so-called auxiliary ellipsoid whose semiaxes $\tilde{a}$, b, and $\tilde{c}$ are related to the semiaxes of the ellipsoid under study a, b, and c by the relationships: $\tilde{a} = a$, $b = ac/b$, $\tilde{c} = c$. Introducing an auxiliary ellipsoid can significantly shorten the notation on the left-hand sides of Eqs. (2) and (3) which contain cumbersome expressions for elliptic integrals. We note a simple property of the auxiliary ellipsoid. The prolate ellipsoid of revolution ($a > b = c$) corresponds to the oblate auxiliary ellipsoid of revolution ($\tilde{a} = \tilde{b} > \tilde{c}$).

We shall analyze this system of Eqs. (2) and (3). It has a solution in the form of an oblate ellipsoid of revolution ($a = b > c$) for any values of the field. In this case, Eq. (2) is converted into an identity and Eq. (3) determines the implicit dependence of the eccentricity $e = \sqrt{1 - (c/a)^2}$ of the ellipsoid on the Bond number in the form

$$\begin{aligned}\mathrm{Bo} = 4\pi\left(\frac{1}{\mu - 1} + n_1\right)^2 \\ \times \frac{e^2}{(1 - e^2)^{1/3}} \frac{1 + e^2 - \tilde{n}_1(3 + e^2)}{-1 + e^2 + n_1(3 - 2e^2)}.\end{aligned} \tag{4}$$

Equation (4) describes continuous flattening of the droplet with increasing field, beginning from its spherical shape in the absence of the field. This solution is only stable for comparatively low magnetic permeabilities of the magnetic fluid. From $\mu = 5.08$ another solution appears in the form of a triaxial ellipsoid which exists in a bounded range of magnetic field amplitudes G_A and G_B and corresponding Bond numbers. Having appeared, this solution is always energetically more favorable than the spheroidal one. The scenario for spheroid–triaxial ellipsoid–disk transitions of the droplet also depends on μ. Figure 2 gives the curve for the droplet semiaxis ratio a/b calculated using Eqs. (2) and (3) in the plane of the field for $\mu = 5.94$. It can be seen that both transitions are the result of a soft bifurcation: the dependence of the deviation of a/b from unity is continuous in terms of the Bond number. Near the bifurcation points this dependence obeys a square-root law: $a/b - 1 \propto \sqrt{\mathrm{Bo} - \mathrm{Bo}_A}$ and $\propto \sqrt{\mathrm{Bo}_B - \mathrm{Bo}}$, respectively. A further increase in μ is accompanied by a change in the type of transitions from soft to hard. It follows from the solution of system (2) and (3) that this change in the type of bifurcation should be observed for $\mu = 11$ in weak fields and $\mu = 14$ in strong fields.

Figure 3 gives calculated dependences of the droplet semiaxis ratio a/b near the bifurcation points for a magnetic fluid having the permeability $\mu = 15$. A characteristic feature of the behavior of the curve in weak and strong fields is the appearance of the unstable branches AC and EF which correspond to the maximum of the droplet energy. Consequently, hysteresis of the droplet shape occurs: as the field increases, a spheroidal droplet is abruptly elongated to form a triaxial ellipsoid in the field G_A (shown by the dashed line AD in Fig. 3) whereas the reverse transition with decreasing field also takes place abruptly but at lower values of the field G_C. The behavior of the droplet in strong fields is similar: as the amplitude of the field increases, the droplet shape undergoes a hard bifurcation near point F (curve FK). The result for the disk–triaxial ellipsoid transition which takes place with decreasing field was unexpected. It can be seen from Fig. 3 that this transition should be accompanied by a change in the type of bifurcation: initially a soft bifurcation takes place near point B (curve BE in Fig. 3) and as the field decreases further, this is replaced by a hard bifurcation (curve EG). We

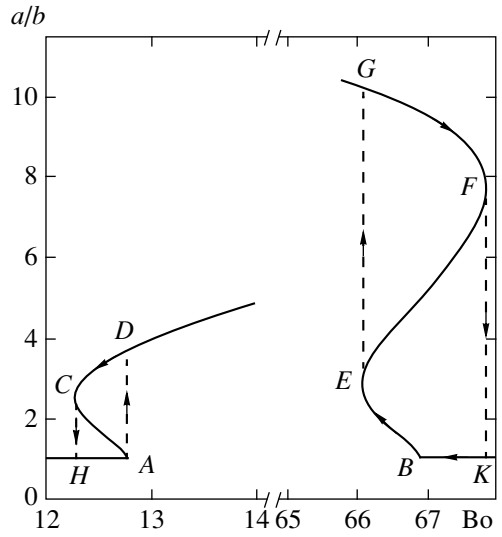

Fig. 3. Droplet semiaxis ratio a/b in the plane of the field as a function of the Bond number near the bifurcation points for $\mu = 15$.

were unable to record this change in the type of bifurcation experimentally because of the narrow range of fields $G_E - G_B$ which is comparable with the sensitivity limit of our equipment. Thus, it is impossible to state at the present time whether this type of transition does in fact occur or whether it is an artifact of the proposed model. In our experiments using concentrated magnetic fluids having permeabilities of 16.6 and 20.4 we reliably established that the droplet shape changes abruptly in accordance with the theoretical analysis and that hysteresis exists in weak and strong fields. For concentrated magnetic fluids the experimental values of the ratio a/b were 10–50% (depending on the field) lower than those calculated using Eqs. (2) and (3). The main reason for this deviation is the appearance of magnetization saturation effects for the concentrated magnetic fluids in strong fields. The magnetization curves $M(H)$ depend strongly on the disperse composition of the magnetic fluid and outside the initial (linear) section they are highly individual for different magnetic fluids. Thus, we neglected these effects in the theoretical part of this study.

We shall now analyze Fig. 1 which gives the calculated stability diagram of the droplet shapes in addition to the experimental data. The stability boundaries *A*, *B*, *C*, and *F* correspond to similar bifurcation points in Fig. 3 obtained for various permeabilities μ. The line described by the bifurcation point *E* at which the bifurcation regime changes (see Fig. 3) lies very close to line *B* and is not shown in Fig. 1. It can be seen that the experimental and theoretical data show good agreement except for the case of concentrated magnetic fluids and strong fields for which the saturation effects are appreciable.

Hence, droplets of magnetic fluid in a rapidly rotating magnetic field are either converted continuously from a spheroid to a disk or via a triaxial ellipsoid stage as the amplitude of the field increases. Does the life of the droplet end with this? It would seem not. As the field increases further, a set of needles (up to several tens) forms around the perimeter of the increasingly oblate droplet, converting the droplet into a "starfish." An experimental and theoretical study of the disk-shaped droplet stage in strong fields was made in [8].

ACKNOWLEDGMENTS

The authors are extremely grateful to A.F. Pshenichnikov for useful observations and discussions of the results. This work was supported by the Russian Foundation for Basic Research (project no. 98-01-00182).

REFERENCES

1. H. Lamb, *Hydrodynamics* (Dover, New York, 1945; Gostekhizdat, Moscow, 1947).
2. T. G. Wang, A. V. Anilkumar, C. P. Lee, and K. C. Lin, J. Fluid Mech. **276**, 389 (1994).
3. K. Ohsaka and E. H. Trinh, Phys. Rev. Lett. **84**, 1700 (2000).
4. M. I. Shliomis, Usp. Fiz. Nauk **112**, 427 (1974) [Sov. Phys. Usp. **17**, 153 (1974)].
5. A. V. Lebedev and K. I. Morozov, Pis'ma Zh. Éksp. Teor. Fiz. **65**, 150 (1997) [JETP Lett. **65**, 160 (1997)].
6. J.-C. Bacri and D. Salin, J. Phys. Lett. **43**, 649 (1982).
7. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media* (Nauka, Moscow, 1992; Pergamon, New York, 1984).
8. J.-C. Bacri, A. Cebers, and R. Perzynski, Phys. Rev. Lett. **72**, 2705 (1994).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 1033–1045.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1193–1206.

SOLIDS
Electronic Properties

Self-Oscillations in Semiconductor Superlattices

Yu. A. Romanov and Yu. Yu. Romanova*

Institute of Physics of Microstructures, Russian Academy of Sciences, Nizhni Novgorod, 603600 Russia
**e-mail: jul@ipm.sci-nnov.ru*

Received January 24, 2000

Abstract—An investigation is made of nonlinear oscillations of the field and current in semiconductor superlattices driven by strong terahertz radiation. Regimes of periodic, quasi-periodic, and stochastic self-oscillations are determined and mechanisms for their formation are discussed. It is shown that the self-oscillation spectra are many-valued functions of the external field amplitude and the static field in them is either absent, weak, or fractionally quantized. Previously predicted states of self-induced superlattice transparency and dynamic electron localization are destroyed as a result of the evolution of dissipative and parametric instabilities and can only be observed in transient processes whose duration decreases with increasing electron concentration. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A semiconductor superlattice is a single-crystal structure whose chemical composition varies periodically in space at distances (1–10 nm) greater than the periods of the crystal lattices of its constituent materials. This structure exhibits an additional periodic (superlattice) potential which leads to splitting of the Brillouin quasi-momentum zones and the allowed electron energy bands in the initial homogeneous materials into a set of relatively narrow (10^4–10^6 cm^{-1}) Brillouin minizones and narrow (10^{-3}–10^{-1} eV) allowed and forbidden energy minibands [1–3]. As a result of the small dimensions of these minibands the superlattice becomes a highly nonlinear and unstable medium even in relatively weak electric fields (10^2–10^4 V/cm) [4]. In general, some contribution to the electron conductivity is made by intraminiband electron motion, interminiband transitions within a single band, interband transitions, and processes containing two or three of these types of motion simultaneously. We shall confine ourselves to the single-miniband approximation. The conditions for its validity will be given below. Significantly the corresponding nonlinear conductivities vary nonmonotonically (exhibit oscillatory behavior) as the amplitude and frequencies of the fields acting in the superlattice increase.

The clearest manifestation of the nonmonotonic behavior (oscillatory behavior) of the intraminiband rf conductivity of the superlattice can be found in the self-induced [5], induced [6], and selective transparencies [4] which essentially consist of the following. Under the action of an rf harmonic field $E(t) = E_1\cos(\omega_1 t)$ ($\omega_1\tau \gg 1$, τ is the relaxation time of the electron momentum distribution), a nonlinear electric current appears in the superlattice [4]

$$j(t) \approx 2j_0\left\{J_0(g_1)\sum_{\mu=1}^{\infty}J_{2\mu-1}(g_1)\sin[(2\mu-1)\omega_1 t] + \frac{1}{g_1\omega_1\tau}[1-J_0^2(g_1)]\cos(\omega_1 t)\right\}, \quad (1)$$

where j_0 = const, $g_1 = eE_1d/\hbar\omega_1$, e is the electron charge, d is the superlattice period, and $J_n(x)$ are Bessel functions. It can be seen from (1) that for discrete values of the ratio of the amplitude of this field to the frequency for which $J_0(g_1) = 0$, the macroscopic polarization of the electron gas disappears and the superlattice behaves almost as a linear dielectric possessing the permittivity of the main crystal lattice and weak (but extremal!) nonlinear absorption. This is self-induced transparency. It is found [6] that if the superlattice has been converted to a state of transparency using one rf harmonic field, it will also be transparent for another rf harmonic field (i.e., it will not be polarized by this field) if the amplitude of this field is not too large and the harmonic frequencies of the fields are not very similar. This is induced transparency. Unlike induced and self-induced transparency, selective transparency implies the alternate disappearance of various harmonics of the nonlinear current as the amplitudes or frequencies of the fields acting in the superlattice vary. In a harmonic field the nth harmonic of the current [see (1)] disappears when $J_n(g_1) = 0$.

1063-7761/00/9105-1033$20.00

In [7] the authors reported the experimental observation of self-induced transparency in a superlattice driven by terahertz laser radiation. However, since the authors only studied the third harmonic of the current and it is difficult to convert the external field to the field inside the superlattice, the conclusions of this study cannot be considered to be definitive. Moreover, in fields with $J_0(g_1) = 0$ dynamic electron localization [8] and collapse of its quasienergy minibands occurs [9]. This implies that the electron translational motion is completely converted into vibrational and the maximum possible buildup of oscillations occurs (since the width of the energy miniband is finite and the electron velocity in it is limited), i.e., the energy exchange between the field and the isolated electrons is most efficient. Thus, when $J_0(g_1) = 0$ the absorption of the harmonic field has a maximum which is reflected in the dissipative current surge in (1). The electrons can transfer the energy acquired from the harmonic field not only to the lattice but also partly to other fields (including the static field), amplifying them. The appearance of a multifrequency field in the superlattice destroys the self-induced transparency which can only exist in the superlattice under certain conditions (which will be discussed below). Thus, it is possible that the authors of [7] observed selective transparency rather than self-induced transparency. No observations of induced transparency have yet been reported in the literature.

Manifestations of the nonmonotonic behavior of the dissipative conductivities (including inversion of their sign) in a superlattice include alternating energy exchange between harmonic fields [6] and between harmonic and static fields [10], including absolute negative conductance (dc negative conductance caused by the absorption of rf field energy) [4, 5, 10], spontaneous generation of a static field and a static current [12] by a harmonic field [4, 11], resonant superheterodyne amplification and explosive parametric instability of electromagnetic waves [13, 14]. Absolute negative conductance was observed experimentally in [15], alternating energy exchange between static and harmonic fields was observed in [16], but we are not aware of any reported observations of spontaneous static field generation in a superlattice.

Another manifestation of the nonmonotonic behavior of nonlinear superlattice conductivities is the many-valued dependence of the intralattice field on the external field. The need to allow for this many-valued behavior (multistability) which leads to hysteresis effects was evidently first noted in [17] in a study of the penetration of a transverse electromagnetic wave into a nondissipative superlattice. This was also discussed in [18] where the authors studied the possibility of observing self-induced transparency experimentally in a superlattice.

The behavior of a superlattice in a given internal electric field with different time dependences has been studied fairly comprehensively (see, for example, [4, 19] and the literature cited there). In order to study superlattice behavior in external fields, at first glance it is sufficient merely to make a suitable many-valued change in the field amplitudes. This approach was used in [17, 18] to study multistable states in a nondissipative superlattice. However, this approach is not always correct even for studying steady states. A correct approach must allow for the linear and nonlinear resonances of the electrodynamic system, the inertia of the coupling between the field inside the superlattice and the external field, dissipative and parametric instabilities of the fields. As a result of the combined manifestation of these factors, a superlattice located in a strong external harmonic field is generally converted to self-oscillatory regimes of which there may be several. The nature of the transient processes and the final state of the system depend on the electron concentration, the amplitude and frequency of the external field, the parameters of the external circuit, and in general, on the initial conditions and the rate of switching on the external field. Self-oscillation regimes in a superlattice (in [17, 18] these were not analyzed because the approximation of a single-frequency internal field was used) are of considerable interest, particularly in the experimental context. In particular, without taking these regimes into account it is impossible to correctly formulate an experiment to study superlattice transparency.

In the present study we investigate nonlinear oscillations of the field and current observed in one-dimensional superlattices driven by external terahertz radiation or in superlattices connected to an external circuit with a given voltage source. A specific electrodynamic system may be a quasioptic system similar to that studied in [18] in which terahertz laser radiation is fed inside the superlattice using a microwave antenna. The equivalent electric circuit corresponding to that used in the calculations contains a superlattice of thickness Nd (N is the number of periods in the sample) shunted by a resistance R, a voltage source $V = V_0 \cos(\omega_1 t)$, whose amplitude is determined by the intensity of the laser field, an external capacitance C_1, which takes into account the substrate surrounding the medium and the series-connected capacitance of the antenna, and also a load impedance r_i, which includes the radiation resistances of the antenna, the contacts, and the voltage source. All the dimensions of the structure are assumed to be small compared with the wavelength. Unlike [12], where stochastic current oscillations were studied, the sample is open-circuit to direct current.

Important characteristics of the self-oscillatory system under study are as follows.

(1) The same external harmonic field is a source of self-oscillation energy and traps their frequencies in a specific range of its parameters.

(2) The natural frequencies of the system are the plasma frequency and the frequency of the electron Bloch oscillations. The frequency of the plasma oscillations depends strongly on the amplitudes of the harmonics of the total field in the superlattice (it may even go to zero) while oscillations of the macroscopic quan-

tities at the Bloch frequency (in particular, the current) only occur in transient processes and are damped rapidly.

2. BASIC EQUATIONS

We shall consider an electron superlattice in which electrons only fill the lower miniband having the harmonic dispersion law

$$\begin{aligned}\varepsilon(k) &= \varepsilon_3(k_3) + \varepsilon_\perp(k_\perp)\\ &= \frac{\Delta}{2}[1-\cos(k_3 d)] + \frac{\hbar^2 k_\perp^2}{2m},\end{aligned} \tag{2}$$

where Δ is the width of the miniband, $\varepsilon_3(k_3)$, $\varepsilon_\perp(k_\perp)$, and $\hbar k_{3,\perp}$ are the longitudinal and transverse energies and the components of the electron momentum $\hbar k$, relative to the superlattice axis, respectively, and m is its transverse mass. This miniband is separated from the other minibands by a distance greater than Δ_g which is the characteristic nearest miniband gap. We shall assume that the electric field E having the characteristic frequency ω is uniform and directed along the superlattice axis, and the number of periods in the sample is $N \sim 10$–10^3. A lower constraint is imposed on N by the possibility of introducing the concept of continuous quasimomentum and an upper constraint is imposed by the wavelength of the radiation acting on the superlattice and the approximation of a uniform field in the lattice. (The formation of field domains is neglected in the present study.) We shall assume that the following inequalities are satisfied

$$\begin{aligned}\Delta_g \gg \hbar\tau^{-1}, \quad \Delta \gg \hbar\tau^{-1},\\ \Delta_g \gg \hbar\omega_1, eEd, \quad \Delta \gg eEd.\end{aligned} \tag{3}$$

The first inequality in (3) is required for the appearance of a miniband structure, the second is required for the existence of a specific dispersion law in the miniband (it is not necessary for the validity of the qualitative and even some quantitative results of the present study), the third allows the analysis to be confined to the single-miniband approximation, and the fourth (also not necessary for the validity of the qualitative results) combined with the previous three ensures that the semiclassical description of electron behavior in the miniband is valid. Assuming that these conditions are satisfied, the electron and field behavior in the superlattice will be described by the Boltzmann equation in the τ-approximation,

$$\frac{\partial f(\mathbf{k},t)}{\partial t} + \frac{eE(t)}{\hbar}\frac{\partial f(\mathbf{k},t)}{\partial k_3} = -\frac{f(\mathbf{k},t) - f_0(\mathbf{k})}{\tau}, \tag{4}$$

and the equation of continuity of the total current,

$$\varepsilon_0\frac{dE(t)}{dt} + 4\pi j(t) + \frac{\varepsilon_0 E(t)}{RC_S} = 4\pi j_e(t). \tag{5}$$

Here $f(\mathbf{k}, t)$ and $f_0(\mathbf{k})$ are the field-perturbed and equilibrium electron functions, $E(t)$ and $j(t)$ are the electric field and electron current density in the superlattice, respectively, $C_S = \varepsilon_0/4\pi Nd$ is the linear capacitance of the superlattice, ε_0 is the permittivity in the absence of electrons, $j_e(t)$ is the external current density which depends on the scheme for connection of the superlattice to the external circuit, and τ = const. In a scheme with a given external field $E_e(t) = E_0\cos(\omega_1 t)$, we have

$$j_e(t) = \frac{\varepsilon_e}{4\pi}\frac{\partial E_e(t)}{\partial t}, \tag{6}$$

where ε_e is the permittivity of the external medium. In a scheme with a given voltage source $V(t) = V_0\cos(\omega_1 t)$ the external current $j_e(t)$ is determined by the equations

$$Sj_e(t) = C_1\frac{dV_1(t)}{dt}, \tag{7}$$

$$\left(1 + C_1 r_i\frac{d}{dt}\right)V_1(t) = V(t) - E(t)Nd, \tag{8}$$

where S is the area of the superlattice cross section and $V_1(t)$ is the voltage drop at the capacitance C_1.

We introduce the complex function

$$\varphi(k_3,t) = \left[\frac{\Delta}{2} - \varepsilon_3(k_3) - \frac{i}{d}\frac{d\varepsilon_3(k_3)}{dk_3}\right]\left(\frac{\Delta}{2} - \langle\varepsilon_3\rangle_0\right)^{-1},$$

where $\langle\varepsilon_3\rangle_0$ is the average equilibrium longitudinal electron energy, multiply the left- and right-hand sides of Eq. (4) by this function, and integrate over $\mathbf{k}$ within the first Brillouin minizone. As a result, for the average longitudinal electron energy

$$\langle\varepsilon_3\rangle = \frac{1}{n}\int\varepsilon_3(k_3) f(\mathbf{k},t)\frac{d^3k}{(2\pi)^3},$$

and the current density

$$j(t) = \frac{e}{\hbar}\int\frac{\partial\varepsilon_3(k_3)}{\partial k_3} f(\mathbf{k},t)\frac{d^3k}{(2\pi)^3},$$

we obtain hydrodynamic equations in the following complex form convenient for investigation:

$$\tau\frac{d\Phi(t)}{dt} + [1 + i\tau\Omega(t)]\Phi(t) = 1, \tag{9}$$

where

$$\Phi(t) \equiv \langle\varphi(k_3,t)\rangle = \frac{\langle\varepsilon_3\rangle - \Delta/2}{\langle\varepsilon_3\rangle_0 - \Delta/2} - i\frac{j(t)}{j_0}, \tag{10}$$

n is the electron concentration, $\Omega(t) = edE(t)/\hbar\omega_1$ is the instantaneous "Bloch" frequency, $j_0 = \hbar\varepsilon_0\omega_0^2/4\pi ed$,

$$\omega_0^2 = \frac{4\pi ne^2d^2}{\varepsilon_0\hbar^2}\left(\frac{\Delta}{2} - \langle\varepsilon_3\rangle_0\right)$$

is the square of the plasma frequency which describes linear oscillations of the superlattice plasma in the absence of external electric fields.

In order to make a qualitative study of self-oscillations of the currents and fields, it is not necessary to introduce the resistances r_i and R since dissipating current harmonics always exist in a superlattice in the nonlinear regime. Dissipation in the superlattice must be taken into account even for $r_i \neq 0$ and finite R. Neglecting this leads to the absence of spontaneous generation of the static field which significantly alters the nature of the nonlinear oscillations in the superlattice. For quantitative calculations we shall assume that the following inequalities are satisfied

$$\omega_1 r_i C_1 \ll 1, \quad \omega_1 R C_S \gg 1,$$
$$\alpha \equiv \frac{r_i\tau}{n\nu\Delta}\sum_\omega \frac{V_\omega^2}{r_i^2 + (\omega C_1)^{-2}} \ll 1, \tag{11}$$

where $\nu = NSd$ is the superlattice volume, V_ω is the amplitude of the voltage harmonic generated over it, and summation is performed over all frequencies ω. The last inequality in (11) implies that the losses to the resistance r_i are small compared with the characteristic losses inside the superlattice and the first two imply that the circuit has a high Q factor. In this case, terms containing r_i and R can be neglected in Eqs. (5) and (8) (the numerical calculations were performed allowing for these terms) and the antenna capacitance can be included in C_S (more accurately, its component parallel to C_S, the series component of the antenna capacitance is included in C_1).

The radiation power of the sample at frequency ω can then be determined using the approximate formula

$$P_\omega = \frac{V_\omega^2 R_r}{2[r_i^2 + (\omega C_1)^{-2}]}, \tag{12}$$

where R_r is the radiation resistance of the sample which forms part of r_i. We shall give typical values of the system parameters taken from [6]: $\omega_1 = 2\pi \times 0.7$ THz, $N = 100$, $d = 100$ Å, $C_S \sim C_1 = 7$ fF, $R_r = 2\ \Omega$, $r_i = 7\ \Omega$, $R = 200\ \Omega$, $n = 10^{17}$ cm^{-3}, $\tau = 2 \times 10^{-12}$ s, and $\Delta = 20$ meV. For these parameters we have $\omega_1 r_i C_1 \approx 0.2$, $\omega_1 R C_S \approx 6$, $V_\omega \approx 0.2$ V (taken from the results of the numerical calculations presented below), $\alpha \approx 0.1$, $P_\omega \sim 10^{-4}$ W; i.e., the inequalities (11) are satisfied and the power emitted by the superlattice is appreciable. At the same time these estimates show that the resistances r_i and R must be taken into account to analyze specific experimental results (at present none are available).

Using the approximation (11) we obtain the following self-consistent system of equations for the electric current $j(t)$, the longitudinal electron energy $n\langle\varepsilon_3\rangle$, and the internal electric field $E(t)$ in terms of dimensionless variables:

$$\omega_1\tau\frac{d\Phi(\tilde t)}{d\tilde t} + [1 + i\omega_1\tau g(\tilde t)]\Phi(\tilde t) = 1, \tag{13}$$

$$\frac{dg(t)}{d\tilde t} = \tilde w\,\mathrm{Im}\,\Phi(t) - V_0\sin\tilde t, \tag{14}$$

where $\tilde t = t\omega_1$, $g(t) = \Omega(t)/\omega_1$ is the dimensionless voltage incident on a single period of the superlattice,

$$\tilde V_0 = \frac{eE_0 d\varepsilon_1}{\hbar\omega_1\varepsilon_0}, \quad \tilde w = \left(\frac{\omega_0}{\omega_1}\right)^2$$

in a scheme with a given external field and

$$\tilde V_0 = \frac{eV_0}{N\hbar\omega_1(1 + C_S/C_1)}, \quad \tilde w = \frac{\omega_0^2}{\omega_1^2(1 + C_1/C_S)}$$

in a scheme with a given voltage source.

The behavior of the superlattice in a given internal field with an arbitrary time dependence is described by the single Eq. (13). Its general solution has the form

$$\Phi(t) = \psi(t)\left[\Phi(0)\exp\left(-\frac{t}{\tau}\right) + \int_0^t \exp\left(-\frac{t - t_1}{\tau}\right)\psi^*(t_1)dt_1\right], \tag{15}$$

where

$$\psi(t) = \exp\left(-i\int_0^t \Omega(t_1)dt_1\right). \tag{16}$$

The most typical situation is that when the field in the superlattice can be approximately represented as the sum of the static and biharmonic fields:

$$E(t) = E_C + E_1\cos(\omega_1 t + \delta_1) + E_2\cos(\omega_2 t + \delta_2). \tag{17}$$

In accordance with (10) and (15), this field excites the current density in the superlattice

$$j(t) = \sum_{\nu_{1,2} = -\infty}^{\infty} j_{\nu_1\omega_1, \nu_2\omega_2} \times \exp[-i\nu_1(\omega_1 t - \delta_1) - i\nu_2(\omega_2 t - \delta_2)] + \text{c.c.}, \tag{18}$$

$$j_{\nu_1\omega_1,\nu_2\omega_2} = \frac{i}{2}j_0 \times \sum_{\mu_{1,2}=-\infty}^{\infty} \frac{J_{\mu_1}(g_1)J_{\mu_1+\nu_1}(g_1)J_{\mu_2}(g_2)J_{\mu_2+\nu_2}(g_2)}{1+i(\Omega_C+\mu_1\omega_1+\mu_2\omega_2)\tau}, \quad (19)$$

where $g_{1,2} = \Omega_{1,2}/\omega_{1,2}$, $\Omega_{1,2} = eE_{1,2}d/\hbar$.

If the field E_2 is weak ($g_2 \ll 1$), (18) may be expressed in the form

$$j(t) = \sigma_C(\Omega_C, \Omega_1, \omega_1)E_C + \mathrm{Re}\sum_{n=-\infty}^{\infty} \sigma(\omega_2+n\omega_1; \Omega_C, \Omega_1, \omega_1)E_2 \times \exp\{-i[(\omega_2+n\omega_1)t - n\delta_1 - \delta_2]\} + \mathrm{Re}\sum_{n=1}^{\infty} \sigma(n\omega_1; \Omega_C, \Omega_1, \omega_1)E_1\exp\{-in(\omega_1 t - \delta_1)\}, \quad (20)$$

where

$$\sigma_C(\Omega_C, \Omega_1, \omega_1) = \frac{\sigma_0}{\Omega_C} \times \sum_{\mu=-\infty}^{\infty} \frac{\Omega_C+\mu\omega_1}{1+(\Omega_C+\mu\omega_1)^2\tau^2}J_\mu^2(g_1) \leq \frac{\sigma_0}{2\Omega_C\tau}, \quad \sigma_0, \quad (21)$$

$$\sigma(\omega_2+n\omega_1; \Omega_C, g_1, \omega_1) = \sigma^*(-\omega_2-n\omega_1; \Omega_C, \Omega_1, \omega_1) = i\frac{\varepsilon_0\omega_0^2}{8\pi}\sum_{\nu=-\infty}^{\infty} J_\nu(g_1)J_{\nu+n}(g_1) \times \left\{\frac{1}{[1+i\tau(\nu\omega_1+\Omega_C)][\omega_2-\nu\omega_1-\Omega_C+i\tau^{-1}]} + \frac{(-1)^n}{[1+i\tau(\nu\omega_1-\Omega_C)][\omega_2-\nu\omega_1+\Omega_C+i\tau^{-1}]}\right\}, \quad (22)$$

$$\sigma(n\omega_1; \Omega_C, \Omega_1, \omega_1) = i\frac{\varepsilon_0\omega_0^2}{4\pi\Omega_1}\sum_{\mu=-\infty}^{\infty} J_\mu(g_1)J_{\mu+n}(g_1) \times \left\{\frac{1}{1+i(\Omega_C+\mu\omega_1)\tau} - \frac{(-1)^n}{1-i(\Omega_C-\mu\omega_1)\tau}\right\} = [\sigma(\omega_2+(n-1)\omega_1; \Omega_C, \Omega_1, \omega_1) - \sigma(-\omega_2+(n+1)\omega_1; \Omega_C, \Omega_1, \omega_1)]_{\omega_2=\omega_1}, \quad (23)$$

$\sigma_0 = ed\tau/\hbar j_0 = \varepsilon_0\omega_0^2\tau/4\pi$ is the linear static conductivity of the superlattice. The nonlinear conductivity $\sigma(\omega_2 + n\omega_1; \ldots)$ for $n = 0$ is the linear conductivity of the superlattice at frequency ω_2 varied by the harmonic field E_1. It describes nonsynchronous (which does not depend on the phase relationships) interaction of the fields whereas for $n \neq 0$ it describes synchronous interaction of these fields. The conductivities $\sigma(n\omega_1; \ldots)$ describe the generation of n harmonics of the field E_1.

If the frequency of the weak field E_2 is not a multiple or half-multiple of ω_1 ($\omega_2 \neq 0.5n\omega_1$, $n = 1, 2, \ldots$), the static current is completely determined by the nonlinear conductivity $\sigma_C(\Omega_C, \Omega_1, \omega_1)$ and the current at frequency ω_2 is determined by the "linear" conductivity $\sigma(\omega_2; \Omega_C, \Omega_1, \omega_1)$. An important relationship exists between these conductivities:

$$\mathrm{Re}\sigma(\omega_2; \Omega_C, \Omega_1, \omega_1) = \frac{1}{2}\left[\left(1+\frac{\Omega_C}{\omega_2}\right)\sigma_C(\Omega_C+\omega_2, \Omega_1, \omega_1) + \left(1-\frac{\Omega_C}{\omega_2}\right)\sigma_C(\Omega_c-\omega_2, \Omega_1, \omega_1)\right]. \quad (24)$$

In particular, we have

$$\mathrm{Re}\sigma(\omega_2; \Omega_C = 0, \Omega_1, \omega_1) = \sigma_C(\Omega_C = \omega_2, \Omega_1, \omega_1)|_{E_2=0}. \quad (25)$$

In a quadratic approximation with respect to E_2 (required to understand the peculiarities of energy exchange between the fields), the static conductivity and the conductivity at ω_1 are given by

$$\sigma_C(\Omega_C, \Omega_1, \Omega_2, \omega_1) = \left(1-\frac{1}{2}g_2^2\right)\sigma_C(\Omega_C, \Omega_1, \omega_1) + \frac{1}{2}g_2^2\mathrm{Re}\sigma(\omega_2=\Omega_C; \Omega_C=\omega_2, \Omega_1, \omega_1), \quad (26)$$

$$\mathrm{Re}\sigma(\omega_1; \Omega_C, \Omega_1, \Omega_2, \omega_1) = \left(1-\frac{1}{2}g_2^2\right)\mathrm{Re}\sigma(\omega_1; \Omega_C, \Omega_1, \omega_1) + \frac{1}{4}g_2^2\mathrm{Re}[\sigma(\omega_1; \Omega_C+\omega_2, \Omega_1, \omega_1) + \sigma(\omega_1; \Omega_C-\omega_2, \Omega_1, \omega_1)], \quad (27)$$

where

$$\mathrm{Re}\sigma(\omega_1; \Omega_C, \Omega_1, \omega_1) = \frac{2\sigma_0}{g_1\Omega_1}\sum_{\mu=-\infty}^{\infty} \frac{\mu(\Omega_C+\mu\omega_1)}{1+(\Omega_C+\mu\omega_1)^2\tau^2}J_\mu^2(g_1) = \frac{2\sigma_0}{\Omega_1^2}\sum_{\mu=-\infty}^{\infty} \frac{(\Omega_C+\mu\omega_1)^2}{1+(\Omega_C+\mu\omega_1)^2\tau^2}J_\mu^2(g_1) \quad (28)$$

$$-2\left(\frac{\Omega_C}{\Omega_1}\right)^2\sigma_C(\Omega_C, \Omega_1, \omega_1)$$

is the real part of the nonlinear conductivity of the superlattice at frequency ω_1 in the field (17) for $E_2 = 0$. It can be seen from (28) that the rf conductivity of the superlattice can become negative only in the presence of a relatively high positive static conductivity. However, we stress that for $\omega_1 = \Omega_C$ this is positive for any amplitudes E_1 at variance with the statement in [20].

3. INSTABILITY OF SUPERLATTICE TRANSPARENCY STATES. LOW ELECTRON CONCENTRATIONS

At low electron concentrations ($\omega_0/\omega_1 \ll 1$), the given internal field approximation is a good approximation. In this approximation we shall first consider the interaction between harmonic and static fields having arbitrary amplitudes described by the conductivities (21), (23), and (28). In accordance with (20), (21), and (28) the absorbed energy in the field $E_C + E_1\cos(\omega_1 t)$ is given by

$$\overline{jE} = \sigma_C E_C^2 + \frac{1}{2}\mathrm{Re}\sigma(\omega_1; \Omega_C, \Omega_1, \omega_1)E_1^2$$

$$= \sigma_0\left\{\frac{E_C^2}{1 + (\Omega_C\tau)^2} + 2\left(\frac{\hbar}{ed}\right)^2\sum_{\mu=1}^{\infty}J_\mu^2(g_1)\right. \qquad (29)$$

$$\left.\times\frac{\Omega_C^2 + (\mu\omega_1)^2}{[1 + (\Omega_C + \mu\omega_1)^2\tau^2][1 + (\Omega_C - \mu\omega_1)^2\tau^2]}\right\}.$$

The first term in (29) describes the losses in a single static field. The second term is always positive and consequently the application of a harmonic field always leads to an additional increase (small for $\omega_1\tau \gg 1$) in the total electromagnetic losses in the superlattice. In the resonant regions $\Omega_C = \mu_0\omega_1 \pm \tau^{-1}$ ($\mu_0 = 1, 2, \ldots$; $\omega_1\tau \gg 1$) the conductivities $\sigma_C(\Omega_C, \Omega_1, \omega_1)$ and $\mathrm{Re}\sigma(\omega_1; \Omega_C, \Omega_1, \omega_1)$ are extremal, depend weakly on τ, and, in accordance with (21), (28), weakly dissipative resonant energy exchange takes place between the fields, described by

$$j_C E_C = -\frac{1}{2}\mathrm{Re}\sigma(\omega_1; \Omega_C, \Omega_1, \omega_1)E_1^2 + O\left(\frac{1}{\omega_1\tau}\right)$$

$$\approx \mu_0\hbar\omega_1\frac{j_0}{ed}J_{\mu_0}^2(g_1)\frac{(\Omega_C - \mu_0\omega_1)\tau}{1 + (\Omega_C - \mu_0\omega_1)^2\tau^2}. \qquad (30)$$

Relation (30) shows that in these resonant regions as it passes between neighboring wells in the superlattice, an electron moving in the opposite direction to the static field, acquires the energy $\hbar\Omega_C$ and has the probability $J_{\mu_0}^2(g_1)$ of emitting μ_0 quanta of the harmonic field E_1 ($\Omega_C > \mu_0\omega_1$) or absorbing them and is shifted by a period of the superlattice with respect to the static field, overcoming the potential barrier of height $\hbar\Omega_C$ ($\Omega_C < \mu_0\omega_1$). As a result of these transitions, only a relatively small ($\sim\hbar\tau^{-1}$) excess energy is transferred to the lattice so that [see (29) and (30)] $j_C E_C \gg \overline{jE}$. As in stimulated Raman light scattering in equilibrium media [see, for example, [21]) and in accordance with the conservation law energy is transferred to the field having the lower of the frequencies ω_1 and Ω_C.

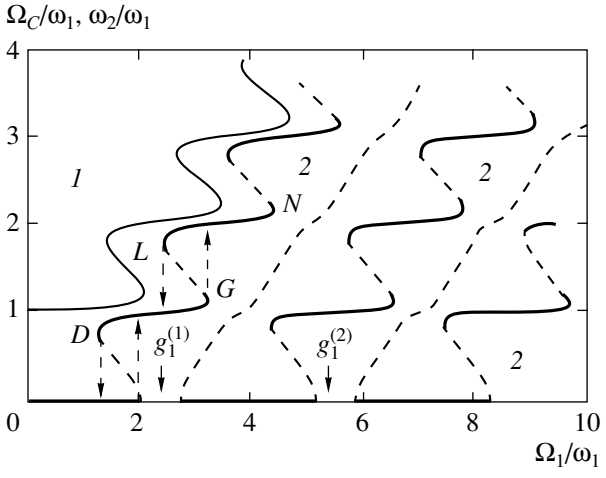

Fig. 1. Regions of negative conductance for the fields E_1 (*1*) and E_C (*2*) for $E_2 = 0$ and for E_2 (*2*) for $E_C = 0$; $\omega_1\tau = 4$, $g_1^{(1)} = 2.41$, $g_1^{(2)} = 5.5$.

Figure 1 shows regions of negative values of the static [Eq. (21)] and rf [Eq. (28)] conductivities for $\omega_1\tau = 4$. An abrupt change in the energy exchange between the fields in the resonant regions described approximately by Eqs. (30) can be clearly seen. The boundary curves of regions *2* correspond to current-free ($j_C = 0$) states where the solid heavy sections are stable, which are determined by the conditions

$$\sigma_C(\Omega_C = 0, \Omega_1, \omega_1) > 0$$

or

$$\sigma_C(\Omega_C, \Omega_1, \omega_1) = 0, \quad \frac{d\sigma_C(\Omega_C, \Omega_1, \omega_1)}{d\Omega_C} > 0, \qquad (31)$$

the dashed and thin solid sections on the abscissa are unstable with respect to quasi-static fluctuations of the field. In the unstable state any arbitrarily small fluctuation of E_C increases with time and the superlattice is transferred to an upper or lower stable state with respect to the quasi-static perturbations with $E_C \neq 0$. An important characteristic of the boundary curves is the presence of regions (E_1, ω_1) which can have two or more stable states with different values of E_C (multistability). Thus, hysteresis occurs in the dependence of the stable values of $E_C(g_1)$. Figure 1 shows examples of these dependences for infinitely slowly varying g_1. The

arrows indicate the direction of change in the amplitude of the harmonic field. For $\omega_1\tau \gg 1$ (and low electron concentrations!) the dc voltage formed in the superlattice for $j_C = 0$ is

$$V_C = E_C L \approx \frac{\hbar\omega_1 L}{e d} n, \quad n = 0, \pm 1, \pm 2, \ldots \qquad (32)$$

(L is the superlattice length), i.e., it has a pronounced stepped character, is a multiple of the frequency of the harmonic field (integer quantized), and does not depend on the superlattice material. (If the electron dispersion law departs from harmonic, the quantization of the static field becomes fractional.) At the nth stable step the nonlinear current is $j(t) \sim J_n(g_1) \neq 0$, i.e., no self-induced transparency occurs, it being shifted toward stronger fields where $\Omega_C \approx n\omega_1$, $j_C \neq 0$ but (see Fig. 1) is also unstable. Numerical calculations show that for $\omega_1\tau \sim 1$ dc voltage jumps only occur with decreasing g_1, they are smooth and significantly smaller than (32). For $\omega_1\tau \ll 1$ no current-free dc voltage occurs.

This reasoning applies to the soft regime for the excitation of dc voltage. Hard regimes may also occur, for example, when $g_1 = 4$ (see Fig. 1) which require fairly appreciable ($\sim\hbar\omega_1/ed$) initial fluctuations of the static field.

Since the relationship (25) exists between the nonlinear static and linear rf conductivities in the superlattice, apart from the substitution $\Omega_C \longrightarrow \omega_2$ the regions of absolute negative conductance shown in Fig. 1 are also regions of negative rf conductance for the weak field E_2. Using this relationship, for weakly dissipative resonant energy exchange between the harmonic fields E_1 and E_2 in the regions $\omega_2 \approx \mu_0\omega_1 \pm \tau^{-1}$ for $E_C = 0$ we obtain from (27) and (30)

$$\overline{j(t)E_1(t)} \approx -\overline{j(t)E_2(t)} + \frac{\hbar j_0}{ed\tau}[1 - J_0^2(g_1)], \qquad (33)$$

$$\overline{j(t)E_2(t)} \approx \pm\frac{j_0}{ed}\frac{g_2^2}{4}\hbar\omega_2\left[J_{\mu_0}^2(g_1) + O\left(\frac{1}{\omega_1\tau}\right)\right]. \qquad (34)$$

Relation (34) can be obtained directly from (22) for $n = 0$. The second term in (33) describes absorption of the strong field E_1 for $E_2 = 0$. It has been noted that this has a maximum in regions of dynamic localization, i.e., when $J_0(g_1) = 0$ [4]. The direction of electromagnetic energy transfer corresponds to stimulated Compton light scattering in equilibrium media, i.e., energy is transferred to the field having the lower of the frequencies $\mu_0\omega_1$ and ω_2. In addition to dc voltage jumps (or instead of these) plasma or other rf (determined by the parameters of the external circuit) oscillations will also be excited in the superlattice. Several scenarios are possible depending on which process takes place faster, generation of a dc voltage or plasma (or other) oscillations. For large g_1 (and high electron concentrations) competition between these processes at the stage of the transient process can substantially lengthen its duration and even lead to stochastic oscillations (see Section 4). In general, the states to which the superlattice may be transferred under the influence of an external harmonic field are nonlinear oscillations whose spectrum may contain a static field (zeroth harmonic). Several such states may exist which is manifest in particular in the many-valued (and hysteresis) dependences of the static field and the current and radiation spectra of the superlattice on the external field amplitude. For the coefficient of reflection of the laser radiation and the generation of its third harmonic this many-valued behavior (even neglecting dissipative and parametric instabilities and static field generation, which significantly alter its character) was observed in [17, 18].

Quite clearly at very low electron concentrations in the regions $\sigma_C(\Omega_C, \Omega_1, \omega_1) < 0$ the transition to a state with finite dc voltage is the determining factor. The initial stage of this transition is described by the dispersion equation

$$\varepsilon(\omega_2) = \varepsilon_0 + i\frac{4\pi}{\omega_2}\sigma(\omega_2; \Omega_C, \Omega_1, \omega_1) = 0. \qquad (35)$$

In the region of initial dynamic localization ($E_C = 0$, $J_0(g_1) \approx 0$) and for $|\omega_2| < \omega_1$ we obtain from (22) and (35)

$$\omega_2 \approx 2i\omega_0^2\tau F(g_1, \omega_1\tau), \qquad (36)$$

where

$$F(x, y) = \sum_{\mu=1}^{\infty} J_\mu^2(x)\frac{\mu^2 y^2 - 1}{(\mu^2 y^2 + 1)^2}. \qquad (37)$$

For $\omega_1\tau > 2$ we have $\omega_2 \approx 0.5i(\omega_0/\omega_1)^2\tau^{-1}$. Consequently, at low electron concentrations the characteristic time for the loss of self-induced transparency and transition of the superlattice to a state with finite E_C is

$$\Delta t \gg \tau, \frac{2\pi}{\omega_1},$$

and for high concentrations ($\omega_0 \sim \omega_1$)

$$\Delta t \sim \tau. .$$

By way of example Fig. 2 gives time evolutions of the current, average field, and current spectra in the superlattice in a constant external harmonic field with $\tilde{w} = 0.05$, $\tilde{V}_0 = 2.4$, $\omega_1\tau = 10$. It can be seen that the superlattice is transferred to a state of self-induced transparency where all the current harmonics are small, fairly rapidly within a few periods of the field ($\sim\tau$). Then, slowly (over approximately 200 periods), accelerating appreciably for $\Omega_C(t) \longrightarrow \omega_1$ in accordance with the resonant increase in energy exchange between the static and harmonic fields noted above [formula (30)], the self-induced transparency is destroyed (aperiodic instability) and the superlattice is transferred to the state with $\Omega_C \approx -\omega_1$, relatively large current ($\sim J_1(g_1)$),

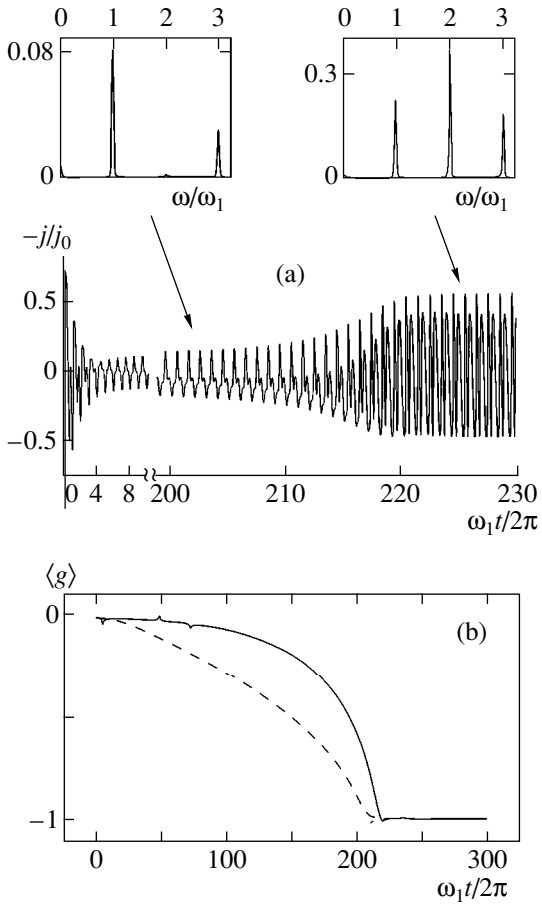

Fig. 2. Time evolutions of the (a) current, (b) average field, and (insets) current spectrum of the superlattice in a constant external harmonic field with $\tilde{w} = 0.05$, $\tilde{V}_0 = 2.4$, $\omega_1\tau = 10$.

and a significantly different spectral composition (the second harmonic which was absent before the loss of self-induced transparency becomes the largest). Quite clearly for short laser pulses (of duration less than 200 periods in this example) self-induced transparency will not have time to be destroyed and is thus observable. For long laser pulses a voltage which chaotically changes sign (and in general magnitude) from one pulse to another will appear at the ends of the superlattice. A similar effect was observed experimentally in bulk GaAs [22]. At high concentrations it is difficult to observe self-induced transparency because it is destroyed rapidly.

The slow loss of self-induced transparency accompanied by the generation of a quantized (for $\omega_1\tau \gg 1$) static field [see (38)] is described by Eq. (14) averaged over the period of the high-frequency oscillations. Thus, at low concentrations we have

$$t = -\frac{1}{4\pi}\int_{E_C^0}^{E_C} \frac{dE_C}{\sigma_C(\Omega_C, \Omega_1, \omega_1)E_C}, \tag{38}$$

where E_C^0 is the initial fluctuation of the static field. Retaining only the resonant terms in (38) we obtain

$$\frac{\omega_0^2}{\omega_1^2}J_1^2(g_1)t \approx \omega_1\Omega_C\tau^2\left(1-\frac{\Omega_C}{2\omega_1}\right) - \ln\left(1-\frac{\Omega_C}{\omega_1}\right) + \text{const}. \tag{39}$$

This dependence is plotted in Fig. 2 (dashed curve) and shows good qualitative agreement with the exact solution (solid curve).

The appearance of a static field significantly alters the pattern of energy exchange between the harmonic fields E_1 and E_2. According to (21) and (24), the absorbed field energy E_2 is given by

$$\frac{1}{2}\sigma(\omega_2; \Omega_C, \Omega_1, \omega_1)E_2^2 = \frac{1}{4}E_2^2\frac{\sigma_0}{\omega_2}\sum_{\mu=-\infty}^{\infty} J_\mu^2(g_1) \times\left[\frac{\Omega_C+\mu\omega_1+\omega_2}{1+(\Omega_C+\mu\omega_1+\omega_2)^2\tau^2} - \frac{\Omega_C+\mu\omega_1-\omega_2}{1+(\Omega_C+\mu\omega_1-\omega_2)^2\tau^2}\right]. \tag{40}$$

It is deduced from (40) that weakly dissipative energy exchange between the fields and resonant amplification of the field E_2 occurs in the following frequency ranges: (a) $\omega_2 = \mu_0\omega_1 + \Omega_C - \tau^{-1}$; (b) $\omega_2 = \mu_0\omega_1 - \Omega_C - \tau^{-1}$; (c) $\omega_2 = \Omega_C - \mu_0\omega_1 - \tau^{-1}$. In all three cases the corresponding losses are half of the value (21) because the absorption (emission) of electromagnetic field quanta as an electron moves parallel and in the opposite direction to the static field is nonequivalent. In case (a) [(b)] the electron absorbs μ_0 quanta of the field E_1 and as it propagates in the direction opposite to the field (parallel to the field) per superlattice period it emits the quantum $\hbar\omega_2$ and a relatively small fraction of energy ($\sim\hbar\tau^{-1}$) is transferred to the lattice. In case (c) the electron propagates in the opposite direction to the static field and emits the quantum $\hbar\omega_2$ and μ_0 quanta $\hbar\omega_1$. Case (b) is interesting when $\omega_2 = \Omega_C \approx 0.5(\mu_0\omega_1 - \tau^{-1})$. For odd μ_0 the additional ($\sim E_2^2$) energy acquired from the field E_1 is transferred equally to the fields E_C and E_2. For even μ_0 it follows from (26) that the initial losses of the static field are of the order $(8/5)J_{\mu_0/2}^2(g_1) - J_{\mu_0}^2(g_1)$ and may be positive. This case indicates that radiation may be generated at the Stark frequency. However, the energy for this is drawn from the harmonic field rather than the static one, especially as our electrical circuit has no dc voltage source. The role of the static field reduces to creating an initial fluctuation of the oscillation at the Stark frequency (transient process) and halving its maximum growth rate. In the absence of a strong harmonic field no amplification can take place at the Stark frequency, as we have noted. This instability channel may be important (see below) for the onset of self-oscillations with a doubled period and fractionally

quantized static field. After the superlattice has been transferred to a state having a finite dc voltage (e.g., at steps *DG* and *LN*, see Fig. 1) the regions of negative conductivity for the weak field E_2 change substantially. According to (24), they are determined by the system of equations

$$\sigma_C(\Omega_C, \Omega_1, \omega_1) = 0, \quad \frac{d\sigma_C(\Omega_C, \Omega_1, \omega_1)}{d\Omega_C} > 0,$$
$$(\Omega_C + \omega_2)\sigma_C(\Omega_C + \omega_2, \Omega_1, \omega_1) \qquad (41)$$
$$+ (\Omega_C - \omega_2)\sigma_C(\Omega_C - \omega_2, \Omega_1, \omega_1) = 0.$$

An analysis shows that these regions have no zero frequency but half-multiple frequencies do exist. At quasistatically stable steps with $\Omega_C \approx \Omega_C^0 = \mu_0\omega_1$ ($\mu_0 = 0, 1, 2, \ldots$) for $\omega_1\tau \gg 1$, $\omega_1 \gg \mathrm{Re}\,\omega_2 \gg \mathrm{Im}\,\omega_2$ from (22) and (35) we have for the plasma oscillations

$$\omega_2^2 \approx \omega_0^2 J_{\mu_0}^2(g_1) - 4i\pi\omega_2 \mathrm{Re}\sigma(\omega_2; \Omega_C^0, \Omega_1, \omega_1). \quad (42)$$

The frequency ω_2 can lie in a region of positive or negative values of the conductivity. In the first case, no oscillations can occur and the dc voltage obtained from Eq. (31) remains the same if we neglect other resonant frequencies of the system and the possibility of hard excitation of nonlinear plasma oscillations. In the second case, these oscillations will grow until steady-state values of their amplitudes and frequencies are established. (However chaos is possible.) In this case, the superlattice conductivity at frequency ω_1 and the dc voltage vary. At low electron concentrations these variations are small.

An analysis of the processes of establishment and destruction of self-induced transparency has revealed the following.

(1) The self-induced transparency of the superlattice has a dissipative nature. It occurs as a result of electron collisions for discrete values of the harmonic field amplitude in the superlattice. In the approximation of constant electron relaxation time it occurs in the same fields as dynamic localization and collapse of the electron quasienergy minibands. However, this does not imply that these effects are identical. The inaccuracy of the identification commonly made in the literature was indicated in [9].

(2) Dynamic electron localization is responsible for the establishment of absolute negative conductance which is one of the main reasons for the loss of self-induced transparency.

4. HIGH ELECTRON CONCENTRATIONS. SELF-OSCILLATIONS OF CURRENT AND VOLTAGE

At high electron concentrations in the superlattice not only the amplitudes of several current harmonics may be of the same order of magnitude (see, e.g., Fig. 2), but also the fields created by them. An important role in the formation of the field inside the superlattice is played by processes of decay and merging of oscillations which did not appear at low electron concentrations. (Spontaneous generation of a static field also occurs at low concentrations.) Processes involving harmonics and subharmonics of the external field (which is also promoted by the frequency locking effect) and processes of weakly dissipative resonant energy exchange between fields appear. In particular, parametric generation of harmonics without any initial fluctuation plays an important role [13]. This has a hybrid character and consists of ordinary harmonic generation with the frequency $n_1\omega_1$ ($n_1 = 1, 2, \ldots$) followed by its continuous degenerate parametric amplification. At the linear stage this process is described by three current terms (20) with $n = 0$, $-2n_1$ (from the first sum) and $n = n_1$ (from the second sum). A significant role is also played by coupled nondegenerate processes involving the decay and merging of oscillations of the type

$$n_1\omega_1 = \alpha_2\omega_1 \pm \alpha_3\omega_1, \qquad (43)$$

where $n_1 = \alpha_2 + \alpha_3 = 1, 2, \ldots$, and $\alpha_{2,3}$ are positive numbers commensurable with n_1. (If $E_C = 0$, then n_1 only has even values as in the previous case.) Then the following values of $\alpha_{2,3}$ and its corresponding processes are the most important.

(1) $\alpha_{2,3}$ are positive integers. These are the processes of independent parametric generation of the harmonics $\alpha_2\omega_1$ and $\alpha_3\omega_1$ noted above supplemented by nondegenerate processes of parametric amplification linking them [decay and merging (43) with different n_1]. At the linear stage these processes are described by current terms from the first sum (20) with $\omega_2 = \alpha_{2,3}\omega_1$ and $n = -n_1$ for the decay process ["+" sign in (43)] and $n = \pm n_1$ for the merging process ["–" sign in (43)]. An important factor is that initial fluctuations are also not required for these processes. As a result of their evolution two types of nonlinear oscillations are formed: (a) oscillations containing only odd harmonics; (b) oscillations containing a static field, even, and odd harmonics. The static field is either quantized [see (32)] or weak. In the first case, it occurs as a result of spontaneous oscillation, in the second case as a result of detection. For clarity we shall analyze the case when the nonlinear oscillation can be approximately described by the biharmonic field (17) with $\omega_2 = n_0\omega_1$ ($n_0 = 1, 2, \ldots$) and $g_2 \ll 1$. In this case, the steady-value of the field E_C is given by

$$\sigma_C(\Omega_C, \Omega_1, \omega_1)E_C + \Delta j_C = 0, \qquad (44)$$

where in accordance with (18), (19) the detection current is

$$\Delta j_C = \frac{\sigma_0}{n_0\omega_1}\sum_{\mu=-\infty}^{\infty}\frac{\Omega_C + \mu\omega_1}{1 + (\Omega_C + \mu\omega_1)^2\tau^2}J_\mu(g_1) \qquad (45)$$
$$\times [J_{\mu-n_0}(g_1) - J_{\mu+n_0}(g_1)]E_2\cos(n_0\delta_1 - \delta_2).$$

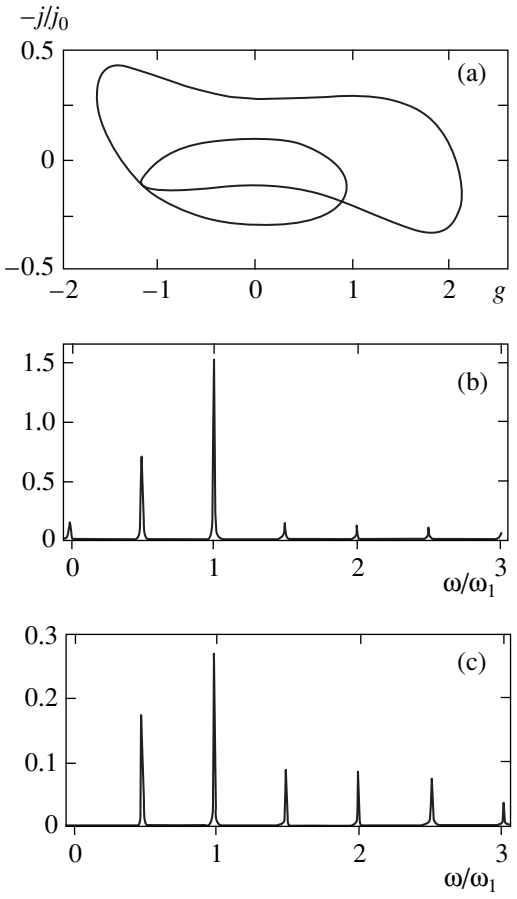

Fig. 3. Period doubling regime: (a) projection of the phase portrait on the current–superlattice voltage plane; (b) superlattice voltage spectrum; (c) current spectrum; $\omega_1\tau = 10$, $\tilde{w} = 2$, $\tilde{V}_0 = 1$.

For even n_0 the current Δj_C is symmetric with respect to E_C whereas for odd n_0 it is asymmetric. It can be seen from (21), (44), and (45) that the steady-state values of E_C only depend on the electron concentration in terms of the amplitudes and phases of the fields. For $\omega_1\tau \gg 1$, $\Omega_C \ll \omega_1$, and even n_0 we obtain from (44)

$$E_C \approx \frac{2A_{n_0}(g_1)}{n_0(\omega_1\tau)^2 J_0^2(g_1)} E_2 \cos(n_0\delta_1 - \delta_2), \qquad (46)$$

where

$$A_\nu(x) = \sum_{n=-\infty}^{\infty}{}' n^{-1} J_n(x) J_{n+\nu}(x)$$
$$= \frac{1}{x}[(2\nu-1)A_{\nu-1}(x) - 2J_0(x)J_{\nu-1}(x) + 2J_{\nu-1}(0)] - A_{\nu-2}(x), \qquad (47)$$
$$A_0 = 0, \quad A_1(x) = x^{-1}[1 - J_0^2(x)].$$

It can be seen from (46) and (47) that the detection field is not quantized and oscillates with increasing E_1.

(2) $\alpha_{2,3}$ are positive half-integers. This case only differs from case 1 by the absence of ordinary generation of the harmonics $\alpha_{2,3}\omega_1$ (parametric amplification remains) so that initial fluctuations are required. The evolution of these processes leads to the generation of period-doubled self-oscillations. The degenerate case $\alpha_2 = \alpha_3$ is either achieved for $E_C \neq 0$ or in the hard excitation regime. (The corresponding nonlinear current is $\sim E_2^3$.) However for $\mathrm{Re}\,\sigma(\omega_2; \Omega_C, \Omega_1, \omega_1) < 0$ a regime involving two-stage soft excitation of the half-integer harmonic may also occur for $E_C = 0$. At the first (linear) stage the subharmonic is excited as a result of dissipative instability. At the second (nonlinear) stage it is amplified as a result of the induced decay of $4\alpha_2$ quanta of the field E_1 into four quanta of the field E_2.

Figure 3 shows a projection of the phase portrait on the current–voltage plane and the voltage and current spectra for steady-state period-doubled self-oscillations with $\tilde{V}_0 = 1$, $\tilde{w} = 2$, and $\omega_1\tau = 10$. The most important characteristics of the curves are: the appearance of a weak static field $\Omega_C \ll \omega_1$ ($\neq n\omega_1$!), even harmonics, subharmonics $\omega_1/2$ with a relatively large amplitude and the combination harmonics $(n + 1/2)\omega_1$. In this case the period doubling process of the nonlinear oscillations begins with the nondegenerate parametric decay $2\omega_1 = 1/2\omega_1 + 3/2\omega_1$. The superlattice does not enter the region of absolute negative conductance and thus the dc voltage is caused by the nonresonant detection effect and is not quantized. We also observed period-doubled self-oscillations in which the ratio Ω_C/ω_1 had values close to 1.0; 1.5; 2.0; 2.5. The fractional and integer quantization of the static field was associated with the resonant terms in the expressions for the static (21) and rf (40) conductivities containing the factors

$$\frac{\left(\Omega_C - \sum_i \omega_i\right)\tau}{1 + \left(\Omega_C - \sum_i \omega_i\right)^2 \tau^2}$$

(ω_i are the frequencies of the field harmonics in the superlattice), and corresponding to weakly dissipative energy exchange between the fields. These terms also determine the characteristics of the current–voltage characteristics of the superlattice in a laser field, which differ qualitatively from the Shapiro steps in Josephson junctions [23]. The experimental results of [16] confirm this.

(3) $\alpha_{2,3} = (2n-1)/3$, excluding integer $\alpha_{2,3}$. Specific examples are the coupled decay $2\omega_1 = (1/3)\omega_1 + (5/3)\omega_1$ and merging $2\omega_1 + (1/3)\omega_1 = (7/3)\omega_1$ of the oscillations. As in case 2, a hard regime involving a

degenerate decay process of the type $(2n + 1)\omega_1 = 3\omega_2$ ($n = 0, 1, 2, \ldots$) is possible. The interrelated infinite set of these processes forms a period-trebled self-oscillation which contains the set of frequencies $(2n + 1)\omega_1/3$. Generally, it contains no fixed bias.

(4) $\alpha_{2,3} = 2n/3$, $n = 1, 2, \ldots$, excluding integer $\alpha_{2,3}$. Specific examples are the decay $2\omega_1 = (2/3)\omega_1 + (4/3)\omega_1$ and merging $2\omega_1 + (2/3)\omega_1 = (8/3)\omega_1$ of oscillations. The interrelated infinite set of these processes forms another period-trebled self-oscillation which contains the set of frequencies $2n\omega_1/3$, including a static field and even harmonics (in addition to the odd harmonics which are always present). Generally this oscillation is unstable. As a result of the presence of E_C, forced decay processes of the type $\omega_1 = (1/3)\omega_1 + (2/3)\omega_1$ and merging of oscillations of the type $(1/3)\omega_1 + (4/3)\omega_1 = (5/3)\omega_1$ lead to its coupling with oscillation of the previous type. As a result a complex period-trebled oscillation is established containing the harmonics $n\omega_1/3$ ($n = 0, 1, 2, \ldots$). Figure 4 shows corresponding dependences for $\tilde{V}_0 = 1.25$, $\tilde{w} = 2$, $\omega_1\tau = 10$ for a period-trebled self-oscillation of the first type. This oscillation contains the harmonics $(2n + 1)\omega_1/3$ and no static field and even harmonics occur. Its appearance is initiated by the decay $2\omega_1 = (1/3)\omega_1 + (5/3)\omega_1$. For a superlattice with a low carrier concentration (for example, for $\tilde{w} = 0.05$) the other parameters being the same, no period trebling occurs and the voltage is almost harmonic (the amplitude of the third harmonic of the internal field is only 0.15 of the fundamental one). In addition to the self-oscillation shown in Fig. 4, we also observed complex self-oscillations containing the subharmonics $n\omega_1/3$, even harmonics, and a quantized static field.

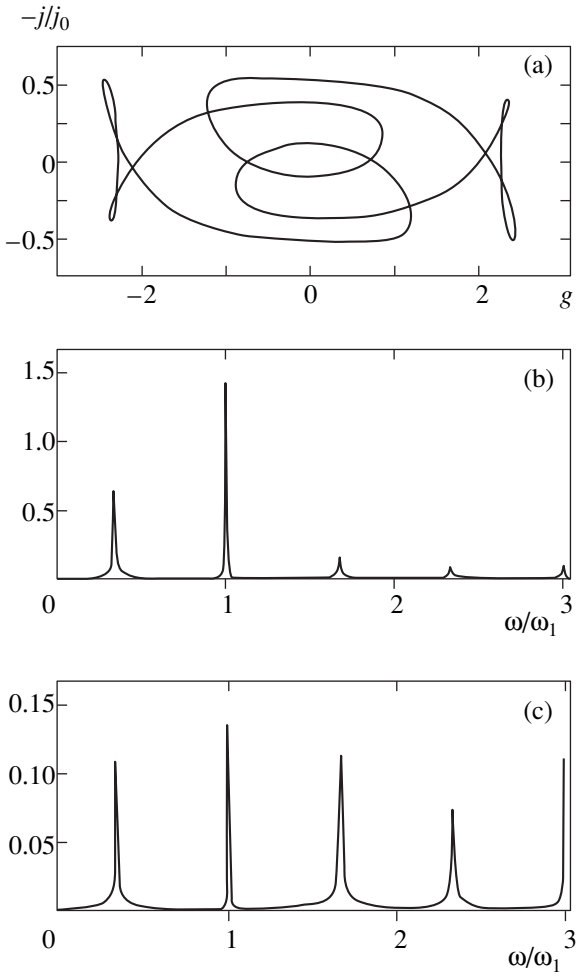

Fig. 4. Period trebling regime, as Fig. 3 but $\tilde{V}_0 = 1.25$.

As a result of the weakly dissipative resonant energy exchange between the field with the maximum growth rates oscillation decay and merging processes (43) with $\alpha_{2,3} = n_{2,3} \pm (\omega_1\tau)^{-1}$ ($n_{2,3} = 0, 1, 2, \ldots$) take place at the nonlinear stage. Since the corresponding frequencies are incommensurable, these processes lead to the establishment of quasi-periodic oscillations (beats) in the system. As we know, at the nonlinear stage the oscillation frequencies vary and the beats may be converted into periodic self-oscillations, for example, as a result of frequency locking.

The appearance of incommensurable frequencies is also a path for the evolution of chaos. Another scenario for its evolution (most commonly used for $\omega\tau = 4$, as has been shown by the numerical calculations) is the systematic period doubling of the self-oscillations (Feigenbaum scenario). Numerical investigations have revealed another path for the evolution of chaos via period trebling and narrow-band (bounded) chaos. Figure 5 shows the time evolution of the period-averaged voltage on the superlattice, the projection of the phase portrait on the current–voltage plane, and the voltage and current spectra of the stochastic self-oscillation for $\tilde{V}_0 = 2.05$, $\tilde{w} = 4.5$, $\omega_1\tau = 10$. (Stochastic self-oscillations for a superlattice with nonzero static current were considered in [12] but unfortunately, without discussing the mechanisms for their occurrence. The conclusion reached by the authors that only integer quantization of the static current occurs is not always correct.) A characteristic feature of this self-oscillation is a temporally chaotic transition between two quasi-steady states with $\langle g \rangle = \pm 1$.

To conclude this section, Fig. 6 shows the regions of existence of various types of nonlinear oscillations on the plane of the parameters $\tilde{w}$ and $\tilde{V}_0$ in a superlattice with $\tilde{w} = 10$; nonlinear oscillations with the period of the external field (unshaded unnumbered regions), subharmonic periodic self-oscillations (numbered regions which give the ratio of the periods of the self-oscillation and the external field), quasi-periodic (vertical shading) and stochastic (horizontal shading) self-oscillations. The absence of stochastic oscillations for large $\tilde{w}$ is caused by the strong screening of the external field

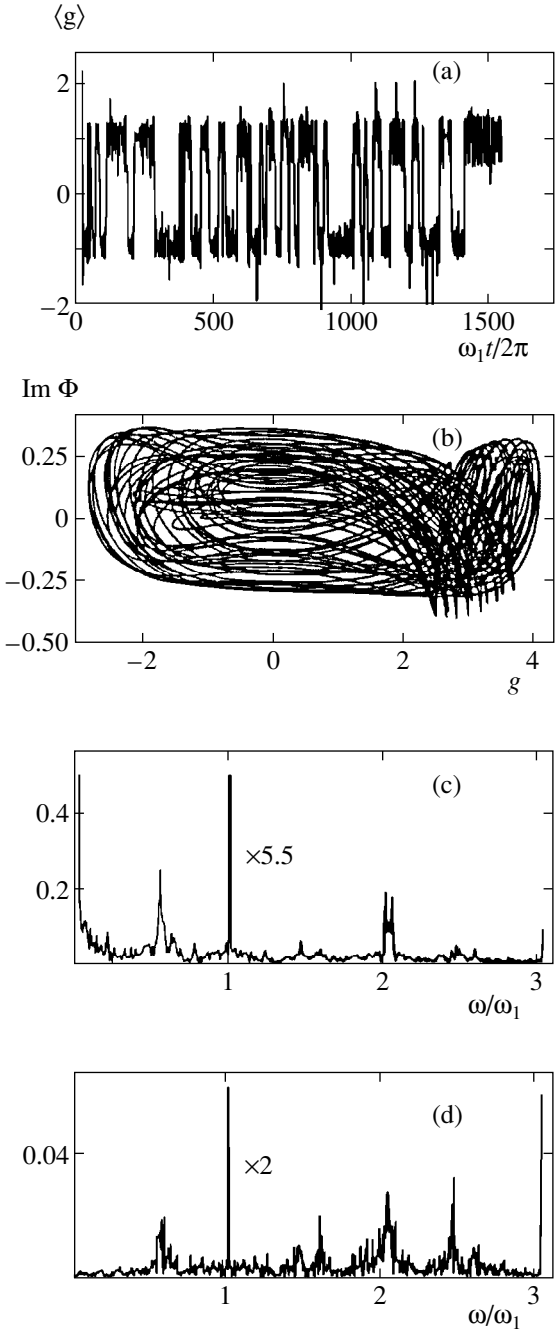

Fig. 5. Stochastic oscillations: (a) time evolution of the period-averaged external field of the superlattice voltage ($\langle g \rangle$); (b), (c), and (d) correspond to, respectively, (a), (b), and (c) in Fig. 3; $\tilde{w} = 4.5$, $\tilde{V}_0 = 2.05$.

whereas for large $\tilde{V}_0$ it is attributable to the smallness of all the current harmonics as a result of the frequent Bragg reflections of an electron from the miniband boundaries. This pattern is not complete. It has been noted that in general for given $\tilde{V}_0$ several steady states exist. In particular, studies under various initial conditions have shown that in regions of chaos it is also possible to have purely periodic self-oscillations.

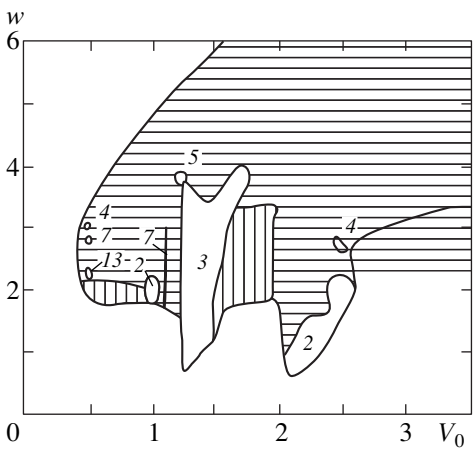

Fig. 6. Regions of existence of various types of nonlinear oscillations in a superlattice with $\omega_1\tau = 10$.

5. CONCLUSIONS

(1) The spectra of the nonlinear steady-state oscillations of the current and voltage in a superlattice are many-valued functions of the external field amplitude.

(2) Under the action of a strong rf harmonic field in a superlattice having a low electron concentration multistable steady states are formed with a zero static current and integer-quantized static field.

(3) In a superlattice with a high electron concentration periodic self-oscillations appear, containing harmonics and subharmonics of the external field and also quasi-periodic and stochastic oscillations. The period of the self-oscillations varies nonmonotonically with increasing electron concentration, frequency, and amplitude of the external field. The static field in the rf self-oscillations is either weak or fractionally quantized according to their period.

(4) States of self-induced transparency are unstable and can only be observed in a superlattice with a low electron concentration when this is exposed to pulsed action and in transient processes.

(5) In long (in particular, planar) superlattices periodic and nonperiodic dissipative structures may appear with traveling and standing domains of rf and static fields. This leads to the appearance of a diffraction pattern in the fields reflected and transmitted by the sample.

Effects similar to those analyzed may also be observed in double-quantum-well and double-quantum-dot structures.

ACKNOWLEDGMENTS

This work was supported by INTAS-RFBR (project no. 95-0615) and the Russian Interbranch Scientific-

Technical Program "Physics of Sold-State Nanostructures" (project no. 99-1129).

REFERENCES

1. L. V. Keldysh, Fiz. Tverd. Tela (Leningrad) **4**, 2265 (1962) [Sov. Phys. Solid State **4**, 1658 (1962)].
2. L. Esaki and R. Tsu, IBM J. Res. Dev. **14**, 61 (1970); P. Lebwohl and R. Tsu, J. Appl. Phys. **41**, 2664 (1970).
3. M. I. Ovsyannikov, Yu. A. Romanov, V. N. Shabanov, and R. G. Loginova, Fiz. Tekh. Poluprovodn. (Leningrad) **4**, 2225 (1970) [Sov. Phys. Semicond. **4**, 1919 (1970)].
4. Yu. A. Romanov, Opt. Spektrosk. **33**, 917 (1972); in *Multilayer Semiconductor Structures and Superlattices*, Ed. by A. M. Belyantsev and Yu. A. Romanov (Inst. Prikl. Fiz. Ross. Akad. Nauk, Gorki, 1984), p. 63.
5. A. A. Ignatov and Yu. A. Romanov, Fiz. Tverd. Tela (Leningrad) **17**, 3388 (1975) [Sov. Phys. Solid State **17**, 2216 (1975)]; A. A. Ignatov and Yu. A. Romanov, Phys. Status Solidi **73**, 327 (1976).
6. L. K. Orlov and Yu. A. Romanov, Fiz. Tverd. Tela (Leningrad) **19**, 726 (1977) [Sov. Phys. Solid State **19**, 421 (1977)].
7. M. C. Wanke, A. G. Markelz, K. Unterrainer, *et al.*, in *Physics of Semiconductors*, Ed. by N. Scheffter and R. Zimmerman (World Scientific, Singapore, 1996), p. 1791.
8. O. N. Dunlap and V. M. Kenkre, Phys. Rev. B **34**, 3625 (1986); Phys. Lett. A **127**, 438 (1988).
9. M. Holthaus, Z. Phys. B **89**, 251 (1992); Phys. Rev. Lett. **69**, 351 (1992); M. Holthaus and D. Hone, Phys. Rev. B **47**, 6499 (1993).
10. Yu. A. Romanov, L. K. Orlov, and V. P. Bovin, Fiz. Tekh. Poluprovodn. (Leningrad) **12**, 1665 (1978) [Sov. Phys. Semicond. **12**, 987 (1978)].
11. Yu. A. Romanov, Fiz. Tverd. Tela (Leningrad) **21** (3), 877 (1979) [Sov. Phys. Solid State **21**, 513 (1979)].
12. K. N. Alekseev, E. H. Cannon, J. C. McKinney, *et al.*, Phys. Rev. Lett. **80**, 2669 (1998); K. N. Alekseev, G. P. Berman, D. K. Campbell, *et al.*, Phys. Rev. B **54**, 10625 (1996).
13. Yu. A. Romanov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **23**, 617 (1980).
14. L. K. Orlov and Yu. A. Romanov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **23**, 1421 (1980); **25**, 570 (1982); **25**, 702 (1982).
15. B. J. Keay, S. Zenner, S. J. Allen, *et al.*, Phys. Rev. Lett. **75**, 4102 (1995).
16. K. Unterrainer, B. J. Keay, M. C. Wanke, *et al.*, Phys. Rev. Lett. **76**, 2973 (1996).
17. A. W. Ghosh, A. V. Kuznetsov, and J. W. Wilkins, Phys. Rev. Lett. **79**, 3494 (1997).
18. A. W. Ghosh, M. C. Wanke, S. J. Allen, and J. W. Wilkins, Appl. Phys. Lett. **74**, 2164 (1999).
19. F. G. Bass, A. A. Bulgakov, and A. P. Tetervov, *High-Frequency Properties of Semiconductors with Superlattices* (Nauka, Moscow, 1989).
20. V. V. Pavlovich and É. M. Épshteĭn, Fiz. Tekh. Poluprovodn. (Leningrad) **10**, 2001 (1976) [Sov. Phys. Semicond. **10**, 1196 (1976)].
21. N. Bloembergen, *Nonlinear Optics: a Lecture Note and Reprint Volum* (W. A. Benjamin, New York, 1965; Mir, Moscow, 1966).
22. T. Ya. Banis, I. V. Parshelyunas, and Yu. K. Pozhela, Fiz. Tekh. Poluprovodn. (Leningrad) **5**, 1990 (1971) [Sov. Phys. Semicond. **5**, 1727 (1971)].
23. B. D. Josefson, Phys. Lett. **1**, 251 (1962); S. Shapiro, Phys. Rev. Lett. **11**, 80 (1963).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 1046–1058.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1207–1221.

SOLIDS
Electronic Properties

Symmetry Transformations and Reciprocity Relations in the Theory of Multicomponent Dielectric Media

Yu. P. Emets
Institute of Electrodynamics, National Academy of Sciences of Ukraine, Kiev, 03680 Ukraine
e-mail: emets@irpen.kiev.ua
Received April 13, 2000

Abstract—Characteristics of the formation of a dipole electric field are established for multicomponent fiber composites having a regular distribution of inclusions in the matrix. It is shown that these media are characterized by symmetry transformations of the average electric fields. These symmetry transformations yield reciprocity relations for the effective parameters. Various forms of these relations are given and their physical interpretation is presented. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In the theory of inhomogeneous media an important part is played by reciprocity relations which establish a general relationship between the local characteristics of the inhomogeneities and the effective characteristics of the medium as a whole. These relations are valid for many stochastic, polycrystalline, and matrix mixtures which may have a complex structure. An explicit form of these reciprocity relations was established by the Keller theorem [1]. Keller considered an unbounded matrix medium having a regular packing of uniform cylindrical inclusions whose centers are positioned at the nodes of a rectangular lattice. In cross section this medium is two-dimensional and using the conjugate properties of harmonic functions, Keller put forward an elegant proof of the theorem for a mixture of two isotropic components. The matrix model of an inhomogeneous medium considered by Keller was first proposed by Rayleigh [2] who developed a method of calculating the electric fields in this medium and made analytic calculations of its effective parameters. The results obtained by Rayleigh confirm the reciprocity relations. They are also satisfied by inhomogeneous media having a hexagonal structure [3] and matrix media containing two-layer cylindrical inclusions [4, 5]. In all these cases the technique of calculating the effective parameters proposed by Rayleigh is used. The Keller theorem can also be applied to one-dimensional, stratified media.

Balagurov [6] showed that the reciprocity relations are valid for various forms of inclusions and arbitrary spatial configurations of these inclusions. Extremely general proofs of the Keller theorem were put forward by Fokin [7], Schulgasser [8], Mendelson [9], and other authors. In these studies the authors considered all possible real structures of randomly inhomogeneous and regular two-dimensional media with discrete or continuously varying local material characteristics.

Dykhne found that the reciprocity relations for two-component media follow from the symmetry transformations satisfied by spatially averaged physical fields (inhomogeneous conducting media in a static electric field were studied [10]). In order to derive the symmetry transformation Dykhne [10] considered a cell model of a two-phase material in which the entire volume is covered with a system of closed nonoverlapping cells in each of which the properties are constant and the cells are distributed statistically uniformly and isotropically in space. Miller [11] proposed and studied a model of an inhomogeneous material with a similar structure in the theory of elasticity. If it is additionally assumed in the cell model that the geometry of all the cells is the same for both phases, we obtain an inhomogeneous medium with a checkerboard structure. For this model the electric field inside the cells can be calculated exactly and the average characteristics of the material can be calculated [12]. It was shown that the symmetry transformation of the spatially averaged fields depends on the direction of the external field in the system; the Dykhne transformation corresponds to the case when the external field is directed along the diagonal of the squares [12]. Dykhne transformations were used to obtain various interesting results in the theory of the galvanomagnetic properties of two-dimensional two-component systems [13–15].

It should be noted that reciprocity theorems are also known in other fields of physics and mechanics: for example, in acoustics, in electrodynamics, in elasticity theory (Maxwell and Betti theorem), and in theories of electric circuits, electromechanical systems, electroacoustics, and other fields.

Reciprocity theorems are important in the theory of inhomogeneous media because, having a high degree of generality, they can be used to monitor the correctness and accuracy of the calculations of effective parameters in numerical calculations. They also sim-

1063-7761/00/9105-1046$20.00

plify the calculation of components of the material tensor.

This mainly applies to two-component materials. Multicomponent systems having more than two phases have been very little studied. These media are interesting from the point of view of the theory and possible applications since combinations of many components with different characteristics can give a composite unique properties. However, calculations of the effective parameters of multiphase media involve overcoming serious mathematical complexities. In this case, unfortunately, the classical methods of potential theory developed by Rayleigh [2] cannot be applied to regularly inhomogeneous media having a periodic structure. These difficulties are associated with the need to solve complex boundary-value conjugation problems for multiply connected regions.

In the present study we establish symmetry transformations and determine reciprocity relations for two-dimensional multicomponent dielectric media. The model of a three-component matrix medium is used as the initial model to prove the reciprocity theorem. These results are then generalized to multicomponent dielectrics. The problem is discussed with reference to dielectric media. However, the mathematical methods used and the results obtained can also be applied to study similar problems in theories of thermal conductivity, diffusion, hydrodynamics, elasticity, magnetostatics, and electrical conductivity.

2. PROPERTIES OF A DIPOLE ELECTRIC FIELD

We shall analyze a two-dimensional three-component dielectric medium with a regular periodic structure. Figure 1 shows the doubly periodic repeating element of a spatially unbounded matrix medium. The periodic parallelogram contains two types of cylindrical inclusions having permittivities ε_2 and ε_3. The matrix has the permittivity ε_1. The cylindrical inclusions are parallel, their radii generally differ, and consequently the concentrations of inclusions in the material are not the same.

It is assumed that no free charges are present in the dielectric and the electric field is described by electrostatic equations with linear coupling between the induction vector **D** and the electric field vector **E**:

$$\nabla \cdot \mathbf{D} = 0, \quad \nabla \times \mathbf{E} = 0, \quad \mathbf{D} = \varepsilon \mathbf{E}. \tag{1}$$

In the cross section perpendicular to the axes of the cylindrical inclusions the electric field is two-dimensional and the equations are the same as the Cauchy–Riemann equations. This allows us to convert to the plane of the complex variable z and introduce the complex values of the electric field

$$\begin{aligned} D(z) &= D_x - iD_y, \\ E(z) &= E_x - iE_y \quad (z = x + iy). \end{aligned} \tag{2}$$

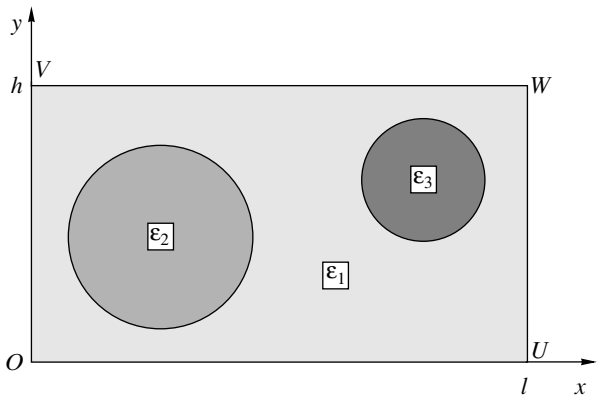

Fig. 1. Periodic cell of a dielectric material with a regular structure. The material is reinforced with cylindrical fibers of two different types.

In the isolated periodic cell the electric field is determined by dipole–dipole interactions between inclusions in the cell and between these and other inclusions in the system. For the analysis and subsequent calculations this field is conveniently expressed in the following generalized form:

$$\begin{aligned} E_1(z) &= E_0 + E_0 \exp(i\alpha_k) F_1(\Delta_{12}^{2k}, \Delta_{13}^{2k}, \Delta^k, z) \\ &\quad - \tilde{E}_0 G_1(\Delta_{12}^{2k-1}, \Delta_{13}^{2k-1}, z), \\ E_2(z) &= (1 + \Delta_{12})[E_0 + E_0 \exp(i\beta_k) F_2(\Delta_{12}^{2k}, \Delta^k, z) \\ &\quad - \bar{E}_0 G_2(\Delta_{12}^{2k-1}, z)], \\ E_3(z) &= (1 + \Delta_{13})[E_0 + E_0 \exp(i\gamma_k) F_3(\Delta_{13}^{2k}, \Delta^k, z) \\ &\quad - \bar{E}_0 G_3(\Delta_{13}^{2k-1}, z)]. \end{aligned} \tag{3}$$

Here $E_1(z)$ is the electric field in the matrix; $E_2(z)$ and $E_3(z)$ are the electric fields in inclusions having the permittivities ε_2 and ε_3, respectively; E_0 is the uniform external electric field (the bar above E_0 denotes complex conjugation); α_k, β_k, and γ_k are the angles between the x axes and lines connecting the centers of the interacting inclusions. The dimensionless parameters Δ_{12}, Δ_{13}, and Δ are given by

$$\begin{aligned} \Delta_{12} &= \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 + \varepsilon_2}, \quad \Delta_{13} = \frac{\varepsilon_1 - \varepsilon_3}{\varepsilon_1 + \varepsilon_3}, \\ \Delta &= \Delta_{12}\Delta_{13} \quad (-1 \le \Delta_{12}, \Delta_{13}, \Delta \le 1). \end{aligned} \tag{4}$$

The functions $E_0 \exp(\ldots) F_m(\ldots)$ and $\bar{E}_0 G_m(\ldots)$, $m = 1, 2, 3$ correspond to the nonuniform parts of the electric field in the components; in fact, they are infinite sums of induced dipoles; $k = 1, 2, \ldots$ is the running index in these sums. The function $E_0 \exp(\ldots) F_m(\ldots)$ combines all the dipoles whose moments depend on the direction of the external field E_0 and on the position of the inclusions in the system with which the isolated

inclusion in a particular cell interacts. It is important to note that the moments of this set of dipoles have an even dependence on the parameters Δ_{12} and Δ_{13}. All the remaining dipoles are assigned to the function $\bar{E}_0 G_m(\ldots)$. The moments of this group of dipoles are directed along the vector of the external electric field $\mathbf{E} = \bar{E}_0$ and have an odd dependence on the dimensionless parameters Δ_{12} and Δ_{13}.

Hence, the dipole electric field in this inhomogeneous medium may be represented by two components. These are selected according to the following criteria. All dipoles having an even dependence on the parameters Δ_{12} and Δ_{13} are assigned to one component and all the other dipoles are assigned to the other. The latter have an odd dependence on these parameters.

These properties of the electric field are characteristic features of the formation of dipole fields in multicomponent matrix media. They are used extensively in the following calculations. These properties of a dipole electric field can be determined explicitly by solving a model problem involving the interaction of two dielectric cylinders having arbitrary permittivities located in an external uniform electric field. An exact solution of this problem has been obtained [16, 17] and can be used to calculate the electric field in a two-dimensional three-component matrix medium having a low concentration of cylindrical inclusions [18]. The results of calculations of the field for close-packed inclusions in this system are presented in Section 4 of the present article. The analytic expressions for the electric field presented in [18] and in Section 4 are fully consistent with the generalized representation of the electric field in the form (3).

At very low concentrations of inclusions, when the radii of the cylindrical bodies in the matrix are extremely small compared with the characteristic dimensions of the periodic cell, $r_1, r_2 \ll l, h$ and the inclusions are separated by a large distance $d \gg r_1, r_2$, the relative influence of inclusions in the system can be neglected. In this medium the electric field in the periodic cell may be represented as the sum of the fields of isolated cylindrical bodies. In this case, the field inside the cylinders is uniform whereas in the outer region of the inclusions it is the sum of the external uniform field and the field of the induced linear dipoles positioned on the axes of the cylindrical bodies. We have

$$\begin{aligned} E_1(z) &= E_0 - \frac{\bar{E}_0 \Delta_{12} r_1^2}{(z-a)^2} - \frac{\bar{E}_0 \Delta_{13} r_2^2}{(z-b)^2}, \\ E_2(z) &= (1+\Delta_{12})E_0, \\ E_3(z) &= (1+\Delta_{13})E_0, \end{aligned} \tag{5}$$

where a and b are the coordinates of the centers of the inclusions on the z plane.

It can be seen that the formulas (5) are consistent with the representation of the electric field in the generalized form (3). They can be assumed to be the zeroth approximation for calculating the electric field in a medium containing interacting inclusions.

3. SYMMETRY TRANSFORMATIONS

In order to determine the effective parameters of the inhomogeneous medium, we need to calculate the spatially averaged values of the electric field. The averaging

$$\langle \ldots \rangle = \frac{1}{S}\int_S (\ldots) ds, \tag{6}$$

is performed over the area S which is naturally taken to be the area of the periodic parallelogram (see Fig. 1) which is the basic region of a medium having a doubly periodic inhomogeneity structure.

By definition, this composite consists of components with locally isotropic materials. However, on a scale much larger than the dimensions of the averaging cell the composite may acquire anisotropic properties. Anisotropy of the macroscopic permittivity occurs either as a result of the geometric shape of the cell or as a result of the layered distribution of inclusions of different types. In these cases, averaging the material equation (1) gives

$$\langle \mathbf{D} \rangle = \hat{\varepsilon}_{\text{eff}} \langle \mathbf{E} \rangle, \tag{7}$$

where the permittivity is described by the effective tensor

$$\hat{\varepsilon}_{\text{eff}} = \{\varepsilon_{\text{eff}xx}, \varepsilon_{\text{eff}yy}\}, \tag{8}$$

reduced to the principal axes.

The averaging cell can be selected such that for fixed directions of the electric field the opposite sides of the periodic parallelogram in a regular-structure composite coincide with the equipotential and field lines. This property of the periodic cell can simplify the field averaging procedure, by replacing calculation of the integrals over its area by calculation of the contour integrals along its boundary lines.

Let us assume that the external electric field is initially directed along the x-axis: $E_0 = E_{0x}$. Then, two sides of the periodic parallelogram OV and UW are located on the equipotentials and the other two OU and UW coincide with the field lines (Fig. 1). In this case, the component $\varepsilon_{\text{eff}\,xx}$ of the tensor $\hat{\varepsilon}_{\text{eff}}$ may be determined from

$$\langle D \rangle_x = \varepsilon_{\text{eff}xx} \langle E \rangle_x, \tag{9}$$

where

$$\langle E \rangle_x = \frac{1}{l}\int_0^l \text{Re}E(x)dx, \quad \langle D \rangle_x = \frac{\varepsilon_1}{h}\int_0^h \text{Re}E(iy)dy. \tag{10}$$

Integrating the expressions (3) gives

$$\begin{aligned}\langle E\rangle_x &= \langle E_1\rangle_x + \langle E_2\rangle_x + \langle E_3\rangle_x,\\ \langle D\rangle_x &= \varepsilon_1\left[\langle E_1\rangle_x + \frac{1-\Delta_{12}}{1+\Delta_{12}}\langle E_2\rangle_x + \frac{1-\Delta_{13}}{1+\Delta_{13}}\langle E_3\rangle_x\right].\end{aligned} \tag{11}$$

The last expression is written allowing for the relationships

$$\frac{\varepsilon_2}{\varepsilon_1} = \frac{1-\Delta_{12}}{1+\Delta_{12}}, \quad \frac{\varepsilon_3}{\varepsilon_1} = \frac{1-\Delta_{13}}{1+\Delta_{13}}. \tag{12}$$

In order to find the component $\varepsilon_{\text{eff}\,yy}$ of the tensor $\hat{\varepsilon}_{\text{eff}}$, the external electric field must be directed along the y-axis: $E_0 = -iE_{0y}$ (without loss of generality we can assume that $E_{0x} = E_{0y}$). Now in the periodic parallelogram the sides OU and VW coincide with the equipotentials and the sides OV and UW are positioned on the field lines. For the assumed conditions the component $\varepsilon_{\text{eff}\,yy}$ is determined from

$$\langle D\rangle_y = \varepsilon_{\text{eff}\,yy}\langle E\rangle_y. \tag{13}$$

Here the average fields are calculated using the integrals:

$$\begin{aligned}\langle E\rangle_y &= \frac{1}{h}\int_0^h \mathrm{Im}E_y(iy)dy,\\ \langle D\rangle_y &= \frac{\varepsilon_1}{l}\int_0^l \mathrm{Im}E(x)dx.\end{aligned} \tag{14}$$

Substituting Eqs. (3) into Eqs. (14), we obtain

$$\begin{aligned}\langle E\rangle_y &= \langle E_1\rangle_y + \langle E_2\rangle_y + \langle E_3\rangle_y,\\ \langle D\rangle_y &= \varepsilon_1\left[\langle E_1\rangle_y + \frac{1-\Delta_{12}}{1+\Delta_{12}}\langle E_2\rangle_y + \frac{1-\Delta_{13}}{1+\Delta_{13}}\langle E_3\rangle_y\right].\end{aligned} \tag{15}$$

It is found that the average fields given by Eqs. (11) and (15) are not independent. As will be shown below by means of specific calculations, the following relations exist between them:

$$\begin{aligned}\langle D(\Delta_{21}, \Delta_{31})\rangle_x &= \varepsilon_1\langle E(\Delta_{12}, \Delta_{13})\rangle_y,\\ \langle D(\Delta_{21}, \Delta_{31})\rangle_y &= \varepsilon_1\langle E(\Delta_{12}, \Delta_{13})\rangle_x.\end{aligned} \tag{16}$$

where in accordance with Eqs. (4) $\Delta_{21} = -\Delta_{12}$ and $\Delta_{31} = -\Delta_{13}$. Note that Eqs. (16) are valid under the condition assumed above $E_{0x} = E_{0y}$.

Using the complex representation of the electric field (2), Eqs. (16) may be expressed as a single relationship:

$$\langle D(\Delta_{21}, \Delta_{31})\rangle = i\varepsilon_1\langle \bar{E}(\Delta_{12}, \Delta_{13})\rangle. \tag{17}$$

Here we recall that the bar over the function E denotes complex conjugation. In vector form the relationships (16) are written as:

$$\langle \mathbf{D}(\Delta_{21}, \Delta_{31})\rangle = -\varepsilon_1\hat{T}\langle \mathbf{E}(\Delta_{12}, \Delta_{13})\rangle, \tag{18}$$

where $\hat{T}$ is a two-dimensional tensor having the transverse components:

$$\hat{T} = \begin{Vmatrix} 0 & 1 \\ 1 & 0 \end{Vmatrix}. \tag{19}$$

The minus sign appeared in (18) because the external field was expressed in the complex form $E_0 = E_{0x} - iE_{0y}$; the vector $\mathbf{E}_0$ corresponds to the conjugate of E_0: $\mathbf{E}_0 = \bar{E}_0$.

The equality (16) can also be expressed as follows:

$$\begin{aligned}\langle D(\Delta_{12}, \Delta_{13})\rangle_x &= \varepsilon_1\langle E(\Delta_{21}, \Delta_{31})\rangle_y,\\ \langle D(\Delta_{12}, \Delta_{13})\rangle_y &= \varepsilon_1\langle E(\Delta_{21}, \Delta_{31})\rangle_x.\end{aligned} \tag{20}$$

It is quite clear that Eqs. (16) and (20) are equivalent. From Eqs. (20) we obtain a relation equivalent to (17)

$$\langle D(\Delta_{12}, \Delta_{13})\rangle = i\varepsilon_1\langle \bar{E}(\Delta_{21}, \Delta_{31})\rangle, \tag{21}$$

and thus in vector form we have

$$\langle \mathbf{D}(\Delta_{12}, \Delta_{13})\rangle = -\varepsilon_1\hat{T}\langle \mathbf{E}(\Delta_{21}, \Delta_{31})\rangle. \tag{22}$$

The equalities (16) and other Eqs. (17)–(22) derived from them determine the symmetry transformations of the average values of the electric field in three-component matrix dielectrics having anisotropic properties of the effective parameters.

If an inhomogeneous material is on average isotropic, it is characterized by a scalar effective permittivity which is now determined from

$$\langle \mathbf{D}\rangle = \varepsilon_{\text{eff}}\langle \mathbf{E}\rangle. \tag{23}$$

In this case, the symmetry transformations have the form

$$\langle D(\Delta_{21}, \Delta_{31})\rangle_{x,y} = \varepsilon_1\langle E(\Delta_{12}, \Delta_{13})\rangle_{x,y}, \tag{24}$$

or in equivalent notation

$$\langle D(\Delta_{12}, \Delta_{13})\rangle_{x,y} = \varepsilon_1\langle E(\Delta_{21}, \Delta_{31})\rangle_{x,y}. \tag{25}$$

The relationships (24) and (25) can be expressed in the vector form:

$$\langle \mathbf{D}(\Delta_{21}, \Delta_{31})\rangle = \varepsilon_1\langle \mathbf{E}(\Delta_{12}, \Delta_{13})\rangle, \tag{26}$$

$$\langle \mathbf{D}(\Delta_{12}, \Delta_{13})\rangle = \varepsilon_1\langle \mathbf{E}(\Delta_{21}, \Delta_{31})\rangle. \tag{27}$$

The validity of the symmetry transformations (16) for these periodic media is demonstrated using the following assumptions.

(1) The electric field in a three-component matrix medium is generally represented in the form (3) where the functions $E_0\exp(i\alpha_k)F_n(\ldots)$ and $E_0G_n(\ldots)$, $m = 1, 2, 3,$

are expressed by infinite sums of dipoles where each element is given by

$$E_k(z) = \frac{p_k}{(z-a_k)^2}, \quad p_k = p_{xk} + ip_{yk}, \tag{28}$$

where p_k are the moments of the linear dipoles and a_k are the coordinates of the dipoles on the z plane. The dipole moments generally depend on the external field E_0, the parameters Δ_{12}, Δ_{13}, Δ, and geometric factors, and some also depend on the angle α_k [see, e.g., Eq. (5)].

(2) Calculation of the average values of the field can be reduced to integrating and then summing the dipole fields. In the calculations of the averages $\langle D\rangle_x$ and $\langle E\rangle_y$ the corresponding integrals (10) and (14) of the real and imaginary parts of the field for the same limits of integration only differ in respect of the signs. This can easily be established by integrating expression (28). However, after integrating the function $E_0\exp(i\alpha_k)F_n(\ldots)$ and summing the results, the value obtained in two cases has the same sign. This is attributable to the presence of the factor $\exp(i\alpha_k)$ as a result of which the values at complex-conjugate points are summed in a periodic system so that the sign of the integral becomes irrelevant.

Ultimately, by averaging the functions containing the parameter Δ to any power and the parameters Δ_{12} and Δ_{13} to even powers, we obtain the same value for the imaginary and real parts. A similar operation for functions containing the parameters Δ_{12} and Δ_{13} gives values which differ only in respect of the signs. This rule is also conserved in calculations of the average $\langle D\rangle_x$ where the factors (12) must be taken into account.

All this reasoning also applies to calculations of the averages $\langle D\rangle_y$ and $\langle E\rangle_x$.

It has thus been demonstrated that symmetry transformations (16) exist for matrix three-component media having a periodic structure. In this case, the shape of the inclusions is unimportant and their concentration in the matrix may be arbitrary. In the following section the correctness of the symmetry transformations is confirmed by specific calculations.

The relations (16)–(27) correspond to the symmetry transformations of three-component matrix media. In particular cases when the inclusions have the same characteristics $\varepsilon_3 = \varepsilon_2$ ($\Delta_{13} = \Delta_{12}$) or the permittivity of one of the inclusions is the same as the permittivity of the matrix, e.g., $\varepsilon_3 = \varepsilon_1$ ($\Delta_{13} = 0$), Eqs. (16)–(27) will determine the symmetry transformations of two-component media. The equalities (16) then have the form

$$\begin{aligned}\langle D(\Delta_{21})\rangle_x &= \varepsilon_1\langle E(\Delta_{12})\rangle_y,\\ \langle D(\Delta_{21})\rangle_y &= \varepsilon_1\langle E(\Delta_{12})\rangle_x.\end{aligned} \tag{29}$$

In complex and vector forms these may be expressed as:

$$\begin{aligned}\langle D(\Delta_{21})\rangle &= i\varepsilon_1\langle \bar{E}(\Delta_{12})\rangle,\\ \langle \mathbf{D}(\Delta_{21})\rangle &= -\varepsilon_1\hat{T}\langle \mathbf{E}(\Delta_{12})\rangle,\end{aligned} \tag{30}$$

where the tensor $\hat{T}$ was defined above, see (19).

Instead of Eqs. (20), we now have

$$\begin{aligned}\langle D(\Delta_{12})\rangle_x &= \varepsilon_1\langle E(\Delta_{21})\rangle_y,\\ \langle D(\Delta_{12})\rangle_y &= \varepsilon_1\langle E(\Delta_{21})\rangle_x,\end{aligned} \tag{31}$$

or briefly in the complex and vector representations:

$$\langle D(\Delta_{12})\rangle = i\varepsilon_1\langle \bar{E}(\Delta_{21})\rangle, \tag{32}$$

$$\langle \mathbf{D}(\Delta_{12})\rangle = -\varepsilon_1\hat{T}\langle \mathbf{E}(\Delta_{21})\rangle. \tag{33}$$

In accordance with Eqs. (26) and (27), the symmetry transformations for two-component isotropic media have the form

$$\langle \mathbf{D}(\Delta_{21})\rangle = \varepsilon_1\langle \mathbf{E}(\Delta_{12})\rangle, \tag{34}$$

$$\langle \mathbf{D}(\Delta_{12})\rangle = \varepsilon_1\langle \mathbf{E}(\Delta_{21})\rangle. \tag{35}$$

Symmetry transformations in the form (30) were obtained by Dykhne [10] in a study of statistically isotropic randomly inhomogeneous media.

4. EXACT ANALYTIC SOLUTIONS

In order to show that the symmetry transformations are satisfied explicitly, we shall give the results of calculations of various exactly solvable problems. Unfortunately, there are only a few models of inhomogeneous media which can be studied analytically. These are generally composite materials having a regular structure. Media having these properties can be studied comparatively easily. Despite various assumptions usually used in problems and assumptions made to simplify the calculations, the solutions obtained are of great importance for the theory of inhomogeneous media.

4.1. Three-Component Matrix Medium Having a Low Concentration of Unidirectional Cylindrical Fibers

Figure 2 shows a fragment of an unbounded inhomogeneous medium in the transverse cross section to the fibers. The centers of the cross sections of two different types of fibers are distributed uniformly at the nodes of a square lattice. Overall, a composite material having this structure possesses isotropic properties and its effective permittivity is determined from Eq. (23). The local electric field in this material was calculated in [18].

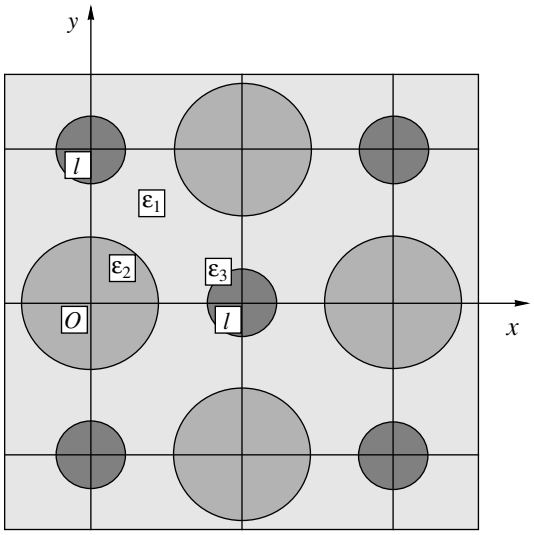

Fig. 2. Fragment of a three-component dielectric material having a square distribution of unidirectional fibers. Case of an isotropic medium.

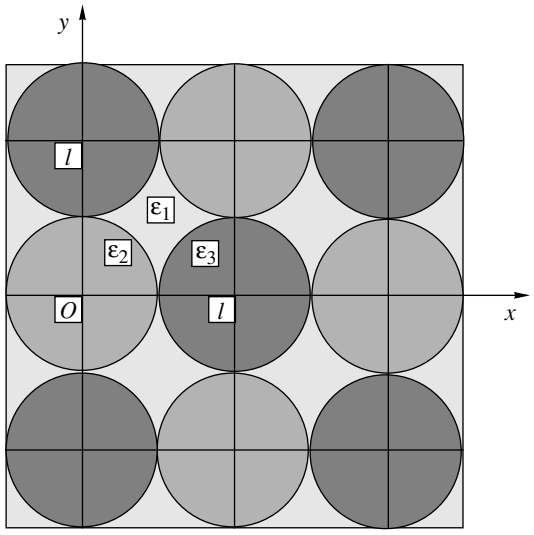

Fig. 3. Fragment of a dielectric material having close-packed fibers of two different types. Isotropic medium.

When the electric field E_0 is directed along the x-axis, $E_0 = E_{0x}$, calculations of the average fields in the composite give

$$\begin{aligned}\langle D\rangle_x &= \varepsilon_1[1 - \Delta_{12}s_1/2 - \Delta_{13}s_2/2\\ &+ \Delta_{12}^2A_1 + \Delta_{13}^2A_2 + \Delta(B_1 + B_2)],\\ \langle E\rangle_x &= 1 + \Delta_{12}s_1/2 + \Delta_{13}s_2/2\\ &+ \Delta_{12}^2A_1 + \Delta_{13}^2A_2 + \Delta(B_1 + B_2).\end{aligned} \tag{36}$$

The expressions (36) are written in terms of the relative units E/E_0 and D/ε_0E_0 where ε_0 is the electric constant. In Eqs. (36) s_1 and s_2 are the concentrations of inclusions having permittivities ε_2 and ε_3, respectively. The parameters A_n and B_n (n = 1, 2) are functions of the inclusion radii r_1 and r_2 or, which is equivalent, functions of the inclusion concentrations s_1 and s_2 ($s_n = \pi r_n^2/l^2$, l is the linear dimension of a square cell, its side). Explicit expressions for the parameters A_n and B_n are given in [18].

It can be seen that Eqs. (36) satisfy the symmetry transformations (24) of inhomogeneous isotropic media.

4.2. Three-Component Matrix Medium with Close-Packed Cylindrical Inclusions

In the plane perpendicular to the axes of the inclusions, as in the previous example, the medium is considered to be a two-dimensional inhomogeneous medium with isotropic properties. Figure 3 shows a fragment of a composite with square-packed cylindrical fibers. The electric field in this material can be calculated using methods described in [16–18]. It should be noted that for a medium having an extremely high concentration of inclusions, as in this case, the field calculations must be made with a fairly high degree of approximation, retaining a considerable number of terms in the asymptotic expansions.

Finally, for a fixed direction of the external field $E_0 = E_{0x}$ the average values of the electric field in the material are determined by the expressions (for dimensionless quantities)

$$\begin{aligned}\langle D\rangle_x &= \varepsilon_1\Bigg\{2 - \frac{(\Delta_{12} + \Delta_{13})s}{2}\\ &+ \sum_{k=1}^{\infty}\{(\Delta_{12}^{2k} + \Delta_{13}^{2k})(C_k + D_k)\\ &- (\Delta_{12}^{2k+1} + \Delta_{13}^{2k+1})(C_k + D_{k+1})\\ &+ \Delta^k[2(A_k + B_k) - (\Delta_{12} + \Delta_{13})(A_k + B_{k+1})]\}\Bigg\},\\ \langle E\rangle_x &= 2 + \frac{(\Delta_{12} + \Delta_{13})s}{2}\\ &+ \sum_{k=1}^{\infty}\{(\Delta_{12}^{2k} + \Delta_{13}^{2k})(C_k + D_k)\\ &+ (\Delta_{12}^{2k+1} + \Delta_{13}^{2k+1})(C_k + D_{k+1})\\ &+ \Delta^k[2(A_k + B_k) + (\Delta_{12} + \Delta_{13})(A_k + B_{k+1})]\}.\end{aligned} \tag{37}$$

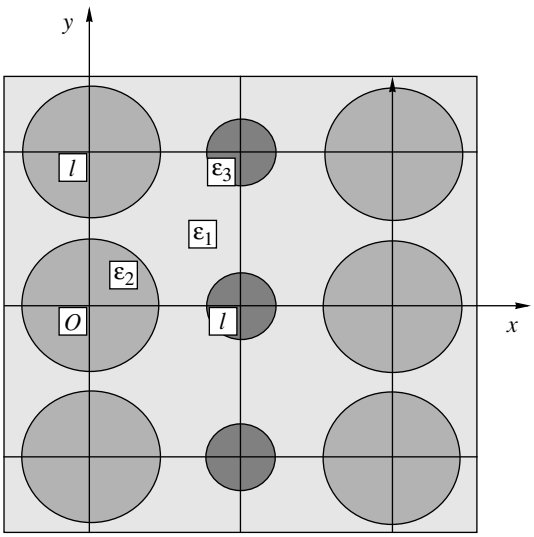

Fig. 4. Anisotropic dielectric material containing unidirectional fibers of two different types.

Here s is the total concentration of inclusions, $s = \pi/4$. The numerical coefficients of the infinite sums A_k, B_k, C_k, and D_k are given by

$$A_k = \frac{4(2k+1)^2}{(2k+1)^4 - 16k^4},$$

$$B_k = \frac{4(2k-1)^2}{(2k-1)^4 - 16k^4},$$

$$C_k = \frac{16\lambda^{2k+1}(1-\lambda^{2k+1})^2}{(1-\lambda^{2k+1})^4 + \lambda^2(1-\lambda^{2k})^4}, \qquad (38)$$

$$D_k = \frac{16\lambda^{2k+1}(1-\lambda^{2k-1})^2}{(1-\lambda^{2k})^4 + \lambda^2(1-\lambda^{2k-1})^4},$$

$$\lambda = 3 - 2\sqrt{2}.$$

It is easily confirmed that the average electric fields in the material determined by Eqs. (37) and (38) satisfy the symmetry transformation (24).

4.3. Anisotropic Three-Component Medium Containing Unidirectional Cylindrical Fibers

Figure 4 shows a fragment of an inhomogeneous medium in the transverse cross section to the fibers. The centers of two types of fibers are located at the nodes of a square lattice. Macroscopically this material has anisotropic properties. The anisotropy of the effective permittivity occurs as a result of the grouping of inclusions of each type into rows oriented along the y axis; the rows alternate periodically in the direction of the x-axis.

The tensor of the effective permittivity $\hat{\varepsilon}_{\text{eff}}$ is now determined from (7). The electric field in the medium can be calculated using the methods indicated in Section 4.2. At low concentrations of inclusions the expressions for the average fields in the material have the following form (in arbitrary units):

for $E_0 = E_{0x}$

$$\begin{aligned} \langle D\rangle_x &= \varepsilon_1[1 - \alpha(\Delta_{12}s_1 + \Delta_{13}s_2) \\ &- \Delta_{12}^2\Psi(r_1) - \Delta_{13}^2\Psi(r_2) - \Delta\Phi(r_1, r_2)], \\ \langle E\rangle_x &= 1 + \beta(\Delta_{12}s_1 + \Delta_{13}s_2) \\ &+ \Delta_{12}^2\Psi'(r_1) + \Delta_{13}^2\Psi'(r_2) + \Delta\Phi'(r_1, r_2); \end{aligned} \qquad (39)$$

for $E_0 = -iE_{0y}$

$$\begin{aligned} \langle D\rangle_y &= \varepsilon_1[1 - \beta(\Delta_{12}s_1 + \Delta_{13}s_2) \\ &+ \Delta_{12}^2\Psi'(r_1) + \Delta_{13}^2\Psi'(r_2) + \Delta\Phi'(r_1, r_2)], \\ \langle E\rangle_y &= 1 + \alpha(\Delta_{12}s_1 + \Delta_{13}s_2) \\ &- \Delta_{12}^2\Psi(r_1) - \Delta_{13}^2\Psi(r_2) - \Delta\Phi(r_1, r_2), \end{aligned} \qquad (40)$$

where α and β are numerical coefficients characterizing the geometric structure of the material ($\alpha + \beta = 1$, $\alpha = 0.7045$); s_1 and s_2 are the concentrations of inclusions (in periodic structures $s_1 = \pi r_1^2/l^2$, $s_2 = \pi r_2^2/l^2$, where r_1 and r_2 are the inclusion radii and l is the dimension of the square cell). To simplify the notation, explicit expressions for the functions Φ, Ψ and Φ', Ψ' are not given here, which does not influence the essence of this problem. Without any loss of generality we can assume that $E_{0x} = E_{0y}$.

In this case, the average values of the electric field in an inhomogeneous material (39) and (40) satisfy relations in the form (16), as can be established by direct checking, or in the equivalent forms (17)–(22) which determine the symmetry transformations of anisotropic media.

4.4. Three-Component Stratified Media

It is interesting to note that media having a one-dimensional periodic structure (stratified media) also satisfy the symmetry transformations of anisotropic media (16)–(22). In fact, let us assume that a composite material consists of periodically alternating layers of the same dimensions having permittivities ε_1, ε_2, and ε_3. The layers are oriented along the y axis. The average fields in this material are determined elementarily. We have

for $E_0 = E_{0x}$

$$\begin{aligned} \langle D\rangle_x &= \varepsilon_1 E_{0x}, \\ \langle E\rangle_x &= \frac{1}{3}\frac{3 - \Delta_{12} - \Delta_{13} - \Delta}{1 - \Delta_{12} - \Delta_{13} + \Delta}E_{0x}; \end{aligned} \qquad (41)$$

for $E_0 = -iE_{0y}$

$$\langle D\rangle_y = \frac{\varepsilon_1}{3}\frac{3+\Delta_{12}+\Delta_{13}-\Delta}{1+\Delta_{12}+\Delta_{13}+\Delta}E_{0y}, \qquad \langle E\rangle_y = E_{0y}, \tag{42}$$

where, as in previous cases, we can assume that $E_{0x} = E_{0y}$.

All the examples considered above confirm the validity of the symmetry transformations which are satisfied by the average values of the electric field in two-dimensional three-component matrix and stratified media. Naturally these symmetry transformations should also be satisfied for two-component media. In order to illustrate this, two typical models of two-dimensional two-component inhomogeneous media are considered below: a model of a two-phase material having a checkerboard structure with rectangular cells and a model of a matrix medium reinforced with parallel cylindrical fibers having the centers of their cross sections positioned at the nodes of a square lattice (Rayleigh model).

4.5. Two-Dimensional Two-Component Inhomogeneous Medium with a Doubly Periodic Inhomogeneity Structure

Figure 5 shows a fragment of the inhomogeneous medium. In this case we consider the "checkerboard" model of an inhomogeneous medium with rectangular cells. A material with this structure broadly possesses anisotropic properties. In particular cases, by changing the geometry of the cells we can obtain a doubly periodic system with square cells which corresponds to an isotropic medium or we can go over to a one-dimensional structure (stratified medium).

An inhomogeneous medium having this structure serves as a theoretical model to study composites having a critical component composition. This in fact implies that in a quasi-steady-state electric field when the permittivity has a complex value, a metal–insulator phase transition can take place in an inhomogeneous medium.

The electric field in this composite material can be calculated analytically using a calculation technique proposed in [12]. The local electric field in neighboring cells *OUUW* and *OUW'U'* (see Fig. 5) is given by the formulas

$$E_1(z) = AE_{11}(z) + BE_{12}(z), \qquad E_2(z) = AE_{21}(z) + BE_{22}(z), \tag{43}$$

where A and B are real constants. The particular solutions $E_{pq}(z)$, $p, q = 1, 2$ in Eq. (43) have the following expressions:

$$E_{11}(z) = \exp\left[\frac{i\pi}{2}\left(\frac{1}{2}-\gamma\right)\right]X(z),$$

$$E_{12} = E_{11}^{-1}(z),$$

$$E_{21}(z) = \sqrt{\frac{\varepsilon_2}{\varepsilon_1}}\exp\left[\frac{i\pi}{2}\left(\frac{1}{2}+\gamma\right)\right]X(z), \tag{44}$$

$$E_{22} = \frac{\varepsilon_2}{\varepsilon_1}E_{11}^{-1}(z),$$

where

$$X(z) = \left[\frac{1-\mathrm{cn}(2u)}{1+\mathrm{cn}(2u)}\right]^{\gamma} \quad \left(u = \frac{Kz}{l}\right), \qquad -\frac{1}{2} < \gamma = \frac{1}{\pi}\arcsin\Delta_{12} < \frac{1}{2}. \tag{45}$$

In Eq. (45) $\mathrm{cn}(2u)$ is the elliptic cosine, K and K' are complete elliptic integrals of the first kind with the moduli $k(l/h)$ and $k'(l/h)$,

$$K/K' = l/h, \tag{46}$$

l and h are the linear dimensions of the rectangular cell (see Fig. 5).

The average electric fields in the inhomogeneous material are determined by the formulas

for $E_0 = E_{0x}$ $(A = B)$

$$\langle D\rangle_x = \frac{\varepsilon_1}{h}\int_0^h \mathrm{Re}E_1(iy)dy, \qquad \langle E\rangle_x = \frac{1}{l}\int_0^l \mathrm{Re}E_1(x)dx; \tag{47}$$

for $E_0 = -iE_{0y}$ $(A = -B)$

$$\langle D\rangle_y = \frac{\varepsilon_1}{l}\int_0^l \mathrm{Im}E_1(x)dx, \qquad \langle E\rangle_y = \frac{1}{h}\int_0^h \mathrm{Im}E_1(iy)dy. \tag{48}$$

Integration in Eqs. (47) and (48) gives

$$\langle D\rangle_x = \frac{\varepsilon_1\pi A}{\sqrt{2}\sqrt{1+\Delta_{12}}}\frac{F'}{K'}, \qquad \langle E\rangle_x = \frac{\pi A}{\sqrt{2}\sqrt{1-\Delta_{12}}}\frac{F}{K}; \qquad \langle D\rangle_y = \frac{\varepsilon_1\pi A}{\sqrt{2}\sqrt{1+\Delta_{12}}}\frac{F}{K}, \qquad \langle E\rangle_y = \frac{\pi A}{\sqrt{2}\sqrt{1-\Delta_{12}}}\frac{F'}{K'}. \tag{49}$$

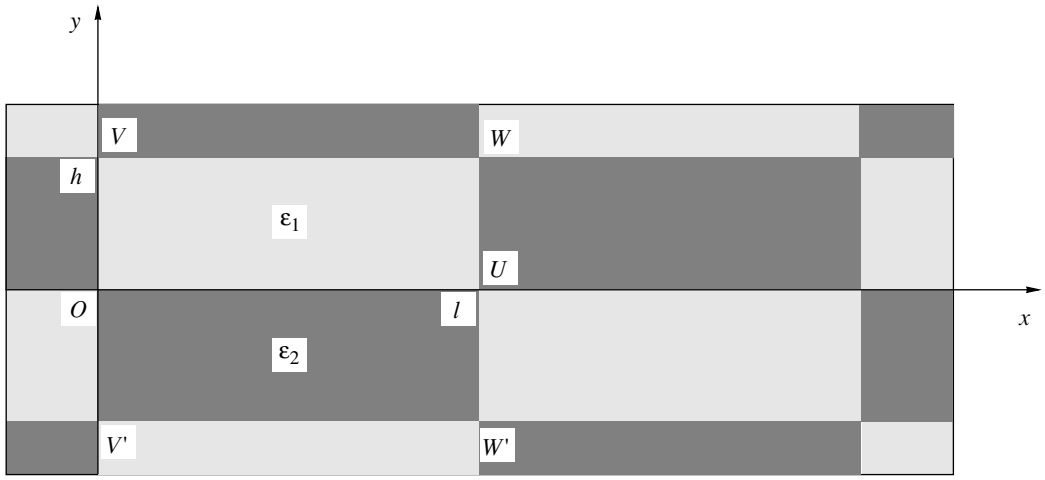

Fig. 5. Two-dimensional dielectric media having a doubly periodic distribution of rectangular cells.

Here F and F' are hypergeometric functions:

$$F \equiv F\left(\frac{1}{2}+\gamma, \frac{1}{2}-\gamma; 1; k^2\right),$$
$$F' \equiv F\left(\frac{1}{2}+\gamma, \frac{1}{2}-\gamma; 1; k'^2\right). \tag{50}$$

Bearing in mind the sign of the parameter γ determined by formula (45)

$$\gamma(\Delta_{12}) = -\gamma(\Delta_{21}), \tag{51}$$

and the symmetric properties of hypergeometric functions relative to the parameters $1/2 + \gamma$ and $1/2 - \gamma$, we can confirm that expressions (49) satisfy the symmetry transformations of the average electric fields (29)–(33) in anisotropic materials.

It is important to note that Eqs. (49) correspond to a strictly exact solution of the problem and were obtained without any approximations in the calculation process.

In the particular case when $l = h$ and the cells are square, the medium becomes broadly isotropic. We then have

$$k = k' = \frac{1}{\sqrt{2}}, \quad K = K' = \frac{1}{4\sqrt{\pi}}\Gamma^2\left(\frac{1}{4}\right),$$
$$F = F' = \sqrt{\pi}\left[\Gamma\left(\frac{3}{4}+\frac{\gamma}{2}\right)\Gamma\left(\frac{3}{4}-\frac{\gamma}{2}\right)\right]^{-1}, \tag{52}$$

where $\Gamma(\ldots)$ is a gamma function. It follows from Eqs. (49) that

$$\langle D\rangle_{x,y} = \varepsilon_1\frac{2\sqrt{2}\pi^2 A}{\sqrt{1+\Delta_{12}}}$$
$$\times\left[\Gamma^2\left(\frac{1}{4}\right)\Gamma\left(\frac{3}{4}+\frac{\gamma}{2}\right)\Gamma\left(\frac{3}{4}-\frac{\gamma}{2}\right)\right]^{-1},$$
$$\langle E\rangle_{x,y} = \frac{2\sqrt{2}\pi^2 A}{\sqrt{1-\Delta_{12}}} \tag{53}$$
$$\times\left[\Gamma^2\left(\frac{1}{4}\right)\Gamma\left(\frac{3}{4}+\frac{\gamma}{2}\right)\Gamma\left(\frac{3}{4}-\frac{\gamma}{2}\right)\right]^{-1}.$$

In this case the inhomogeneous medium is integrally isotropic and its electric fields averaged over space satisfy the symmetry transformations (34) and (35).

If we now go to the limit $h/l \longrightarrow \infty$ in Eqs. (49) assuming that l is fixed, the initial doubly periodic structure becomes singly periodic, a layered two-component medium having the same compositions (more accurately we obtain a medium containing semilayers). In this case, the parameter k tends to zero. In the limit $k = 0$ and

$$K = \pi/2, \quad F = 1. \tag{54}$$

For $h/l \longrightarrow \infty$ we can use the following asymptotic representations of the functions $F'(k)$ and $K'(k)$:

$$F'(k) \approx \Gamma(1)\left[\Gamma\left(\frac{1}{2}+\gamma\right)\Gamma\left(\frac{1}{2}+\gamma\right)\right]^{-1}\ln\left(\frac{1}{k^2}\right)$$
$$= \frac{4}{\pi}\sqrt{1-\Delta_{12}}\ln\left(\frac{1}{k}\right), \tag{55}$$
$$K'(k) \approx \ln\left(\frac{1}{k}\right).$$

Taking into account Eqs. (54) and (55), the average values of the electric field (49) in a two-component material having periodically alternating semilayers

have the form

$$\begin{aligned}\langle D\rangle_x &= \varepsilon_1\sqrt{2}A\sqrt{1-\Delta_{12}},\\ \langle E\rangle_x &= \frac{\sqrt{2}A}{\sqrt{1-\Delta_{12}}},\\ \langle D\rangle_y &= \frac{\varepsilon_1\sqrt{2}A}{\sqrt{1+\Delta_{12}}},\\ \langle E\rangle_y &= \sqrt{2}A\sqrt{1+\Delta_{12}}.\end{aligned} \tag{56}$$

In this case, the inhomogeneous medium is broadly anisotropic and Eqs. (56) correspond to symmetry transformations in the form (29)–(33).

This system is also characterized by symmetry transformations of the local fields where these transformations depend on the direction of the external field. We shall demonstrate this.

Let us assume that the external electric field $E_0 = E_{0x} - iE_{0y}$ has an arbitrary direction in the system. The constants A and B in expressions (43) are determined for a rectangular cell step from the integral equalities

$$E_{0x} = \frac{1}{l}\int_0^l \mathrm{Re}E_1(x)dx, \quad E_{0y} = \frac{1}{h}\int_0^h \mathrm{Im}E_1(y)dy. \tag{57}$$

Integrating in Eqs. (57), we obtain

$$\begin{aligned}E_{0x} &= \frac{\pi(A+B)}{2\sqrt{2}\sqrt{1-\Delta_{12}}}\frac{F}{K},\\ E_{0y} &= \frac{\pi(A-B)}{2\sqrt{2}\sqrt{1-\Delta_{12}}}\frac{F'}{K},\end{aligned} \tag{58}$$

where all the parameters are the same as in Eqs. (49). From the Eqs. (58) we find the values of the constants A and B:

$$\begin{aligned}A &= \frac{\sqrt{2}}{\pi}\sqrt{1-\Delta_{12}}\left(\frac{K}{F}\Delta_{0x} + \frac{K'}{F'}E_{0y}\right),\\ B &= \frac{\sqrt{2}}{\pi}\sqrt{1-\Delta_{12}}\left(\frac{K}{F}\Delta_{0x} - \frac{K'}{F'}E_{0y}\right).\end{aligned} \tag{59}$$

Taking into account the property of the function $X(z)$,

$$X(z) = \bar{X}(\bar{z}), \tag{60}$$

we obtain from Eqs. (43)–(45)

$$E_1^2(z) + \frac{\varepsilon_2}{\varepsilon_1}\bar{E}_2^2(\bar{z}) = 4AB, \tag{61}$$

which in expanded form becomes

$$\begin{aligned}&E_1^2(z) + \frac{\varepsilon_2}{\varepsilon_1}\bar{E}_2^2(\bar{z})\\ &= \frac{8}{\pi^2}(1-\Delta_{12})\left[\left(\frac{K}{F}E_{0x}\right)^2 - \left(\frac{K'}{F'}E_{0y}\right)^2\right].\end{aligned} \tag{62}$$

This is the symmetry transformation of the local fields. The transformation (62) establishes the relationship between the field intensities at congruent points in a doubly periodic system. It is inhomogeneous since it contains a right-hand side.

If the direction of the external field E_0 is fixed so that

$$\frac{E_{0x}}{E_{0y}} = \frac{FK'}{F'K}, \tag{63}$$

or, which amounts to the same thing,

$$\frac{E_{0x}}{E_{0y}} = \frac{h}{l}\frac{F}{F'}, \tag{64}$$

the local-field symmetry transformation (62) becomes homogeneous:

$$E_1^2(z) + \frac{\varepsilon_2}{\varepsilon_1}\bar{E}_2^2(\bar{z}) = 0. \tag{65}$$

The relationship (65) may be written as the linear equality

$$E_1(z) = \pm i\sqrt{\frac{\varepsilon_2}{\varepsilon_1}}\bar{E}_2(\bar{z}). \tag{66}$$

The presence of two signs in Eq. (66) indicates that there are in fact two directions of the external field each corresponding to one of two transformations. Thus, rotational symmetry exists.

For a square cell we have $l = h$ which, according to (52), means that $F = F'$ and instead of (64) we have

$$E_{0x} = E_{0y}. \tag{67}$$

Consequently, the homogeneous symmetry transformation (66) is also valid in a system with square cells if the external field is directed along the diagonal of the squares.

It is found that the local-field symmetry transformation (66) is also satisfied for one-dimensional (layered) structures if the external electric field has the following direction in the system:

$$\frac{E_{0x}}{E_{0y}} = \frac{1}{2\sqrt{1-\Delta_{12}}}. \tag{68}$$

This can be confirmed by going to the limit $h/l \longrightarrow \infty$ and using the asymptotic Eqs. (54) and (55).

Thus, the homogeneous local-field symmetry transformation in the form (66) is valid for singly and doubly periodic structures with equal concentrations of components. In this case, the external field has a fixed

direction which generally depends on the geometric and physical parameters of the system.

An interesting property of these systems is associated with the transformation (66), i.e., the energy of the electric field in inhomogeneous media is distributed uniformly over the phases.

In fact, Eq. (66) may be expressed as:

$$D_1(z) = \pm i\sqrt{\frac{\varepsilon_1}{\varepsilon_2}}\overline{D}_2(\bar{z}). \tag{69}$$

Multiplying the left- and right-hand sides of Eqs. (66) and (69) and performing a preliminary complex conjugation operation in (66) or (69), we obtain

$$W_1(z) = W_2(\bar{z}). \tag{70}$$

It follows from (70) that at congruent points in periodic systems the electric field energy is the same so that the overall field energy is distributed equally between the phases:

$$\langle W_1\rangle = \langle W_2\rangle. \tag{71}$$

It should be stressed that this statement only holds for fixed directions of the electric field when linear homogeneous symmetry transformations are satisfied. These conditions are satisfied for singly periodic and for doubly periodic two-component structures with equal phase concentrations.

4.6. Two-Component Media Containing Unidirectional Cylindrical Fibers

A model of an inhomogeneous medium containing cylindrical fibers of the same type positioned at the nodes of a rectangular lattice was proposed and investigated by Rayleigh when studying optical effects [2]. The electric field in the composite was calculated using methods from classical potential theory which fully take into account the relative influence of the fibers on each other. This allows the physical fields to be calculated for any concentration of inclusions in the material.

The method put forward by Rayleigh has been developed in many studies, mainly involving two-component matrix media [3–5]. In [3], for example, this method was used to calculate the electric field in a composite having a square configuration of cylindrical fibers. The asymptotic solution in the third approximation has the form

$$\begin{aligned}\langle D\rangle_x &= \varepsilon_1\{1-\Delta_{12}s-\Delta_{12}^2 s[a+(b+c)s^4]\\ &\quad -\Delta_{12}^3 bs^9+\Delta_{12}^4 bcs^{16}\},\\ \langle E\rangle_x &= 1+\Delta_{12}s-\Delta_{12}^2 s[a+(b+c)s^4]\\ &\quad +\Delta_{12}^3 bs^9+\Delta_{12}^4 bcs^{16},\end{aligned} \tag{72}$$

where $s = \pi(r/l)^2$ is the concentration of inclusions (r is the fiber radius, l is the spacing of the square lattice); a, b, and c are numerical constants ($a = 0.3058$, $b = 1.40296$, $c = 0.0134$).

This material is integrally isotropic and its average electric fields determined by formulas (72) satisfy the symmetry transformations (34) and (35).

Similar expressions for the average fields can be obtained from the solutions derived previously for the models described in Sections 4.1–4.3. In fact, if we assume in Eqs. (36), (37), and (39) that all the fibers have the same radii $r_1 = r_2$, and permittivities $\varepsilon_2 = \varepsilon_3$ ($\Delta_{12} = \Delta_{13}$), under these conditions the average values of the electric field will correspond to a matrix medium reinforced with fibers of the same type having its axes at the nodes of a square mesh. As we can easily confirm, these also satisfy the symmetry transformations (34) and (35).

Hence, all the analytic solutions given clearly show that the average fields of two-dimensional two- and three-component media satisfy symmetry transformations in accordance with the general theory. We note that because of the known analogy between the macroscopic properties of various two-dimensional structures, many average characteristics of periodic systems can naturally be extended to various classes of randomly inhomogeneous media [6, 7, 10, 11].

5. RECIPROCITY RELATIONS

Various consequences follow from the symmetry transformations which establish the general properties of the average values of the electric field in inhomogeneous media. The most important of these are associated with the reciprocity relations of the effective parameters. These relations may be obtained as follows.

We multiply the left- and right-hand sides of the second Eq. (16) and the first Eq. (20):

$$\begin{aligned}&\langle D(\Delta_{21},\Delta_{31})\rangle_y\langle D(\Delta_{12},\Delta_{13})\rangle_x\\ &= \varepsilon_1^2\langle E(\Delta_{12},\Delta_{13})\rangle_x\langle E(\Delta_{21},\Delta_{31})\rangle_y.\end{aligned} \tag{73}$$

Using the averaged material Eq. (7) we obtain

$$\varepsilon_{\mathrm{eff}xx}(\Delta_{12},\Delta_{13})\varepsilon_{\mathrm{eff}yy}(\Delta_{21},\Delta_{31}) = \varepsilon_1^2, \tag{74}$$

where in accordance with (4) we have $\Delta_{21} = -\Delta_{12}$ and $\Delta_{31} = -\Delta_{13}$.

Similarly, multiplying the left- and right-hand sides of the first Eq. (16) and the second Eq. (20) gives

$$\begin{aligned}&\langle D(\Delta_{21},\Delta_{31})\rangle_x\langle D(\Delta_{12},\Delta_{13})\rangle_y\\ &= \varepsilon_1^2\langle E(\Delta_{12},\Delta_{13})\rangle_x\langle E(\Delta_{21},\Delta_{31})\rangle_y.\end{aligned} \tag{75}$$

From this it follows that

$$\varepsilon_{\mathrm{eff}xx}(\Delta_{21},\Delta_{31})\varepsilon_{\mathrm{eff}yy}(\Delta_{12},\Delta_{13}) = \varepsilon_1^2. \tag{76}$$

These Eqs. (74) and (76) are equivalent. They determine the reciprocity relations of the effective parame-

ters of two-dimensional three-component media. These relations are extremely general and play an important role in the theory of inhomogeneous structures.

If an inhomogeneous material possesses broadly isotropic properties, instead of two Eqs. (74) and (76) we have

$$\varepsilon_{\text{eff}}(\Delta_{12}, \Delta_{13})\varepsilon_{\text{eff}}(\Delta_{21}, \Delta_{31}) = \varepsilon_1^2. \tag{77}$$

In the particular case when a two-dimensional inhomogeneous medium consists of only two components, ε_1 and ε_2, Eqs. (74) and (76) have the form

$$\begin{aligned} \varepsilon_{\text{eff}xx}(\Delta_{12})\varepsilon_{\text{eff}yy}(\Delta_{21}) &= \varepsilon_1^2, \\ \varepsilon_{\text{eff}xx}(\Delta_{21})\varepsilon_{\text{eff}yy}(\Delta_{12}) &= \varepsilon_1^2. \end{aligned} \tag{78}$$

The equality (77) becomes

$$\varepsilon_{\text{eff}}(\Delta_{12})\varepsilon_{\text{eff}}(\Delta_{21}) = \varepsilon_1^2. \tag{79}$$

Strictly, Eqs. (78) and (79) follow from the symmetry transformations for two-component media (29) and (31) and their derivation is exactly the same as the procedure used to derive (74), (76), and (77).

The reciprocity relations (78) and (79) are usually written in the form:

$$\varepsilon_{\text{eff}xx}(\varepsilon_1, \varepsilon_2)\varepsilon_{\text{eff}yy}(\varepsilon_2, \varepsilon_1) = \varepsilon_1\varepsilon_2 \tag{80}$$

for anisotropic media and

$$\varepsilon_{\text{eff}}(\varepsilon_1, \varepsilon_2)\varepsilon_{\text{eff}}(\varepsilon_2, \varepsilon_1) = \varepsilon_1\varepsilon_2 \tag{81}$$

for isotropic materials. Here the first argument implies the permittivity of the matrix and the second implies that of the inclusions. Both forms of writing (78), (79) and (80), (81) are equivalent, as was shown in [12].

Reciprocity relations in the form (80) and (81) were established by Keller [1] who adopted the Rayleigh model of an inhomogeneous medium to prove the relevant theorem. Balagurov showed [6] that the results of the theorem are valid for less stringent constraints on the shape of the inclusions and their position in the composite. Generalizations of the Keller theorem proposed by Mendelson [9], Fokin [7], Schulgasser [8], and other authors extended this theorem to a very broad class of two-component media in terms of their possible geometric structure.

The reciprocity relations can be assigned a simpler form if we introduce the tensor of the relative effective permittivity

$$\hat{\Delta}_{\text{eff}} = \{\Delta_{\text{eff}xx}, \Delta_{\text{eff}yy}\}, \tag{82}$$

whose components are related to the components of the tensor $\hat{\varepsilon}_{\text{eff}}$ by

$$\begin{aligned} \Delta_{\text{eff}xx} &= \frac{\varepsilon_1 - \varepsilon_{\text{eff}xx}}{\varepsilon_1 + \varepsilon_{\text{eff}xx}}, \\ \Delta_{\text{eff}yy} &= \frac{\varepsilon_1 - \varepsilon_{\text{eff}yy}}{\varepsilon_1 + \varepsilon_{\text{eff}yy}} \\ (-1 < \Delta_{\text{eff}xx}, \Delta_{\text{eff}yy} < 1). & \end{aligned} \tag{83}$$

Inverting Eqs. (83), we obtain

$$\frac{\varepsilon_{\text{eff}xx}}{\varepsilon_1} = \frac{1 - \Delta_{\text{eff}xx}}{1 + \Delta_{\text{eff}xx}}, \quad \frac{\varepsilon_{\text{eff}yy}}{\varepsilon_1} = \frac{1 - \Delta_{\text{eff}yy}}{1 + \Delta_{\text{eff}yy}}. \tag{84}$$

Substituting Eqs. (84) into Eqs. (78) gives

$$\begin{aligned} \Delta_{\text{eff}xx}(\Delta_{12}, \Delta_{13}) &= \Delta_{\text{eff}yy}(\Delta_{21}, \Delta_{31}), \\ \Delta_{\text{eff}xx}(\Delta_{21}, \Delta_{31}) &= -\Delta_{\text{eff}yy}(\Delta_{12}, \Delta_{13}). \end{aligned} \tag{85}$$

For isotropic media, Eq. (79) gives

$$\Delta_{\text{eff}}(\Delta_{12}, \Delta_{13}) = -\Delta_{\text{eff}}(\Delta_{21}, \Delta_{31}). \tag{86}$$

In the form (85) and (86) the reciprocity relations appear as odd functions of their arguments.

The reciprocity relations and symmetry transformations can have a simpler and clearer physical interpretation if we introduce the concept of reciprocal media [6]. With reference to three-component media this term has an extended meaning.

Let us assume that all the physical parameters of the reciprocal medium relative to the initial medium are denoted by a tilde: $\tilde{\varepsilon}_1$, $\tilde{\varepsilon}_2$, $\tilde{\varepsilon}_3$, $\tilde{\Delta}_{12}$, $\tilde{\Delta}_{13}$, $\tilde{E}(\tilde{\Delta}_{12}, \tilde{\Delta}_{13})$, $\tilde{D}(\tilde{\Delta}_{12}, \tilde{\Delta}_{13})$, …. A double tilde returns this medium to the initial one.

By definition the reciprocal medium differs from the initial one by the substitution of parameters Δ_{12}, $\Delta_{13} \longrightarrow \Delta_{21}, \Delta_{31}$ with the initial geometry conserved. It therefore follows that

$$\tilde{\Delta}_{12} = \Delta_{21}, \quad \Delta_{13} = \Delta_{31}, \tag{87}$$

which is equivalent to the equalities

$$\varepsilon_1\tilde{\varepsilon}_1 = \varepsilon_2\tilde{\varepsilon}_2, \quad \varepsilon_1\tilde{\varepsilon}_1 = \varepsilon_3\tilde{\varepsilon}_3 \tag{88}$$

Here we have two relations and three indeterminate parameters: $\tilde{\varepsilon}_1$, $\tilde{\varepsilon}_2$ and $\tilde{\varepsilon}_3$. Thus, we need an additional condition. For example, if we assume that in the initial medium and in the medium reciprocal to it the matrix has the same permittivity $\tilde{\varepsilon}_1 = \varepsilon_1$, from Eqs. (73) we obtain quite specific values of the permittivities of the additional phases in the reciprocal medium expressed in terms of the parameters of the initial medium:

$$\tilde{\varepsilon}_2 = \varepsilon_1^2/\varepsilon_2, \quad \tilde{\varepsilon}_3 = \varepsilon_1^2/\varepsilon_3. \tag{89}$$

It is found that the average energies of the reciprocal media are the same. From Eqs. (17) and (21) we obtain

$$\text{Re}(\langle D\rangle\langle\bar{E}\rangle) = \text{Re}(\langle\tilde{D}\rangle\langle\tilde{\bar{E}}\rangle), \tag{90}$$

or in vector form

$$\langle \mathbf{D} \rangle \cdot \langle \mathbf{E} \rangle = \langle \tilde{\mathbf{D}} \rangle \cdot \langle \tilde{\mathbf{E}} \rangle. \quad (91)$$

The components of the average electric field of the reciprocal media are interrelated by the following:

$$\begin{aligned} \langle \tilde{E} \rangle_x &= \mp \varepsilon_1^{-1} \varepsilon_{\text{eff}yy} \langle E \rangle_y, \\ \langle \tilde{E} \rangle_y &= \mp \varepsilon_1^{-1} \varepsilon_{\text{eff}xx} \langle E \rangle_x, \\ \langle \tilde{D} \rangle_x &= \pm \varepsilon_1 \varepsilon_{\text{eff}yy}^{-1} \langle D \rangle_y, \\ \langle \tilde{D} \rangle_y &= \pm \varepsilon_1 \varepsilon_{\text{eff}xx}^{-1} \langle D \rangle_x, \\ \tan\tilde{\phi} \tan\phi &= \varepsilon_{\text{eff}xx}/\varepsilon_{\text{eff}yy}, \end{aligned} \quad (92)$$

where ϕ and $\tilde{\phi}$ are the angles between the average values of the field E_x, E_y and $\tilde{E}_x$, $\tilde{E}_y$, respectively.

One advantage of the concept of reciprocal systems is the obvious simplification of the study. The reciprocity relations (74), (76), and (77), for example, can be written concisely as:

$$\begin{gathered} \varepsilon_{\text{eff}xx} \tilde{\varepsilon}_{\text{eff}yy} = \varepsilon_1^2, \quad \tilde{\varepsilon}_{\text{eff}xx} \varepsilon_{\text{eff}yy} = \varepsilon_1^2, \\ \varepsilon_{\text{eff}} \tilde{\varepsilon}_{\text{eff}} = \varepsilon_1^2. \end{gathered} \quad (93)$$

The reduced notation is particularly convenient for describing multicomponent media.

The form of the symmetry transformations of the average values of the electric field and the form of the reciprocity relations for the effective parameters remain the same for an arbitrary number of components, only the number of arguments of the corresponding functions changes. In fact, let us assume that a composite having this structure consists of n components and the matrix having the permittivity ε_1 contains cylindrical inclusions having the permittivities $\varepsilon_2, \varepsilon_3, \ldots, \varepsilon_n$ in a doubly periodic configuration. This medium is characterized by the parameters $\Delta_{12}, \Delta_{13}, \ldots, \Delta_{1n}$. The average values of the field of its reciprocal medium will be functions of the parameters $\Delta_{21}, \Delta_{31}, \ldots, \Delta_{n1}$. These properties follow naturally from repeating the previous scheme for calculating the electric field, in a multicomponent fiber composite the calculations of the field are also based on summing the pair interactions of cylindrical inclusions and locally conserve their dipole nature.

6. CONCLUSIONS

Studies of the characteristics of multicomponent media are attracting interest in many applications in the physics and mechanics of inhomogeneous polarizable materials. However, a systematic study of these materials is held back by various serious factors. Among these mention may be made of the large number of constituent elements, the complexity of the internal geometry, and the scarcity of exactly solvable models. In this respect, multicomponent media are considerably inferior to two-component, two-dimensional composites whose theory is based on the powerful mathematical tool of the functions of a complex variable.

Subsequently, general theorems in the theory of inhomogeneous media, i.e., reciprocity relations, symmetry transformations, and energy relations, may play an important role in obtaining specific solutions. Then, it may be possible to check the correctness of the empirical approaches and the accuracy of the approximate calculations more effectively.

REFERENCES

1. J. B. Keller, J. Math. Phys. **5**, 548 (1964).
2. Lord Rayleigh, Philos. Mag. **34**, 87 (1892).
3. W. T. Perrins, D. R. McKenzie, and R. C. McPhedran, Proc. R. Soc. London, Ser. A **369**, 207 (1979).
4. N. A. Nicorovici, R. C. McPhedran, and G. W. Milton, Proc. R. Soc. London, Ser. A **442**, 599 (1993).
5. R. D. Manteufel and N. E. Todreas, Int. J. Heat Mass Transf. **37**, 647 (1994).
6. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **81**, 665 (1981) [Sov. Phys. JETP **54**, 355 (1981)].
7. A. G. Fokin, Usp. Fiz. Nauk **166**, 1071 (1996).
8. K. Schulgasser, Int. Commun. Heat Mass Transfer **19**, 639 (1992).
9. K. S. Mendelson, J. Appl. Phys. **46**, 4740 (1975).
10. A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 110 (1970) [Sov. Phys. JETP **32**, 63 (1971)].
11. M. N. Miller, J. Math. Phys. **10**, 1988 (1969).
12. Yu. P. Emets, Zh. Éksp. Teor. Fiz. **96**, 701 (1989) [Sov. Phys. JETP **69**, 397 (1989)].
13. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **85**, 568 (1983) [Sov. Phys. JETP **58**, 331 (1983)].
14. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **108**, 2202 (1995) [JETP **81**, 1200 (1995)].
15. Yu. P. Emets, *Electrical Properties of Composites with Regular Structure* (Naukova Dumka, Kiev, 1986).
16. Yu. P. Emets and Yu. V. Obnosov, Prikl. Mekh. Tekh. Fiz., No. 1, 21 (1990).
17. Yu. P. Emets and Yu. P. Onofrichuk, IEEE Trans. Dielectr. Electr. Insul. **3**, 87 (1996).
18. Yu. P. Emets, Zh. Éksp. Teor. Fiz. **114**, 1121 (1998) [JETP **87**, 612 (1998)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 1059–1065.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1222–1229.

SOLIDS
Electronic Properties

Generalized Model of Recombination in Inhomogeneous Semiconductor Structures

S. V. Bulyarskiĭ * and N. S. Grushko
Ulyanovsk State University, Ulyanovsk, 432700 Russia
**e-mail: bsv@ulsu.ru*
Received May 22, 2000

Abstract—A generalized expression is obtained for the recombination velocity in structures with spatially separated electrons and holes. One stage of this process is tunneling. Under certain assumptions the general model yields the Shockley–Read recombination model and the tunneling recombination model. It is shown that an induced recombination effect may be observed in tunnel-coupled regions. An expression is obtained for the current–voltage characteristic of a surface-barrier diode in which tunneling recombination takes place. The theoretical results are compared with an experiment carried out using surface-barrier structures formed on gallium arsenide and thin As_2Se_3 films. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In addition to perfect crystal semiconductors, inhomogeneous materials in which the electrons and holes may be spatially separated and recombination takes place by induced tunneling to overcome the potential barriers are also being studied and used. These materials particularly include compensated semiconductors [1] in which fluctuations of the band potential occur as a result of an uncorrelated impurity distribution. As the degree of compensation tends to unity, the amplitude of the fluctuations is comparable with the band gap. Fluctuations may also occur as a result of variations in the composition of the solid solutions. This category may include a broad class of glassy and amorphous materials [2]. Quantum-well structures are playing an increasing role in electronics. For example, these structures have been used to fabricate highly efficient light-emitting devices based on wide-gap III–V compounds [3, 4]. A spatially nonuniform distribution of free carriers and recombination centers also forms in these structures which may lead to new physical effects.

Recombination in spatially inhomogeneous structures has been investigated in various studies [5–7]. However, the formulas for the recombination velocity in these studies have a particular character expressed for the current–voltage characteristics and are not analyzed. At the same time, the sections of the current–voltage characteristic attributable to recombination in the space charge region carry useful information on the properties of recombination centers [8–11] which is not generally utilized. In the present study we obtain a generalized expression for the recombination velocity in structures with tunnel-coupled regions and for the current–voltage characteristic of these structures. We show that the Shockley–Read model [12] emerges as a particular case from the results of the present study. In addition, methods of obtaining the parameters of recombination centers are obtained.

2. THEORETICAL ANALYSIS

The recombination scheme is shown schematically in Fig. 1. According to this, the semiconductor has two regions which for various reasons contain different concentrations of free carriers and localized states which may serve as recombination centers. These regions are separated by a thin tunnel-transparent layer. Each region has recombination centers whose energies are distributed in accordance with some law which is generally unknown. Carrier recombination may take place independently in each region. We shall discuss the properties of the model in greater detail. We shall consider a quasi-equilibrium steady-state problem in which both free and bound carriers have their own steady-state concentrations at each point in space. For various reasons, primarily

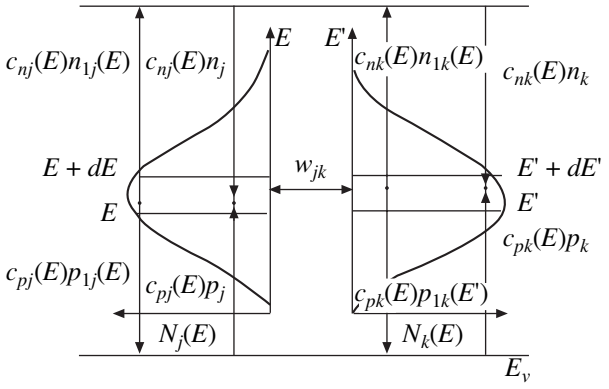

Fig. 1. Diagram of electron transitions in the generalized recombination model.

1063-7761/00/9105-1059$20.00

because of the spatially nonuniform distribution of the electric potential, these concentrations have different values in each of the coupled regions. However, since quasi-equilibrium is established in the system (since injection and generation take place), the free carriers of each type form a common subsystem. The electron and hole concentrations generally differ. The change in their concentrations is made up of the changes in concentrations in all the coupled regions. The energy distribution of the traps is determined by the physical characteristics of the semiconductor and the conditions of formation of the structure. In neighboring coupled regions these may differ.

In order to calculate the recombination velocity, we give the electron and hole recombination velocity bearing in mind, as has been noted, that this is determined by the recombination velocities in all the coupled regions:

$$\begin{aligned}\frac{\partial n}{\partial t} = \int_E \{&-c_{nj}(E)n_j N_j(E)[1-f_j(E)] \\ &- c_{nk}(E)n_k N_k(E)[1-f_k(E)] \\ &+ c_{nj}(E)n_{1j}(E)N_j(E)f_j(E) \\ &+ c_{nk}(E)n_{1k}(E)N_k(E)f_k(E)\}dE,\end{aligned} \tag{1}$$

$$\begin{aligned}\frac{\partial p}{\partial t} = \int_E \{&-c_{pj}(E)p_j N_j(E)f_j(E) \\ &- c_{pk}(E)p_k N_k(E)f_k(E) \\ &+ c_{pj}(E)p_{1j}(E)N_j(E)[1-f_j(E)] \\ &+ c_{pk}(E)p_{1k}(E)N_k(E)[1-f_k(E)]\}dE,\end{aligned} \tag{2}$$

where $c_{nj,k}(E)$ ($c_{pj,k}(E)$) is the coefficient of electron (hole) capture by localized states in the range between E and $E + dE$ in regions j and k: $n_{j,k}$ ($p_{j,k}$) is the density of the electron concentration at the bottom of the conduction band (at the corresponding percolation level) or the hole concentration (at the top of the valence band or at the corresponding percolation level); $n_{1i,1k}(E) = N_c\exp[-(E_c - E)/kT]$ is a parameter characterizing the rate of electron emission; $p_{1i,1k}(E) = N_v\exp[-(E - E_v)/kT]$ is a parameter characterizing the rate of hole emission; E_c is the energy of the bottom of the conduction band (the corresponding percolation level); E_v is the energy of the top of the valence band (the corresponding percolation level); $N_{j,k}(E)$ are the energy density distributions of the localized states in the jth and kth regions; $f_{j,k}(E)$ is the probability of electron occupancy of the localized states.

In the steady state the recombination velocities of the electrons and holes are the same. Equating (1) and (2), we find the relationship between the electron occupancy functions of the traps in regions j and k:

$$\begin{aligned}f_j(E) = \{&[c_{pj}(E)p_{1j}(E) + c_{nj}(E)n_j]N_j(E) \\ &+ c_{pk}(E)p_{1k}(E)N_k(E) - [t_{nk}(E) + t_{pk}(E)] \\ &\times N_k(E)f_k(E)\}\{[t_{nj}(E) + t_{pj}(E)]N_j(E)\}^{-1},\end{aligned} \tag{3}$$

where $t_{nj,k}(E) = c_{nj,k}(E)[n_{j,k} + n_{1j,1k}(E)]$; $t_{pj,k}(E) = c_{pj,k}(E)[p_{j,k} + p_{1j,1k}(E)]$.

In order to find each distribution function separately, we need another equation. We obtain this equation by applying the steady-state condition for the process to the rate of change of the electron concentration at the traps in one of the regions. This rate may be expressed in the following form:

$$\begin{aligned}\frac{\partial n_{tj}}{\partial t} = &\int_E c_{nj}(E)n_j N_j(E) + c_{pj}(E)p_{1j}N_j(E) \\ &- [t_{nj}(E) + t_{pj}(E)]N_j(E)f_j(E)dE \\ &- \iint_{E\,E'} w_{LR}(E,E')N_j(E)f_j(E)N_k(E') \\ &\times [1-f_k(E')]dEdE' + \iint_{E\,E'} w_{RL}(E,E')N_j(E) \\ &\times [1-f_j(E)]N_k(E')f_k(E')dEdE'.\end{aligned} \tag{4}$$

In accordance with the usual algorithm for calculating tunneling transitions [13], we shall assume that tunneling takes place without any change in energy. We shall assume that the tunneling probability depends only on the overlap integral [1, 2]:

$$w(E) = w_{LR}(E) = w_{RL}(E) = \nu\exp(-2r/a), \tag{5}$$

where ν is the frequency of attempts to overcome the potential barrier, which is equal to the characteristic phonon frequency; $a = \hbar/2mE$ is the localization radius [1]; r is the average hopping length which is equal to the average distance between traps which in turn is determined by their concentration: $r = 1/\sqrt[3]{N}$. In addition, we introduce the notation

$$N_{k,j} = \int_E N_{k,j}(E)dE.$$

We equate to zero the rate of change in the electron concentration at the trap (4) and obtain another expression linking the occupancy functions of the traps in different regions:

$$\begin{aligned}f_j(E) = &[c_{nj}(E)n_j N_j(E) + c_{pj}p_{1j}(E) \\ &+ w(E)N_j N_k(E)f_k(E) \\ &\times \{[t_{nj}(E) + t_{pj}(E)]N_j(E) + w(E)N_j(E)N_k\}^{-1}.\end{aligned} \tag{6}$$

By jointly solving Eqs. (3) and (6), we obtain a relationship between the occupancy functions and the trap parameters and free carrier concentrations in neighboring coupled regions of space:

$$f_j(E)N_j(E) = \frac{C_{pnj}T_{pnk} + w(E)N_j(C_{pnj} + C_{pnk})}{T_{pnk}T_{pnj} + w(E)N_kT_{pnk} + w(E)N_jT_{pnj}}, \quad (7)$$

$$f_j(E)N_k(E) = \frac{C_{pnk}T_{pnj} + w(E)N_k(C_{pnj} + C_{pnk})}{T_{pnk}T_{pnj} + w(E)N_kT_{pnk} + w(E)N_jT_{pnj}}, \quad (8)$$

where $C_{pnj,k} = [c_{nj,k}(E)n_{j,k} + c_{pj,k}(E)p_{1j,1k}]N_j(E)$; $T_{pnj,k} = t_{pnj,k}(E) + t_{pnj,k}(E)$.

On analyzing formulas (7) and (8), we can note that the degree of occupancy of the traps depends on their parameters and on the concentration of free carriers in the two coupled regions. We substitute (7), (8) into (1) or (2) and obtain a formula for the recombination velocity in the two coupled regions:

$$R = R_j + R_k + R_{jk}, \quad (9)$$

where

$$R_j = \int_E \{c_{nj}(E)c_{pj}(E)(n_i^2 - p_jn_j) \times [T_{pnk} + w(E)N_j]N_j(E)\} \times [T_{pnk}T_{pnj} + w(E)N_kT_{pnk} + w(E)N_jT_{pnj}]^{-1}dE,$$

$$R_k = \int_E \{c_{nk}(E)c_{pk}(E)(n_i^2 - p_kn_k) \times [T_{pnj} + w(E)N_j]N_j(E)\} \times [T_{pnk}T_{pnj} + w(E)N_kT_{pnk} + w(E)N_jT_{pnj}]^{-1}dE,$$

$$R_{jk} = \int_E \{w(E)[N_jN_k(E)D_{jk} + N_kN_j(E)D_{kj}]\} \times [T_{pnk}T_{pnj} + w(E)N_kT_{pnk} + w(E)N_jT_{pnj}]^{-1}dE,$$

$$D_{jk} = t_{nj}(E)c_{pk}(E)p_{1k}(E) - t_{pj}(E)c_{nk}(E)n_k,$$

$$D_{kj} = t_{nk}(E)c_{pj}(E)p_{1j}(E) - t_{pk}(E)c_{nj}(E)n_j,$$

Like the occupancy function, the recombination velocity is determined by all the parameters of the coupled system. Below we consider some particular cases of the generalized recombination model.

3. SHOCKLEY–READ MODEL

This well-known model is obtained from Eq. (9) if we assume that first, recombination takes place via discrete energy levels, i.e.,

$$N_{j,k}(E) = N_{tj,k}\delta(E - E_{tj,k}),$$

where $E_{tj,k}$ are the activation energies of the traps, $N_{tj,k}$ are their concentrations, and second, the coefficients of capture by the traps in one region are zero, i.e., none exist. Applying these conditions ($c_{nk} = c_{pk} = 0$), we obtain from (9)

$$R = \frac{c_{nj}c_{pj}(p_jn_j - n_j^2)N_{tj}}{c_{nj}(n_{1j} + n_j) + c_{pj}(p_{1j} + p_j)}. \quad (10)$$

Formula (10) is exactly the same as the expression for the recombination velocity via a trap in the jth region of the semiconductor structure, which is the main result of the study [12]. (Here we assume that $p_1n_1 = n_i^2$). For this case the formula for the current–voltage characteristic of a forward-biased p–n junction in the recombination region was obtained and analyzed in [8–11]. For the case of several recombination centers the resultant current is the sum of the recombination currents through each recombination center:

$$I_r = \sum_{m=1}^{g} \frac{qSw(U)c_{nm}c_{pm}n_i^2\left[\exp\left(\frac{qU}{2kT}\right) - 1\right]N_{tm}}{2n_i\sqrt{c_{nm}c_{pm}}\exp\left(\frac{qU}{2kT}\right) + c_{nm}n_{1m} + c_{pm}p_{1m}} \times \frac{2kT}{q(V_d - U)}, \quad (11)$$

where w is the width of the space charge region, g is the number of doubly charged recombination centers which participate in the recombination process at the same time.

An indication of the presence of complex recombination processes involving several recombination levels may be the nonmonotonic behavior of the differential slope coefficient of the current–voltage characteristic which is defined as

$$\beta = \frac{q}{kT}\left(\frac{d\ln j_r}{dU}\right)^{-1} = \frac{qj_r}{kT}\left(\frac{dj_r}{dU}\right)^{-1}. \quad (12)$$

As the voltage across the p–n junction varies, β varies between 1 and 2, having a value of 1 when the voltage is low. As the voltage increases, the value of β tends toward 2. The derivative

$$\frac{d\beta}{dU} = \beta^2\frac{q}{2kT} \times \frac{n_i\sqrt{c_nc_p}(c_nn_1 + c_pp_1)\exp\left(\frac{qU}{2kT}\right)}{\left[2n_i\sqrt{c_nc_p}\exp\left(\frac{qU}{2kT}\right) + c_nn_1 + c_pp_1\right]^2} \quad (13)$$

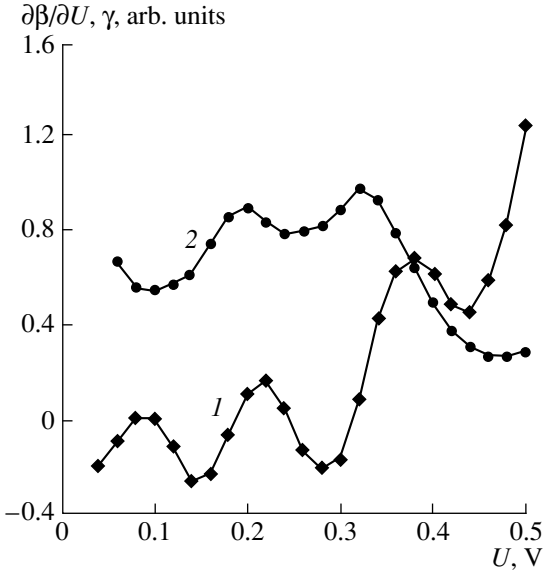

Fig. 2. Differential coefficients of the current–voltage characteristics of silicon-based *p–n* junctions: (*1*) $\partial\beta/\partial U$; (*2*) γ.

has extrema. The maxima on the curve $d\beta/dU = f(U)$ are achieved at voltages U_0 which can be used to find the deep-level activation energy:

$$E_t = \frac{E_g - qU_0}{2} + \delta, \quad \delta = \frac{kT}{2}\ln\left[\frac{1}{4}\frac{c_n}{c_p}\frac{N_c}{N_v}\right]. \qquad (14)$$

Similar information may be obtained using a different differential coefficient:

$$\gamma = \frac{\partial R_{np}}{\partial U}\frac{2kT}{q}\frac{1}{R_{np}}, \qquad (15)$$

where R_{np} is given by

$$R_{np}(U) = \sum_k R_{npk}$$

$$= \sum_m \frac{c_{nm}c_{pm}n_i N_{tm}\left[\exp\left(\frac{qU}{2kT}\right) + 1\right]}{2n_i\sqrt{c_{nm}c_{pm}}\exp\left(\frac{qU}{2kT}\right) + n_{1m}c_{nm} + p_{1m}c_{pm}}.$$

Under certain assumptions we have

$$\frac{d\gamma}{dU} = -\frac{q}{2kT}\frac{1}{R_{np}}\sum_m \frac{R_{npm}\exp\left(\frac{q(U-U_{0m})}{2kT}\right)}{\left[\exp\left(\frac{q(U-U_{0m})}{2kT}\right) + 1\right]^2}. \qquad (16)$$

A function of the type (16) is the sum of bell-shaped humps having minima at the points U_{0m}. The amplitude of each hump depends on the contribution of the specific deep level to the total recombination current. Assuming that the deep level lies above the center of the band gap (i.e., $c_{pm}p_{1m} \ll c_{nm}n_{1m}$), we obtain

$$E_{tnm} = \frac{E_g - qU_{0m}}{2} + \delta_m, \quad \delta_m = \frac{kT}{2}\ln\left(\frac{1}{4}\frac{c_n}{c_p}\frac{N_c}{N_v}\right).$$

The last value may be considered to be a systematic error arising from a lack of knowledge of the capture coefficients.

The differential coefficients can easily be determined experimentally. Figure 2 gives these values measured for silicon *p–n* junctions. They exhibit well-defined extrema which can be used to determine the activation energies of the recombination centers. The results of measurements using the current–voltage characteristic agree with independent measurements made by the thermally stimulated capacitance method. In order to determine the ratio of the capture coefficients we need to make additional temperature measurements.

4. INDUCED RECOMBINATION

The phenomenon of induced recombination involves a change in the degree of occupancy of the traps in one region under the influence of recombination fluxes taking place via traps in another region coupled to the first by means of tunneling processes. Thus, as a result of tunnel coupling, traps in the first region transfer charge to traps in the second region and conversely. This changes the recombination fluxes in both regions. Since this phenomenon is accompanied by charge transfer of levels, it may be called tunneling charge transfer. This interesting new phenomenon is also described by formula (9). However, it can be demonstrated more conveniently by analyzing the expressions for the trap occupancy probability (7) and (8). We shall assume that no tunneling takes place between regions j and k ($w \longrightarrow 0$). In addition, we shall confine ourselves to the case of a single discrete level in the band gap and then from (7) and (8) we obtain the occupancy probability

$$f_{j,k} = \frac{c_{nj,k}n_{j,k} + c_{pj,k}p_{1j,1k}}{t_{nj,k} + t_{pj,k}}N_{j,k}. \qquad (17)$$

We shall normalize the functions (7) and (8) to $f_{j,k}$ (17) and analyze the result. Figure 3 gives a modeling variant when $c_{nk} < c_{pk}$. For the tunneling probabilities $w < 10^{-12}$ cm^3 s^{-1} the normalized probability functions for trap occupancy in both regions are unity. This is the case of weak coupling and recombination takes place independently in both regions. At high tunneling probabilities the values of the normalized functions begin to differ from unity. The traps influence each other. At low injection levels we find $f_j > 1$. Then this value becomes less than one and $f_k > 1$. We note that at high injection levels tunnel coupling is suppressed. This is expressed as a tendency of both normalized functions to go to unity. Similar behavior is observed when $c_{nk} > c_{pk}$. In

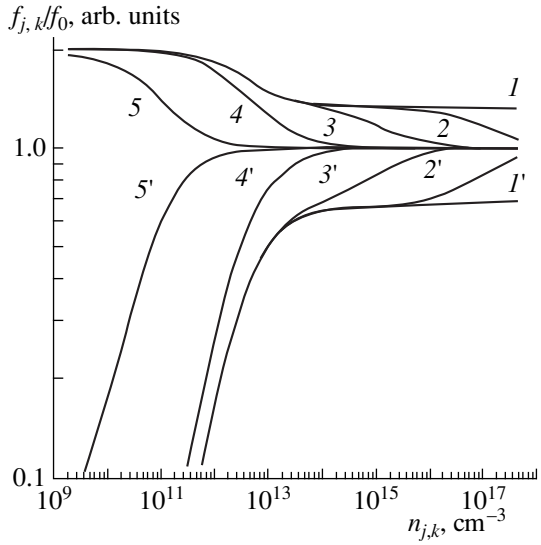

Fig. 3. Charge transfer between traps associated with induced recombination phenomena. Curves *1*–*5* give the functions f_k, and curves *1*'–*5*' give f_j. The modeling was performed using formula (9) for the following values of the tunneling probabilities (in cm^3 s^{-1}): (*1*) 10^{-2}, (*2*) 10^{-4}, (*3*) 10^{-6}, (*5*) 10^{-10}.

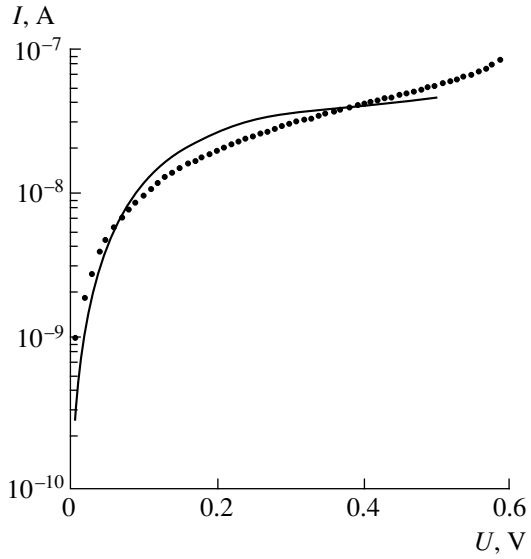

Fig. 4. Tunneling recombination in GaAs structures; the circles give the experimental results and the solid curves give the modeling using Eq. (17).

this case, however, f_i is greater than one over the entire range of injection levels.

5. TUNNELING RECOMBINATION

As we have noted, in amorphous and glassy semiconductors the electrons and holes are spatially separated. Recombination can only take place when one of the stages of the process is tunneling, hence the name of the model. This model also follows from Eq. (9). For this we need to assume that in one region traps are only exchanged with the electron percolation level and in the other region, with the hole percolation level. We shall assume that $c_{pj} = 0$ and $c_{nk} = 0$, and then we obtain from (9)

$$\begin{aligned} R = \int_E & [w(E)N_jN_k(E)c_{nj}(E)c_{pk}(E)p_{1k}(E) \\ & - w(E)N_kN_j(E)c_{nj}(E)n_jc_{pk}(E)p_k] \\ & \times [t_{pk}(E)t_{nj}(E) + w(E)N_k(E)t_{pk}(E) \\ & + w(E)N_jt_{nj}(E)]^{-1}dE, \end{aligned} \tag{18}$$

which agrees with the conclusions reached in [7]. This model has an important property. This is that the recombination velocity saturates at low probabilities of tunneling between the coupled regions. These current–voltage characteristics are also observed in compensated (Fig. 4) and glassy (Fig. 5) semiconductors.

In order for recombination to be observed in metal–semiconductor contacts, inversion conductivity must occur in the layer adjacent to the contact. In this case, the potential barrier at the interface with the metal should be more than half the mobility gap. We shall assume that a certain constant bias voltage U is applied to the contact. The electrons and holes have a quasi-equilibrium distribution in the space charge region. Their concentration at the percolation levels may be obtained

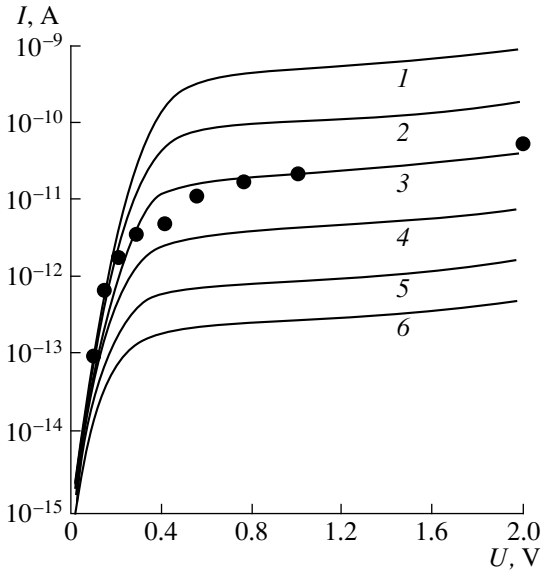

Fig. 5. Current–voltage characteristic of a metal–As_2Se_3 contact: the circles give the experimental results and the curves give the calculations using Eq. (17) for tunneling probabilities (in cm^3 s^{-1}): 9.4×10^{-20} (*1*), 1.9×10^{-20} (*2*), 4×10^{-21} (*3*), 7.7×10^{-22} (*4*), 1.7×10^{-22} (*5*), and 2×10^{-23} (*6*).

by multiplying the concentration of the appropriate carriers at the boundaries of this region with the Boltzmann factor, which allows for the change in the potential $\phi(x)$ near the contact. In the one-dimensional case we obtain

$$n(x) = n(0)\exp\left[-\frac{e(U_k - U) - \phi(x)}{kT}\right], \qquad p(x) = p(d)\exp\left[-\frac{\phi(x)}{kT}\right], \tag{19}$$

where $n(0)$ is the concentration of free electrons at the interface with the metal $n(0) = N_c\exp(-\varphi_{bn}/kT)$; $p(d)$ is the concentration of free holes at the boundary of the space charge region in the bulk of the semiconductor (is the same as the bulk concentration of free holes p_0); d is the width of the space charge region of the metal–semiconductor contact; U_k is the contact potential difference; $eU_k = \varphi_{bp} - E_f$, where φ_{bp} is the height of the potential barrier for holes at the metal–semiconductor contact, and E_f is the Fermi level energy. The instantaneous value of the potential energy $\phi(x)$ is measured from the corresponding percolation level in the bulk of the semiconductor. The recombination current density (9) can be determined by integrating the recombination velocity (9) over the space charge region taking into account the expression for the free carrier concentration (19).

The dependence of the recombination velocity on the coordinate is a bell-shaped function having almost exponential wings. This means that it can be calculated using the method of steepest descents [8]:

$$j_r = e\int_0^d R(x)dx \approx 2R_{\max}\left(\frac{1}{kT}\frac{d\phi(x)}{dx}\right)^{-1} = \frac{2kTR_{\max}}{eF}, \tag{20}$$

where $R_{\max}$ is the maximum recombination velocity; F is the average electric field strength in the contact.

In order to integrate (20) we need to find the instantaneous value of the potential energy corresponding to the maximum recombination velocity. We can change the sequence of integration over the coordinate and energy. As a result it is easier to search for an extremum. Having performed a standard procedure [8], after various transformations we find the free carrier concentrations at the point of maximum recombination velocity:

$$p = n_i\exp\left(-\frac{E_f}{2kT}\right) \times\sqrt{\frac{c_n(E)}{c_p(E)}\left[\frac{c_p(E)p_1(E) + wN}{c_n(E)n_1(E) + wN}\right]}\exp\left(\frac{eU}{2kT}\right), \tag{21}$$

$$n = n_i\exp\left(-\frac{E_f}{2kT}\right) \times\sqrt{\frac{c_p(E)}{c_n(E)}\left[\frac{c_n(E)n_1(E) + wN}{c_p(E)p_1(E) + wN}\right]}\exp\left(\frac{eU}{2kT}\right), \tag{22}$$

Taking this into account, the expression for the current density of the current–voltage characteristic has the form

$$j_r = \frac{2kTd(U)}{U_k - U}\int_E\left\{c_n(E)c_p(E) \times\left[n_i^2\exp\left(\frac{eU}{kT}\right) - n_1(E)p_1(E)\right]w(E)NN(E)\right\} \times\{t_n(E)t_p(E) + w(E)N[t_n(E) + t_p(E)]\}^{-1}dE. \tag{23}$$

Expression (23) yields the result [14] whereby the saturation of the current–voltage characteristic can be attributed to limitation of the transmitting capacity of the tunnel channel ($j_r \propto wN^2$) [15]. Allowing for (5), we obtain a formula for the saturation current density:

$$j_r = \frac{2kTd(U)}{U_k - U}N^2\nu\exp\left(-\frac{2}{a^3\sqrt{N}}\right). \tag{24}$$

Thus, the saturation current can be used to calculate the concentration of traps involved in the tunneling recombination.

We shall assume that the energy distribution of the traps is Gaussian. For low dispersions, when this distribution is close to discrete, we can estimate the energy of the distribution center from the maximum of the reduced recombination velocity. In this approximation we have

$$R_{np} = \frac{j_r}{n_i^2}\frac{2kTd(U)}{U_k - U} = \frac{c_nc_p[\exp(eU/2kT) + 1]wN^2}{t_n(E_t)t_p(E_t) + wN[t_n(E_t) + t_p(E_t)]}. \tag{25}$$

We shall perform a standard procedure to search for the maximum. To be specific we shall assume that the level, for example, lies in the upper half of the band. In this case, we have $n_1 > n_i > p_1$. Under these assumptions the voltage at the maximum of the reduced recombination velocity is related to the energy of the distribution center by the following relationship:

$$E_t = 0.5E_g - qU_m + kT\ln\left[\left(\frac{m_n^*}{m_p^*}\right)^{3/2}\frac{wN}{c_pn_i}\right], \tag{26}$$

$m_{n(p)}^*$ is the effective density-of-state mass of the conduction band (valence band).

Formulas (24) and (26) simplify the procedure for modeling the current–voltage characteristic whose results are given in Figs. 4 and 5 for samples having different degrees of ordering. These show fairly good agreement with the experimental results. Other parameters required for the modeling were determined from the results of independent experiments which are not given in detail because of the restricted size of the article although they are mentioned below.

In GaAs the energy of the distribution center determined using formula (26) was equal to the energy of the EL2 trap which was checked independently by capacitive methods. The parameters of this trap are well known and they were used for the modeling. The dispersion of the distribution was calculated from the broadening of the luminescence spectra of these samples. For the glassy semiconductor the trap energy and the dispersion were calculated using the results of measurements of the space-charge-limited currents. Thus, there were no fitting parameters for GaAs and one, the capture coefficient, for the glassy semiconductor.

The model of recombination in surface-barrier diodes developed in the present study is common to a fairly wide range of semiconductors with different degrees of ordering.

REFERENCES

1. B. I. Shklovskiĭ and A. L. Efros, *Electronic Properties of Doped Semiconductors* (Nauka, Moscow, 1979; Springer-Verlag, New York, 1984).
2. N. F. Mott and E. A. Davis, *Electronic Processes in Non-Crystalline Materials* (Clarendon Press, Oxford, 1979; Mir, Moscow, 1982).
3. F. T. Vas'ko and V. I. Pipa, Zh. Éksp. Teor. Fiz. **115**, 1337 (1999) [JETP **88**, 738 (1999)].
4. Yu. A. Aleshchenko, I. P. Kazakov, V. V. Kapaev, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 194 (1999) [JETP Lett. **69**, 207 (1999)].
5. R. Rentzch and I. S. Shlimac, Phys. Status Solidi A **43**, 231 (1977).
6. R. A. Street, Adv. Phys. **30**, 593 (1981).
7. S. D. Baranovskiĭ, V. G. Karpov, and B. I. Shklovskiĭ, Zh. Éksp. Teor. Fiz. **94** (3), 278 (1988) [Sov. Phys. JETP **67**, 588 (1988)].
8. S. V. Bulyarskiĭ, N. S. Grushko, A. I. Somov, and A. V. Lakalin, Fiz. Tekh. Poluprovodn. (St. Petersburg) **31**, 1146 (1997) [Semiconductors **31**, 983 (1997)].
9. S. V. Bulyarskiĭ, N. S. Grushko, and A. V. Lakalin, Fiz. Tekh. Poluprovodn. (St. Petersburg) **32**, 1193 (1998) [Semiconductors **32**, 1065 (1998)].
10. S. V. Bulyarskiĭ, M. O. Vorob'ev, N. S. Grushko, and A. V. Lakalin, Fiz. Tekh. Poluprovodn. (St. Petersburg) **33**, 733 (1999) [Semiconductors **33**, 668 (1999)].
11. S. V. Bulyarskiĭ, N. S. Grushko, and A. V. Lakalin, Zavod. Lab., No. 7, 25 (1997).
12. W. Shocley and W. T. Read, Phys. Rev. **87**, 835 (1952).
13. G. E. Pikus, *Basic Theory of Semiconductors Devices* (Nauka, Moscow, 1965).
14. N. S. Grushko, in *Critical Technology and Fundamental Problems of Physics of Condensed Medium* (Ul'yanov. Gos. Univ., Ul'yanovsk, 1999), p. 81.
15. N. S. Grushko, Uch. Zap. – Ul'yanovsk. Gos. Univ., Ser. Fiz., No. 2 (5), 51 (1998).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 1066–1069.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1230–1233.

SOLIDS
Electronic Properties

Mesoscopic Fluctuations of the Josephson Current of an *S–I–S* Junction with Weak Structural Disorder

V. Ya. Kirpichenkov
South-Russia State Technical University, Novocherkassk, 346400 Russia
e-mail: kirp@srstu.novoch.ru
Received June 14, 2000

Abstract—It is shown that in the range of tunneling resonance energies of an S–I–S (superconductor–insulator–superconductor) junction with weak (low impurity concentrations) structural disorder in the I-layer, the average critical current and the magnitude of its mesoscopic fluctuations are determined by tunneling along quantum resonance-percolation trajectories. For a "small" junction situated in a parallel magnetic field at temperature $T = 0$ conditions for smallness of the mesoscopic fluctuations are obtained and an estimate is made of the range of parameters in which the resonance mechanism for supercurrent propagation predominates. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Random spatial inhomogeneities of the tunnel transparency in the junction plane caused, for example, by random inhomogeneities of its thickness appreciably influence the properties of a tunnel S–I–S junction in a magnetic field parallel to the junction plane, leading in particular to a change in the dependence of the critical current on the magnetic field [1, 2].

Here we consider that the inhomogeneities of the tunnel transparency of the I-layer are caused by quantum resonance-percolation trajectories [3, 4], i.e., paths of resonant tunneling along chains of approximately equidistant impurities, formed randomly in an I-layer with weak structural disorder and connecting the opposite S-edges of the junction. In the tubes of resonant transparency along these chains the coefficient of electron transmission is $D \sim 1$ whereas outside these the coefficient is exponentially small which leads to strong spatial fluctuations of the transparency in the resonant energy range. An important difference between this model and the model [1] is that fluctuation formations such as the quantum resonance-percolation trajectories at low impurity concentrations determine not only the magnitude of the fluctuations in the range of tunnel resonance energies but also the magnitude of the average critical current in an ensemble of "like" samples which is considerably higher than the critical current of an impurity-free junction.

2. MODEL. BASIC EQUATIONS

We shall consider a tunnel junction model in the form of an S–I–S sandwich located at $T = 0$ K in a uniform external magnetic field $(0, H_y, 0)$ parallel to the junction plane and comprising two identical solid superconductors separated by a planar insulator layer of fairly small thickness L_x and area $S = L_yL_z$ containing identical electron-attracting impurities for which the single-impurity local level is ε_0 and the radius of localization of the state at this level is $a^{-1} = (U_0 - \varepsilon_0)^{-1/2}$ ($\hbar^2/2m = 1$). The regular (unperturbed by impurities) barrier potential of the I layer is $U_0 = \text{const} > \mu$, where μ is the Fermi level of the junction. Over the layer volume $V = L_xS$ N impurities are distributed macroscopically uniformly with the density $n = N/V$. The Fermi level μ is situated near ε_0, i.e., within the energy spectrum of the resonant tunnel transparency of the disordered I-layer. The junction is assumed to be "small": $L_z \ll \lambda_J$ (λ_J is the Josephson depth of penetration of the magnetic field), which means that the self-produced magnetic field of the Josephson current can be neglected [5].

In order to calculate the Josephson current the superconducting order parameters in the S-edges $\psi_{1,2} = \Delta\exp(i\varphi_{1,2})$ are assumed to be constant, as is usually the case [2, 6], but are perturbed by the presence of weak tunnel coupling between the edges. The characteristic energy width γ of the tunnel resonance which is significant for this value $|\mu - \varepsilon_0|$ satisfies the relationship $\Delta \ll \gamma \ll \mu$.

Under these conditions the Josephson current through the junction may be expressed in the form [7]

$$J(\mu - \varepsilon_0, \Phi, \Gamma_N) = \int\limits_{(S)} j(\mu - \varepsilon_0, \Gamma_N, \boldsymbol{\rho}) \times \sin\left[\Delta\varphi + \frac{2\pi\Phi}{\Phi_0}\frac{z}{L_z}\right] d^2\rho, \quad (1)$$

where $j(\mu - \varepsilon_0, \Gamma_N, \boldsymbol{\rho})$ is the current density, $\boldsymbol{\rho} = (y, z)$, $\Delta\varphi = \varphi_2 - \varphi_1$, Φ and Φ_0 are the magnetic flux through the contact and the magnetic flux quantum, respec-

tively, $\Gamma_N = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\}$ is a random configuration of N impurities in the I-layer, $N \gg 1$.

We write the Josephson current density in the form

$$j(\mu - \varepsilon_0, \Gamma_N, \boldsymbol{\rho}) = \frac{\pi\Delta}{2e}\int D(\mu - \varepsilon_0, \mathbf{q}, \boldsymbol{\rho}, \Gamma_N)\frac{d^2q}{(2\pi)^2}, \quad (2)$$

where [4] $D(\mu - \varepsilon_0, \mathbf{q}, \boldsymbol{\rho}, \Gamma_N)$ is the tunnel transparency of the I-layer for electrons of energy μ having the fixed transverse momentum component $\mathbf{q}$ at the "entrance" to the barrier and the fixed transverse coordinate $\boldsymbol{\rho}$ at the "exit," and integration over q is performed in the range $0 \le q^2 \le \mu$.

Substituting (2) into (1), we obtain

$$J(\mu - \varepsilon_0, \Phi, \Gamma_N) = \frac{\pi\Delta}{2e}G(\mu - \varepsilon_0, \Phi, \Gamma_N), \quad (3)$$

where the form of the function $G(\mu - \varepsilon_0, \Phi, \Gamma_N)$ is clear from the substitution.

The objects of the calculations are

$$\langle J\rangle = \frac{\pi\Delta}{2e}\langle G\rangle, \quad \langle(\delta J)^2\rangle = \left(\frac{\pi\Delta}{2e}\right)^2\langle(\delta G)^2\rangle, \quad \langle(\delta G)^2\rangle = \langle G^2\rangle - \langle G\rangle^2, \quad (4)$$

where

$$\langle G\rangle = \frac{1}{\Delta\Gamma_N}\int\limits_{\{\Gamma_N\}} G(\mu - \varepsilon_0, \Phi, \Gamma_N)d\Gamma_N, \quad (5)$$

$$\langle G^2\rangle = \frac{1}{\Delta\Gamma_N}\int\limits_{\{\Gamma_N\}} G^2(\mu - \varepsilon_0, \Phi, \Gamma_N)d\Gamma_N, \quad (6)$$

$$\Delta\Gamma_N = V^N, \quad d\Gamma_N = d\mathbf{r}_1 d\mathbf{r}_2 \ldots d\mathbf{r}_N.$$

At energies μ close to ε_0 the phase space of the impurity system $\{\Gamma_N\}$ is factorized as a set of resonant and nonresonant regions and the main contribution to the averages (5) and (6) is made by the resonant regions corresponding to quantum resonance-percolation trajectories [3, 4]. Expressions (5) and (6) are then reduced to the form

$$\langle G\rangle = \sum_{m=1}^{N}\int_{(S)} d^2\rho_m \times \int\limits_{\mathscr{L}/m}^{\infty} p_m(u)G_m^{\mathrm{res}}(\mu - \varepsilon_0, u, \Phi, \boldsymbol{\rho}_m)du, \quad (7)$$

$$\langle G^2\rangle = \sum_{m=1}^{N}\int_{(S)} d^2\rho_m \times \int\limits_{\mathscr{L}/m}^{\infty} p_m(u)[G_m^{\mathrm{res}}(\mu - \varepsilon_0, u, \Phi, \boldsymbol{\rho}_m)]^2 du, \quad (8)$$

where

$$G_m^{\mathrm{res}} = \iint G_m^{res}(\mu - \varepsilon_0, \mathbf{q}, \boldsymbol{\rho} - \boldsymbol{\rho}_m, u) \times \sin\left[\Delta\varphi + \frac{2\pi\Phi}{\Phi_0}\frac{z}{L_z}\right]\frac{d^2q}{(2\pi)^2}d^2\rho. \quad (9)$$

Here [4]

$$p_m(u) = a^2c^m\exp(-cm\pi u^3)\left[2u^2\left(\frac{mu}{\mathscr{L}} - 1\right)\right]^{m-1} \quad (10)$$

is the probability (per unit area) of formation of an m-impurity quantum resonance-percolation trajectory with the dimensionless "step" $u = 2la$ ($2l$ is the distance between neighboring impurities in the chain), $c = na^{-3}$ is the dimensionless impurity concentration, $\mathscr{L} = aL_x$ is the dimensionless thickness of the I-layer,

$$D_m^{\mathrm{res}}(\mu - \varepsilon_0, \mathbf{q}, \boldsymbol{\rho} - \boldsymbol{\rho}_m, u) = 4\sigma_0\frac{k_q}{k} \times \exp\left\{-\frac{2a^2|\boldsymbol{\rho} - \boldsymbol{\rho}_m|^2}{u} - \frac{uq^2}{2a^2} - \frac{(\mu - \varepsilon_0)^2}{\gamma^2(u)}\right\}, \quad (11)$$

where

$$4\sigma_0 = a^2k^2\pi^{-4}(a^2 + k^2)^{-2}, \quad k^2 = \mu, \quad k_q^2 = \mu - q^2,$$

$\boldsymbol{\rho}_m$ is the transverse coordinate of the mth (last) impurity in the chain,

$$\gamma(u) = 4a^2u^{-1}e^{-u} \quad (12)$$

is the energy width of the resonant transparency zone along the quantum resonance-percolation trajectory.

Substituting (9)–(11) into (7) and (8), calculating the integrals contained therein, and finding the extrema of the expressions obtained with respect to $\Delta\varphi$, we obtain the averages of the critical current and its fluctuations:

$$\langle J_c\rangle = \frac{\pi\Delta}{2e}\langle G_c\rangle, \quad \langle(\delta J_c)^2\rangle = \left(\frac{\pi\Delta}{2e}\right)^2\langle(\delta G_c)^2\rangle, \quad (13)$$

where

$$\langle G_c(\mu - \varepsilon_0, \Phi)\rangle = S\left[\sum_{m=1}^{N} g_m^{\mathrm{res}}(\mu - \varepsilon_0)\right]\left|\frac{\Phi_0}{\pi\Phi}\right|\sin\left(\frac{\pi\Phi}{\Phi_0}\right)\Bigg|, \quad (14)$$

$$\langle(\delta G_c(\mu-\varepsilon_0,\Phi))\rangle = S\left[\sum_{m=1}^{N} f_m^{res}(\mu-\varepsilon_0)\right]\frac{1}{2}\left[1+\frac{\Phi_0}{2\pi\Phi}\sin\left(\frac{2\pi\Phi}{\Phi_0}\right)\right], \quad (15)$$

$$g_m^{res}(\mu-\varepsilon_0) = \sigma_0\int_{\mathscr{L}/m}^{\infty} p_m(u)\exp\left[-\frac{(\mu-\varepsilon_0)^2}{\gamma^2(u)}\right]du, \quad (16)$$

$$f_m^{res}(\mu-\varepsilon_0) = \sigma_0^2\int_{\mathscr{L}/m}^{\infty} p_m(u)\exp\left[-\frac{2(\mu-\varepsilon_0)^2}{\gamma^2(u)}\right]du. \quad (17)$$

The integrals (16) and (17) can be calculated by the method of steepest descents although in order to determine the nature of the dependence of $\langle G_c\rangle$ (14) and $\langle(\delta G_c)^2\rangle$ (15) on $\mu-\varepsilon_0$ it is sufficient to note that

(a) as a result of the exponentially fast decrease in $\gamma(u)$ with increasing u (12), the main contribution to the integrals (16) and (17) occurs near their lower limits;

(b) for

$$|\mu-\varepsilon_0| > \gamma_m = \gamma\left(\frac{\mathscr{L}}{m}\right) = \frac{4a^2}{\mathscr{L}/m}\exp\left(-\frac{\mathscr{L}}{m}\right)$$

these integrals are exponentially small. Consequently for given m the main contribution to (16) and (17) is made by the slightly winding quantum resonance-percolation trajectories close to the shortest, having the step $u_m = \mathscr{L}/m$ and therefore the largest (for given m) energy width γ_m.

Thus, Eqs. (14) and (15) [allowing for (10), (16), and (17)] give the tunnel conductance $\langle G_c\rangle$ of a disordered junction in a magnetic field and its fluctuation $\langle(\delta G_c)^2\rangle$ as a series in powers of the concentration c whose mth term gives the contribution of m-impurity quantum resonance-percolation trajectories to these values and as a function of the argument $\mu-\varepsilon_0$ is a Gaussian curve having a maximum at $\mu=\varepsilon_0$, characteristic width $\sim\gamma_m$, and height $\sim c^m$.

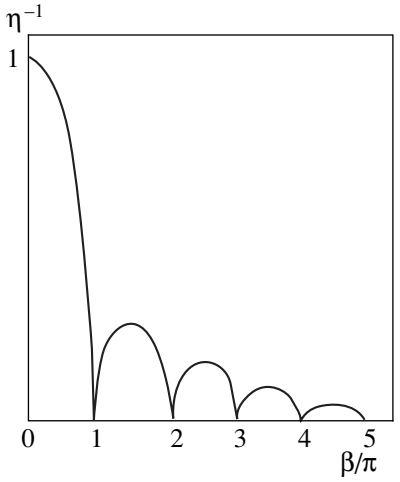

Figure.

3. ESTIMATES

We shall consider the "higher" single-impurity ($m=1$) tunnel resonance for $c \ll 1$ which for $|\mu-\varepsilon_0| < \gamma(\mathscr{L})$ makes the main contribution to $\langle G_c\rangle$ and $\langle(\delta G_c)^2\rangle$. In this case, (14) and (15) have the form ($\beta=\pi\Phi/\Phi_0$)

$$\langle G_c(\mu-\varepsilon_0,\beta)\rangle^2 = \langle G_c(\mu-\varepsilon_0,0)\rangle^2\beta^{-2}\sin^2\beta, \quad (18)$$

$$\langle(\delta G_c(\mu-\varepsilon_0,\beta))^2\rangle = \langle(\delta G_c(\mu-\varepsilon_0,0))^2\rangle \times\frac{1}{2}[1+(2\beta)^{-1}\sin 2\beta], \quad (19)$$

where

$$\langle G_c(\mu-\varepsilon_0,0)\rangle^2 = \eta^2\sigma_0^2\exp\left[-\frac{2(\mu-\varepsilon_0)^2}{\gamma^2(\mathscr{L})}\right], \quad (20)$$

$$\langle(\delta G_c(\mu-\varepsilon_0,0))\rangle^2 = \eta\sigma_0^2\exp\left[-\frac{2(\mu-\varepsilon_0)^2}{\gamma^2(\mathscr{L})}\right], \quad (21)$$

$$\eta = ca^2S\exp(-c\pi\mathscr{L}^3). \quad (22)$$

The condition for strong suppression of fluctuations

$$\langle G_c\rangle^2 \gg \langle(\delta G_c)^2\rangle,$$

yields the constraint:

$$\eta^{-1} \ll 2\beta^{-2}\sin^2\beta[1+(2\beta)^{-1}\sin 2\beta]^{-1}. \quad (23)$$

The curve $\langle G_c\rangle^2 = \langle(\delta G_c)^2\rangle$ which arbitrarily separates the regions of strong and weak fluctuations on the plane (η^{-1},β) is shown qualitatively in the figure. The envelope of the maxima behaves as β^{-2}, the region of strong fluctuations is located above the curve, and for $\beta=n\pi$ no region of weak fluctuations exists. We shall give numerical estimates of the range of parameters where this tunnel resonance may appear.

We shall estimate the range of concentrations c in which the conductance $\langle G_c(\mu-\varepsilon_0,0)\rangle$ is considerably higher than the conductance G_0 of the "empty" (without impurities) junction for $|\mu-e_0| < \gamma(\mathscr{L})$:

$$\langle G_c\rangle \gg G_0 = 4\pi^3a^2S\sigma_0\mathscr{L}^{-1}\exp(-2\mathscr{L}). \quad (24)$$

Substituting (20) into (24), we obtain

$$c\exp(-c\pi\mathscr{L}^3) \gg 4\pi^3\mathscr{L}^{-1}\exp(-2\mathscr{L}). \quad (25)$$

Bearing in mind that for low-resistivity tunnel junctions [2] $\exp(-2\mathscr{L}) \sim 10^{-6}$ ($\mathscr{L}\approx 7$), we obtain from (25)

$$10^{-5} \ll c \ll 10^{-2}. \quad (26)$$

Assuming that $a^2 \sim k^2 = \mu \sim 10$ eV ($\sim 10^5$ K) we obtain from (12) the estimate of the resonance width:

$$\gamma(\mathcal{L}) \sim 10^{-2} \text{ eV} \quad (\sim 10^2 \text{ K}). \tag{27}$$

In accordance with this particular model $\Delta \ll \gamma(\mathcal{L})$ and consequently

$$\Delta \sim 10^{-3} \text{ eV} \ (\sim 10 \text{ K}), \tag{28}$$

and the temperature should satisfy the condition $T \ll \Delta$ and therefore $T \sim 1$ K. The junction area S required to suppress the fluctuations for a given value of β is estimated from (23). For example, for $\beta = 0$ it follows from (22) and (23) that

$$S \gg c^{-1} \exp(c\pi\mathcal{L}^3) a^{-2}. \tag{29}$$

Thus, for $c \sim 10^{-3}$ and $\mathcal{L} \approx 7$ we obtain $S \gg 10^3 a^{-2}$ where a^{-1} is the radius of localization of the state at an impurity, which is of the order of magnitude of the interatomic distance.

It can therefore be concluded that in the range of parameters indicated above, in the presence of resonant impurities having local levels in the band $|\mu - \varepsilon_0| < \gamma(\mathcal{L})$ the Josephson current and its mesoscopic fluctuations should be determined by the existence of quantum resonance-percolation trajectories with $m = 1$ in the disordered I-layer.

REFERENCES

1. I. K. Yanson, Zh. Éksp. Teor. Fiz. **58**, 1497 (1970) [Sov. Phys. JETP **31**, 800 (1970)].
2. I. O. Kulik and I. K. Yanson, in *Josephson Effect in Superconducting Tunnel Structures* (Nauka, Moscow, 1970), Par. 20.
3. I. M. Lifshits and V. Ya. Kirpichenkov, Zh. Éksp. Teor. Fiz. **77**, 989 (1979) [Sov. Phys. JETP **50**, 499 (1979)].
4. V. Ya. Kirpichenkov, Zh. Éksp. Teor. Fiz. **116**, 1048 (1999) [JETP **89**, 559 (1999)].
5. A. A. Abrikosov, in *Fundamentals of the Theory of Metals* (Nauka, Moscow, 1987; North-Holland, Amsterdam, 1988), Par. 22.4.
6. L. G. Aslamazov and M. V. Fistul', Zh. Éksp. Teor. Fiz. **83**, 1170 (1982) [Sov. Phys. JETP **56**, 666 (1982)].
7. A. Barone and G. Paterno, in *Physics and Applications of the Josephson Effect* (Wiley, New York, 1982; Mir, Moscow, 1984), Par. 4.2.

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 1070–1078.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1234–1244.

NONLINEAR PHYSICS

Nonsteady-State Self-Interaction in a Medium with Striction Nonlinearity

N. A. Zharova, A. G. Litvak, and V. A. Mironov

Institute of Applied Physics, Russian Academy of Sciences, Nizhni Novgorod, 603600 Russia

Received June 16, 1999

Abstract—An analytic and numerical investigation is made of the self-focusing of a wave beam allowing for the inertia of the nonlinear response of the medium described by an acoustic type of equation. Some characteristics of the dynamics of self-interaction of the wave fields are analyzed in the paraxial optics approximation and the self-similar structures and space-time instability of a plane wave are considered. The stages of instability buildup, structure formation, and the establishment of a steady state are studied numerically. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

With the successful generation of strong short electromagnetic pulses, theoretical studies of the self-interaction of wave beams under conditions when the inertia of the nonlinear response of the medium plays an important role have become a topical issue. We have recently seen the publication of theoretical studies devoted to the self-interaction of wave beams in media having the commonest mechanisms for relaxation of the nonlinear response: Kerr, diffusion [1, 2], and acoustic [3, 4]. The spatial evolution of these beams is described by the same nonlinear Schrödinger equation. The transition regime to the steady-state pattern depends strongly on the type of relaxation equation for the nonlinear response of the medium to the action of an electromagnetic field. For the case of material coupling described by a first-order equation with respect to time, which corresponds to the Kerr and diffusion mechanisms of nonlinearity relaxation, a steady-state pattern is established as a result of the formation of singularities of the compressible filament or traveling focus type and the subsequent structural instability of these formations [1, 2]. For striction nonlinearity, whose relaxation is described by an acoustic equation, the establishment of a steady-state self-consistent distribution is associated with the excitation of an intermediate wave propagating into the nonlinear medium [3, 4]. This conclusion was reached for axisymmetric wave beams and consequently applies to radiation whose power is of the order of (higher than) the critical self-focusing power. It can be predicted that dynamic structures of the compressible filament or traveling focus type will play an important role in the evolution of wave beams having powers considerably higher than the critical self-focusing power, as in other mechanisms for relaxation of the nonlinear response.

In the present study we make an analytic and numerical investigation of the nonsteady-state self-focusing of wave beams in the presence of striction nonlinearity. We find self-similar structures over a wide range of supercriticality parameters and analyze characteristics of the space–time instability of a plane wave and the dynamics of self-interaction of the wave field in the paraxial approximation under acoustic relaxation of the nonlinear response. A numerical investigation of the nonsteady-state interaction of wave beams was based on a two-dimensional model system which keeps the main features of the initial equations. This allowed us to make a detailed study of the transition processes under conditions of structural instability of the wave fields.

2. BASIC EQUATIONS. SELF-SIMILAR STRUCTURES

We shall consider nonsteady-state self-focusing processes of an electromagnetic wave beam in a medium where the dominant mechanism of nonlinearity is striction nonlinearity associated with redistribution of the density of the medium under the action of a ponderomotive force. An example of such a medium in which we neglect the contribution of other nonlinearities may be a strongly ionized plasma at moderate electromagnetic radiation intensities. In this case, the wave beam self-interaction process may be described by a well-known system of equations consisting of the nonlinear Schrödinger equation for the scalar envelope of the wave beam field $\psi(\mathbf{r}, t)$ and the wave equation for the perturbations of the density n. We shall consider the case when the time taken for propagation of a wave through the region occupied by the nonlinear medium is substantially shorter than the characteristic relaxation time of the nonlinearity, i.e., the time taken for sound to pass through the transverse dimension of the beam. The field distribution in the medium can then be considered to be quasi-steady-state, i.e., the time derivative $\partial\psi/\partial t$ in the nonlinear Schrödinger equation can

be neglected. As a result, the initial system of equations can be expressed in the form

$$i\frac{\delta\psi}{\partial z} + \Delta_\perp\psi - n\psi = 0, \quad (1)$$

$$\frac{\partial^2 n}{\partial t^2} - \Delta_\perp n = \Delta_\perp F(|\psi|^2). \quad (2)$$

Here we shall also neglect time dispersion effects since these are negligible for the pulses of interest to us, having durations comparable with the sound relaxation time. The scale invariance of these equations means that they can be written for various nonlinear media [3, 5], including a plasma [4, 5] in the dimensionless form given. Here $\Delta_\perp = \partial^2/\partial x^2 + \partial^2/\partial y^2$ is the transverse Laplacian, and the field ψ is normalized to the field characteristic for nonlinear effects in which the steady-state perturbation of the refractive index varies substantially. Here we also introduce the new coordinates $z = kz$, $\mathbf{r}_\perp = k\mathbf{r}_\perp$, and the time $t = k_0 v_s t$, where k is the wave number and v_s is the velocity of sound (in a plasma the ion sound velocity). By introducing the arbitrary function $F(|\psi|^2)$ we can describe a wide range of situations from the simplest dependence $F(|\psi|^2) = |\psi|^2$ corresponding to cubic nonlinearity to nonlinearity saturation effects. We shall subsequently analyze the space–time dynamics using the system (1), (2) with the following boundary and initial conditions:

$$\psi = \psi_0(x, y, t) \text{ at } z = 0, \quad n(t = 0) = 0. \quad (3)$$

In the simplest case $F(|\psi|^2) = |\psi|^2$ after the establishment of steady-state cubic nonlinearity ($n = -|\psi|^2$) the problem reduces to the well-studied nonlinear Schrödinger equation with cubic nonlinearity. From the theory of steady-state self-focusing we know that if the power of the wave beam entering a nonlinear medium ($z = 0$) exceeds a critical value called the critical self-focusing power

$$P = \int |\psi|^2 dr^3 > P_{cr},$$

the evolution of any distribution inevitably ends in the formation of a singularity [6–9]. Near the focus $z \approx z_0$ the axisymmetric distribution of the field evolves in accordance with the self-similar law [9]

$$a \propto \sqrt{\frac{z - z_0}{\ln\ln(z_0 - z)}}$$

(a is the transverse dimension of the beam); the energy flux trapped in the singularity is exactly P_{cr}. If the beam power is considerably higher than the critical self-focusing power, the wave beam separates into beams of critical power in the transverse direction [6, 5, 8].

In the nonsteady-state regime a common class of solutions for media having nonlinear response inertia are solutions in the form of jets which are homogeneous along z and compressible with time (homogeneous wave channels), along which the trapped electromagnetic field propagates. Substituting into the initial equations

$$F = |\psi|^2, \quad \psi = \frac{1}{a(t)}\Phi(\xi)\exp\left(\frac{iz}{a^2}\right),$$
$$n = \frac{1}{a^2(t)}N(\xi), \quad \xi = \frac{r}{a}, \quad a = t_0 - pt, \quad (4)$$

we can easily obtain the following equations for the self-similar functions:

$$\Delta\xi\Phi - (1 + N)\Phi = 0, \quad (5)$$

$$p^2\xi^2 N_{\xi\xi} + 6p^2\xi N_\xi + 6p^2 N - \Delta_\xi N = \Delta_\xi\Phi^2, \quad (6)$$

where Δ_ξ is the Laplacian in cylindrical coordinates. A numerical analysis of the system of Eqs. (5), (6) shows that a localized self-similar solution exists for any value of the parameter p characterizing the rate of collapse. For $p \longrightarrow 0$ a transition takes place to the limit of a homogeneous waveguide channel ("Townes" mode) when the coupling between the perturbations of the medium and the field is quasi-steady-state. High values of p ($p \gg 1$) correspond to the highly nonsteady-state case when the term containing the time derivative plays a determining role in the initial equations (1) and (2). In the supersonic regime the power $P = \int\Phi^2 d\xi$ channeled in the dynamic jet depends on the rate of compression. The rate of compression increases with increasing parameter p as $p \propto \sqrt{P}$. It is important that the power stored in the filament in the dynamic regime may be significantly higher than the critical self-focusing power in a system with steady-state nonlinearity ($p \approx 0$).

At this point we note the existence of a broader class of self-similar structures corresponding to traveling singularities of the focus type $a \sim z/v - t + t_0$. It can be seen that these are described by a system of equations similar to (5) and (6). These equations contain the square of the focus velocity v and thus, unlike other types of nonlinearity relaxation (Kerr, diffusion) [1, 2], in this particular case the direction of motion of the singularity cannot be determined from the condition of localization of the solution. Using the generalized lens transformation

$$\psi = \frac{1}{a(z, t)}\Phi(\xi, z, t)\exp\left(i\frac{a_z}{4a}r^2\right),$$
$$n = \frac{1}{a^2}N(\xi, z, t), \quad \xi = \frac{r}{a}, \quad (7)$$

where a is an as yet arbitrary function of the arguments z, t, yields the following equation for the field:

$$ia^2\frac{\partial\Phi}{\partial z} + \Delta_\xi\Phi - \left(N + \frac{a^2 a_{zz}}{4}\xi^2\right)\Phi = 0. \quad (8)$$

From this it can be seen that for a traveling-focus self-similar structure ($a \sim z/v - t + t_0$), the last (lens) term goes to zero ($a_{zz} = 0$). In this case, the self-similarity is exact (complete) and the dominant eigenmode of equation (8) is strictly localized.

Hence, an analysis of the self-similar solutions reveals a new possibility for the formation of a singularity other than steady-state self-focusing, in which a power considerably higher than the critical value is localized. However, the existence of suitable self-similarities still does not guarantee the feasibility of the corresponding solutions for arbitrary initial and boundary conditions of the problem. A broad class of self-similar solutions tends to indicate a wide range of dynamics for the establishment of a steady-state distribution in the system (1), (2) and especially in the highly supercritical regime.

3. PARAXIAL OPTICS APPROXIMATION

The self-similar structures obtained naturally give some idea of the problem but a simpler and more convenient method of analyzing the self-interaction of wave beams is the paraxial optics approximation. The accuracy of the results obtained using this approximation, particularly in the case of the nonsteady-state self-interaction being considered is obviously low. This is because the lens formed under the action of the ponderomotive force, is in principle aberrational, especially in the dynamic regime when acoustic perturbations of the medium density are excited. In this case, however, the paraxial approximation gives a qualitatively correct description of the system behavior. For example, when a homogeneous waveguide channel forms in a pulsed radiation field [3], a comparison between the gain factors calculated using equations of the type (1) and (2) and using the aberration-free approximation shows that neglecting aberrations merely yields slightly exaggerated values of the field on the axis of the system.

We shall consider a system of Eqs. (1) and (2) for $F(|\psi|^2) = |\psi|^2$ in the paraxial optics approximation (in the so-called aberration-free approximation) where it is assumed that the field distribution in a wave beam propagating in a nonlinear medium remains Gaussian

$$\psi = \frac{\sqrt{P}}{a}\exp\left(-\frac{r^2}{2a^2} + i\frac{a_z}{4a}r^2\right), \tag{9}$$

having the width a which depends on z and t, and the distribution of the refractive index perturbation in the axial region is approximated by a parabola

$$n = n_0(z, t) + n_2(z, t)r^2. \tag{10}$$

As a result we have the standard equation for the beam width in the paraxial approximation of self-focusing theory

$$\frac{\partial^2 a}{\partial z^2} = \frac{4}{a^3} - 4n_2 a. \tag{11}$$

The self-consistent evolution of the dependence $n_2(z, t)$ is determined by the material Eq. (2) in the axial region ($r \approx 0$). The normal method of calculating n_2 is as follows [10]. Integrating the material Eq. (2) for a Gaussian wave beam, we can easily find expressions for n and consequently the value of the coefficient

$$n_2 = \frac{1}{2}\frac{\partial^2 n}{\partial r^2} \quad (r = 0)$$

of the quadratic term in the parabolic approximation of the perturbation of the refractive index n. As a result, the self-consistent equation to describe the behavior of the wave beam width in the aberration-free approximation is integrodifferential [10].

We shall subsequently adopt a different approach. We shall assume that the parabolic approximation of the refractive index (10) is a consequence of the expansion of $n = n_0\exp(-r^2/a^2)$ for $r = 0$. The existence of additional coupling ($n_2 = -n_0/a^2$) allows us to obtain the oscillator equation from the material equation (2) for n_2:

$$\frac{\partial^2 n_2}{\partial t^2} + \frac{8n_2}{a^2} = \frac{8P}{a^6}. \tag{12}$$

Note that as a result of this approach we have obtained a simpler and clearer system of equations to describe the space–time evolution of the wave beam (11) and (12) instead of an integrodifferential Eq. [10].

Naturally the equations in the aberration-free approximation (11) and (12) have the same self-similar substitution as the initial system (1) and (2):

$$a = t_0 - pt. \tag{13}$$

From (11) and (12) we can then obtain the following relationship for the rate of collapse:

$$p = \sqrt{\frac{2(P-1)}{5}}. \tag{14}$$

For $P \gg 1$ we then find the value of $p = \sqrt{2P/5}$ which is the same as the similar value for the parameter of the self-similar structure (5) and (6). For a power of the order of the critical self-focusing power $P_{cr} \approx P \approx 1$ the rate of collapse is low and equation (12) may be predicted to have oscillator properties. Obviously, these properties should be manifest most clearly at the entrance to the interaction zone ($z = 0$) where the width of the collimated wave beam ($a(z = 0) = a_0$) varies negligibly.

We give the expression for the square of the wave beam width as a series:

$$a^2 = a_0^2 + \frac{\partial^2 a^2}{\partial z^2}(z = 0, t)\frac{z^2}{2} + \frac{\partial^4 a^2}{\partial z^4}(z = 0, t)\frac{z^4}{24}. \quad (15)$$

It is then easy to obtain an equation describing the motion of the maximum coordinate of the field ($da/dz = 0$):

$$z_{\max}^2 = -6\frac{\partial^2 a^2}{\partial z^2}/\frac{\partial^4 a^2}{\partial z^4}. \quad (16)$$

The spatial derivatives of the square of the width are calculated for $z = 0$ where a_0 remains constant in time [having renormalized the variables in the system (11) and (12), this can be taken as unity]. This allows us to obtain explicit expressions for them. For example, from (11) we can easily obtain

$$\frac{\partial^2 a^2}{\partial z^2} = 8(1 - n_{20}), \quad (17)$$

where the second derivative of the refractive index perturbation at the boundary ($z = 0$) is described by

$$\frac{\partial^2 n_{20}}{\partial t^2} + 8n_{20} = 8P. \quad (18)$$

Integrating (18) for the case where the field is "switched on" instantaneously ([$P(t \geq 0) = \text{const}$], we find

$$n_{20} = P[1 - \cos(2\sqrt{2}t)]. \quad (19)$$

Similarly we can obtain an expression for $\partial^4 a^2/\partial z^4$ ($z = 0$, t). This is fairly cumbersome and thus we shall not give it here in its entirety. For $P \approx 1$ and times when the numerator in (16) vanishes ($n_{20} = 1$) we find $\partial^4 a^2/\partial z^4 \approx 80$. Finally we obtain

$$z_{\max}^2 \approx -\frac{12}{5}\{1 - P[1 - \cos(2\sqrt{2}t)]\}.$$

From this it can be seen that at time

$$t \approx (4P)^{-1/2} \quad (20)$$

the first maximum of the field appears near the boundary ($z \approx 0$) and propagates into the nonlinear medium as it is amplified. The rate of emergence of the field maximum from the linear focal region is approximately unity.

Figure 1 shows the time dependence of the position $z_{\max}$ of the field maximum on the axis obtained in the paraxial approximation, i.e., by a numerical simulation of the system of Eqs. (11) and (12) for near-critical values of the power. The three curves plotted in Fig. 1 correspond to the powers $P_1 = 1.2$, $P_2 = 1.4$, $P_3 = 1.6$ A characteristic feature of the motion of the field maximum is

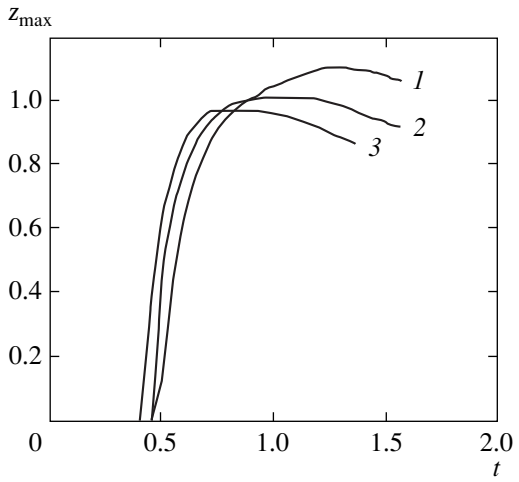

Fig. 1. Time dependence of the position $z_{\max}$ of the field maximum on the beam axis obtained using the paraxial approximation. Curves *1*, *2*, and *3* correspond to powers $P_1 = 1.2$, $P_2 = 1.4$, and $P_3 = 1.6$, respectively.

that it exhibits a rapid shift from the origin $z = 0$, beginning at $t \approx 0.5$ [see expression (20)], to the point $z \approx 1$ and then moves slowly back to the boundary $z = 0$ at the velocity $d(z_{\max})/dt \sim 0.1$. Note that the motion of the maximum into the nonlinear medium is unlikely to be caused by the motion of the beam focal point since the beam width at the field maximum point $a_{\min}$ differs negligibly from the value at the entrance $a_{\min} \sim a(z = 0) = 1$. However, the motion of the maximum back to the boundary $z = 0$ is characterized by low values of the width $a_{\min}$ and high values of the concentration perturbation $n_{2\max}$ and may be considered to be the motion of the beam focus. This last stage of the beam evolution in the paraxial approximation realizes a solution of the traveling focus type.

The displacement of the region of maximum field into the nonlinear medium is typical of the initial stage of nonsteady-state self-interaction for various mechanisms of relaxation of the nonlinear response. A characteristic feature of the excitation of acoustic motion in a medium is the periodic formation of regions of wave beam constrictions (field maxima) near $z = 0$ and their subsequent motion into the nonlinear medium [3, 4]. Typical parameters of the process are the same as the values obtained below in a study of the resonant space–time instability of a plane wave. In terms of dimensional variables the velocity of the field maximum in a wave beam having the characteristic width a_0 is $v_M \approx c_s k_0 a_0$, where c_s is the velocity of sound and k_0 is the wave number of the electromagnetic wave. The period of the inhomogeneity generation is of the order of magnitude of $T \approx a_0/c_s$. Estimating the collapse time from (11) and (12), we can easily establish that the number of overoscillations is proportional to $(P - 1)^{-1/2}$. It is also clear that these effects are suppressed appreciably

for pulses having a leading edge duration τ_f greater than the oscillation period ($\tau_f \gg 2\pi$).

Another conclusion can also be drawn from equation (17). A spatially bounded pulse of duration τ and width a_0 is trapped in the self-focusing process if the power P exceeds the threshold value:

$$P > a_0^2/4\tau^2. \tag{21}$$

Calculations for pulses of different shape based on an integrodifferential equation were presented in [10]. The results of the calculations show that the threshold power remains almost constant $P \approx 1$ when the parameter a_0/τ varies between zero and one and then increases in accordance with the relationship obtained above proportionately as $(a_0/\tau)^2$.

4. SPACE–TIME INSTABILITY OF A WAVE BEAM

Singularities in the dynamics of transverse separation of a wave beam having a power higher than the critical value for self-focusing can be investigated most easily by studying the stability of a homogeneous plane wave. The role of various transverse perturbation scales in the formation of a nonlinear structure is determined by the corresponding instability growth rates. This formulation of the problem has recently been actively used to study the behavior dynamics of abrupt inhomogeneities ("hot spots") against the background of a continuous wave field distribution [11, 12]. We shall consider some singularities in the dynamics of self-interaction associated with acoustic relaxation of the nonlinear response.

Assuming that $\psi = \Phi_0 + q\cos(\mathbf{k}\cdot\mathbf{r}_\perp)$, $n = p\cos(\mathbf{k}\cdot\mathbf{r}_\perp)$, ($|q| \ll \Phi_0$) we linearize the initial system of equations. As a result, we obtain the following equations for the real part of the field perturbation $q_r = \mathrm{Re}(q)$ and p [11–13]:

$$\frac{\partial^2 q_r}{\partial z^2} + k^4 q_r + k^2 p\Phi_0 = 0, \tag{22}$$

$$\frac{\partial^2 p}{\partial t^2} + k^2 p + k^2 \cdot 2q_r\Phi_0 = 0. \tag{23}$$

The system of Eqs. (22) and (23) reflects singularities of the nonsteady-state self-interaction accompanying acoustic relaxation of the nonlinear response of the medium. It is a system of coupled equations for space (22) and time oscillators (23). Unlike [11–14], we shall consider a special class of resonant perturbations

$$q_r,\ p \sim A,\ h\exp(ikt \pm ik^2 z), \tag{24}$$

corresponding to waves traveling away from (minus) or toward (plus) the boundary. This type of perturbation will clearly evolve as fast as possible. The equations for the slowly varying amplitudes of the resonant perturbations have the form

$$i\frac{\partial A}{\partial z} = \pm\frac{\Phi_0}{2}h, \tag{25}$$

$$i\frac{\partial h}{\partial t} = -k\Phi_0 A. \tag{26}$$

An unstable solution of these equations with the initial condition $q_r(z, t = 0) = A_0$, q_r can only be obtained for waves traveling away from the boundary [plus sign in Eq. (25)]. This solution has the form

$$A = A_0 I_0(\sqrt{2kzt}\Phi_0), \tag{27}$$

$$h = \frac{iA_0\sqrt{2kt}}{\sqrt{z}} I_1(\sqrt{2kzt}\Phi_0), \tag{28}$$

and thus, at the boundary $z = 0$ we have

$$\psi(z = 0, t) \approx \Phi_0 + A_0\cos(kt), \tag{29}$$

i.e., the field exhibits weak ($A_0 \ll \Phi_0$) periodic time ($T = 2\pi/k$) modulation.

For waves traveling toward the boundary the modified Bessel functions are replaced by the ordinary functions J_0 and J_1 and consequently no instability occurs.

From the condition that the envelope (27), (28) increases slowly on the characteristic space and time scale of the wave perturbation we can find that the range of validity of the solution (27), (28) is achieved for $k \gg \Phi_0$. These perturbations are the fastest growing.

Small perturbations on the propagation path of the wave field increase exponentially with time. As a result the phase of the wave perturbation travels from the boundary of the nonlinear medium $z = 0$ into this medium at the velocity $1/k$ while its amplitude increases in space and time. Note that the space–time dynamics of the perturbation envelope [the dependence of the function $I_0(\sqrt{2kzt}\Phi_0)$ on z and t] negligibly influences the motion of the maxima of the field modulus $|\psi|$ so that this motion takes place almost at the same velocity as the phase motion: $dz_{\max}/dt \approx 1/k$. An important characteristic of the dynamics is the presence of a phase shift $\pi/2$ between the perturbations of the field (27) and the refractive index of the medium (28). This can be seen in the numerical calculations of the nonsteady-state self-focusing in a medium exhibiting acoustic relaxation of the nonlinear response [3, 4]. For the velocity of a resonant perturbation under conditions of space–time instability we have $v \approx v_s k_0 q_r$. This value is the same as the propagation velocity of wave-like perturbations in the paraxial optics approximation and in the numerical calculations [3, 4].

Space–time instability on nonresonant scales was investigated in [13]. This instability evolves as in Kerr and diffusion relaxation of the nonlinear response [15]. The growth rate of this instability increases as the per-

turbation wave number k increases. In a layer of bounded length the instability reaches the nonlinear stage of evolution initially at the rear boundary and then the region of the nonlinear regime expands at a velocity of the order of $z/2t$ toward the incident radiation. In the intermediate region between the front boundary and the boundary of the nonlinear regime for the nonresonant instability the space–time dynamics is determined by the evolution of resonant instability if its corresponding growth rate is higher than the nonresonant instability growth rate. This is achieved under the condition

$$zk\Phi_0^2 \gg t, \tag{30}$$

i.e., in fairly strong pump fields and for large k.

5. NUMERICAL INVESTIGATION OF NONSTEADY-STATE SELF-FOCUSING

It is difficult to make a numerical investigation of a nonsteady-state three-dimensional problem in terms of the variables x, y, z, and t for times considerably greater than the nonlinearity relaxation time even using powerful computers. In order to identify the fundamental characteristics of nonsteady-state self-interaction under conditions of structural instability of the wave fields, we decided to use the following (model) system of equations:

$$\begin{aligned} i\frac{\partial\psi}{\partial z} + \frac{\partial^2\psi}{\partial} - n\psi &= 0, \\ \frac{\partial^2 n}{\partial t^2} + \Gamma\frac{\partial n}{\partial t} - \frac{\partial^2 n}{\partial x^2} &= -\frac{\partial^2}{\partial x^2}\frac{|\psi|^4}{1+\alpha|\psi|^4}, \end{aligned} \tag{31}$$

where Γ is an operator describing the damping of acoustic perturbations.

This system of equations for $\alpha = 0$, $\Gamma = 0$ is exactly the same as the initial system (1), (2) in the sense of the laws of asymptotic behavior of the field near a singularity of the same type. Here we used a well-known procedure in the theory of wave collapse (see, for example [16]) which reduces the dimensions of the problem but retains the main physical dependences and scalings. If we consider equation (1) with an arbitrary steady-state power nonlinearity $n = |\psi|^{2s}$ and various dimensions of the transverse Laplacian d, it is easy to show that the power of the steady-state localized nonlinear mode $\psi = \psi(r)\exp(-ihz)$ and the corresponding self-similar solutions is independent of its width (amplitude) for $sd = 2$. For $s = 1$, $d = 2$ we have a standard Townes mode and collapse of beams having powers exceeding this critical power. A similar situation is achieved in the two-dimensional case (x, z) $d = 1$ for $s = 2$, i.e. when $n = |\psi|^4$. This means that the model system of Eqs. (31) can be used to investigate beam collapse with the formation of a singularity and beam decay into filaments as a result of the evolution of self-focusing instability, i.e., to some extent it can simulate the self-interaction of real three-dimensional beams. Saturation of the nonlinearity ($\alpha \neq 0$) is introduced to confine the field near the foci.

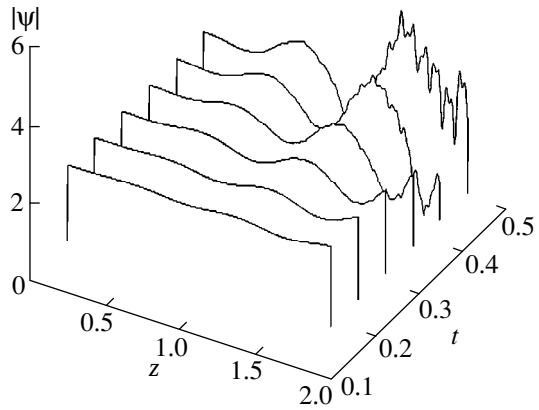

Fig. 2. Dependence of the modulus of the field on the axis of the system (the function $|\psi|(x = 0, z)$) on the longitudinal coordinate z for several successive times obtained as a result of a numerical simulation of the process of space–time instability of a plane wave with the boundary conditions $\psi(z = 0) = 2 + 0.02\cos(kx)$ and the transverse wave number $k = \pi$. As far as the time $t = 0.5$ the solution is only characterized by a shift of the maxima (see Fig. 3) and an increase in the spatial modulation and may be described using the approximate analytic solution $\psi \approx \Phi_0 + aI_0(\Phi_0(2kzt)^{1/2})\exp(ikt - ik^2z)$. From this time the solution becomes essentially nonlinear (the perturbations ψ become of the order of Φ_0).

The system of Eqs. (31) was integrated numerically using a fast Fourier transformation. The number of harmonics $nx = 512$ was selected to obtain a fairly good description of the smallest-scale spatial inhomogeneities. The boundary conditions were assumed to be periodic in x, and the calculation accuracy was monitored from the conservation of the integrals of the initial system. We shall first consider the nonlinear stage of resonant instability, then the transition process in the evolution of space–time instability and the establishment of a steady-state field distribution. This formulation of the problem corresponds to a "broad" wave beam having above-critical power. The evolution of the wave field was investigated using the same boundary distribution of the field

$$z = 0, \quad \psi = \Phi_0 + q_r\cos(kx), \tag{32}$$

as in [1, 2] in order to illustrate singularities associated with the acoustic mechanism for relaxation of the nonlinear response. The self-interaction dynamics of a localized field distribution were studied separately.

(1) The process of space–time instability of a plane wave was investigated numerically on a spatial interval having the transverse dimension $L_x = 2$ and longitudinal dimension $L_z = 1.75$. The homogeneous field distribution $\Phi_0 = 2$ was simulated by a small perturbation having the amplitude $q_r = 0.02$ and transverse wave number $k = \pi$. Figure 2 shows the field distribution on the axis of the system ($|\psi|(x = 0, z)$) at various times. The initial stage of evolution of the small perturbations is charac-

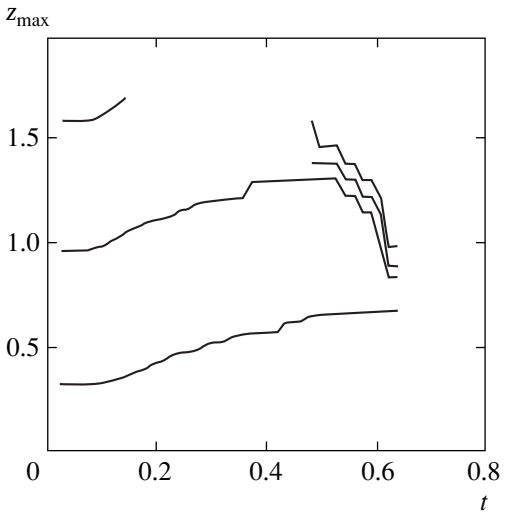

Fig. 3. Time evolution of the z coordinates of the four maxima of the function $|\psi|$ closest to the entrance to the nonlinear medium ($z = 0$). This dependence is the result of a numerical simulation of the process of space–time instability of a plane wave having the boundary conditions $[(\psi(z = 0) = 2 + 0.02\cos(kx)]$ and the transverse wave number $k = \pi$. The motion of the maxima at times $t < 0.4$ takes place at almost constant velocity.

terized by an increase in the modulation of the field. Its amplitude increases in space and time. In addition, the perturbation wave travels into the nonlinear medium, as is clearly demonstrated in Fig. 3 which gives the time dependence of the coordinates of the four maxima of the field modulus closest to the entrance to the nonlinear system (the point $z = 0$). For times $t < 0.4$ the z coordinates of the field maxima increase almost linearly with time [as for the approximate analytic solution (27)] and the spatial structure of the field always resembles (27). We note that this "similarity" of the behavior of the numerical and analytic solutions at the initial stages of the evolution is manifest despite a difference in the boundary conditions [for the analytic solution (27) the field at the boundary varies periodically with time (29) whereas in our problem $\psi(z = 0, t) = \text{const}$. At $t \approx 0.4$ the motion of the maxima becomes slower and is accompanied by the appearance of small-scale spatial modulation of the field which is subsequently amplified. Quite clearly, the field distribution becomes essentially nonlinear at these times (the amplitude of the perturbations is of the same order of magnitude as the amplitude of the unperturbed plane wave) and can no longer be described using the approximate analytic solution (27). At times $t > 0.5$ the second maximum of the field begins to move back toward the entrance to the nonlinear medium, which is very reminiscent of the time behavior of the coordinate of the beam maximum $z_{max}(t)$ obtained in the paraxial approximation (see Fig. 1). The overall space–time pattern of the initial stage of the process corresponds to the evolution of resonant instability. By the time $t \approx 0.6$ the spectrum of the field is significantly enriched in spatial harmonics and the system goes over to a nonlinear regime characterized in that the first maximum remains in its original position and its amplitude increases.

(2) A numerical investigation of the competing behavior of many spatial harmonics was made allowing for saturation of the nonlinearity ($\alpha \neq 0$). At the boundary we defined a field distribution of the type (32) having the average $\Phi_0 = 2$ modulated by the sum of several spatial harmonics having different phases and small amplitudes. This formulation of the problem can be used to analyze the dynamics of the self-interaction of wave beams having powers considerably higher than the critical value for self-focusing. The results are plotted in Fig. 4. Resonant space–time instability of a plane wave similar to that shown in Fig. 2 develops initially. This stage concludes with the formation of a "homogeneous" collapsing waveguide channel. Over its entire lifetime the channel looks corrugated. The corrugation is initially dynamic and is determined by the acoustic relaxation of nonlinearity (Fig. 4a) and is then associated with a nonlinearity saturation effect (Figs. 4b, 4c). Note that by modifying the spectrum of the initial perturbations (32) we did not succeed in achieving a smooth homogeneous channel or a traveling focus regime as in the Kerr or diffusion relaxation of the nonlinear response [1, 2].

The concluding stage involving the establishment of a steady-state pattern of wave field separation takes place as in other types of nonlinearity relaxation: structural instability develops initially, leading to perturbation of the dynamic turbulence. This stage of nonlinear interaction of electromagnetic radiation with matter accompanying Kerr relaxation of the nonlinear response of a medium was observed recently in a numerical investigation of processes for three-dimensional wave beams [17]. Short-lived localized turbulent structures of this type have recently been increasingly frequently described in the literature as flickers [11, 12]. It is important to note that the ensuing fairly chaotic spatial distribution of the field (sometimes having the outline of a "homogeneous" channel) is then gradually expelled from the boundary into the nonlinear medium (Figs. 4d, 4e) and a completely steady-state pattern shown separately in Fig. 4f forms in this region. This is naturally the same as the field distribution in other types of relaxation of the nonlinear response. However, unlike the Kerr [1] and diffusion [2] mechanisms of nonlinearity relaxation, in this case the steady-state distribution is established without any dissipation of the acoustic motion of the medium (2). The time of establishment is of the same order of magnitude as that for Kerr nonlinearity and is 1.5 orders of magnitude greater than the characteristic time for the diffusion mechanism of nonlinearity. This increase is evidently attributable to the strong "damping" of sound in the diffusion regime. The agreement between the times of establishment for the Kerr and acoustic mechanisms of relaxation is apparently arbitrary. At this point it is better to search for the reason for the appreciable increase in the time of establish-

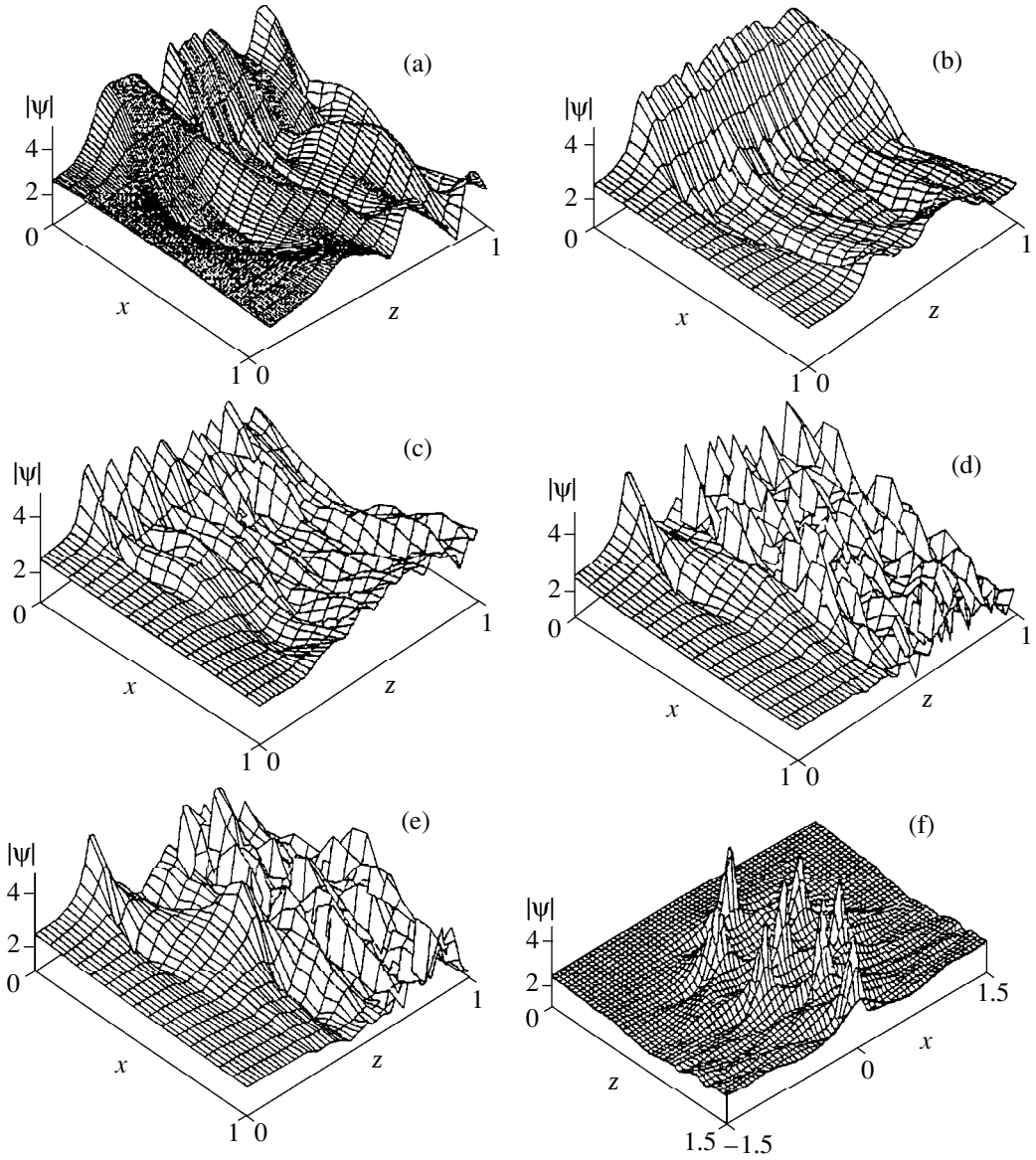

Fig. 4. Time evolution of the spatial structure of the field ψ: (a) $t = 0.843$, (b) $t = 1.56$, (c) $t = 1.875$, (d) $t = 21.875$, (e) $t = 59.375$, and (f) steady-state pattern.

ment of the steady-state pattern for Kerr nonlinearity which is more than an order of magnitude greater than the characteristic relaxation time of the nonlinear response.

(3) When the wave beam power exceeds the critical value by a moderate amount, the establishment of a steady-state pattern takes place most smoothly. In order to suppress the waves reflected from the boundaries of the calculation range, we introduced the term $i\beta(x)\psi$ in Eq. (31), which corresponds to smooth damping of the field near the boundaries of the calculation range. The calculations were made for a Gaussian beam $\psi = A\exp(-x^2/a^2)$ collimated at the entrance to the nonlinear medium ($z = 0$). For $A \approx 2$ a self-consistent distribution of the perturbations of the field and the medium was established as a result of the excitation of an intermediate wave near $z = 0$, i.e., the self-interaction dynamics had the same character as in [3, 4]. As A increased, some structural characteristics were observed.

A region of strong field forms initially and then aberrational distortions cause this region to expand mainly into the nonlinear medium and form a fairly extended homogeneous channel. Instability leads to decay of the quasi-homogeneous channel. The inhomogeneities observed as a result of this process are initially distributed regularly along the axis of the system and then their distribution varies uncontrollably until two regions of maximum field remain. The appearance

of some turbulence in the interaction zone is most likely attributed to weak reflection from the boundaries of the calculation interval. For $A > 3$ we did not succeed in completely suppressing this reflection and as a result the reliability of the numerical simulation was reduced appreciably.

At the concluding stage in the establishment of a steady-state pattern the "peak" closest to the boundary remains constant while the second peak moves into the nonlinear medium. As a result a steady-state pattern is established where the wave beam splits into two slightly diverging beams beyond the first focus and these then converge and form a region of strong field on the axis of the system.

6. CONCLUSIONS

This investigation has shown that there is a significant difference in the dynamics of the nonsteady-state self-interaction of wave beams having powers of the order of (higher than) and considerably higher than the critical self-focusing power. In the first case, the steady-state pattern is established in accordance with the scenario proposed in [3]. An intermediate wave is excited near the boundary of the nonlinear medium and moves into the medium, forming a steady-state self-interaction pattern.

The main result of this study relates to the long-term evolution of wave beams having powers significantly higher than critical. Three stages can be identified in this evolution.

Initially a self-similar distribution in the form of a homogeneous collapsing wave channel is formed and then the evolution of structural (longitudinal and transverse) beam instability (see Fig. 4d) leads to the appearance of turbulence in the nonlinear medium (uniform filling of the interaction zone with dynamic inhomogeneities of the field and the density, see Fig. 4d). Nonlinear diffraction of incoming radiation at these inhomogeneities is accompanied by the slow expulsion of the region of dynamic turbulence toward large z. As a result, a steady-state ("laminar") filamentation structure of the field distribution is established in a time almost two orders of magnitude longer than the relaxation time of the striction nonlinearity (the time taken for sound to pass through the transverse dimension of the wave beam).

Note that this scenario for the establishment of a steady-state pattern (the formation of speckle inhomogeneities and their subsequent expulsion from the front boundary of the nonlinear medium) is fairly typical for high supercriticalities and is achieved in the Kerr and diffusion mechanisms for the relaxation of the nonlinear response [1, 2]. In our particular "conservative" case of acoustic relaxation of the nonlinearity this process is slower.

ACKNOWLEDGMENTS

The authors are grateful to E.I. Rakova and A.A. Balakin for assistance with formulating the study. This work was supported by the Russian Foundation for Basic Research (project nos. 98-02-17205 and 99-02-16399).

REFERENCES

1. A. G. Litvak, V. A. Mironov, and A. M. Sergeev, Phys. Scr., T **30**, 57 (1990).
2. A. G. Litvak, V. A. Mironov, and A. M. Sergeev, in *Nonlinear Waves 3*, Ed. by A. V. Gaponov-Grekhov, M. I. Rabinovich, and J. Engelbrecht (Springer-Verlag, New York, 1990), p. 240.
3. N. E. Andreev, L. M. Gorbunov, S. V. Tarakanov, *et al.*, Phys. Fluids B **5**, 1986 (1993).
4. N. E. Andreev, L. N. Gorbunov, A. I. Zykov, *et al.*, Zh. Éksp. Teor. Fiz. **106**, 1676 (1994) [JETP **79**, 905 (1994)].
5. Y. R. Shen, *Principles of Nonlinear Optics* (Wiley, New York, 1984; Nauka, Moscow, 1989).
6. A. G. Litvak, in *Reviews of Plasma Physics*, Ed. by M. A. Leontovich (Atomizdat, Moscow, 1980; Consultants Bureau, New York, 1986), Vol. 10.
7. S. N. Vlasov, V. A. Petrishchev, and V. I. Talanov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **14**, 1353 (1971).
8. S. N. Vlasov and V. I. Talanov, *Self-Focusing of Waves* (Inst. Prikl. Fiz. Ross. Akad. Nauk, Nizhni Novgorod, 1997).
9. G. M. Fraĭman, Zh. Éksp. Teor. Fiz. **88**, 390 (1985) [Sov. Phys. JETP **61**, 228 (1985)].
10. E. L. Kerr, Phys. Rev. A **4**, 1195 (1971).
11. R. L. Berger, B. F. Lasinski, T. B. Kaiser, *et al.*, Phys. Fluids B **5**, 2243 (1993).
12. H. A. Rose, Phys. Plasmas **2**, 2216 (1995).
13. M. Lontano, A. M. Sergeev, and A. Cardinali, Phys. Fluids B **1**, 901 (1989).
14. L. M. Gorbunov and A. S. Shirokov, Kvantovaya Élektron. (Moscow) **12**, 146 (1985); S. V. Tarakanov, Fiz. Plazmy **14**, 487 (1988) [Sov. J. Plasma Phys. **14**, 288 (1988)].
15. E. V. Vanin, V. A. Mironov, E. A. Pyan'kina, *et al.*, Fiz. Plazmy **17**, 821 (1991) [Sov. J. Plasma Phys. **17**, 480 (1991)].
16. V. E. Zakharov, A. G. Litvak, E. I. Rakova, *et al.*, Zh. Éksp. Teor. Fiz. **94** (5), 107 (1988) [Sov. Phys. JETP **67**, 925 (1988)].
17. M. Mlejnek, M. Kolesik, J. V. Moloney, and E. M. Wright, Phys. Rev. Lett. **83**, 2938 (1999).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 845–858.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 979–995.

GRAVITATION, ASTROPHYSICS

Search for Astro-Gravitational Correlations

V. N. Rudenko*, A. V. Gusev, V. K. Kravchuk, and M. P. Vinogradov**

Sternberg Astronomical Institute, Moscow State University, Moscow, 119899 Russia

**e-mail: rvn@sai.msu.ru*

***e-mail: avg@sai.msu.ru*

Received June 5, 2000

Abstract—A special arrangement of a gravity-wave experiment, in which the noise background of the gravity detector is investigated near time markers corresponding to the detection of astrophysical events accompanying neutron or gamma bursts, is studied. A general algorithm is developed for analyzing the traces for the case of resonant solid-state detectors. The efficiency of the algorithm is demonstrated in a reanalysis of old data concerning the "neutron-gravity correlation" effect associated with the explosion of the SN1987A Supernova. Modifications of the algorithm for searching for gamma-gravity correlations are proposed. *© 2000 MAIK "Nauka/Interperiodica".*

1. RADIATION EFFECTS OF RELATIVISTIC ASTROPHYSICAL EVENTS

The conventional arrangement of a gravity-wave experiment searching for random bursts of gravitational radiation from astrophysical sources presupposes detection of the coincidence of signals from two or more spatially separated gravity detectors. For good isolation of the detectors this method has been considered for a long time to be the only method for establishing the global nature of the detected signal [1]. In the automatic mode events have been detected only by Weber during his first experiments using uncooled resonant detectors, located in Chicago and Maryland [2]. Subsequently, a number of groups have searched for coincidences not in real time but rather a posteriori, i.e., by analyzing the electronic traces obtained using antennas. The last experiment of this kind, using cryogenic antennas Explorer and Allegro, is described in [3]. At the present time the a posteriori method is actually the only real possibility of searching for coincidences, since there are no systems for synchronized coupling between several antennas, forming a "worldwide network" (automatic search for events in real time would possibly be organized in the LIGO project [4] on two large interferometric antennas after they come on line).

In the present paper a different form of a gravity-wave experiment is examined. The idea is to search for weak perturbations of the noise background of a gravity detector which are correlated with certain astrophysical events, such as neutrino and gamma bursts [5–7]. This method has attracted attention because of the recognition that the last stages of stellar evolution involve an explosion of a Supernova, merging of binary stars, collapse of single stars, and so on, traditionally viewed as sources of gravity pulses, should be accompanied also by neutrino and, very likely, gamma radiation. Generally speaking, this means that the detection of neutrino and gamma bursts, using appropriate detectors, determines the time intervals near which it makes sense to search for a perturbation of a gravity detector. The advantages of this approach over the conventional method are, in the first place, that the observation time is shorter and the sources are identified with greater certainty and, in the second place, that there is a potential possibility of accumulating weak signals. The latter is especially important because currently existing gravity detectors lack adequate sensitivity.

As regards radiation scenarios, it should be noted that the neutrino bursts produced by collapsing stars at the end of their evolution have been studied quite well theoretically [8–10]. According to theory, the total energy released in the form of neutrinos (of all flavors) is of the order of $0.1M_{\odot}c^2$ with time scales of several seconds (2–20 s). This radiation can be detected (primarily on account of the inverse β-decay reaction) if the source is not too far from the Earth (10–100 kpc). The corresponding programs "Collapse search" (or "Supernova Watcher") are being conducted by all research groups which have suitable liquid scintillation detectors [11, 12] or water Cherenkov detectors [13, 14]. Moreover, a neutrino flux from a Supernova was probably detected during the burst of the SN1987A Supernova [15, 16]. However, such programs are search for collapsing stars in our and neighboring galaxies, i.e., the expected average frequency of events is three events per 100 yr. Even the large detector Super Kamiokande with an effective mass an order of magnitude greater than the mass of other setups makes it possible to detect in principle 150 neutrinos per year from the Large Magellanic Cloud (LMC) and only one from the Andromeda Nebula [17]. It is unlikely that neutrinos will be detected from the Supernova in the Virgo cluster

(15–20 Mpc), which is viewed as the main source of gravity signals. Thus, the search for correlations between the noise background of neutrino and gravity detectors is limited by the condition of low frequency of events (3–10 events per 100 years). Hence there is virtually no possibility of increasing the signal/noise ratio by integrating the signals. At the same time, the expected amplitude of an individual gravity pulse can be relatively large, up to 10^{-18} from a source located at the center of the Galaxy (when describing the amplitude in units of the perturbation of the metric).

Another astrophysical phenomenon is more attractive but less well determined—gamma bursts, concerning whose nature only hypotheses exist at the present time [18]. The main favorable feature of this phenomenon is its relatively high frequency: on the average one event per day. The high energy of the detected gamma bursts (up to $0.1M_{\odot}c^2$) together with the short time scales means that relativistic stars are the source. Two main ideas concerning the nature of the gamma bursts are being discussed so far. The first idea examines their possible galactic origin taking account of the fact that the sources, fast pulsars, are distributed not only over the galactic disk but they also occur in the halo [19]. The second (and more popular) idea is associated with the cosmological nature of gamma bursts, which are a result of catastrophic processes with relativistic stars in distant galaxies (the above-mentioned collapses, merging of binary stars, explosions of Supernova) [20]. Thus, both scenarios are concerned with objects which are also typical sources of bursts of gravity waves. The following is known about their possible intensity.

Galactic pulsars can produce only weak gravity-wave bursts as a result of "starquakes" with a corresponding perturbation of the metric at the Earth of the order of 10^{-24}–10^{-23} [21] for a source at the center of the galaxy. In [22, 23] it is conjectured that closer pulsars, near 100 pc, can give the observed frequency of gamma events (as a consequence of "star quakes") ≈5 per month. In this case the expected amplitude of the corresponding "close" gravity-wave bursts will be 10^{-22}–10^{-21}. In the cosmological scenario, if binary stars with black holes are considered, the predicted frequency of gravity-wave bursts can reach 30 events per year with a metric perturbation amplitude 10^{-21} near 50–100 Mpc [24, 25]. These estimates were made assuming that only 10^{-4} of the rest mass of a star can be converted into gravitational radiation. A more optimistic value of the conversion factor 10^{-2} is given in other works [26, 27]; this should increase the perturbation amplitude up to 10^{-20}.

It is well known that the new observations obtained on the BerroSAX satellite paired with the Keck II telescope on the identification of several gamma bursts from distant galaxies, for which the red shift $z \approx 0.8$–3.4 [28], confirm the cosmological nature of at least some of the gamma events. However, more distant events have also been detected together with these strongly distant sources (1–10 Gpc). The burst GRB980425 can be tentatively attributed to an optical object, such as Supernova burst, located at a distance of 40 Mpc (z = 0.08) [29]. Although it is not entirely clear how gamma radiation passes through the shell of a Supernova or how the merging of black holes can produce a gamma burst, the energetics of the observed events definitely requires scenarios associated with relativistic catastrophes (an energy estimate for GRB971214 gives $\sim 2 \times 10^{53}$ ergs, which, in general, is greater than the typical energy from the explosion of a Supernova or from the merging of binary neutron stars 10^{51} ergs [30]. Models with binary black holes or a rapidly rotating massive black hole with accretion (so-called "hypernova") are required [31]).

An adequate network of detectors for implementing programs searching for astro-gravitational correlations already exists. There are four operating neutrino telescopes and two cosmic satellites (CGRO (BATSE) and BeppoSAX), which provide a list of "productive" astrophysical events. The key question concerns the sensitivity of the gravitational detectors, which are in a continuous-observation mode. Only the supercryogenic resonance detectors Nautilus (INFN, Frascati) and Auriga (NFN, Legnaro) can attain a sensitivity of 10^{-21} for detecting short bursts $\sim 10^{-3}$ s [32]. The "burst" sensitivity of the previously mentioned helium–temperature detectors Allegro and Explorer is 6×10^{-19}, i.e., 2.5 orders of magnitude lower than the desired value. Nonetheless, it should be kept in mind that their "effective sensitivity" increases to 10^{-21} for longer events (up to 1 s) and also if equivalent accumulation of short signals is possible (an explanation is given in [33]).

Actually, the first works searching for astro-gravitational correlations are the work of the PTM collaboration on the investigation of a correlation of the traces of the neutron scintillator under Mont Blanc and the gravitational detectors in Rome and Maryland [11, 15, 34, 35]. Here a "signal accumulation algorithm" optimized for Gaussian noise of gravitational detectors was proposed. The same problem in application to astrophysical events with gamma bursts has been studied in recent publications [36, 37]. The authors of [36] performed a numerical experiment simulating the flux of gamma bursts from cosmological objects with prescribed spatial distribution in order to estimate the number of events for which the "accumulated gravitational signal" should exceed the detection threshold (i.e., it is detected) for cryogenic resonant antennas. The accumulation rule (the correlation search algorithm), however, was not specified, and the hypothesis adopted concerning the rate of growth of the signal/noise ratio as $\sqrt{N}$, where N is the number of events taken into account, is valid only for a noiseless gravitational antenna.

In [37] the check of the hypothesis that "gamma-gravitational coupling" exists was studied in application to the mutual cross-correlation function of two separated interferometric antennas in the LIGO project.

The question of the difference of the average values of the cross-correlation variable, measured in the vicinity of and away from gamma bursts, was solved on the basis of Student's *t* criterion. The reliability of the "difference" argument in this case actually increases as $\sqrt{N}$, but the argument for the "optimality" of the chosen "cross-correlation variable" remains outside the framework of the paper. In addition to the theoretical works, the first experimental attempts have already been made to observe the "gamma–gravitational correlations." Thus, a search for correlations between BATSE events and noise in the Explorer cryogenic detector was reported in [38, 39]: there was no effect at the amplitude level $(2–3) \times 10^{-18}$.

The solution of the problem of optimal accumulation of weak gravitational signals associated with astrophysical events depends on our knowledge of their structure, arrival time, and so on. The theory does not offer us a large choice for the forms of the gravitational signal. In most cases the signal at the maximum intensity level can be represented in the form of short pulses with a carrier frequency 10^2–10^3 Hz [21]. The simultaneous description of the gravitational, neutrino, and gamma radiation has been developed in detail for only a few scenarios of relativistic events. The most definite model is the model of merging of a relativistic binary star, where the gravity-wave burst is definitely expected first (on an inwardly spiraling trajectory) and then, after merging has occurred, neutrino and gamma bursts can appear [40]. More complicated scenarios are presented, for example, in [21, 26, 27, 41], where the processes leading to the formation of a neutron star and the interaction of stellar residues are considered as multistage gravitational collapse. A multistage scenario is also proposed for the collapse of a massive star with large initial angular momentum [21]. The matter can be prevented from falling toward the center by reflected shock waves, fragmentation, merging or emergence of some of the fragments, and so on. In principle, gravity, electromagnetic, and neutrino bursts can be generated at each stage of these scenarios. A detailed description and a strict temporal arrangement does not yet exist for them.

The brief review of the situation with a new trend in gravity-wave experiments makes it possible to formulate the problem addressed in this paper. Traces of gravitational, neutrino, and gamma detectors are available. It is necessary to check the presence of any relationship between these data. A simple comparison does not give the desired result because the arrival time of signals of different nature is unknown, because of the relative insensitivity of the detectors [34, 42], and so on. The data must be analyzed using the methods of optimal filtering, keeping in mind information about the noise background of the receivers, the structure of the signal, and other information [43]. The specific goal is to determine the optimal algorithm for searching for the correlation of neutrino (or gamma) events with perturbations of the resonance gravity detector. The noise statistics of detectors of this type is determined more accurately than the noise of a gravity antenna with free masses. After a possible algorithm is formulated, the application of this algorithm is considered for reanalysis of old data concerning the "neutrino–gravity correlation effect," noted in the PTM collaboration during the burst of the Supernova SN1987A. As is well known, the authors of [15, 34, 35, 44] did not find a clear astrophysical interpretation of the correlations which they observed; our reanalysis sheds light on the reason for these difficulties. An extension of the correlation algorithm to the case of gamma events with a complicated structure is discussed in the last section of this paper.

2. MAXIMUM LIKELIHOOD ALGORITHM FOR DETECTION OF AN INCOHERENT GRAVITY-WAVE PACKET

In accordance with the astrophysical picture, we shall examine a model of a signal with the conventional name "incoherent pulse packet." The gravity-wave perturbation is given as a random group of short pulses with two main parameters: the arrival time τ_i and the amplitude a_i. The intervals between individual pulses can vary over wide limits in accordance with the frequency and nature of astrophysical events. The form of an individual pulse is ignored, and the duration $\hat{\tau}$ is assumed to be short, including only several periods of the carrier frequency ω, so that

$$\omega\hat{\tau} \sim 1, \quad \omega \sim \omega_0 \pm 1/\hat{\tau},$$

where ω_0 is the central frequency of the pulse spectrum. The random background is determined by the noise of the gravity resonance antenna. A modern antenna consists of a cooled solid-state detector, an electromechanical transducer (as a reading device), an amplifier, and a preliminary filtering circuit with a bounded frequency band $\Delta\omega \le \hat{\tau}^{-1}$, which is implemented by a "difference unit" or a Wiener–Kolmogorov filter, and so on. The simplest approximation in describing the noise of such a system is a Gaussian model (assuming perfect acoustic, seismic, and electric insulation of the antenna).

Let us consider the problem of optimal detection. The output signal $x(t)$ of the antenna is an additive mixture of the noise $\xi(t)$ and the signal $S(t)$, which is a random sequence of short pulses s_k, i.e.,

$$x(t) = \lambda S(t) + \xi(t), \quad S(t) = \sum_k s_k(t). \qquad (1)$$

It is assumed that $\xi(t)$ is stationary Gaussian noise with spectral density $W(\omega)$, determined by the antenna structure; $\lambda = (1, 0)$ is a parameter that determines the presence or absence of a signal. An individual pulse in the

sequence $S(t)$ can be represented in a complex form as follows:

$$\begin{aligned} s_k(t) &= \mathrm{Re}[\tilde{s}_k(t)\exp(j\omega_0 t)], \\ \tilde{s}_k(t) &= a_k\tilde{H}(t-t_k)\exp(j(\Theta_k-\omega_0 t_k)). \end{aligned} \tag{2}$$

The symbols $\tilde{s}_k(t)$ and $\tilde{H}(t)$ introduced here are, respectively, the complex amplitude of the "gravity bursts" and the pulse characteristic (Green's function) of a linear antenna

$$\begin{aligned} H(t) &= \mathrm{Re}[\tilde{H}(t)\exp(j\omega_0 t)] \\ &= H_0(t)\cos[\omega_0 t + \psi(t)] \end{aligned}$$

with resonance frequency ω_0.

Expression (2) contains a third signal parameter—the initial phase Θ_k. The amplitude parameter a_k, if it is small, does not have a large effect on the structure of the data processing algorithm. The effect of the two other parameters—the initial phase and the arrival time—is significant. The arrival time of the gravity bursts, according to the "astro-gravitational correlation hypothesis," should be localized near the astrophysical events:

$$t_k = t_{ak} + \tau, \quad k = [1, 2, \ldots, n]. \tag{3}$$

Here t_{ak} are the markers of the "astrophysical events," whose total number is n in the observation interval $[0, T]$, and τ is an unknown shift between the "astrophysical" and "gravitational" events. The admissible value of this shift must be limited a priori by some interval $(\tau_{\min}, \tau_{\max})$ which is assumed to be given. The algorithm for optimal processing of the output data can be synthesized on the basis of the maximum likelihood (ML) principle. According to this principle, it is possible to construct a variable (which is a function of the output signal $x(t)$) maximizing which can give the highest probability for detecting the a priori signal according to the realization $x(t)$ taking account of the existing a priori information on the observed interval $[0, T]$. For a signal on a background of Gaussian noise the solution is known: the ML variable z is proportional to the logarithm of the functional of the likelihood ratio $\Lambda[x]$ [43, 45]:

$$\Lambda[x] = \left\langle \exp\left[\int_0^T x(t)u(t)dt - \frac{1}{2}\int_0^T S(t)u(t)dt\right]\right\rangle, \tag{4}$$

where the reference function $u(t)$ is the solution of the integral equation

$$\int_0^T K_\xi(t-\tau)u(\tau)d\tau = S(t), \quad 0 < t < T, \tag{5}$$

where $K_\xi(t)$ is the correlation function of the process $\xi(t)$. The brackets $\langle\ldots\rangle$ denote statistical averaging over the signal parameters.

We introduce one more assumption about the signal: the pulses in a packet are quite sparse and cannot overlap with one another in time. This agrees with the idea that the frequency of relativistic events is relatively low and makes it possible to represent the functional of the ratio of the likelihood in a factorized form:

$$\Lambda[x] = \prod_{k=1}^{n}\Lambda_k[x], \tag{6}$$

where Λ_k is the ratio of the likelihood of an isolated kth pulse. In Eqs. (4) and (5) $u_k(t)$ must be substituted for $u(t)$ and $s_k(t)$ must be substituted for $S(t)$. Then, finally, the ML variable can be represented as

$$Z = \sum_{k=1}^{n}\ln\Lambda_k[x] = \sum_{k=1}^{n} z_k, \quad z_k = \ln\Lambda_k[x]. \tag{7}$$

Equations (4) and (5), rewritten for an individual pulse $s_k(t)$, and Eq. (7) represent a general solution obtained by the maximum likelihood method for an incoherent packet of pulses against the background of Gaussian noise. For practical applications, the reference function $u_k(t)$ must be found in an explicit form, after which the value z_k is calculated.

Since the correlation time of the detector noise (just as the duration of a pulse) is much shorter than the observation interval T, the upper limit of integration in Eq. (5) can be replaced by infinity. Then we obtain the ratio

$$u_k(\omega) \approx \frac{s_k(\omega)}{N_\xi(\omega)}, \tag{8}$$

where $u_k(\omega)$, $s_k(\omega)$, and $N_\xi(\omega)$ are the Fourier transforms of $u_k(t)$, $s_k(t)$, and $K_\xi(t)$, respectively. Using Parseval's equation

$$\int_{-\infty}^{\infty} a(t)b(t)dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} a(\omega)b^*(\omega)d\omega.$$

Equation (4) can be rewritten for an individual pulse in the following form:

$$\begin{aligned} \Lambda_k[x] \approx \left\langle \exp\left[\mathrm{Re}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} x(\omega+\omega_0)\frac{\tilde{s}_k^*(\omega)}{N_\xi(\omega+\omega_0)}d\omega\right.\right.\right. \\ \left.\left.\left. - \frac{1}{8\pi}\int_{-\infty}^{\infty}\frac{|\tilde{s}_k^*(\omega)|^2}{N_\xi(\omega+\omega_0)}d\omega\right)\right]\right\rangle. \end{aligned} \tag{9}$$

Here $\tilde{s}_k(\omega)$ is the Fourier transform of the complex amplitude $\tilde{s}_k(t)$. After substituting Eq. (2) into Eq. (9),

the latter becomes

$$\Lambda_k[x] \approx \left\langle \exp\left[a_k \mathrm{Re}\left(\exp(-j\chi_k)\frac{1}{2\pi} \right.\right.\right.$$
$$\times \int_{-\infty}^{\infty} x(\omega+\omega_0)\frac{\tilde{H}^*(\omega)}{N_\xi(\omega+\omega_0)}\exp(j\omega t_k)\,d\omega \qquad (10)$$
$$\left.\left.\left. - a_k^2\frac{1}{8\pi}\int_{-\infty}^{\infty}\frac{|\tilde{H}(\omega)|^2}{N_\xi(\omega+\omega_0)}d\omega \right)\right]\right\rangle,$$

where

$$\tilde{H}(\omega) \longleftrightarrow \tilde{H}(t), \quad \chi_k = \omega_0 t_k - \Theta_k.$$

We now express Eq. (10) in terms of the antenna output variable $\tilde{y}(t)$, which is a result of the passage of the input signal $x(t)$ (1) through an optimal filter with transfer function

$$K_{\mathrm{opt}} = \frac{H^*(\omega)}{N_\xi(\omega)}\exp(-j\omega t_0),$$

where t_0 is the time delay of the filter:

$$\Lambda_k[x] \approx \langle \exp\{a_k \mathrm{Re}[e^{j\psi_k}\tilde{y}(t_k)] - a_k^2\sigma^2/2\}\rangle. \qquad (11)$$

Here

$$\psi_k = \omega_0(t_0 + t_k) - \Theta_k,$$
$$\sigma^2 = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{|K_{\mathrm{opt}}(\omega)|^2}{N_\xi(\omega)}d\omega$$

is the variance of the output noise.

Introducing the amplitude of the output signal

$$A_k = |\langle \tilde{y}(t_k)\rangle| = a_k\sigma^2,$$

we obtain the final expression for the kth likelihood ratio:

$$\Lambda_k[x] \approx \left\langle \exp\left\{ \frac{A_k}{\sigma^2}\mathrm{Re}[\exp(j\psi_k)\tilde{y}(t_k)] - \frac{A_k^2}{2\sigma^2} \right\}\right\rangle. \qquad (12)$$

This formula contains the unknown pulse parameters A_k and Θ_k. The uncertainty can be removed by using the "generalized form of the ML algorithm" [45], where the unknown parameters are replaced by the "maximum likelihood estimates" $\hat{A}_k$ and $\hat{\Theta}_k$, which are taken as the solution of the following extremal equations:

$$\frac{\partial z_k}{\partial A_k} = 0, \quad \frac{\partial z_k}{\partial \Theta_k} = 0, \qquad (13)$$

where

$$z_k = \ln\Lambda_k[x].$$

A direct calculation leads to the following results.

(a) The ML estimate of the amplitude is identical to the envelope of a narrow-band output process $R(t)$ of the antenna:

$$\hat{A}_k^2 = \{\mathrm{Re}[\exp(j\hat{\psi}_k)\hat{y}(t_k)]\}^2 = |y(t_k)|^2 = R^2(t_k). \qquad (14)$$

Then

$$z_k = \frac{\hat{A}_k^2}{2\sigma^2} = \frac{R^2(t_k)}{2\sigma^2}. \qquad (15)$$

(b) The ML estimate of the unknown shift τ between the time marker of the "astrophysical" event t_{ak} and the arrival time of the "gravity" signal t_k is determined by the position of the maximum of the function $z_k(t_{ak}+\tau)$ as a function of τ.

The results (a) and (b) correspond to the assumption that the values of the signal parameters are unknown but are not random. In a different approach the initial phase Θ_k is treated as a random variable uniformly distributed on the interval $[0, 2\pi]$. Statistical averaging gives

$$\langle \Lambda_k[x]\rangle = \exp[-(\hat{A}_k^2/2\sigma^2)]I_0(\hat{A}_k R(t_k)/\sigma^2), \qquad (16)$$

where the amplitude $\hat{A}_k$ is the solution of the equation

$$\hat{A}_k = R(t_k)\frac{I_1[\hat{A}_k R(t_k)/\sigma^2]}{I_0[\hat{A}_k R(t_k)/\sigma^2]} \qquad (17)$$

(I_0 and I_1 are modified Bessel functions). In this case the ML variable reduces to the form

$$z_k = [\ln I_0(\hat{A}_k R(t_k)/\sigma^2) - \hat{A}_k^2/2\sigma^2]. \qquad (18)$$

The solution of Eq. (17) is shown in Fig. 1. It is obvious that the difference in estimates (14) and (17) is substantial, when the amplitude of the signal is small ($A_k < \sigma$). For amplitudes $A_k \geq 2\sigma$, the estimates are essentially

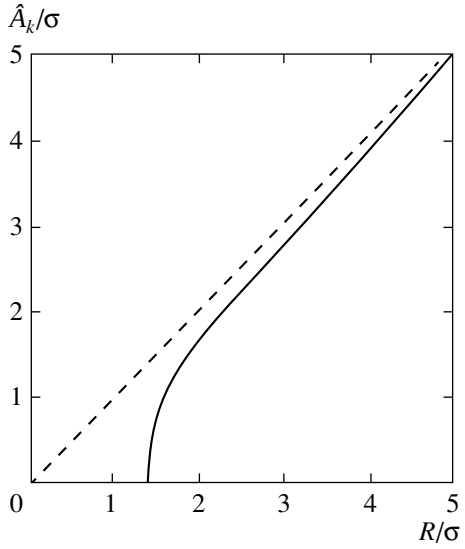

Fig. 1.

identical and it is recommended that the value of the envelope $R(t_k)$ be taken as the ML estimate of A_k; the difference between variables (15) and (18) is likewise negligible.

Thus, for signals that are not too weak, in the model of an incoherent packet of pulses against the background of narrow-band Gaussian noise the ML algorithm recommends using in Eq. (7) the variable

$$Z = \sum_{k=1}^{n} (R^2(t_k)/2\sigma^2), \tag{19}$$

which is a sum of the squared measurements of the envelope which are shifted relative to the astrophysical event by τ (3); the summation extends on an interval of observations which includes a posteriori n events. Next, similarly to Eq. (13), the absolute maximum of Z can be found in the space of shifts τ, i.e., it is possible to switch to the "absolute maximum variable"

$$Z_{\max} = \max_{\tau} Z(\tau), \quad \tau \in [\tau_{\min}, \tau_{\max}]. \tag{20}$$

The value of τ_{opt} which gives the maximum of $Z(\tau)$ must be taken as the ML estimate of the time shift between the astrophysical event and the gravitational signal (in our simple approach the shift is assumed to be the same for all events—this is the "uniformity of events" hypothesis). As noted above, in the statistical model there are no indications for the limits of variation of τ; they must be found from additional considerations.

Under conditions of insufficient a priori information, the gravitational data in the presence of a list of astrophysical events is processed, naturally, on the basis of the Neiman–Pearson strategy. After forming the variables $Z_{\max}$ it must be compared with the threshold, which depends on the statistics of the variable. Crossing a threshold will signify "presence of a signal" with reliability $(1 - \alpha)$, where α is the probability of a "false alarm." This strategy presupposes that the statistics of $Z_{\max}$ is known; this can be found from theory or empirically, i.e., by studying the noise characteristics of the output realization of an antenna.

3. STATISTICAL PROPERTIES OF THE VARIABLES EMPLOYED

Maximum likelihood data processing employs three variables: the squared envelope $R^2(t)$ (14); the sum of the squared readings Z (19) taken near the astrophysical events in the interval $[0, T]$; and, the maximum value of this sum $Z_{\max}$ (20), corresponding to the optimal time shift. The statistics of each of these three variables can be calculated analytically, using the hypothesis that the noise $\xi(t)$ of the gravitational detector is Gaussian. Experiment shows that for uncooled antennas this assumption is close to reality for thresholds whose energy is not too high.

The formulas (19) and (20) were written in a dimensionless form. For comparison with experiment, it is helpful to have also relations where the variables studied are expressed in degrees Kelvin.

3.1. Statistics of the Squared Amplitude Readings

The thermal oscillations of the resonance solid-state detector can be described by a narrow-band Gaussian random process

$$x(t) = A(t)\cos\omega_0 t - B(t)\sin\omega_0 t$$

with slowly varying quadratures $A(t)$ and $B(t)$. Their correlation function is

$$k(\tau) = \sigma_0^2 \exp(-\gamma|\tau|),$$

where

$$\sigma_0^2 = kT_0/m\omega_0^2$$

is the Brownian variance,

$$\gamma = \omega_0/2Q$$

is the relaxation constant (m, T_0, and Q are, respectively, the mass, absolute temperature, and mechanical Q of the detector).

After pre-filtering (the differentiating unit, the Weiner–Kolmogorov filter, and so on) $x(t)$ becomes a narrow-band process inside a limited frequency band $\Delta\omega$ with the squared envelope

$$R^2 = (\Delta A)^2 + (\Delta B)^2.$$

This quantity is proportional to the variation of the energy $E_t(t)$ of the detector over a time $\Delta t = \Delta\omega^{-1} \ll \gamma^{-1}$:

$$E(t) = m\omega_0^2 R^2(t)/2, \quad \langle E(t)\rangle = kT_0 2\gamma\Delta t.$$

The correlation function of the corresponding variation of the quadratures ΔA or ΔB is of the type

$$k_\Delta(\tau) = \sigma^2\rho(\tau),$$

where

$$\rho = \begin{cases} 1 - (|\tau|/\Delta t), & |\tau| \le \Delta t \\ 0, & |\tau| > \Delta t. \end{cases}$$

The variance σ^2 is related with the Brownian variance σ_0^2 via the effective noise temperature T_e of the detector:

$$\sigma^2 = \frac{kT_e}{m\omega_0^2} = \sigma_0^2\frac{T_e}{T_0}, \quad T_e = T_0(2\gamma\Delta t).$$

The correlation function of the energy variation

$$K(\tau) = \langle E(t)E(t+\tau)\rangle - \langle E(t)\rangle^2$$

has the form

$$K(\tau) = 4(kT_e)^2\rho^2(\tau). \tag{21}$$

The correlation coefficient $\rho^2(\tau)$ in this formula decreases to zero on time scales Δt. In the discrete representation, $E(t) \longrightarrow E(t_k)$, the readings are independent for $(t_{k+1} - t_k) \geq \Delta t$.

3.2. Statistics of the Sum of the Squared Amplitude Readings

The next variable which of interest is the sum of the squared amplitude readings taken at the moments of the astrophysical events (19). For convenience, this variable can be normalized by dividing by the total number of events in the observed interval $[0, T]$. The new variable $C = Z/n$ is proportional to the "selected average value" of the energy variations

$$\bar{E} = \frac{1}{n}\sum_{k=1}^{n} E(t_k)$$

on the observable interval at the sampling moments of the astrophysical events:

$$C = \frac{Z}{n} = \frac{1}{nkT_e}\sum_{k=1}^{n} E(t_k) = \frac{1}{kT_e}\tilde{E}. \qquad (22)$$

If the number of events in sum (22) is greater than 30, then according to the central limit theorem the variable C has asymptotically a Gaussian distribution with average value $\langle C(t)\rangle = 1$ or $\langle \bar{E} \rangle = kT_e$. The correlation function

$$K_c = \langle C(t_1, t_2 \ldots t_n) C(t_1 + \tau, t_2 + \tau, \ldots t_n + \tau)\rangle - \langle C(t_1, t_2 \ldots t_n)\rangle^2$$

has the form

$$K_c(\tau) = \frac{1}{n^2}\left[n\rho^2(\tau) + \sum_{i=1}^{n}\sum_{k=1,\, k\neq i}^{n} \rho^2(t_i - t_k + \tau)\right],$$

i.e., there is a principal maximum in the range $0 \leq \tau \leq \Delta t$, decreasing parabolically, and there is a series of side maxima at the points $\tau = (t_i - t_k)$. This leads to a specific determination and calculation of the "correlation time" for the variable C. Assuming that the sequence of astrophysical events is a Poisson flux of pulses, $\rho_c(\tau)$ can be put into the form (provided that $\Delta t \leq \tau \leq T$)

$$K_c(\tau) = \frac{1}{n}\left[\rho^2(\tau) + \pi^{-1}(n-1)\frac{\Delta t}{T}\left(1 - \frac{|\tau|}{T}\right)\right]. \qquad (23)$$

The total number n of events in the observation time T is not expected to be too large, $n(\Delta t/T) \ll 1$ and the correlation time for $C(\tau)$ is bounded $|\tau_c| \ll T$. Then, Eq. (23) simplifies:

$$K_c(\tau) \approx \frac{\rho^2(\tau)}{n} \longrightarrow K_{\bar{E}}(\tau) = \frac{\rho^2(\tau)(kT_e)^2}{n}. \qquad (24)$$

The variance of the variable C (just as $\bar{E}$) is less than the variance of R^2 by the factor $1/n$, which corresponds to the properties of the sum of independent readings.

3.3. Statistics of the Absolute Maximum

Studying the normalized sum of the squared amplitude readings, instead of $Z_{\max}$ (20) we must consider

$$C_{\max} = \max_{\tau} Z(\tau)/n.$$

The search for a maximum is made by varying the time shift on an a priori prescribed interval (20). Let the time shift be discretized with step δt. Then there is a combination of L values

$$\{C(t_{ak} + m\delta t)\}, \quad m = 1, 2 \ldots L,$$
$$L = \frac{\tau_{\max} - \tau_{\min}}{\delta t} = \frac{\Delta\tau}{\delta t}.$$

If the variable C is Gaussian, the distribution $C_{\max}$ can be found in the literature. Specifically, the Cramer formula can be used [46]. This formula represents the statistics of $C_{\max}$ in terms of the auxiliary random variable ξ:

$$C_{\max} - \langle C\rangle \approx \sqrt{\frac{1}{n}}\left[\sqrt{2\ln\mu(\Delta\tau)} + \xi/\sqrt{2\ln\mu(\Delta\tau)}\right] \qquad (25)$$

with probability density

$$\omega(\xi) = e^{-\xi}\exp(-e^{-\xi}), \qquad (26)$$

where the average value $\langle\xi\rangle = 0.577$ and the variance is $\sigma_\xi^2 = \pi^2/6$. The formulas (25) and (26) are valid asymptotically when

$$L = \Delta\tau/\delta t \longrightarrow \infty.$$

The parameter μ in Eq. (25) depends on the variation of the time shift $\Delta\tau = \tau_{\max} - \tau_{\min}$ and the second derivative of the correlation coefficient $\rho^2(\tau)$ of the variable C (24) at the point $\tau = 0$:

$$\mu(\Delta\tau) = \frac{1}{2\pi}\Delta\tau\sqrt{-2\rho''(0)},$$
$$\rho''(0) = \left[\frac{d^2\rho^2(\tau)}{d\tau^2}\right]_{\tau=0}. \qquad (27)$$

The calculation of the value of $\rho''(0)$ for Markov processes is always nontrivial. In our case the estimate can be obtained in terms of the approximation using Owen functions [47]:

$$\mu(\Delta\tau) = \frac{1}{\pi}\frac{\Delta\tau}{\delta t}\sqrt{\frac{1-\rho^2}{1+\rho^2}}, \quad \rho^2 = \left(1 - \frac{\delta t}{\Delta\tau}\right)^2. \qquad (28)$$

Formulas (25), (26), and (28) solve the problem of calculating the "case probability" where $C_{\max}$ is greater than a certain chosen threshold C_{th}.

4. NEUTRINO–GRAVITY CORRELATION EFFECT FROM SN1987A

We shall now consider the application of the ML algorithm to the neutrino–gravity correlation effect, represented by a series of works performed by the PTM collaboration (INFN, Rome University La Sapienza and Rome University Tor Vergata; CNR, Institute of Cosmogeophysics (Turin); University of Maryland (Washington) and Institute for Nuclear Research, Russian Academy of Sciences (Moscow)) [48, 49]. The effect consists in "observing a significant correlation" between the generalized noise background of uncooled gravity detectors in Rome and Maryland and the background neutrino scintillator under Mont Blanc during the burst from the Supernova SN1987A.

The difficulty of direct astrophysical interpretation of the effect lies in the two-order of magnitude deficit of sensitivity of the uncooled detectors with respect to the average estimate of the energy of gravitational radiation from a collapsing star in the LMC [48] (although speculations on the uncertainties of the source parameters could possibly reduce the deficit to a negligible quantity). Subsequently, a series of investigations was performed to elucidate the nature of the effect, for which the "case probability" was estimated by the PTM group to be extremely small, 10^{-6} [35]. Specifically, a correlation was shown between the "neutrino background" of the "collapse" program and the background radiation in other channels [50]; the correlation of the noise of the gravitational detectors with the seismic background [51]. In addition, the dynamics of the directional pattern of the complex antenna from detectors in Rome and Maryland [52] was followed in order to check its orientation toward the LMC; finally, a number of ideas with "new physics" were examined (see the examples in [53]). However, a definite model for explaining the correlation was not obtained. Numerical simulation of the fluxes of neutrino and gravitational data showed that the "*νg*-correlation effect" could be the result of statistical fluctuations with a correct estimate of the probability of an event [54]. However, the PTM group did not accept this criticism, since in their opinion the authors of [54] did not work with real experimental data [55].

The results of our analysis of the real data, kindly provided by the PTM collaboration, are presented below. The method described in Section 2 is used in parallel with the PTM method for comparison.

4.1. Method and Results of the PTM Group

The databank consisting of general traces of the gravity detectors in Rome and Maryland is limited by the time interval from 12:00 UT on February 22 to 06:00 UT on February 23.

For this time interval there exists a trace of the random noise of the LSD neutrino scintillator according to the channel of the "Supernova Watcher" program. All data are presented in digital form. The digitizing step for the gravitational traces $\Delta t = 1$ s corresponds to the correlation time of the output noise ($\Delta\omega = 1/\Delta t$ is the frequency band of the filter). The times of the neutrino events were read with an accuracy of 0.01 s (for technical reasons there are no data after 07:00 UT: the detector in Maryland did not operate). The neutrino traces contain a feature near 2:52 UT on February 23: five neutrino pulses in the form of a dense close packet with a small Poisson probability. These neutrinos were distinguished independently and before information about the optical observation of the Supernova arrived.

The analysis of the PTM group consisted of using auxiliary statistics, constructed from the gravitational data. This is the sum of the variations of the energy of two detectors taken at the moments of the neutrino events, normalized to a number of these events. The motivation was due to the meaning of C as being proportional to the correlation function between the readings of the variations of the energy of the gravitational detector and the neutrino events. The use of the total energy of the detectors in Rome and Maryland was termed "the good excitation method" [54] (the two detectors are considered to be a single, combined, gravitational antenna in a general gravity-wave perturbation; if the frequencies of the detectors are different, the combined antenna collects energy from different spectral components of the gravity-wave pulse). The PTM group operated directly with the dimensional variable

$$C(\tau) = \frac{1}{n}\sum_{k=1}^{n}[E_R(t_k+\tau) + E_M(t_k+\tau)], \qquad (29)$$

which corresponds to $\bar{E}$ in our notation (22) (for convenience in comparing the results we retain the notation of the PTM group in this section; the summation in Eq. (29) is performed with the normalizing factor ε in order to introduce the amplitude correction reflecting the difference of the noise temperatures of both detectors).

The main result obtained by the PTM group is that the variable C reaches its maximum value $C_{\exp} = 72.5$ K in a two-hour window around 2:52 UT with the total number of "neutrino markers" $n = 96$ and time shift -1.2 s. The "case probability" of such an event estimated by the PTM group by a Monte Carlo simulation of the flux of neutrino events was extremely small. This result was presented on two types of plots. Figure 2a shows the relative number of cases where the simulated ("artificial") quantity $C(\tau = -1.2\text{ s})$ exceeded the experimentally observed value $C_{\exp}$, as a function of the position of the two-hour segment on the entire observation interval. The second plot shows the relative number of cases calculated for the distinguished two-hour interval around 2:52 UT (the "anomalous neutrino packet" interval) as a function of the time shift varying near its "optimal" value -1.2 s (see Fig. 2b). Both types of plots demonstrate a "special feature" of the observed data:

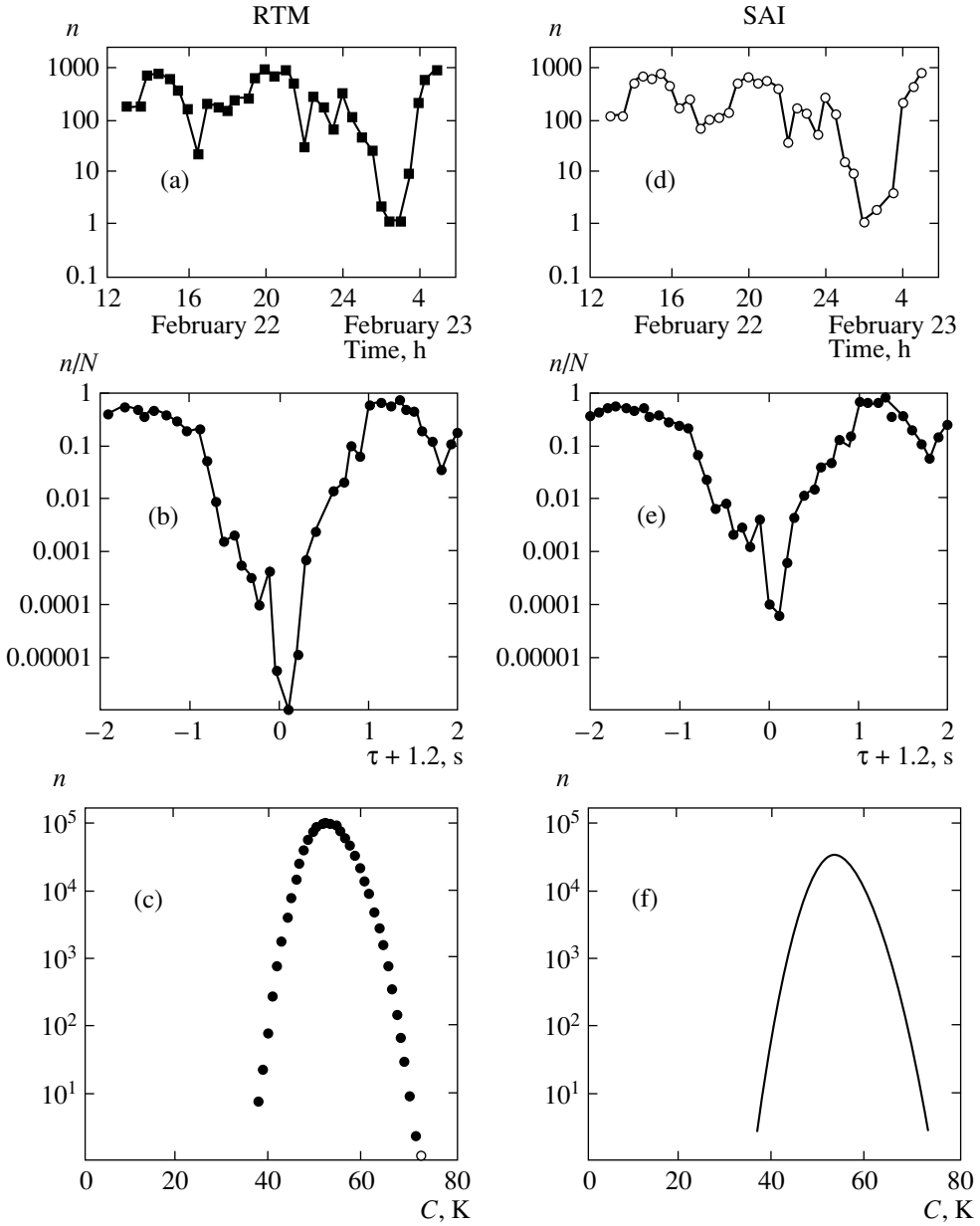

Fig. 2.

the plots contain deep dips at the points 2:52 UT (Fig. 2a) and −1.2 s (Fig. 2b). According to [11, 15, 34, 35], the presence of these dips signifies detection of a rare event. Its probability was estimated by two methods.

The first method employed the binomial formula $p = m/n$, where m is the number of cases where $C \geq C_{exp}$, and n is the total number of values of C employed. The second method consisted of empirical construction of the statistics of C using a simulation of the neutrino events (each simulated neutrino series gives a definite value of C—a single point in the empirical distribution). The differential distribution for C, taken from [35], is shown in Fig. 2c, which also shows the value of C_{exp}. Both methods give the probability $p = 10^{-3}$ for the effect shown in Fig. 2a, and $p = 10^{-6}$ for the effect shown in Fig, 2b. The PTM group considered the last result as detection of the "anomalous correlation" between the gravitational and neutrino data near 2:52 UT according to universal time—the so-called "neutrino-gravitational correlation" effect associated with the burst of SN1987A.

4.2. Method and Results of the Sternberg Astronomical Institute Group

The theory of the method which we used in the reanalysis of the observational data is described in Sections 2 and 3. The method is close to that of the PTM

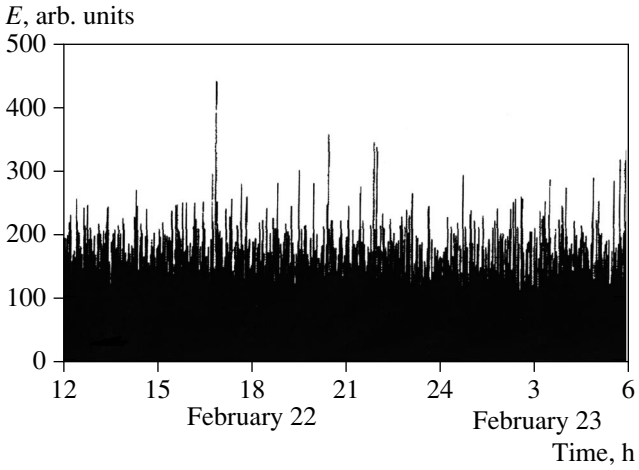

Fig. 3.

group, the difference being that the variable $C_{\max}$ is used instead of C as the "sufficient statistics."

4.2.1. Gaussian estimates. As shown in Section 3, in the Gaussian approximation ($\xi(t)$ is a normal process) all required probability estimates can be calculated and predicted. The experimental data support the "Gaussian nature" for uncooled detectors (specifically, see [35]). The noise temperatures T_e are known for the detectors in Rome $T_e = T_R = 28.6$ K and in Maryland $T_e = T_M = 22.1$ K. Hence the normalizing factor in (29) is

$$\varepsilon \approx T_M/T_R \approx 0.77,$$

which is quite close to the estimate obtained by the PTM group ($\varepsilon = 0.75$). The estimates of the average value and the variance for the variable C in the "grid excitation" method in dimensional form ($C \longrightarrow \bar{E}$) (29), using Eqs. (21) and (24), are

$$\begin{gathered} k^{-1}\langle C\rangle = (T_R + T_M) \approx 51 \text{ K}, \\ k^{-1}\sqrt{K_{\bar{E}}(0)} = \sqrt{[(T_R)^2 + (T_M)^2]/n} \approx 3.7 \text{ K}, \\ n = 96. \end{gathered} \tag{30}$$

Thus, near 2:52 UT the theory gives a Gaussian distribution of the variable C centered at the point 51 K with effective width 3.7 K.

Next, following the ML algorithm, we must consider the absolute maximum of the variable C as a function of τ. The average value of the random variable $C_{\max}$ calculated using Eqs. (25) and (26) becomes, after statistical averaging,

$$\langle C_{\max}\rangle = \langle C\rangle + \sqrt{K_{\bar{E}}(0)}\left[\sqrt{2\ln\mu} + \frac{\langle\xi\rangle}{\sqrt{2\ln\mu}}\right]. \tag{31}$$

Here $\langle\xi\rangle = 0.577$. The "time shift parameters" played the main role in estimating μ: the interval of variation $\Delta\tau$ and the step δt. The ML algorithm places no limitations on these parameters; recommendations can be taken only from physical arguments. We chose $\Delta\tau = \mp 100$ s and $\delta t = 0.01$ s in accordance with the specific nature of the experimental data (see explanation below). Then we have, according to the formula (28),

$$\rho^2 \approx 0.1, \quad \mu \approx 6.46.$$

Substituting these numbers into the expression (31) gives a numerical estimate of the average value $\langle C_{\max}\rangle$ in the time zone of interest to us: $\langle C_{\max}\rangle \approx 65$ K (with $k = 1$).

The formulas (25)–(28) also make it possible to determine the width of the distribution of $C_{\max}$ and to calculate the probability of a "false alarm" (or "case probability") for any value of C that could appear in the experiment, i.e., to solve completely the problem of statistical estimates. However, since all these estimates depend on the characteristic "shift times," we shall postpone this until we perform an empirical analysis of the statistics of the observational data in a manner similar to the empirical method used by the PTM group.

4.2.2. Empirical analysis. An empirical analysis is good in that it does not employ a priori hypotheses about the distribution law of the data. At the same time, the problem of extracting the statistical properties of the observed variables on the basis of only one realization of the random process is strongly "ill-posed," and the extraction errors can then be sufficiently large to make the method ineffective. Hence each step in an empirical analysis must be monitored for possible errors. Actually, the procedure of the ML algorithm is a delicate filtering process with an attempt to discover a weak signal that is strongly covered with noise. Figure 3 shows as an illustration the trace of the output data of the gravitational detector in Rome for February 22 and 23, 1987 (a computer reconstruction of the digital data is shown). It is obvious that it is impossible to extract the signal from this background without using a special procedure.

Our repeated analysis also revealed the presence of features in the behavior of the variable C near the marker 2:52 UT 23 February 1987. Indeed, it reaches a maximum (on the neutrino events in two-hour intervals) with a value of 72.5 K. The plots of the two main tests—the temporal evolution of C in two-hour intervals with a fixed shift of -1.2 s (Fig. 2d) and the change in C for variations of the shift in the range ± 2 s (near 2:52 UT) (Fig. 2e)—are essentially identical to its corresponding histograms obtained by the PTM group.

Estimates of the "case probability" for these events, which we made using the method of the PTM group do not fundamentally differ from their estimates. Thus, the binomial formula gives

$$p = m/n \approx 10^{-3}$$

(compare with Section 4.1; see [33] for the validity of using the binomial formula here). The empirical differential distribution constructed (using the neutrino

events simulated by the Monte Carlo method) of the C statistics (Fig. 2f) is also identical to the corresponding distribution obtained by the PTM group (Fig. 2c). It is centered at the point 52 K and agrees well with the theoretical prediction (30) 51 ± 3.7 K. For this distribution the experimental value $C_{exp} = 72.7$ K, falling "deep in the right-hand wing," has a random realization probability of 10^{-5}–10^{-6}.

Thus, we have confirmed all results obtained by the PTM collaboration in [11, 15, 34, 35] on the basis of the method which they employed. However, it should be stipulated that the reliability of reconstructing the distributions by the empirical method improves on the "wings of the distribution" (ill-posed problem). Consequently, the empirical estimates at the level "vanishingly small probabilities" cannot be interpreted absolutely, but more like a qualitative tendency.

4.2.3. Criticism of the PTM-group results. The key argument of the criticism of the results obtained by the PTM collaboration is the assertion that the variable C cannot be sufficient statistics in the experiment being considered. Consequently, neither the "binomial formula" nor the empirical distribution of C can give the correct estimate of the probability of a "random event." The problem is that the estimates made above did not take into account the fact that the time shift τ between the "neutrino" and "gravitational" events was varied (to find the optimal value, equal to -1.2 s) in order to maximize C. This must influence the probability of a random event.

In the general ML algorithm (see Section 2) the effect of a "shift adjustment" is taken into account by switching to the variable "absolute maximum" C_{max}. It is this distribution that should be used for calculating the statistical errors. In an empirical analysis, the construction of this distribution on the interval of the "observed correlation" is done as follows: the Monte Carlo method is used to simulate the "neutrino events" (a series of time markers) with the total number $n = 96$; next, the shift τ is chosen so that the variable C assumes its maximum value $C = C_{max}$. This is one point in the distribution of C_{max}. The procedure is repeated, giving the set of values of C_{max} required for a histogram.

Figure 4 displays such a distribution of C_{max} together with the distribution of the variable C and the experimentally observed value $C_{exp} = 72.5$ K.[1] The two characteristic times, and the interval and step of variation of the time shift which are required for the construction were chosen from a feature of the experimental data. The interval for the variations of τ cannot be greater than the characteristic time between the Poisson neutrino events. Otherwise, the additional "neutrinos" from neighboring two-hour intervals will be "captured." For an average value between the pulses of the

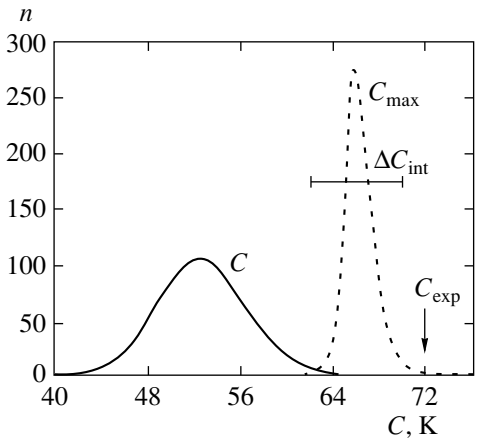

Fig. 4.

LSD detector in 70–80 s we took at $\Delta\tau = \pm 100$ s. The step for varying the time shift was taken to be the shortest of the experimental time scales, i.e., the accuracy of the "neutrino readings" $\delta t = 0.01$ s.

As expected, the empirical distribution of C_{max} is shifted to the right relative to the C curve, closer to the experimental value $C_{exp} = 72.5$ K. Its center lies in the zone 65–66 K, which agrees well with the prediction of the theory in the Gaussian approximation (65 K; see Section 4.2.1). An estimate of the probability of "randomly obtaining the result C_{exp} " taking account of this distribution increases to 10^{-4}–10^{-3}. This is much greater than the estimate presented by the PTM group (10^{-6}), but it is still small enough to consider the correlation between the gravitational and neutrino events as an objective fact.

However, another source of errors, which is present in this experiment (though, in principle, not necessary) interferes with this. It is due to the different discretization times of the gravitational and neutrino data: 1 s for readings of the gravitational detector and 0.01 s for the neutrino events. This necessarily requires interpolation in order to search for adequate gravitational readings (the values of the energy variations) when the "neutrino marker" falls between the gravitational markers. Interpolation can be performed by different methods, including an optimal method [56], but it still introduces an error, which must be taken into account when calculating the statistical parameters and, specifically, when determining the position of the center of the distribution of C_{max}. The calculations performed in [56] showed that the interpolation error is

$$\Delta C_{int} \approx 4.6 \text{ K}.$$

This "corridor" is shown in the plot (Fig. 4) near the center of the distribution of C_{max}. If the center of C_{max} is shifted to the right-hand edge of the corridor, the "probability of a random event" for C_{exp} reduces to 0.01 and greater. Such numbers are no longer "distinguished"

[1] The construction of the distribution of C_{max} requires much more computer time than the distribution of the variable C. Consequently, its histogram contains only 10^3 points.

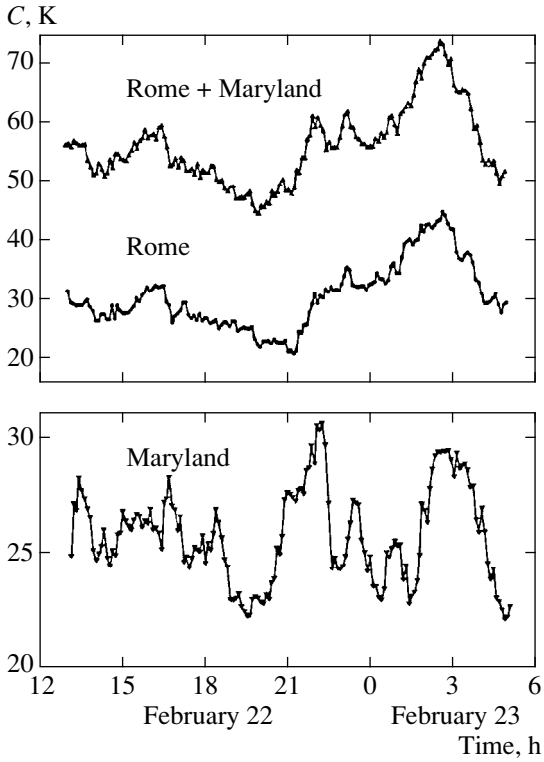

Fig. 5.

for measurements with random processes and do not presuppose the existence of any extra corridor event.

In summary, it follows from our reanalysis that the existing experimental data are inadequate to make a reliable conclusion about the detection "of significant *vg*-correlation" during the burst of the Supernova SN1987A.

5. SPECIAL FEATURES OF THE ML ALGORITHM IN SEARCHING FOR GAMMA–GRAVITATIONAL CORRELATIONS

The direct application of the ML algorithm to the gamma bursts encounters the problem of adequately determining the time marker of an astronomical event. This is due to the large diversity of durations and structures of gamma bursts. The "trigger time" presented in the BATSE catalog can have a very remote relation with the real phase of the active operation of a "gamma emitter" for pulses with a complex configuration. For this reason, it is reasonable to attempt to test adaptive algorithms in which the "time marker of a gamma event" is determined taking into account the structure of the pulse. The concept of structure and duration for gamma bursts is itself ambiguous and is a subject of debate. For our purposes, the most attractive approach is probably that of [57], where the effective duration is determined according to the fraction of the emitted energy, starting with the most intense bin in the pulse (a bin is the number of photons of fixed energy in a fixed count time). The pulse is represented by its temporal structure $G_i(t)$, located under the threshold energy level in accordance with the chosen fraction of the total energy (50, 80%, and so on).

The contribution of the variations of the energy of the gravitational background, corresponding to a gamma burst with number i, in the average selective variation (22) can be represented by the integral

$$E_i(\tau) = \int_{-\infty}^{\infty} W_i(\tau_j)E(\tau_j - \tau)d\tau_j, \qquad (32)$$

where $W_i(\tau_j)$ is the probability density of the arrival time of a gamma burst (in other words, the probability that the time τ_j must be chosen as the "effective arrival time" for the pulse i).

We note that the function $W_i(\tau_j)$ determines the time window in which integral (32) is different from zero. This window approximately corresponds to the duration of a gamma burst; it is assumed that it is much shorter than the interval between bursts. The unknown shift τ between the gravitational and gamma events, just as in the "case with a neutrino," is assumed to be independent of the number of the burst (uniformity of sources) and limited by some integral Δ.

For a "multipeak" gamma burst the following procedure can be phenomenologically proposed for estimating the function $W_i(\tau_j)$ and integral (32). Let τ_{ik} be the time position of a local maximum k in the gamma burst with number i, whose amplitude exceeds the threshold G_c, $G_{ik} \geq G$. Then there exists a set of random "arrival moments" of the gamma burst:

$$\tau_{i*} \longrightarrow \tau_{i1}, \tau_{i2}, \ldots \tau_{ik}, \ldots,$$

where $k = 1, 2, \ldots m_i$ is the total number of local maxima of the function $G_i(t)$. The probability density $W_i(\tau_j)$ can be represented by a discrete series of coefficients p_{ik}, proportional to the relative amplitudes of the maxima:

$$W_i(\tau_j) = c_i \sum_{k=1}^{m_i} p_{ik}\delta(\tau_j - \tau_{ik}), \qquad (33)$$

where

$$p_{ik} = \frac{G_{ik}}{\sum_{il}^{m_i} G_{il}} \leq 1, \quad c_i = 1.$$

Ultimately, the variable "average of the selective variation" of the energy of the gravitational detector will be represented by the formula

$$C(\tau) = \sum_{k=1}^{m_i} p_{ik}E(\tau_{ik} - \tau). \qquad (34)$$

Expression (34) extends the simpler Eq. (22) to the case of gamma bursts as astronomical events in the search for astro-gravitational correlations. As in the case with neutrinos, the shift τ is not determined within the statistics and requires astrophysical arguments. The transition to the variable $C_{\max}$ on the basis of the ML algorithm occurs just as in the neutrino case. The formulas for the statistical estimates in the Gaussian approximation remain the same.

6. DISCUSSION

The main results of this work are a formulation of the ML algorithm for searching for astro-gravitational correlations and finding the "phenomenon of neutrino–gravitational correlation," observed during the burst of the Supernova 1987A.

The advantages and drawbacks of the proposed algorithm were clearly demonstrated in the analysis of the last event. Its main weakness is that the results of the analysis depend on the unknown time interval between the astrophysical and gravitational events. An effective algorithm requires an a priori estimate of this interval. Attempts to limit the interval by considering the special features of the experimental data (as done in the present work) lead to a solution only within the model adopted. In the general case the estimate of the probability of the presence or absence of a correlation remains undetermined. The most favorable case is one where the value of the shift is known exactly. Then the probability of the observed correlation can be estimated according to the variable C, which is more reliable than an estimate made using $C_{\max}$. However, the latter distribution must be used if the shift is not known in advance.[2]

The shift $\tau = \pm 2$ s used in [48, 49] can be associated only with the expected delay of the neutrino signal (less than 2.7 s), if the neutrino has a rest mass (less than 10 eV), and for the hypothesis that the neutrino and gravity waves are emitted simultaneously. This special model is also very limited.

It is interesting to note that even for a substantial range of the shift $\tau = \pm 100$ s, which was used in our repeated analysis, the probability of a random "correlation" remains small, 10^{-4}–10^{-3}, and only the "interpolation error" made it impossible to confirm the existence of the "effect," described by the PTM group.

An alternative hypothesis for explaining the observed experimental data, in which the presence of a "correlation" is assumed only between the detectors in Rome and under Mont Blanc, caused by the regional seismic activity was advanced in [53]. The contribution of the Maryland detector to the effect was negligible. This is illustrated by the graph of the temporal evolution of the C variable for each detector separately compared with their sum (see Fig. 5). However, evidently, it is difficult to confirm the fact of a seismic correlation between the "Rome–Mont Blanc" detectors numerically.[3]

Returning to the general problem of searching for the "astro–gravitational correlations," we shall make several concluding remarks.

In principle, the ML algorithm presumes accumulation of a signal. However, in the case of "post-detector detection" according to the envelope the accumulation of weak incoherent pulses ($A_k \leq \sigma$) increases the signal/noise ratio in proportion to $n^{1/4}$, i.e., the accumulation efficiency is low. The accumulation procedure leads to the standard law for adding independent random counts ($\propto n^{1/2}$) only for pulses with a large amplitude. However, in this case, large n with reasonable observation times cannot be expected.

There is no doubt that the gravity-wave experiment in the form of the search for "astro–gravitational correlations" has an obvious advantage because the "empty" observational time is smaller. This decreases the probability of a "false alarm" proportional to the factor $n\Delta\tau/T$; but, for the same detection threshold (signal/noise ratio) the probability of "correct detection" increases negligibly, only by a factor of $(1/2)\ln[T/n\Delta\tau]$.

It should be noted that the algorithm developed cannot be transferred without any changes to the case of a laser interferometric antenna on free masses. The problem is that the response of this wide-band detector cannot be represented in a universal form ("response to a δ excitation") and the complicated structure of an individual gravitational pulse must be taken into account. The construction of an optimal algorithm for a packet of such pulses, correlated with astrophysical events, transforms into a multiparameter problem requiring a separate analysis.

ACKNOWLEDGMENTS

We thank the members of the PTM group, first and foremost, Professors E. Pizzella, G. Pallottino, and S. Frasca, for providing the experimental gravitational data and stimulating discussions. We value highly the assistance provided by O. Ryazhskaya (Institute for Nuclear Research, Russian Academy of Sciences) and S. Vernetto for providing the neutrino data from the LSD detector. The authors had many fruitful discussions with K. Postnov and M. Shakura, colleagues from Sternberg Astronomical Institute, Moscow State University. This work was supported in part by the Russian State Subprograms "High Energy Physics" and "Fundamental Nuclear Physics."

[2] We note that the authors of [54] came quite close to this idea, introducing the special parameter q in order to determine how frequently realizations with rare statistical properties occur in the computer simulation.

[3] A correlation between seismic observations and the background of the (R + M) antenna during the explosion of the Supernova SN1987A was previously reported in [51].

REFERENCES

1. J. Weber, Phys. Rev. **117**, 306 (1960).
2. J. Weber, Phys. Rev. Lett. **22**, 1320 (1969); **24**, 276 (1970).
3. P. Astone, P. Bonifazi, G. V. Pallottino, *et al.*, in *General Relativity and Gravitational Physics*, Ed. by M. Cerdonio *et al.* (World Scientific, Singapore, 1994), p. 551.
4. A. Abramovici, W. E. Althouse, R. V. Drever, *et al.*, Science **256**, 281 (1992).
5. P. Michelson, *Gamma Ray Bursts*, ibid., p. 37.
6. G. Modestino and G. Pizzella, Preprint Nota Interna LNF 97/038 IR.
7. A. V. Gusev, V. K. Milyukov, V. N. Rudenko, and M. P. Vinogradov, in *Proceedings of the Second Edorado Amaldi Conference*, Ed. by E. Coccia, G. Veneziano, and G. Pizzella (World Scientific, Singapore, 1998), p. 512.
8. D. K. Nadyozhin and I. V. Otrochenko, Astron. Zh. **57**, 78 (1980) [Sov. Astron. **24**, 47 (1980)].
9. R. Bowers and J. R. Wilson, Astrophys. J. **263**, 366 (1982).
10. H. Bethe, in *Proceedings of the International School of Physics "Enrico Fermi," Varena, 1984* (North Holland, Amsterdam, 1986), Course XCI, p. 181.
11. M. Aglietta, A. Castellina, W. Fulgione, *et al.*, Nuovo Cimento C **9**, 185 (1986).
12. E. Alexeyev, L. N. Alexeyeva, I. V. Krivosheina, *et al.*, Phys. Lett. B **205**, 209 (1988).
13. R. M. Bionta, G. Blewtt, C. V. Braton, *et al.*, Phys. Rev. Lett. **51**, 27 (1983).
14. K. S. Hirata, T. Kajita, M. Koshiba, *et al.*, Phys. Rev. D **38**, 448 (1988).
15. M. Aglietta, G. Badino, G. F. Bologna, *et al.*, Europhys. Lett. **3**, 1315 (1987).
16. E. N. Alekseev, L. N. Alekseeva, V. N. Zakidyshev, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **49**, 480 (1989) [JETP Lett. **49**, 548 (1989)].
17. M. Takita, in *Frontier of Neutrino Astrophysics*, Ed. by Y. Suzuki and K. Nakamura (Universal Academical Press, Tokio, 1993), No. 5.
18. G. J. Fishman and C. A. Meegan, Annu. Rev. Astron. Astrophys. **33**, 415 (1993).
19. B. M. Belli, Astrophys. J. Lett. **479**, L31 (1997).
20. R. Wijers, Nature **393**, 13 (1998).
21. K. S. Thorn, in *Particle and Nuclear Astrophysics and Cosmology in the Next Millenium*, Ed. by E. W. Kolb and R. Peccei (World Scientific, Singapore, 1995), p. 160.
22. G. S. Bisnovatyi-Kogan, Astrophys. J., Suppl. Ser. **97**, 185 (1995).
23. B. V. Komberg and D. A. Kompaneets, Astron. Zh. **74**, 690 (1997) [Astron. Rep. **41**, 611 (1997)].
24. V. M. Lipunov, K. A. Postnov, M. E. Prokhorov, *et al.*, Astron. Astrophys. **298**, 677 (1995).
25. V. M. Lipunov, K. A. Postnov, and M. E. Prokhorov, New Astron. **2**, 43 (1997).
26. M. V. Sazhin, S. D. Ustyugov, and V. M. Chechetkin, Pis'ma Zh. Éksp. Teor. Fiz. **64**, 817 (1996) [JETP Lett. **64**, 871 (1996)].
27. V. S. Imshennik, Pis'ma Astron. Zh. **18**, 489 (1992) [Sov. Astron. Lett. **18**, 194 (1992)].
28. S. R. Kulkarni, S. G. Djorgovski, A. N. Ramaprakash, *et al.*, Nature **393**, 35 (1998).
29. T. J. Galama, P. M. Vreeswijk, J. van Paradijs, *et al.*, Nature **395**, 670 (1998).
30. A. N. Ramaprakash, S. R. Kulkarni, D. A. Frail, *et al.*, Nature **393**, 43 (1998).
31. B. Paczyński, Astrophys. J. Lett. **494**, L45 (1998).
32. P. Astone, G. Barbiellini, M. Bassan, *et al.*, Astropart. Phys. **7**, 231 (1997).
33. A. V. Gusev, V. V. Kulagin, S. I. Oreskin, *et al.*, Astron. Zh. **74**, 287 (1997) [Astron. Rep. **41**, 248 (1997)].
34. M. Aglietta, A. Castellina, W. Fulgione, *et al.*, in *Proceedings of the 23-ICPC, Calgary, 1993*, Vol. 1, p. 69; in *Proceedings of 24-ICPC, Roma, 1995*, Vol. 2, p. 73.
35. M. Aglietta, G. Badino, G. Bologna, *et al.*, Nuovo Cimento C **12**, 75 (1989).
36. M. T. Murphy, J. K. Webb, and I. S. Heng, astro-ph/9911071.
37. J. C. Finn, S. D. Mohanty, and J. D. Romano, gr-qc/9903101.
38. J. Amati, P. Astone, M. Bassan, *et al.*, Astron. Astrophys., Suppl. Ser. **138**, 605 (1999).
39. P. Astone, G. Barbiellini, M. Bassan, *et al.*, Astron. Astrophys., Suppl. Ser. **138**, 603 (1999).
40. P. Meszaros, M. J. Rees, and R. A. Wijers, astro-ph/9808106.
41. A. F. Zakharov, Astron. Zh. **73**, 605 (1996) [Astron. Rep. **40**, 552 (1996)].
42. P. Astone, G. Barbiellini, M. Bassan, *et al.*, Astron. Astrophys. **351**, 403 (1999).
43. C. W. Helstrom, *Statistical Theory of Signal Detection* (Pergamon, New York, 1968).
44. M. Aglietta, A. Castellina, W. Fulgione, *et al.*, Nuovo Cimento C **14**, 171 (1991).
45. H. L. Van, *Trees Detection, Estimation and Modulation Theory* (Wiley, New York, 1968).
46. H. Cramer, *Mathematical Methods of Statistics* (Princeton Univ. Press, Princeton, 1946; Mir, Moscow, 1975).
47. V. I. Tikhonov, *Overshooting of Random Processes* (Nauka, Moscow, 1970).
48. E. Amaldi, P. Bonifazi, M. G. Castellano, *et al.*, Europhys. Lett. **3**, 1325 (1987).
49. G. Pizzella, Nuovo Cimento C **15**, 931 (1992).
50. V. Dadykin, G. T. Zatsepin, E. V. Korol'kova, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **56**, 441 (1992) [JETP Lett. **56**, 426 (1992)].
51. V. K. Kravchuk, V. N. Rudenko, and O. E. Starovoit, Phys. Solid Earth **31**, 780 (1995).
52. E. K. Kuchik and V. N. Rudenko, Astron. Zh. **68**, 732 (1991) [Sov. Astron. **35**, 361 (1991)].
53. V. N. Rudenko and V. K. Kravchuk, Phys. Chem. Earth **8**, 715 (1999).
54. C. Dikson and B. Schutz, Phys. Rev. D **51**, 2644 (1995).
55. S. Frasca, G. Pallottino, and G. Pizzella, INFN (Roma), Nota Interna, No. 1088 (1997).
56. A. V. Gusev, Vestn. Mosk. Univ., Ser. 3: Fiz., Astron., No. 4, 17 (1999).
57. I. G. Mitrofanov, D. S. Anfimov, M. L. Litvak, *et al.*, Astrophys. J. **522**, 1069 (1999).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 859–872.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 996–1011.

GRAVITATION, ASTROPHYSICS

Dynamical Method for Quantizing Gravity and the Problem of Decoherence in Quantum Cosmology

S. N. Vergeles

Landau Institute for Theoretical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia
e-mail: vergeles@itp.ac.ru

Received June 7, 2000

Abstract—A regularized quantum theory of gravity interacting with matter is constructed. The construction is made on the basis of the method of dynamical quantization of generally covariant theories. A solution of the problem of decoherence in quantum cosmology is proposed on the basis of this method. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A new dynamical method for quantizing generally covariant theories has been proposed in a series of works [1–4]. The study of two-dimensional models, such as the two-dimensional bosonic string [5] and two-dimensional gravity interacting with matter [7], from the standpoint of this method has led, in the first place, to anomaly-free quantization of these models[1] and, in the second place, to further elaboration of the method of dynamical quantization itself. Consequently, a formulation of the dynamical quantization method in the interpretation which makes it possible to advance as far as possible in calculations in the theory of two-dimensional gravity is needed. This is done in Sections 2 and 3. The main purpose of the present paper is an attempt to solve the problem of decoherence in quantum cosmology on the basis of the method of dynamical quantization. To this end, quantization of the model of the theory of gravity with a Λ term, interacting with a Dirac field, is done in Section 4. A possible solution of the problem of decoherence in an inflating universe is proposed in Section 5. We show that (in a closed model) since the number of physical degrees of freedom is finite in dynamical quantization in an inflating universe, the quantum fluctuations "die out" with time at an exponential rate. This result is valid to all orders in the Planck scale l_P.

Specific results of the application of the method of dynamical quantization to the above-mentioned two-dimensional theories [5, 7], obtained by explicit constructions and direct calculations, justify the abstract assumptions and axioms on which this method is based.[2]

The ideology and logical scheme of the dynamical method will be expounded below taking account of the experience in quantizing two-dimensional gravity.

The key point in the quantization of two-dimensional gravity was the construction of a complete set of such operators $\{A_n, B_n, \ldots\}$, designated below as $\{A_N, A_N^\dagger\}$, which possess the following properties.

1. The operators A_N and $A_N^\dagger$ are Hermitian conjugates of one another and

$$[A_N, A_M] = 0, \quad [A_N, A_M^\dagger] = \delta_{NM}. \tag{1}$$

2. The set of operators $\{A_N, A_N^\dagger\}$ describes all physical dynamical degrees of freedom of the system.

3. Each operator from the set $\{A_N, A_N^\dagger\}$ commutes with all constraints of the first kind or with the complete Hamiltonian of the theory.

Quantization is performed directly using the operators $\{A_N, A_N^\dagger\}$. This means that the space of physical states is constructed using the operators $\{A_N^\dagger\}$ from the ground state and all operators are expressed in terms of the operators $\{A_N, A_N^\dagger\}$, as well as in terms of the operators describing the gauge degrees of freedom.

Quantization is also performed according to the described scheme on the basis of the dynamical method. However, in the theory of two-dimensional gravity the operators $\{A_N, A_N^\dagger\}$ were constructed explicitly (i.e., they were expressed explicitly in terms of the initial dynamical variables), in more realistic theories this problem is hardly solvable. Consequently, the set of operators $\{A_N, A_N^\dagger\}$ with the properties 1–3 must be introduced axiomatically. Conversely, the properties 1–3 make it possible,

[1] See [6] for a discussion of anomaly-free quantization of two-dimensional gravity from a standpoint close to that of the present author.

[2] The basic assumptions and axioms of the method of dynamical quantization [1–4] appeared before the problem of anomaly-free quantization of two-dimensional gravity was solved.

in principle, to express the initial variables in terms of the convenient operators $\{A_N, A_N^\dagger\}$.

However, in contrast to the two-dimensional theory of gravity, regularization is required in real models of gravity. In the method of dynamical quantization, regularization is done precisely in terms of the operators $\{A_N, A_N^\dagger\}$. As will be shown below, such regularization is natural in generally covariant theories, since it preserves the form of the Heisenberg equations and thereby also the general covariance of the theory.

2. METHOD OF DYNAMICAL QUANTIZATION

Let us consider a generally covariant field theory. Let us assume that in this theory the Hamiltonian in the classical limit is an arbitrary linear combination of constraints of the first kind and there are no constraints of the second kind.

Let $\{\Phi^{(i)}(x), P^{(i)}(x)\}$ be a complete set of fundamental fields of the theory and their canonically conjugate momenta, in terms of which all other physical quantities and fields in the theory are expressed. Here the index (i) enumerates the types of fields. For example, for some (i) these can be either six spatial components of the metric tensor $g_{ij}(x)$ or the scalar field $\phi(x)$ or the Dirac field $\psi(x)$, and so on. The set of fields $\{\Phi^{(i)}(x)\}$ is a complete set of the mutually commuting fundamental fields of the theory.

Next, to simplify the notation the index i will be omitted. It can be assumed that the variable x includes, besides the spatial coordinates, the index i also.

The construction of a quantum theory by the dynamical method is based on the following natural assumptions or axioms relative to the structure of the unregularized space F of the physical states of the theory.

Axiom 1. *All states of the theory which are physically meaningful are obtained from the ground state* $|0\rangle$ *using the creation operators* $A_N^\dagger$:

$$|n_1, N_1; \ldots; n_s, N_s\rangle = (n_1! \cdot \ldots \cdot n_s!)^{-1/2} \times (A_{N_1}^\dagger)^{n_1} \cdot \ldots \cdot (A_{N_s}^\dagger)^{n_s}|0\rangle,$$
$$A_N|0\rangle = 0. \quad (2)$$

States (2) *form an orthonormal basis of the space F of physical states of the theory.*

The numbers $n_1, \ldots, n_s$ assume integer values and are called occupation numbers.

Axiom 2. *The set of states* $\Phi(x)\,|n_1, N_1; \ldots; n_s, N_s\rangle$, *where the set of numbers* $(n_1, N_1; \ldots; n_s, N_s)$ *is fixed, contains a superposition of all states of the theory, in which one of the occupation numbers differs in modulus by one and all other occupation numbers equal to the occupation numbers of state* (2).

Here the operators $A_N^\dagger$ and their conjugates A_N possess the standard commutation properties (1). The operators $\{A_N, A_N^\dagger\}$, generally speaking, can be bosonic or fermionic. If the creation and annihilation operators follow the Fermi statistics, then the commutator is taken in Eq. (1). For compact spaces the case of interest to us, we can assume without loss of generality that the index N, enumerating the creation and annihilation operators, belongs to a discrete finite lattice. A norm can be easily introduced in the space of indices N.

Since states (2) are physical, they satisfy the relations

$$\mathcal{H}_T|n_1, N_1; \ldots; n_s, N_s\rangle = 0, \quad (3)$$

where $\mathcal{H}_T$ is the complete Hamiltonian of the theory. Equations (3) follow from the equations

$$\mathcal{H}_T|0\rangle = 0, \quad (4a)$$

$$[\mathcal{H}_T, A_N^\dagger] = 0, \quad [\mathcal{H}, A_N] = 0. \quad (4b)$$

We call attention to the fact that the commutation relations (4b) are a consequence of the general covariance of the theory. In other theories a set of operators with properties (4b) exhausting the physical degrees of freedom of the system may not exist.

It should be noted that both axioms introduced above can be somewhat altered depending on the properties of the dynamical system under consideration. Here the simplest variant of a formulation of the axioms of the dynamical quantization method is presented. Ultimately, only the following assumptions need be satisfied.

(a) There exists a set of operators $\{A_N, A_N^\dagger\}$, which exhaust the physical degrees of freedom of the system and satisfy the commutation relations (4b) or a weakened variant of the commutation relations

$$[\mathcal{H}_T, A_N^\dagger] = \lambda_N A_N^\dagger. \quad (5)$$

(b) The matrix constructed from the elements $[A_M, A_N^\dagger]$, $[A_M, A_N]$, and $[A_M^\dagger, A_N^\dagger]$ is invertible.

In contrast to problem (2)–(4) considered here, in the exactly solvable case of two-dimensional gravity [7] we had constraints on the occupation numbers. The absence of any constraints on the occupation numbers means physically that the phenomenon of "not flying out" is absent in the theory. Thus, the axioms 1 and 2 together with Eqs. (1)–(4) refer to systems where the quanta of the fundamental fields can exist as stable particles. Conversely, if Eqs. (5) are satisfied, then in order for Eqs. (3) to be satisfied the following constraint on the occupation numbers is necessary (compare with the analogous constraint in [7]):

$$\sum_N n_N \lambda_N = 0.$$

In the following, axioms 1 and 2 and Eqs. (4) are assumed to be satisfied. We present certain consequences of axioms 1 and 2.

Let $|N\rangle = A_N^\dagger|0\rangle$. It follows from axiom 2 that

$$\Phi(x)|N\rangle = \phi_N(x)|0\rangle + |N; \Phi(x)\rangle, \quad \langle 0|N; \Phi(x)\rangle = 0, \tag{6}$$

and the linearly independent fields $\phi_N(x)$ do not depend on the operators A_M or $A_M^\dagger$:

$$[\phi_N(x), A_M] = 0, \quad [\phi_N(x), A_M^\dagger] = 0. \tag{7}$$

If $\Phi(x)$ is a real field, then it follows from the axioms and Eqs. (6) that the following expansion is valid:

$$\Phi(x) = \sum_N (A_N\phi_N(x) + A_N^\dagger\phi_N^*(x)) + \phi(x), \tag{8}$$

where the field $\phi(x)$ does not contain the operators A_N and $A_N^\dagger$ to the first power. For a complex Dirac field ψ Eq. (8) becomes

$$\psi(x) = \sum_N A_N\psi_N(x) + \chi(x), \tag{9}$$

where $\{A_N, A_N^\dagger\}$ are fermionic operators. If there is no fermionic condensate, then the fermionic field $\chi(x)$ depends on the fermionic operators $\{A_N, A_N^\dagger\}$ to powers no less than cubic.

Since the operators $\{A_N, A_N^\dagger\}$ are conserved, the entire dynamics of the fields $\Phi(x)(\psi(x))$ is contained in the variables $\phi_N(x)$ and $\phi(x)(\psi_N(x), \chi(x))$.

The preceding constructions make it possible to examine a regularization of the theory. Thus far our exposition was formal, since the singularity of the theory was not taken into account.

The importance of commutation relations (1) and (4b) lies in the fact that any set of pairs of operators $\{A_N, A_N^\dagger\}''$ can be viewed as a set of constraints of the second kind. This makes it possible to perform a regularization as follows.

We single out a finite set of pairs of annihilation and creation operators $\{A_N, A_N^\dagger\}'$ and enumerate them so that $|N| < N_0$. Since the physical information is contained in the wave functions $\phi_N(x)$, this set is actually determined by the choice of set of linearly independent wave functions $\{\phi_N(x)\}'$ which corresponds to the set of operators $\{A_N, A_N^\dagger\}'$. The choice of functions in the set $\{\phi_N(x)\}'$ is determined by the physical conditions of the problem. For example, if the x space is a torus, then periodic traveling waves, whose wave numbers are bounded in modulus, can be taken as the wave functions of this set at a given moment in time. All creation and annihilation operators, except for those chosen, i.e., the operators with $|N| > N_0$, are set equal to zero:

$$A_N = 0, \quad A_N^\dagger = 0, \quad |N| > N_0. \tag{10}$$

We shall prove a theorem, which is important for our method, that gives meaning to the entire dynamical quantization scheme.

Theorem. *The imposition of the constraints of the second kind* (10) *does not change the form of the Heisenberg equations, preserving their classical form.*

Proof. Let $|\mathcal{M}'\rangle$, $|\mathcal{N}'\rangle$, … be basis vectors (2), constructed using a bounded set of operators $\{A_N, A_N^\dagger\}'$, and let F' denote the Fock space with these basis vectors. The imposition of constraints (10) means that the space of physical states F is limited up to the regularized subspace $F' \subset F$. For any operator A in the regularized theory, only the matrix elements of the form $\langle\mathcal{M}'|A|\mathcal{N}'\rangle$ are considered. Consequently, the matrix elements, corresponding to constraints (10), of the quantum Dirac brackets of the operators A and B can be represented in the form (an asterisk denotes a Dirac bracket)

$$\langle\mathcal{M}'|[A, B]^*|\mathcal{N}'\rangle = \sum_{\mathcal{L}'}(\langle\mathcal{M}'|A|\mathcal{L}'\rangle\langle\mathcal{L}'|B|\mathcal{N}'\rangle - \langle\mathcal{M}'|B|\mathcal{L}'\rangle\langle\mathcal{L}'|A|\mathcal{N}'\rangle). \tag{11}$$

By definition of the quantum Dirac brackets the operators A_N and $A_N^\dagger$ with $|N| > N_0$, contained in the operators A and B from Eq. (11), are assumed to be zero after normal ordering. The commutator $[A, B]$ is formally distinguished from the Dirac bracket (11) by the fact that when the matrix elements $\langle\mathcal{M}'|[A, B]|\mathcal{N}'\rangle$ are calculated according to a formula similar to (11), the summation extends over all intermediate states (2). Let us assume that the operator B is diagonal in basis (2) and does not depend on the operators A_N and $A_N^\dagger$ with $|N| > N_0$. Then it is evident from Eq. (11) that

$$\langle\mathcal{M}'|[A, B]^*|\mathcal{N}'\rangle = \langle\mathcal{M}'|[A, B]|\mathcal{N}'\rangle. \tag{12}$$

It now remains to note that all occupation number operators $n_N = A_N^\dagger A_N$ commute with the complete Hamiltonian. Moreover, as a result of the commutation relations (4b) the Hamiltonian does not depend on the operators A_N and $A_N^\dagger$. Consequently, in Eq. (12) the Hamiltonian $\mathcal{H}_T$ can be substituted for the operator B. This proves the theorem.

There is also a classical variant of the theorem presented above: *the imposition of constraints* (10) *completely preserves the form of the equations of motion. This is expressed by the equality*

$$[\xi, \mathcal{H}_T] = [\xi, \mathcal{H}_T]^*.$$

The last assertion immediately follows from the definition of the Dirac brackets [8] and Eqs. (4b).

Corollary. *The regularized theory is generally covariant.*

Indeed, this follows directly from the theorem proved above. The equations of motion satisfied by the field $\Phi(x)$ in the regularized theory have the same form as the classical equations of motion which are generally covariant. This proves the corollary.

Since, in principle, any number of physical degrees of freedom can be retained in the regularized theory, a perturbation theory can be developed with respect to this number.

We underscore that the Dirac brackets $[\xi, \chi]^*$ of two arbitrary operators ξ and χ, generally speaking, is different from the commutator $[\xi, \chi]$ in the classical and quantum cases.

The following circumstance should be noted. Let the operator field $\mathbb{O}(x; A_N, A_N^\dagger)$ be a normal-ordered series relative to the generators of the Heisenberg algebra $\{A_N, A_N^\dagger\}$:

$$\mathbb{O}(x; A_N, A_N^\dagger) = \mathbb{O}^{(0)}(x) + \sum_N [\mathbb{O}_N^{(-1)}(x)A_N + \mathbb{O}_N^{(+1)}(x)A_N^\dagger] + \ldots. \tag{13}$$

Here the operator fields $\mathbb{O}^{(0)}(x)$, $\mathbb{O}_N^{(\pm 1)}(x)$, and so on do not depend on the generators of the Heisenberg algebra $\{A_N, A_N^\dagger\}$. Then the equality

$$\mathbb{O}(x; A_N, A_N^\dagger) = 0 \tag{14}$$

is equivalent to the system of equalities

$$\mathbb{O}^{(0)}(x) = 0, \quad \mathbb{O}_N^{(\pm 1)}(x) = 0, \ldots. \tag{15}$$

3. AXIOMATIC APPROACH

A more axiomatized scheme of dynamical quantization will now be presented. This scheme, which possibly is less natural, is logically stricter and simplifies the calculations.

This approach is based on the following assumption: *the theory is regularized in a manner so that the axioms are satisfied.*

Axiom 3. *All states of the theory which are physically meaningful are obtained from the ground state* $|0\rangle$ *using the creation operators* $A_N^\dagger$ *with* $|N| < N_0$:

$$|n_1, N_1; \ldots; n_s, N_s\rangle = (n_1! \cdot \ldots \cdot n_s!)^{-1/2} \times (A_{N_1}^\dagger)^{n_1} \cdot \ldots \cdot (A_{N_s}^\dagger)^{n_s}|0\rangle, \quad A_N|0\rangle = 0. \tag{16}$$

States (16) *form an orthonormal basis of the states F' of physical states of the theory.*

Axiom 4. *The dynamical variables* $\Phi(x)$ *transfer state* (16) *with fixed values of the numbers* $(n_1, N_1; \ldots; n_s, N_s)$ *into a superposition of the states of the theory of form* (16)*, containing all states in which one of the occupation numbers is different in modulus by one and all other occupation numbers are identical to those of state* (16).

Axiom 5. *The equations of motion and constraints for the physical fields* $\{\Phi(x), \mathcal{P}(x)\}$ *have the same form, to within the arrangement of the operators, as the corresponding classical equations and constraints*.

Axioms 3 and 4 are analogs of axioms 1 and 2 in the unregularized theory. Axiom 5 replaces the theorem from Section 2. It postulates the correct form of the equations of motion and the constraints in agreement with classical mechanics. Since now the equations of motion no longer need to be derived, the Hamiltonian also becomes unnecessary.

We call attention to the fact that the Heisenberg equations are completely equations of type (14) and (15), while the equations of constraints (4) decompose into two series:

$$\mathcal{H}_T^{(0)}|0\rangle = 0, \tag{17}$$

$$\mathcal{H}_{TN}^{(\pm 1)} = 0, \ldots. \tag{18}$$

Equations (18) should be interpreted as an identity.

Thus, in the formal approach the problem reduces to the following. A finite set of linearly independent functions $\{\phi_N(x)\}'$ and the corresponding set of operator pairs $\{A_N, A_N^\dagger\}'$ are chosen and they are used to construct, according to Eq. (8), a regularized quantum field $\Phi(x)$. Next, the regularized quantum field is substituted into Heisenberg's equations of motion and the equations for the constraints, which are solved in accordance with Eqs. (15), (17), and (18). As a result of such calculations, an explicit expression should be found for the field $\phi(x)$ in Eq. (8) as a normal-ordered series with respect to the generators of the Heisenberg algebra $\{A_N, A_N^\dagger\}$. The physical state is given by an expansion in basis (16) and all possible averages are calculated with respect to this state.

The logical scheme for quantization presented above leaves the following question unanswered: how are Eqs. (17) solved? Here we propose a solution using a method which in [7] was termed the second method of quantization.

We shall assume below that all equations weaker than Eqs. (17) are satisfied:

$$\langle 0|\mathcal{H}_T^{(0)}|0\rangle = 0. \tag{19}$$

We underscore that in Eq. (19) the average extends only over the gauge degrees of freedom of the system; the

operator $\mathscr{H}_T^{(0)}$ depends on the gauge degrees of freedom but not on the physical degrees of freedom $\{A_N, A_N^\dagger\}$, which follows from Eq. (18).

We assume next that the ground state $|0\rangle$ is a specific state with respect to the gauge degrees of freedom. This means the following.

Let field (8) be such that all gauge degrees of freedom are explicitly singled out in it in a linear approximation. In the theory of gravitation such a field is the gravitational field (the metric tensor or tetrad), but the matter field is not. We shall represent the field $\phi(x)$, which is the second term on the right-hand side of Eq. (8), in the form

$$\begin{aligned}\phi(x) &= \phi_0(x) + \phi_\parallel^{(1)}(x) + \phi'(x),\\ \phi_\parallel^{(1)}(x) &= \sum_n (a_n\phi_{\parallel n}(x) + a_n^\dagger\phi_{\parallel n}^*(x)).\end{aligned} \tag{20}$$

Here $\phi_0(x)$ and $\phi_{\parallel n}(x)$ are c numbers of the field, and the set of modes $\{\phi_{\parallel n}(x)\}$ forms a complete set of longitudinal modes. In other words, any infinitesimal purely gauge transformation of the field $\Phi(x)$ can be uniquely expanded in a set of orthonormal (in some sense) modes $\{\phi_{\parallel n}(x), \phi_{\parallel n}^*(x)\}$. For the theory of gravity the longitudinal part of the metric tensor is given by the expression $\xi_{i;j} + \xi_{j;i}$. The field $\phi'(x)$ depends on the operators $\{a_n, a_n^\dagger\}$ to powers no less than quadratic. It is natural to assume that the operators $\{a_n, a_n^\dagger\}$ are the generators of the Heisenberg algebra:

$$[a_m, a_n^\dagger] = \delta_{m,n}, \quad [a_m, a_n] = 0. \tag{21}$$

Let us assume that the ground state $|0\rangle$ is coherent with respect to the gauge degrees of freedom:

$$a_n|0\rangle = z_n|0\rangle, \quad \langle 0|a_n^\dagger = \langle 0|z_n^*. \tag{22}$$

Here $\{z_n\}$ are complex numbers. Since all operators $\{a_n, a_n^\dagger, A_N, A_N^\dagger\}$ on the right-hand side of Eq. (8) are by definition normal-ordered, then

$$\begin{aligned}\langle 0|\Phi(x)|0\rangle &= \Phi_{(cl)}(x)\\ &+ \sum_{|N|<N_0} [\phi_N(x)A_N + \phi_N^*(x)A_N^\dagger] + \ldots.\end{aligned} \tag{23}$$

In contrast to Eq. (18), on the right-hand side of Eq. (23) all functions $\Phi_{(cl)}(x)$, $\phi_N(x)$, and so on are c-number functions which depend on the numbers z_n. Note that the collection of numerical functions $\{\phi_{\parallel n}(x), \phi_N(x)\}$ forms a complete independent set of functions in terms of which any variation of the field $\Phi(x)$ can be expanded.

In what follows we shall assume that the Heisenberg equations and the constraints are given in a Lagrangian form. This means that the momentum variables $\mathscr{P}(x)$ are expressed in terms of the coordinate variables $\Phi(x)$ and their derivatives with the aid of the corresponding part of the Heisenberg equations and are substituted into the constraints equations and the remaining Heisenberg equations. As a result, in the theory of gravitation we obtain quantum microscopic Einstein equations and Lagrangian equations for the matter fields. The term "microscopic" in this case means that the energy–momentum tensor in Einstein's equations is clearly expressed in terms of the matter field (scalar, vector, spinor, and so on), and the term "quantum" means that all fields in the Einstein–Lagrange equations are quantized. In what follows we shall call the collection of quantum microscopic equations of motions and constraints in the Lagrangian form briefly as the equations of motion.

The problem of ordering the operator fields in the equations of motion cannot be solved on the basis of such a general analysis. Apparently, this problem is due to the problem of the consistency of the theory, and it can be solved together with a development of an effective computational scheme.

We now average the equations of motion relative to the gauge degrees of freedom. In order to be able to use Eqs. (22) in the equations of motion within the averaging symbol

$$\langle 0|\left\{R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R - \frac{8\pi G}{c^4}T_{\mu\nu}\right\}|0\rangle = 0 \tag{24}$$

the collection of operators $\{a_n, a_n^\dagger\}$ in the braces in Eq. (24) must be normal-ordered. Since the field $\Phi(x)$ [or $g_{\mu\nu}(x)$] is a series in the field $\phi_\parallel^{(1)}(x)$ and its derivatives [see Eq. (20)], as a result of such ordering sums of the form

$$\sum_n \phi_{\parallel n}(x)\phi_{\parallel n}^*(x), \quad \sum_n \phi_{\parallel n}(x)_{,\mu}\phi_{\parallel n}^*(x), \tag{25}$$

and so on arise. In sums (25) the index n, enumerating the gauge degrees of freedom, runs through all its values. It is very important that the physical quantities do not depend on the gauge degrees of freedom. Consequently, regularization with respect to the gauge degrees of freedom is not required. This means that in sums (25) summation indeed extends over all n. Therefore, the sums in Eq. (25) are proportional to integrals of the form

$$\int \frac{d^{(D-1)}k}{|k|}P(|k|, k_i), \tag{26}$$

where D is the dimension of space-time, and $P(|k|, k_i)$ are polynomials with positive powers of $|k|$ and k_i, and

the polynomials in Eq. (26) include, generally speaking, a polynomial of zero degree

$$P_0(|k|, k_i) \equiv 1.$$

Indeed, let us consider the field

$$h(x) = \sum_N (A_N \phi_N(x) + A_N^\dagger \phi_N^*(x)) + \phi_\parallel^{(1)}(x), \qquad (27)$$

which is the first term in the expansion of field (8) in terms of the operators $\{A_N, A_N^\dagger, a_n, a_n^\dagger\}$. In the theory of gravitation the first term in the expansion of the metric tensor with respect to the Planck scale according to Eq. (40) plays the role of field (27). This field is a bosonic massless tensor field in curved space-time, and its kinetic energy has the usual structure for bosonic fields, being a second-order differential operator. Consequently, the quantization procedure, leading to the expansion of field (27) in terms of the modes $\{\phi_N(x), \phi_{\parallel n}\}$, in the dimensional sense is identical to the quantization procedure in scalar field theory. As is well known, in the theory of a massless scalar field in Minkowski space the modes satisfy the formula

$$\phi_{\mathbf{k}}(x) \propto |\mathbf{k}|^{-1/2} \exp(i\mathbf{k} \cdot \mathbf{x}).$$

Since the curvature of space-time plays no role here, it is evident that sums (25) have the form of integrals (26).

In the method of dimensional regularization all integrals (26) and thereby all sums (25) vanish for $D > 2$, i.e., in all theories of gravitation in space-time with dimension greater than 2. This result means that in all theories of gravitation, except for two-dimensional theories, all operators a_n and $a_n^\dagger$ in Eq. (24), even before being normal-ordered, can be replaced by the numbers z_n and z_n^*, respectively, after which the averaging operation in Eq. (24) can be dropped. In two-dimensional generally covariant theories more direct calculations are required when working with gauge degrees of freedom. Fortunately, this is possible (see [7]) because of the kinematic simplicity of two-dimensional theories.

We underscore that the sums of the form

$$\sum_{|N| < N_0} \phi_N(x) \phi_N^*(x)$$

which arise when ordering the operators A_N and $A_N^\dagger$ in Eq. (24), cannot be dropped, since these sums are regularized. The physically measured quantities include the indicated regularized sums, giving the quantum corrections. These quantum corrections arise as a result of normal ordering of the operators describing the physical degrees of freedom. In this manner, Dirac [9] calculated precisely the contribution to the anomalous magnetic moment of the electron and the Lamb shift of the electron levels in the hydrogen atom.

The coefficient functions of the matter fields [e.g., $\psi_N(x)$ in Eq. (9)] also depend on the gauge degrees of freedom $\{a_n, a_n^\dagger\}$. The latter can be replaced within the averaging symbol in Eq. (24) by the numbers $\{z_n, z_n^*\}$. The basis for such a substitution and the resulting limitation on the dimension of space-time remain the same.

Now we can greatly supplement our system of axioms by the following supposition: field (23) is used in axioms 3–5, i.e., the quantized field, averaged with respect to the gauge degrees of freedom. The fields $\Phi_{(cl)}(x)$, $\phi_N(x)$, $\psi_N(x)$, and so on satisfy certain equations which can be obtained uniquely from the Lagrangian equations of motion, if the expansion of the field $\Phi(x)$ in form (23) is substituted into them and then, after normal ordering of the operators $\{A_N, A_N^\dagger\}$, the coefficients of the various powers of the generators of the Heisenberg algebra $\{A_N, A_N^\dagger\}$ are equated to zero. As a result of the indicated normal ordering, a relation arises between the higher order coefficient functions and the lower order coefficient functions in expansion (23). We obtain an infinite chain of equations for the coefficient functions $\{\Phi_{(cl)}(x), \phi_N(x), \psi_N(x), \ldots\}$.

The latter conjecture can be introduced with the aid of the following axiom, replacing axiom 5.

Axiom 5'. *The equations of motion for the quantized fields* (23), *to within the ordering of the quantized fields, have the same form as the corresponding classical equations of motion.*

We note that according to axiom 5' the dynamics of the gauge degrees of freedom in a real theory of gravitation is always semiclassical. At the intuitive level this conjecture can be justified by the fact that for the gauge degrees of freedom there is no potential, and their dynamics are similar to that of a free particle. It is easy to see that the latter becomes classical as time elapses. Indeed, let x and p be Heisenberg coordinate and momentum operators of a free nonrelativistic particle with mass m. Then

$$p = p_0, \quad x = x_0 + \frac{p_0}{m} t,$$

where t is the time, and x_0 and p_0 are constant operators, satisfying the commutation relation $[x_0, p_0] = i\hbar$. It is obvious that if $\langle p_0 \rangle \neq 0$, then as $t \longrightarrow \infty$

$$\frac{\langle x \rangle \langle p \rangle}{|\langle [x, p] \rangle|} \longrightarrow \infty,$$

which means that the dynamics of the free particle is semiclassical.

The system of axioms 3, 4, and 5' gives a definition of the quantized variant of the given generally covariant theory.

We note once again that such a general analysis does not solve the problem of the ordering of operator fields in the equations of motion.

4. DYNAMICAL QUANTIZATION OF GRAVITY

We shall now apply the quantization scheme developed above to the theory of gravitation. Let us consider the theory of gravitation with a Λ term, where the gravitation interacts in a minimal manner with the Dirac field. The action of such a theory has the form

$$S = -\frac{1}{l_P^2}\int d^4x\sqrt{-g}(R+2\Lambda) + \int d^4x\sqrt{-g}\left\{\frac{i}{2}e_a^\mu(\bar\psi\gamma^a\mathscr{D}_\mu\psi - \overline{\mathscr{D}_\mu\psi}\gamma^a\psi) - m\bar\psi\psi\right\}. \quad (28)$$

Here $\{e_a^\mu\}$ is an orthonormalized basis, $g_{\mu\nu}$ is the metric tensor, and $\eta_{ab} = \mathrm{diag}(1, -1, -1, -1)$, so that

$$g_{\mu\nu}e_a^\mu e_b^\nu = \eta_{ab}, \quad R = e_a^\mu e_b^\nu R_{\mu\nu}^{ab},$$

the 2-form of the curvature is given by

$$d\omega^{ab} + \omega_c^a \wedge \omega^{cb} = \frac{1}{2}R_{\mu\nu}^{ab}dx^\mu \wedge dx^\nu,$$

where the 1-form $\omega_b^a = \omega_{b\mu}^a dx^\mu$ is the connectivity in the orthonormalized basis $\{e_a^\mu\}$. The spinor covariant derivative is given by the formula

$$\mathscr{D}_\mu\psi = \left(\frac{\partial}{\partial x^\mu} + \frac{1}{2}\omega_{ab\mu}\sigma^{ab}\right)\psi,$$

$$\sigma^{ab} = \frac{1}{4}[\gamma^a, \gamma^b],$$

γ^a are the Dirac matrices:

$$\gamma^a\gamma^b + \gamma^b\gamma^a = 2\eta^{ab}.$$

Since the Planck constant is assumed to be 1, the parameter l_P is the Planck scale.

We shall write out the equations of motion for system (28). Varying action (28) relative to the connectivity gives the equation

$$\nabla_\mu e_\nu^a - \nabla_\nu e_\mu^a = -\frac{1}{4}l_P^2\varepsilon_{bcd}^a e_\mu^b e_\nu^c \bar\psi\gamma^5\gamma^d\psi \equiv T_{\mu\nu}^a. \quad (29)$$

In deriving the last equation, we employed the equality

$$\gamma^a\sigma^{bc} + \sigma^{bc}\gamma^a = -i\varepsilon^{abcd}\gamma^5\gamma_d. \quad (30)$$

Here ε_{abcd} is the absolutely antisymmetric tensor, where $\varepsilon_{0123} = 1$. The right-hand side of Eq. (29) is the torsion tensor. We can see that including into the theory a Dirac field results in the appearance of torsion.

We note that torsion (29) possesses the property

$$T_{\mu\nu}^\nu \equiv e_a^\nu T_{\mu\nu}^a \equiv 0. \quad (31)$$

Consequently, even though torsion exists in the theory being considered, the torsion tensor is not present in the Dirac theory:

$$(ie_a^\mu\gamma^a\mathscr{D}_\mu - m)\psi = 0. \quad (32)$$

Varying action (28) with respect to the orthonormalized basis gives the Einstein equation, which we write in the form

$$R_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{1}{2}l_P^2\left\{\frac{i}{2}(\bar\psi\gamma^c e_{c(\mu}\mathscr{D}_{\nu)}\psi - e_{c(\mu}\overline{\mathscr{D}_{\nu)}\psi}\gamma^c\psi) - \frac{1}{2}m\bar\psi\psi g_{\mu\nu}\right\}. \quad (33)$$

Here the expression in braces is $(T_{\mu\nu} - (1/2)g_{\mu\nu}T)$, where $T_{\mu\nu}$ is the energy–momentum tensor on the mass shell [i.e., taking account of the equations of motion of matter—in our case, the Dirac equation (31)].

Equations (29), (30), and (33), together with the relations

$$g_{\mu\nu} = \eta_{ab}e_\mu^a e_\nu^b, \quad e_a^\mu e_\mu^b = \delta_a^b$$

form a complete system of classical equations of motion and constraints for system (28).

We now represent the field as a sum of classical and quantum components:

$$g_{\mu\nu} = g_{(cl)\mu\nu} + h_{\mu\nu}, \quad e_\mu^a = e_{(cl)\mu}^a + f_\mu^a. \quad (34)$$

We assume that the fermionic field has no classical component, so that

$$\psi(x) = \sum_{|N|<N_F}(B_N\psi_N^{(+)}(x) + C_N^\dagger\psi_N^{(-)}(x)) + \ldots, \quad (35)$$

where the Fermi creation and annihilation operators satisfy the following anticommutation relations (as usual, only the nonzero relations are written out):

$$\{B_M, B_N^\dagger\} = \{C_M, C_N^\dagger\} = \delta_{M,N}. \quad (36)$$

The complete orthonormal set of fermionic modes $\{\psi_N^{(\pm)}(x)\}$ can be naturally determined as follows. We denote by $\Sigma^{(3)}$ the spacelike hypersurface, defined by the equation $t = \mathrm{const}$, and by $\Sigma_0^{(3)}$ the hypersurface at $t = t_0$. Let the metric in space-time be given by means of the tensor $g_{\mu\nu}$. This metric induces a metric on $\Sigma_0^{(3)}$,

which in the local coordinates x^i, $i = 1, 2, 3$, is represented by the metric tensor ${}^3g_{ij}$. Using the equations

$$
\begin{gathered}
{}^3g_{ij,k} = \gamma^l_{ik}\,{}^3g_{lj} + \gamma^l_{jk}\,{}^3g_{il}, \quad \gamma^k_{ij} = \gamma^k_{ji}, \\
{}^3g_{ij} = -\sum_{\alpha=1}^{3} {}^3e^\alpha_i\,{}^3e^\alpha_j, \quad {}^3e^\alpha_i\,{}^3e^i_\beta = \delta_{\alpha\beta}, \\
\partial_i\,{}^3e^i_\alpha + \gamma^j_{ki}\,{}^3e^k_\alpha + {}^3\omega_{\alpha\beta i}\,{}^3e^j_\beta = 0, \\
{}^3\omega_{\alpha\beta i} = -{}^3\omega_{\beta\alpha i}
\end{gathered}
$$

the connectivity (without torsion) in local coordinate γ^i_{jk} and a spin connectivity ${}^3\omega_{\alpha\beta i}$ are determined on $\Sigma_0^{(3)}$. For a Dirac single-particle Hamiltonian we have

$$
\begin{gathered}
\mathscr{H}_{\mathscr{D}} = -i\,{}^3e^i_\alpha\alpha^\alpha\left(\partial_i + \frac{1}{2}\,{}^3\omega_{\beta\gamma i}\frac{1}{4}[\alpha^\beta, \alpha^\gamma]\right) + m\gamma^0, \\
\alpha^\beta = \gamma^0\gamma^\beta.
\end{gathered}
$$

It is easy to check that in the metric

$$
\langle \psi_M, \psi_N \rangle = \int\limits_{\Sigma_0^{(3)}} d^3x\sqrt{-{}^3g}\,\psi^\dagger_M\psi_N \tag{37}
$$

the operator $\mathscr{H}_{\mathscr{D}}$ is self-conjugate. Consequently, the solution of the problem for the eigenvalues on $\Sigma_0^{(3)}$

$$
\mathscr{H}^{(0)}_{\mathscr{D}}\psi^{(\pm)}_N(x) = \pm\varepsilon_N\psi^{(\pm)}_N(x), \quad \varepsilon_N > 0, \tag{38}
$$

has a complete set of orthonormalized modes in metric (37). The index 0 everywhere means that in the corresponding quantity the fields are taken in the zero approximation with respect to quantum fluctuations.

We note that a one-to-one relation can be established between the positive- and negative-frequency modes by means of the equation

$$
\gamma^0\gamma^5\psi^{(+)}_M = \psi^{(-)}_M.
$$

We call attention to the fact that the scalar product

$$
(\psi_M, \psi_N) = \int\limits_{\Sigma^{(3)}} d^3x\sqrt{-g^{(0)}}\,\psi^\dagger_M\psi_N \tag{39}
$$

is not always the same as the scalar product (37). These scalar products coincide, if the path function $N = 1$, which happens, for example, for the metric

$$
g^{(0)}_{0i} = 0, \quad g^{(0)}_{00} = 1.
$$

The scalar product (39) has the advantage over the scalar product (37) that if the modes $\{\psi^{(\pm)}_N(x)\}$ satisfy the Dirac equation in the zero approximation with respect to quantum fluctuations (which, according to the exposition below, does indeed happen), then the scalar product (39) is conserved in time.

The field $h_{\mu\nu}$ in Eq. (34) can be expanded as follows:

$$
\begin{aligned}
h_{\mu\nu} &= l_P \sum_{|N|<N_0} (h_{N\,\mu\nu}A_N + h^*_{N\,\mu\nu}A^\dagger_N) \\
&+ l_P^2\Bigg\{ \sum_{|N_1|,\,|N_2|<N_0} (h_{N_1,N_2\,\mu\nu}A_{N_1}A_{N_2} \\
&+ h^*_{N_1N_2\,\mu\nu}A^\dagger_{N_1}A^\dagger_{N_2} + h_{N_1|N_2\,\mu\nu}A^\dagger_{N_1}A_{N_2}) \\
&+ \sum_{|N_1|,\,|N_2|<N_F} (h^{F(++)}_{N_1N_2\,\mu\nu}B^\dagger_{N_1}B_{N_2} + h^{F(--)}_{N_2N_1\,\mu\nu}C^\dagger_{N_1}C_{N_2} \\
&+ h^{F(+-)}_{N_1N_2\,\mu\nu}B^\dagger_{N_1}C^\dagger_{N_2} + h^{F(+-)*}_{N_1N_2\,\mu\nu}C_{N_2}B_{N_1})\Bigg\} + \ldots.
\end{aligned} \tag{40}
$$

In Eqs. (34), (35), and (40) the c-number coefficient fields $\psi^{(\pm)}_N$, $g_{(cl)\,\mu\nu}$, $h_{N\,\mu\nu}$, and so on can be expanded in powers of the Planck scale, for example,

$$
g_{(cl)\,\mu\nu} = g^{(0)}_{\mu\nu} + l_p^2 g^{(2)}_{(cl)\,\mu\nu} + \ldots.
$$

Since fields (40) are real, we have

$$
\begin{gathered}
h_{N_1N_2\,\mu\nu} = h_{N_2N_1\,\mu\nu}, \quad h^*_{N_1|N_2\,\mu\nu} = h_{N_2|N_1\,\mu\nu}, \\
h^{F(++)*}_{N_2N_1\,\mu\nu} = h^{F(++)}_{N_1N_2\,\mu\nu}, \quad h^{F(--)*}_{N_2N_1\,\mu\nu} = h^{F(--)}_{N_1N_2\,\mu\nu}.
\end{gathered} \tag{41}
$$

The operators $\{A_N, A^+_N\}$ satisfy the Bose commutation relations (31). A method for choosing the set of functions $\{h_{N\,\mu\nu}\}$ will be discussed below.

According to the dynamical quantization scheme, we must substitute fields (34), (35), and (40) into Eqs. (29) and (32), (33), after which the operators $\{A_N, A^+_N\}$ must be normal-ordered and all coefficients of the various powers of these operators and the Planck scale must be set equal to zero.

Thus, we obtain the first of these equations:

$$
\nabla^{(0)}_\mu e^{(0)a}_\nu - \nabla^{(0)}_\nu e^{(0)a}_\mu = 0, \quad R^{(0)}_{\mu\nu} + \Lambda g^{(0)}_{\mu\nu} = 0. \tag{42}
$$

Here and below all raising and lowering of indices are done with the tensors $g^{(0)}_{\mu\nu}$ and $g^{(0)\mu\nu}$. Thus, in the lowest approximation the fields satisfy the classical equations of motion. In the zeroth approximation we also have a series of equations for the fermionic modes:

$$
(ie^{(0)\mu}_a\gamma^a\mathscr{D}^{(0)}_\mu - m)\psi^{(\pm)}_N = 0. \tag{43}
$$

We now introduce the notation

$$K_{\mu\nu}^{(0)\lambda\rho} = \Big[-\frac{1}{2}\nabla_\sigma^{(0)}\nabla^{(0)\sigma}\delta_\mu^\lambda\delta_\nu^\rho - R_{\mu\nu}^{(0)\lambda\rho} + R_\nu^{(0)\rho}\delta_\mu^\lambda + \nabla_\mu^{(0)}\Big(\nabla^{(0)\lambda}\delta_\nu^\rho - \frac{1}{2}\nabla_\nu^{(0)}g^{(0)\lambda\rho}\Big)\Big] + [\mu \longleftrightarrow \nu] + 2\Lambda\delta_{(\mu}^\lambda\delta_{\nu)}^\rho, \quad (44)$$

$$R_{\mu\nu}^{(0)(2)}(h, h') = \frac{1}{2}[R_{\mu\nu}^{(0)(2)}(h + h', h + h') - R_{\mu\nu}^{(0)(2)}(h, h) - R_{\mu\nu}^{(0)(2)}(h', h')]. \quad (45)$$

It is easily checked that

$$\frac{1}{2}K_{\mu\nu}^{(0)\lambda\rho} = \frac{\delta(R_{\mu\nu} + \Lambda g_{\mu\nu})}{\delta g_{\lambda\rho}}\bigg|_{g_{\mu\nu} = g_{\mu\nu}^{(0)}},$$

where $R_{\mu\nu}^{(0)(2)}(h, h)$ is a quadratic form of the tensor field $h_{\lambda\rho}$, which can be constructed in terms of the second variation of $R_{\mu\nu}$ relative to the metric tensor at the point $g_{\mu\nu}^{(0)}$. We write out the complete form:

$$R_{\mu\nu}^{(0)(2)}(h, h) = \frac{1}{2}(h_\lambda^\rho h_{\rho;\,\mu}^{\lambda}{}_{);\nu} - \frac{1}{2}[h_\sigma^\lambda(h_{\mu;\,\nu}^{\sigma} + h_{\nu;\,\mu}^{\sigma} - h^{;\sigma}_{\mu\nu})]_{;\lambda} + \frac{1}{4}h_{\lambda;\,\rho}^{\lambda}(h_{\mu;\,\nu}^{\rho} + h_{\nu;\,\mu}^{\rho} - h^{;\rho}_{\mu\nu}) - \frac{1}{4}(h_{\rho;\,\nu}^{\lambda} + h_{\nu;\,\rho}^{\lambda} - h^{;\lambda}_{\nu\rho})(h_{\mu;\,\lambda}^{\rho} + h_{\lambda;\,\mu}^{\rho} - h^{;\rho}_{\mu\lambda}).$$

Thus, $R_{\mu\nu}^{(0)(2)}(h, h')$ is a symmetric bilinear form with respect to its arguments $h_{\mu\nu}$ and $h'_{\lambda\rho}$, which in what follows are operator fields (40). Thus, here the problem of ordering the operator fields to lowest order has been solved.

Now we can write out the following relations, which follow from the exact quantum equations with the expansion indicated above. To first order in l_P we have

$$\frac{1}{2}K_{\mu\nu}^{(0)\lambda\rho}h_{N\lambda\rho} = 0. \quad (46)$$

We note that, using Eqs. (42), the operator (44) vanishes on the quantity $(\xi_{\mu\,;\nu} + \xi_{\nu\,;\mu})$. Consequently, the value of the operator (44) on the fields $h_{\mu\nu}$ and

$$h'_{\mu\nu} = h_{\mu\nu} + \xi_{\mu\,;\nu} + \xi_{\nu\,;\mu} \quad (47)$$

coincide for any vector field ξ_μ. This fact is a consequence of the gauge invariance of the theory. Using the indicated gauge invariance, any solution of Eq. (46) can be put into the form

$$\nabla_\nu^{(0)}h_\mu^\nu - \frac{1}{2}\nabla_\mu^{(0)}h_\nu^\nu = 0. \quad (48)$$

In what follows, we shall assume that the field satisfies the gauge condition (48), which is convenient in a number of problems. It is obvious that taking account of the gauge condition (48) the term in parentheses in operator (44) vanishes.

To clarify the question of the normalization of the gravitational modes, we shall employ the following technique. The equation of motion (46) can be obtained with the action

$$S^{(2)} = \int d^4x\sqrt{-g^{(0)}}h^{\mu\nu}K_{\mu\nu}^{(0)\lambda\rho}h_{\lambda\rho}. \quad (49)$$

Hence follows the canonically-conjugate momentum for the field $h_{\mu\nu}$ and the one-time commutation relations:

$$\pi^{\mu\nu} = \sqrt{-g^{(0)}}\nabla^{(0)0}h^{\mu\nu}, \quad [h_{\mu\nu}(x), \pi^{\lambda\rho}(y)] = i\delta_{(\mu}^\lambda\delta_{\nu)}^\rho\delta^{(3)}(x - y). \quad (50)$$

Evidently, in Eq. (50) the fields are free of constraints (48). We represent the field $h_{\mu\nu}$ in the form [compare with the first term in Eq. (40)]

$$h_{\mu\nu}(x) = \sum_N (h_{N\mu\nu}(x)A_N + h_{N\mu\nu}^*(x)A_N^\dagger). \quad (51)$$

The set of operators $\{A_N, A_N^\dagger\}$ forms a Heisenberg algebra (1), and the functions $\{h_{N\,\mu\nu}\}$ satisfy Eqs. (46). Equations (50) and (51) lead to the following relations which reflect the orthonormal nature of the set of the modes:

$$i\int\limits_{\Sigma^{(3)}} d^3x\sqrt{-g^{(0)}}[h_M^{\mu\nu *}\nabla^{(0)0}h_{N\,\mu\nu} - (\nabla^{(0)0}h_M^{\mu\nu *})h_{N\,\mu\nu}] = \delta_{M,N}. \quad (52)$$

In the latter equations the integration extends over any spacelike hypersurface $\Sigma^{(3)}$. As a result of Eqs. (46), integrals (52) indeed do not depend on the hypersurface. It is natural to assume that the gravitational modes satisfy conditions (52). The significance of Eq. (52) is that renormalization of the coefficient functions in expansion (40) is given with its help.

In second order in l_P, we obtain the following equations:

$$\frac{1}{2}K_{\mu\nu}^{(0)\lambda\rho}h_{N_1N_2\,\lambda\rho} = -R_{\mu\nu}^{(0)(2)}(h_{N_1}, h_{N_2}), \quad (53)$$

$$\frac{1}{2}K_{\mu\nu}^{(0)\lambda\rho}h_{N_1|N_2\,\lambda\rho} = -2R_{\mu\nu}^{(0)(2)}(h_{N_1}^*, h_{N_2}), \quad (54)$$

$$\frac{1}{2}K^{(0)\lambda\rho}_{\mu\nu}h^{F(\pm\pm)}_{N_1N_2\lambda\rho} = \pm\frac{i}{4}(\bar{\psi}^{(\pm)}_{N_1}\gamma^c e^{(0)}_{c(\mu}\mathcal{D}^{(0)}_{\nu)}\psi^{(\pm)}_{N_2} - e^{(0)}_{c(\mu}\overline{\mathcal{D}^{(0)}_{\nu)}\psi^{(\pm)}_{N_1}}\gamma^c\psi^{(\pm)}_{N_2}), \tag{55}$$

$$\frac{1}{2}K^{(0)\lambda\rho}_{\mu\nu}h^{F(+-)}_{N_1N_2\lambda\rho} = \frac{i}{4}(\bar{\psi}^{(+)}_{N_1}\gamma^c e^{(0)}_{c(\mu}\mathcal{D}^{(0)}_{\nu)}\psi^{(-)}_{N_2} - e^{(0)}_{c(\mu}\overline{\mathcal{D}^{(0)}_{\nu)}\psi^{(+)}_{N_1}}\gamma^c\psi^{(-)}_{N_2}), \tag{56}$$

$$\frac{1}{2}K^{(0)\lambda\rho}_{\mu\nu}g^{(2)}_{(cl)\lambda\rho} = -\sum_{|N|<N_0}R^{(0)(2)}_{\mu\nu}(h^*_N, h_N) + \frac{i}{4}\sum_{|N|<N_F}(\bar{\psi}^{(-)}_N\gamma^c e^{(0)}_{c(\mu}\mathcal{D}^{(0)}_{\nu)}\psi^{(-)}_N - e^{(0)}_{c(\mu}\overline{\mathcal{D}^{(0)}_{\nu)}\psi^{(-)}_N}\gamma^c\psi^{(-)}_N). \tag{57}$$

It is evident from Eq. (29) that torsion appears in the same order ($\sim l_P^2$). Here, however, we do not write out the corresponding corrections for the connectivity.

We shall now briefly summarize the results obtained.

According to the dynamical method, the quantization of gravity starts with finding a solution of the classical microscopic field equations of motion (for example, the solutions of Eqs. (42) in the example considered above). The classical solution is determined by (or determines) the topology of space-time. Then, using the classical approach, Eqs. (43) and (46), which determine the single-particle modes $\{\psi^{(\pm)}_N, h_{N\mu\nu}\}$, are solved. To solve Eq. (46) the gauge must be fixed, since the operator (44) is degenerate because of the gauge invariance of the theory. At the first step these modes are determined in the zeroth approximation according to the Planck scale, and their normalization is fixed using Eqs. (39) and (52). Given the set of modes $\{\psi^{(\pm)}_N, h_{N\mu\nu}\}$, we can explicitly write out the right-hand sides of Eqs. (53)–(57) and then solve them for the two-particle modes $h_{N_1N_2\mu\nu}$, $h_{N_1|N_2\mu\nu}$, and so on, and find the correction $g^{(2)}_{(cl)\mu\nu}$ which is of second order in l_P to the classical component of the metric tensor. We call attention to the fact that the right-hand side of Eq. (57) arises because the operators must be normal-ordered. The solution of Eq. (57) can be interpreted as a single-loop contribution to the average of the metric tensor with respect to the ground state.

We note that if a nonsymmetric bilinear form were used on the right-hand sides of Eqs. (53)–(57), then the condition that the metric tensor be real would be violated. Consequently, the condition that the metric tensor is real determines the ordering of the operator fields in the equations of motion at least in second order with respect to the operator fields.

It is important that all Eqs. (42), (46), and so on which arise are generally covariant, since they are expansions of generally covariant equations. Thus, the method of dynamical quantization leads to a regularized gauge-invariant theory of gravitation, which contains an arbitrary number of physical degrees of freedom.

We shall now make a remark about the compatibility of Eqs. (53)–(57) and the analogous equations arising in higher orders. Let $h_{\mu\nu}$ be an arbitrary symmetric tensor field and $K^{(0)}$ the operator (44), acting on this vector field. It is easily verified that, using Eqs. (42), we obtain the identity (compare with Eq. (48))

$$\nabla^{(0)}_\nu(K^{(0)}h)^\nu_\mu - \frac{1}{2}\nabla^{(0)}_\mu(K^{(0)}h)^\nu_\nu = 0.$$

Consequently, in order for Eqs. (53)–(57) to be compatible the right-hand sides of these equations must satisfy the same identity. It is easy to see that this is indeed the case. Indeed, Eqs. (53)–(56) are identical to the analogous classical equations arising when nonuniform modes (higher order harmonics) and the subsequent expansion of the classical Einstein equation in powers of the nonlinearity or the Planck length are added to the uniform fields. Hence it follows that each term on the right-hand sides of the "loop" equations of the type (57) likewise satisfy the necessary identity, since these terms have the same form as the right-hand sides of the "nonloop" Eqs.(53)–(56).

We also call attention to the fact that in the method of dynamical quantization it is implicitly assumed that the quantum anomaly is absent in the algebra of the operators of the constraints of the first kind. Consequently, the method of dynamical quantization must be justified in each specific case by concrete calculations, which must be not only mathematically correct but also physically meaningful.

5. ON THE PROBLEM OF DECOHERENCE IN QUANTUM COSMOLOGY

We shall now show how the problem of decoherence in quantum cosmology in a model of the inflating universe can be solved on the basis of the method of dynamical quantization. The solution proposed here is, in the opinion of the present author, quite simple and natural.

We shall briefly formulate the problem of decoherence in quantum cosmology (see [10–15] and references therein).

In Friedmann type models it is natural to take the scale factor of the universe a as the time parameter. According to all present experimental data the quantum fluctuations of a are not observed because they are small. This can be easily explained if there existed an external (with respect to the universe) observer who would perform an experiment as a result of which the wave function of the universe would be reduced to a

state in which the scale factor has a definite value. However, in quantum cosmology the observer is always part of the system and consequently the explanation presented above is not acceptable.

At the present time another idea is generally accepted. According to this idea, nonuniform quantum fluctuations lead to decoherence of the density matrix describing the homogeneous degree of freedom—the scale factor. The meaning of this assertion is as follows. Let us consider the density matrix of the universe and calculate its trace with respect to all nonuniform degrees of freedom. This gives a density matrix $\rho(a, a')$ for the scale factor. Qualitative considerations lead to the following form for this matrix:

$$\rho(a, a') \sim \rho(a - a'). \tag{58}$$

The latter formula means that there is no coherence (decoherence) for the scale factor a. In other words, interference of various values of a does not appear in any measurement.

However, concrete calculations performed in the single-loop approximation encounter serious difficulties associated with the ultraviolet divergences of the theory. These difficulties were overcome in [13, 14]. It was found that conventional regularization itself leads to a physically meaningless result. To obtain a physically acceptable result, additional nonlocal transformations of the fields, which possess a completely different character for bosonic and fermionic fields, are required. In addition, the question of the divergences in the calculations in higher order groups remains open. It is obvious that this question cannot be solved without constructing a systematic quantum theory of gravitation or including the theory of gravitation in a more fundamental theory, for example, string theory.

We shall show at a qualitative level how the problem of decoherence can be solved on the basis of the method of dynamical quantization in a model of an inflating universe. The solution proposed here remains valid when higher order corrections with respect to the Planck scale are taken into account.

First we note that the qualitative arguments in [10] which lead to Eq. (58) in our case indeed make sense, since their validity is implicitly based on the assumption that the number of nonuniform modes, though large, is finite. This situation occurs in the method of dynamical quantization.

We shall briefly reproduce the arguments presented in [10]. We denote by $\{a, x_N\}$ the complete set of commuting variables, where a is the scale factor (59) and x_N are the degrees of freedom of the nonuniform modes. In the single-loop approximation (which corresponds to taking into account only the "single-particle" modes $g_{N\,ij}$ and ψ_N) the wave function has the form

$$\Psi\{a_j; x_N\} = \Psi_0(a) \sum_{|N| < N_0} f_N(a, x_N).$$

Here $\Psi_0(a)$ is the wave function of the minisuperspatial model. According to Hartle and Hawking for small a

$$f_N(a, x_N) = f_N^{(0)}(x_N),$$

where $f_N^{(0)}(x_N)$ is the wave function of the ground state of the corresponding degree of freedom for small a. By definition the density matrix (58) is obtained using the following integral:

$$\rho(a, a') = \Psi_0(a)\Psi_0^*(a') \times \prod_{|N| < N_0} \int dx_N f_N(a, x_N) f_N^*(a', x_N). \tag{*}$$

We shall estimate the integral (∗) using the single integrals

$$\rho_N(a, a') = \int dx_N f_N(a, x_N) f_N^*(a', x_N),$$

which are equal to 1 for small a, a' but decrease rapidly in modulus with increasing a, a', remaining equal to 1 only for $a = a'$. Consequently, for increasing a and $a' \neq a$

$$\prod_{|N| < N_0} \rho_N(a, a') \longrightarrow 0, \quad N_0 \longrightarrow \infty.$$

Thus, we arrive at the formula (58). It is important in this argument that although N_0 is large, it is still finite.

We shall now show how the decoherence problem can be solved on the basis of the method of dynamical quantization.

As is well known, the spatially uniform solution of Eq. (42) has the form

$$g_{00}^{(0)} = 1, \quad g_{0i}^{(0)} = 0, \quad g_{ij}^{(0)} = -a^2(t)\tilde{g}_{ij}, \qquad a(t) = \cosh Ht, \quad H^2 = \frac{1}{3}\Lambda. \tag{59}$$

Here $\tilde{g}_{ij}$ is a positive-definite metric on the sphere $S^3_{H^{-1}}$ with radius H^{-1} in some coordinates and $x^0 = t$. The tilde denotes the corresponding quantities on the sphere $S^3_{H^{-1}}$. The solution (59) describes a universe in the inflation stage. We write out the nonzero components of the connectivity:

$$\Gamma_{ij}^{(0)0} = a\dot{a}\tilde{g}_{ij}, \quad \Gamma_{0j}^{(0)i} = \frac{\dot{a}}{a}\delta_j^i, \quad \Gamma_{jk}^{(0)i} = \tilde{\Gamma}^i{}_{jk}. \tag{60}$$

Hence we have for the components of the Riemann and Ricci tensors

$$R_{\mu\nu\lambda\rho}^{(0)} = -H^2(g_{\mu\lambda}^{(0)}g_{\nu\rho}^{(0)} - g_{\mu\rho}^{(0)}g_{\nu\lambda}^{(0)}), \qquad R_{\mu\nu}^{(0)} = -3H^2 g_{\mu\nu}^{(0)}. \tag{61}$$

We also write out the formula

$$-g^{(0)} = a^6(t)\tilde{g}, \tag{62}$$

which follows from Eq. (59).

It follows from general considerations based on the gauge invariance of the theory that at each point of space (or for each "frequency") only two independent degrees of freedom of the field $h_{\mu\nu}$ (40) remain. It is shown in the Appendix that gauge transformations can be used to obtain

$$h_{N\,0\mu} = 0, \quad h^i_{N\,i} = 0, \quad \nabla^{(0)}_j h^j_{N\,i} = 0. \tag{63}$$

In this case and taking account of Eq. (61), Eq. (46) becomes

$$(\nabla^{(0)}_\lambda \nabla^{(0)\lambda} + 2H^2) h_{N\,ij} = 0. \tag{64}$$

The last equation can be rewritten, using Eqs. (59) and (60), as

$$(\nabla^{(0)}_0)^2 h_{N\,ij} - \frac{1}{\cosh^2 Ht}\tilde{\nabla}_k\tilde{\nabla}^k h_{N\,ij} + 3H(\tanh H)\nabla_0 h_{N\,ij} + 2H^2(1+\tanh^2 Ht)h_{N\,ij} = 0, \tag{65}$$

$$(\nabla^{(0)}_0)^n h_{N\,ij} = a^2\left(\frac{\partial}{\partial t}\right)^n (a^{-2}h_{N\,ij}), \quad n = 1, 2, \ldots.$$

We note that the operator $\tilde{\nabla}_k\tilde{\nabla}^k$ leaves invariant the space of vector fields $\{h_{N\,ij}\}$ satisfying Eqs. (63). This means that

$$\tilde{g}^{ij}\tilde{\nabla}_k\tilde{\nabla}^k h_{N\,ij} = 0, \quad \tilde{\nabla}_j\tilde{\nabla}_k\tilde{\nabla}^k h^j_{N\,i} = 0,$$

if the field $h_{N\,ij}$ satisfies Eqs. (63). In addition, the operator $\tilde{\nabla}_k\tilde{\nabla}^k$ is self-conjugate in the metric

$$\int d\tilde{V}\tilde{g}^{ik}\tilde{g}^{jl}h_{M\,kl}h_{N\,ij}, \tag{66}$$

where $d\tilde{V}$ is an element of volume on the sphere $S^{(3)}_{H^{-1}}$. Consequently, we choose as the set of functions $\{h_{N\,ij}\}$' the set of eigenfunctions, orthonormalized in the metric (66), of the operator $(-\tilde{\nabla}_k\tilde{\nabla}^k)$ with bounded eigenvalues:

$$-\tilde{\nabla}_k\tilde{\nabla}^k h_{N\,ij} = \tilde{\varepsilon}_N h_{N\,ij}, \quad \tilde{\varepsilon}_N < \tilde{\varepsilon}_0. \tag{67}$$

The fields $h_{N\,ij}$ in Eq. (67) satisfy Eqs. (63).

Let

$$\tilde{\varepsilon}_N \gg (H\cosh Ht)^2. \tag{68}$$

Then we obtain, using Eq. (65), the estimate

$$\nabla^{(0)}_0 h_{N\,ij} \sim \sqrt{\tilde{\varepsilon}_N}(\cosh Ht)^{-1}h_{N\,ij}. \tag{69}$$

Since, according to Eq. (59), $h^{ij}_N \sim a^{-4}h_{N\,ij}$, we obtain using Eqs. (62), (69), and (52)

$$\sqrt{\tilde{\varepsilon}_N}H^{-3}(\cosh Ht)^{-2}|h_{N\,ij}|^2 \sim 1,$$

or

$$l_p|h_{N\,ij}| \sim \tilde{\varepsilon}_N^{-1/4} l_p H^{3/2}\cosh Ht. \tag{70}$$

In the opposite case

$$\tilde{\varepsilon}_N \ll (H\cosh Ht)^2 \tag{71}$$

we obtain using Eq. (65)

$$\nabla^{(0)}_0 h_{Nij} \sim H|h_{N\,ij}|. \tag{72}$$

Hence, just as above, we find the estimate

$$l_p|h_{N\,ij}| \sim l_p H\sqrt{\cosh Ht}. \tag{73}$$

Comparing estimates (70) and (73) shows that at the time

$$H\cosh Ht_N \sim \tilde{\varepsilon}_N$$

the regime of temporal evolution of the corresponding mode changes. This change of regime occurs because for $t < t_N$ the wavelength of the mode $h_{N\,ij}$ is less than the so-called event horizon, and the opposite situation occurs for $t > t_N$.

Indeed, the distances on the sphere $S^{(3)}_{H^{-1}}$, denoted by $\tilde{l}$, and on the hypersphere $\Sigma^{(3)}_t$ (the section of the de Sitter space with metric (59) at a fixed time t), denoted as $l(t)$, according to Eq. (59) are related by the relation

$$l(t) = (\cosh Ht)\tilde{l}. \tag{74}$$

Consequently, the wavelength of the mode $h_{N\,ij}$ is of the order of

$$\lambda_N \sim (\cosh Ht)\tilde{\varepsilon}_N^{-1/2}. \tag{75}$$

Hence one can see that Eq. (69) corresponds to

$$\lambda_N \ll H^{-1}, \tag{76}$$

and Eq. (71) corresponds to

$$\lambda_N \gg H^{-1}. \tag{77}$$

By definition, the distance $l(t)$ between two points x_1 and x_2 on $\Sigma^{(3)}_t$ is less than the event horizon R_c if the light signal emitted at the point x_1 reaches the point x_2 some time in the future. According to Eq. (59) for propagation of light

$$dt = a(t)d\tilde{l},$$

and therefore

$$\tilde{l}_{12} = \int_{t_1}^{t_2} \frac{dt'}{a(t')}, \tag{78}$$

where t_2 is the time when the signal reaches the point x_2, and $\tilde{l}$ is the distance on $S^{(3)}_{H^{-1}}$ between x_1 and x_2. Comparing Eq. (78) and Eq. (74) and letting t_2 approach infinity, we find

$$R_c \propto a(t)\int_t^\infty \frac{dt'}{a(t')} \propto H^{-1}. \tag{79}$$

The meaning of the event horizon is that two points on $\Sigma^{(3)}$ separated by a distance greater than R_c cannot exchange any signals during the entire subsequent time.

Using Eqs. (39) and (59) it is easy to obtain the following estimate for fermionic modes:

$$|\psi_N| \propto H^{3/2}(\cosh Ht)^{-3/2}. \tag{80}$$

Before estimating the role of quantum fluctuations, we make the assumption

$$l_p H \lll 1, \quad \lambda_{\min} \sim \nu l_p, \tag{81}$$

where $\lambda_N > \lambda_{\min}$ is the minimum wavelength of the modes under study in an epoch close to the time of observation and ν is a dimensionless number.

We now use the well-known formula

$$\begin{aligned}[\langle (g_{ij} - g_{(cl)\,ij})^2(x)\rangle \langle (g_{kl} - g_{(cl)\,kl})^2(x')\rangle]^{1/2} \\ \geq \frac{1}{2}|\langle [g_{ij}(x), g_{kl}(x')]\rangle|\end{aligned} \tag{82}$$

in the lowest approximation. The averaging in Eq. (82) is performed with respect to a state close to the ground state. In this case the inequality in Eq. (82) is close to saturation and the left-hand side can be estimated by estimating the right-hand side of this inequality. Using Eq. (29), in the lowest approximation for quantum fluctuations we obtain

$$\begin{aligned}[g_{ij}(x), g_{kl}(x')] &= l_p^2 \\ \times \sum_{|N|<N_0} &[h_{N\,ij}(x)h^*_{N\,kl}(x') - h^*_{N\,ij}(x)h_{N\,kl}(x')].\end{aligned} \tag{83}$$

To estimate the right-hand side in Eq. (83), this sum must be divided into two terms.

The first term takes into account the infrared modes (the index i) with wavelengths

$$H^{-1} < \lambda_{N\,i} < H^{-1}\cosh Ht, \tag{84}$$

and the second term takes account of the ultraviolet modes (the index u) with wavelengths

$$\nu l_p < \lambda_{N\,u} < H^{-1}. \tag{85}$$

Let the number of infrared and ultraviolet modes be, respectively, of the order of N_i and N_u, and

$$N_i + N_u = N_0 = \text{const}.$$

Then the sum on the right-hand side of Eq. (83) is represented as a sum of two terms of order

$$\Sigma_i(t) \sim N_i l_p^2 |h_{N\,ij}|_i^2 \sim N_i (l_p H)^2 \cosh Ht, \tag{86}$$

$$\begin{aligned}\Sigma_u(t) &\sim N_u l_p^2 |h_{N\,ij}|_u^2 \\ &\sim (N_0 - N_i)\langle \tilde{\varepsilon}_N^{-1/2}\rangle_u l_P^2 H^3 (\cosh Ht)^2.\end{aligned} \tag{87}$$

Here we employed estimates (73) and (79). Since

$$g^{(0)}_{ij} g^{(0)}_{kl} \sim a^4,$$

quantities (86) and (87), referred to the fourth power of the scale factor $a(t)$, are physically meaningful. We also take account of the fact that the number of infrared modes increases with time, so that (see Eq. (84))

$$N_i \sim \left(\frac{H^{-1}a(t)}{\lambda_{N i\min}}\right)^3 \sim (\cosh Ht)^3. \tag{88}$$

Thus we find

$$\Sigma'_i(t) = \frac{\Sigma_i(t)}{a^4(t)} \sim (l_P H)^2, \tag{89}$$

$$\begin{aligned}\Sigma'_u(t) &= \frac{\Sigma_u(t)}{a^4(t)} \\ &\sim (N_0 - \cosh^3 Ht)\tilde{\varepsilon}_0^{-1/2} l_P^2 H^3 (\cosh Ht)^{-2}.\end{aligned} \tag{90}$$

In the latter formula the fact that

$$\langle \tilde{\varepsilon}_N^{-1/2}\rangle_u \sim \tilde{\varepsilon}_0^{-1/2},$$

where $\tilde{\varepsilon}_0$ is the maximum eigenvalue of the modes $h_{N\,ij}$ [see Eq. (67)], was taken into account.

The estimates obtained lead to the following qualitative conclusion: in a model of the inflating universe the value of the quantum fluctuations decreases exponentially with increasing time. Conversely, as the moment of creation of the universe is approached the role of quantum fluctuations becomes determining.

Indeed, according to modern notions, at the de Sitter stage

$$l_P H \sim 10^{-4}\text{–}10^{-12}. \tag{91}$$

Consequently, the infrared contribution to quantum fluctuations is negligibly small at all stages of inflation. The contribution of ultraviolet quantum fluctuations, according to Eq. (90), decays exponentially with increasing time, and vice versa.

ACKNOWLEDGMENTS

This work was supported by the Program for Support of Leading Scientific Schools, grant no. 00-1596579.

APPENDIX

Using Eqs. (61), Eq. (46) becomes (the index N, enumerating the modes, is dropped here)

$$-\nabla^{(0)}_\lambda\nabla^{(0)\lambda}h_{\mu\nu}+2H^2(g^{(0)}_{\mu\nu}h^\lambda_\lambda-h_{\mu\nu})\;=\;0. \tag{A.1}$$

Let ξ_μ be a vector field satisfying the equation

$$h'_{0\mu}\;=\;h_{0\mu}+\nabla^{(0)}_0\xi_\mu+\nabla^{(0)}_\mu\xi_0\;=\;0 \tag{A.2}$$

in all of space-time.

We obtain using Eqs. (48) and (A.2)

$$\nabla^{(0)}_0\left(\nabla^{(0)}_i\xi^i+\frac{1}{2}h^i_i\right)\;=\;-(\nabla^{(0)}_\lambda\nabla^{(0)\lambda}-3H^2)\xi_0. \tag{A.3}$$

We now substitute into Eq. (A.1) with $\mu=\nu=0$ the expression for h^0_0 from Eq. (A.2) and use

$$\begin{aligned}\nabla^{(0)}_\lambda\nabla^{(0)\lambda}\nabla^{(0)}_\mu\xi_\nu\;&=\;\nabla^{(0)}_\mu\nabla^{(0)}_\lambda\nabla^{(0)\lambda}\xi_\nu\\ &+2H^2g^{(0)}_{\mu\nu}\nabla^{(0)\lambda}\xi_\lambda-2H^2\nabla^{(0)}_\nu\xi_\mu-3H^2\nabla^{(0)}_\mu\xi_\nu,\end{aligned} \tag{A.4}$$

which is valid for any vector field ξ_μ. As a result, we obtain

$$\begin{aligned}&\nabla^{(0)}_0[(\nabla^{(0)}_\lambda\nabla^{(0)\lambda}-3H^2)\xi^0]\\ &\quad=\;-2H^2\left(\nabla^{(0)}_i\xi^i+\frac{1}{2}h^i_i\right).\end{aligned} \tag{A.5}$$

Finally, we substitute h_{0i} from Eq. (A.2) into Eq. (A.1) with $\mu=0$ and $\nu=i$ and use once again Eq. (A.4). We obtain

$$\begin{aligned}&\nabla^{(0)}_0[(\nabla^{(0)}_\lambda\nabla^{(0)\lambda}-3H^2)\xi_i]\\ &=\;-\nabla^{(0)}_i[(\nabla^{(0)}_\lambda\nabla^{(0)\lambda}-3H^2)\xi_0].\end{aligned} \tag{A.6}$$

Now, it is evident from Eqs. (A.3), (A.5), and (A.6) that if

$$(\nabla^{(0)}_\lambda\nabla^{(0)\lambda}-3H^2)\xi_\mu\;=\;0, \tag{A.7}$$

$$\nabla^{(0)}_i\xi^i+\frac{1}{2}h^i_i\;=\;0 \tag{A.8}$$

on some spacelike hypersurface $\Sigma^{(3)}_0$, then Eqs. (A.7) and (A.8) hold in all space-time. This can be attained by making an appropriate choice of the vector field ξ_μ. Indeed, using the shift

$$\xi_\mu\longrightarrow\xi_\mu+\varphi_\mu,$$

where the field φ_μ does not depend on x^0, Eq. (A.2) can be made to hold for all x^0 and Eq. (A.7) can be made to hold on the hypersurface $\Sigma^{(3)}_0$. Indeed, using Eqs. (61), (A.1), (A.2), (A.4), and (A.7) we obtain

$$(\nabla^{(0)}_j\nabla^{(0)j}-2H^2)\left(\nabla^{(0)}_i\xi^i+\frac{1}{2}h^i_i\right)\;=\;0,$$

which is valid on $\Sigma^{(3)}_0$. Hence follows the validity of Eq. (A.8) on the hypersurface $\Sigma^{(3)}_0$.

Let us consider the transformed field

$$h'_{\mu\nu}\;=\;h_{\mu\nu}+\nabla^{(0)}_\mu\xi_\nu+\nabla^{(0)}_\nu\xi_\mu. \tag{A.9}$$

If the field $h_{\mu\nu}$ satisfies Eqs. (A.1) and (48), then as a result of Eq. (A.7) the field $h'_{\mu\nu}$ also satisfies these equations. In addition, according to Eqs. (A.2) and (A.8)

$$h'_{0\mu}\;=\;0,\quad h'^i_i\;=\;0. \tag{A.10}$$

Thus, the compatibility of Eqs. (63) and (64) has been established.

REFERENCES

1. S. N. Vergeles, Zh. Éksp. Teor. Fiz. **102**, 1739 (1992) [Sov. Phys. JETP **75**, 938 (1992)].
2. S. N. Vergeles, Yad. Fiz. **57**, 2286 (1994) [Phys. At. Nucl. **57**, 2196 (1994)].
3. S. N. Vergeles, Zh. Éksp. Teor. Fiz. **110**, 1557 (1996) [JETP **83**, 859 (1996)].
4. S. N. Vergeles, Zh. Éksp. Teor. Fiz. **112**, 132 (1997).
5. S. N. Vergeles, Zh. Éksp. Teor. Fiz. **113**, 1566 (1998) [JETP **86**, 854 (1998)].
6. E. Benedict, R. Jackiw, and H.-J. Lee, Phys. Rev. D **54**, 6213 (1996); D. Cangemi, R. Jackiw, and B. Zwiebach, Ann. Phys. (N.Y.) **245**, 408 (1996); D. Cangemi and R. Jackiw, Phys. Lett. B **337**, 271 (1994); Phys. Rev. D **50**, 3913 (1994); D. Amati, S. Elitzur, and E. Rabinovici, Nucl. Phys. B **418**, 45 (1994); D. Louis-Martínez, J. Gegenberg, and G. Kunstatter, Phys. Lett. B **321**, 193 (1994); E. Benedict, Phys. Lett. B **340**, 43 (1994); T. Strobl, Phys. Rev. D **50**, 7346 (1994).
7. S. N. Vergeles, Zh. Éksp. Teor. Fiz. **117**, 5 (2000) [JETP **90**, 1 (2000)].
8. P. A. M. Dirac, *Lectures on Quantum Mechanics* (Yeshiva Univ., New York, 1964).
9. P. A. M. Dirac, *Lectures on Quantum Field Theory* (Yeshiva Univ., New York, 1967).
10. H. G. Zeh, Phys. Lett. A **116**, 9 (1986).
11. C. Kiefer, Class. Quantum Grav. **4**, 1369 (1987).
12. A. O. Barvinsky, A. Yu. Kamenshchik, C. Kiefer, and I. V. Mishakov, Nucl. Phys. B **551**, 374 (1999).
13. A. O. Barvinsky, A. Yu. Kamenshchik, and C. Kiefer, gr-qc/9901055.
14. C. Kiefer, D. Polarski, and A. A. Starobinsky, Int. J. Mod. Phys. D **7**, 455 (1998).
15. C. Kiefer, J. Lesgourgues, D. Polarski, and A. A. Starobinsky, Class. Quantum Grav. **15**, L67 (1998).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 873–893.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1012–1035.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Yang–Mills Theory in Three Dimensions as a Quantum Gravity Theory¶

D. I. Diakonov[a, b, *] and V. Yu. Petrov[b, **]

[a]*NORDITA, Copenhagen Ø, DK-2100 Denmark*
[b]*Petersburg Nuclear Physics Institute, Russian Academy of Sciences, Gatchina, St. Petersburg, 188350 Russia*
**e-mail: diakonov@nordita.dk*
***e-mail: victorp@thd.pnpi.spb.ru*
Received May 31, 2000

Abstract—We perform the dual transformation of the Yang–Mills theory in three dimensions using the Wilson action on the cubic lattice. The dual lattice is made of tetrahedra triangulating a 3-dimensional curved manifold but which is embedded into a flat 6-dimensional space [for the SU(2) gauge group]. In the continuum limit, the theory can be reformulated in terms of 6-component gauge-invariant scalar fields having the meaning of the external coordinates of the dual lattice sites. These 6-component fields induce a metric and a curvature of the 3-dimensional dual-color space. The Yang–Mills theory can also be rewritten as a quantum gravity theory with the Einstein–Hilbert action but with a purely imaginary Newton constant plus a homogeneous "ether" term. The theory can be formulated in a gauge-invariant and local form without explicit color degrees of freedom. © *2000 MAIK "Nauka/Interperiodica".*

1. LATTICE PARTITION FUNCTION

Although our objective is the continuum theory, we start by formulating the SU(N_c) gauge theory on a cubic lattice. The partition function can be written as an integral over all link variables, which are SU(N_c) unitary matrices U, with the action given by the sum over all plaquettes:

$$\mathcal{Z}(\beta) = \int\prod_{\text{links}} dU_{\text{link}} \exp\left(\sum_{\text{plaquettes}} \frac{\beta(\text{Tr}U_{\text{plaq}} + \text{c.c.})}{2\text{Tr}1} \right), \quad (1)$$

where β is the dimensionless inverse coupling. The unitary matrix U_{plaq} is a product of four-link unitary matrices closing a plaquette.

To go to the continuum limit, one writes

$$U_{\text{link}} = \exp(iaA_\mu^a t^a),$$

where a is the lattice spacing and

$$A_\mu^a t^a = A_\mu$$

is the Yang–Mills gauge potential, with t^a being the gauge group generators normalized to

$$\text{Tr}t^a t^b = \delta^{ab}/2;$$

one then expands $\text{Tr}\,U_{\text{plaq}}$ in the lattice spacing a. For a plaquette lying in the (12) plane, the result is

$$\beta\frac{\text{Tr}U_{\text{plaq}} + \text{c.c.}}{2\text{Tr}1} = \beta\left(1 - a^4\frac{\text{Tr}F_{12}^2}{2\text{Tr}1} + O(a^6)\right), \quad (2)$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu A_\nu]$$

is the Yang–Mills field strength. Summing over all plaquettes, one obtains the partition function of the continuum theory,

$$\mathcal{Z}_{\text{cont}} = \int DA_\mu \exp\left(-\frac{1}{2g_d^2}\int d^d x \text{Tr}F_{\mu\nu}^2\right), \quad (3)$$

with the obvious relation

$$\beta = \frac{2N_c}{a^{4-d}g_d^2}, \quad (4)$$

between the dimensionless lattice coupling β and the SU(N_c) gauge coupling constant in d dimensions g_d^2.

In this paper, we concentrate on the Euclidean SU(2) Yang–Mills theory in three dimensions. In this case, Eq. (4) becomes

$$\beta = \frac{4}{ag_3^2}. \quad (5)$$

The continuum limit of the 3-dimensional Yang–Mills theory given by partition function (1) is obtained if one takes the lattice spacing $a \longrightarrow 0$ and $\beta \longrightarrow \infty$ with their product

$$g_3^2 = \frac{4}{a\beta}$$

kept fixed. This quantity provides the theory with a mass scale. It is widely believed (although not proven

¶This article was submitted by the authors in English.

so far) that the theory possesses two fundamental properties: (1) the average of a large Wilson loop has an area behavior with the string tension proportional to g_3^4 and (2) correlation functions of local operators similar to $F_{\mu\nu}^2$ decay exponentially at large separations, with a "mass gap" proportional to g_3^2.

Our aim is to rewrite partition function (1) in dual variables and to study its continuum limit.

2. THE DUALITY TRANSFORMATION

The general idea is to integrate over the link variables $U_{\rm link}$ in Eq. (1) and to make a Fourier transformation in the plaquette variables $U_{\rm plaq}$. This is done in several steps, i.e., one in each subsection.

2.1. Inserting a Unity into the Partition Function

First of all, we need to explicitly introduce the integration over unitary matrices assigned to the plaquettes, $U_{\rm plaq}$. This is done by inserting a unity for each plaquette into the partition function (1),

$$1 = \prod_{\rm plaquettes} \int dU_{\rm plaq} \delta(U_{\rm plaq}, U_1 U_2 U_3 U_4), \qquad (6)$$

where $U_{1\ldots4}$ are the link variables closing into a given plaquette. The δ-function is understood with the group-invariant Haar measure. The realization of such a δ-function is given by the Wigner D-functions:

$$\delta(U, V) = \sum_{J = 0, \frac{1}{2}, 1, \frac{3}{2}, \ldots} (2J+1) D^J_{m_1 m_2}(U^\dagger) D^J_{m_2 m_1}(V). \qquad (7)$$

This equation is known as the completeness condition for the D-functions [1]. The main properties of the D-functions used in this paper are listed in Appendix A.

Equation (7) should be understood as follows: if one multiplies the right-hand side of Eq. (7) with any function of the unitary matrix U and integrates over the Haar measure dU, the same function of the argument V is obtained:

$$\int dU\, f(U) \delta(U, V) = f(V). \qquad (8)$$

Using the multiplication law for the D-functions (see Appendix A, Eq. (A.3)), one can represent the unity to be inserted for each plaquette in partition function (1) as

$$1 = \int dU_{\rm plaq} \sum_J (2J+1) D^J_{m_1 m_2}(U^\dagger_{\rm plaq}) D^J_{m_2 m_3}(U_1) \times D^J_{m_3 m_4}(U_2) D^J_{m_4 m_5}(U_3) D^J_{m_5 m_1}(U_4), \qquad (9)$$

where $U_{1\ldots4}$ are the corresponding link variables forming the chosen plaquette.

2.2. Integrating over Plaquette Variables

Integrating over the plaquette unitary matrices $U_{\rm plaq}$ now becomes very simple. For each plaquette of the lattice, one has the factorized integrals of the type

$$\int dU_{\rm plaq} \exp\left(\beta \frac{{\rm Tr} U_{\rm plaq} + {\rm Tr} U^\dagger_{\rm plaq}}{2{\rm Tr}1}\right) D^J_{m_1 m_2}(U^\dagger_{\rm plaq}) = \delta_{m_1 m_2} \frac{2}{\beta} I_1(\beta) T_J(\beta), \qquad (10)$$

where $T_J(\beta)$ is the ratio of the modified Bessel functions [2],

$$T_J(\beta) = \frac{I_{2J+1}(\beta)}{I_1(\beta)} \longrightarrow \exp\left[-\frac{2J(J+1)}{\beta}\right] \quad \text{as } \beta \longrightarrow \infty. \qquad (11)$$

The quantity $T_J(\beta)$ is the "Fourier transform" of the Wilson action; because the dynamical variables have the meaning of Euler angles and are therefore compact in the lattice formulation, the Fourier transform depends on discrete values $J = 0, 1/2, 1, 3/2, \ldots$. However, as one approaches the continuum limit ($\beta \longrightarrow \infty$), the essential values of the plaquette angular momenta increase as $J \sim \sqrt{\beta}$ and their discreteness becomes less relevant. Strictly speaking, the continuum limit is achieved at plaquette angular momenta $J \gg 1$.

Here, we make a side remark on this occasion. For a given β, the quantity $T_J(\beta)$ gives the probability that the plaquette momentum J is excited. For the typical value ($\beta = 2.6$) used in lattice simulations (in 4 dimensions), we find that the probabilities of having plaquette excitations with $J = 0, 1/2, 1, 3/2$, and 2 are 56, 29, 11, 3, and 1%, respectively. This means that lattice simulations are actually dealing mainly with $J = 0, 1/2$, and 1 with a tiny admixture of higher excitations. It is important to understand why and how continuum physics is reproduced by lattice simulations with such small values of the plaquette momenta J involved.

Thus, we, obtain the partition function

$$\mathcal{Z} = \left[\frac{2}{\beta} I_1(\beta)\right]^{\text{number of plaquettes}} \sum_{J_P} \prod_{\text{plaquettes}} (2J_P+1) T_{J_P}(\beta) \times \prod_{\text{links } l} \int dU_l D^{J_P}_{m_1 m_2}(U_1) D^{J_P}_{m_2 m_3}(U_2) D^{J_P}_{m_3 m_4}(U_3) D^{J_P}_{m_4 m_1}(U_4), \qquad (12)$$

where $U_{1\text{–}4}$ are link variables forming a plaquette with the angular momentum J_P.

2.3. Integrating over Link Variables

It is difficult to integrate over link variables in Eq. (12) because each link enters several plaquettes. In two dimensions, every link is shared by two plaquettes; hence, one has to calculate integrals of the type

$$\int dU D^{J_1}_{kl}(U) D^{J_2}_{mn}(U^\dagger) = \frac{1}{2J_1+1}\delta_{J_1J_2}\delta_{kn}\delta_{lm} \qquad (13)$$

for all links on the lattice. We consider this case in Section 4.

In three dimensions, every link is shared by four plaquettes; hence, the integral over link variables has the form

$$\int dU D^{J_A}_{m_1m_2}(U) D^{J_B}_{m_3m_4}(U) D^{J_C}_{m_5m_6}(U) D^{J_D}_{m_7m_8}(U), \qquad (14)$$

where $J_{A,B,C,D}$ are angular momenta associated with four plaquettes intersecting at a given link U and $m_{1–8}$ are "magnetic" quantum numbers that must be contracted inside closed plaquettes. In four dimensions, six plaquettes intersect at a given link; however, we do not consider this case here.

The general strategy to calculate link integrals (14) is (i) to separate four D-functions into two pairs according to a certain rule and to decompose the pairs of D-functions in terms of single D-functions using Eq. (A.10), (ii) to integrate the resulting two D-functions using Eq. (13), and (iii) to contract the "magnetic" indices. Since all the "magnetic" indices are eventually contracted, we arrive at the partition function written in terms of the invariant 3nj symbols.

There are several different tactics for dividing four D-functions into two pairs, eventually leading to anything from 6j to 18j symbols. In this paper, we take the route used in [3, 4], leading to a product of many 6j symbols, although one loses certain symmetries on this route, causing later difficulties. The gain, however, is that it is easier to work with 6j symbols than with 12j or 18j symbols. Since important sign factors were omitted in [3, 4] and only the final result was reported, we feel it is necessary to give a detailed derivation in what follows.

In three dimensions, all plaquettes are shared by two adjacent cubes; therefore, it is natural to divide all cubes of the lattice into two classes, which we shall call "even" and "odd," and to attribute plaquettes to even cubes only. We call a cube even if its front lower left corner is a lattice site with even coordinates:

$$(-1)^{x+y+z} = +1.$$

It is called odd otherwise. The even and odd cubes form a 3-dimensional checker board, as illustrated in Fig. 1, where only even cubes are drawn explicitly. The even cubes touch each other through a common edge or link, as do the odd ones among themselves. The even and odd cubes have common faces or plaquettes. All plaquettes are attributed to even cubes only: that is the reason for the division of cubes into two classes. We now consider an even cube (Fig. 2). In the figure A, B, C, D, E, and F denote the six faces of the cube; the numbers from 1 to 12 denote its links or edges; and a, b, c, d, e, f, g, and h denote its 8 vertices or sites. Correspondingly, we denote the plaquette angular momenta by $J_{A–F}$, the link variables by $U_{1–12}$, and the "magnetic" numbers of the D-functions carry indices (a–h) referring to the sites connected by the D-functions.

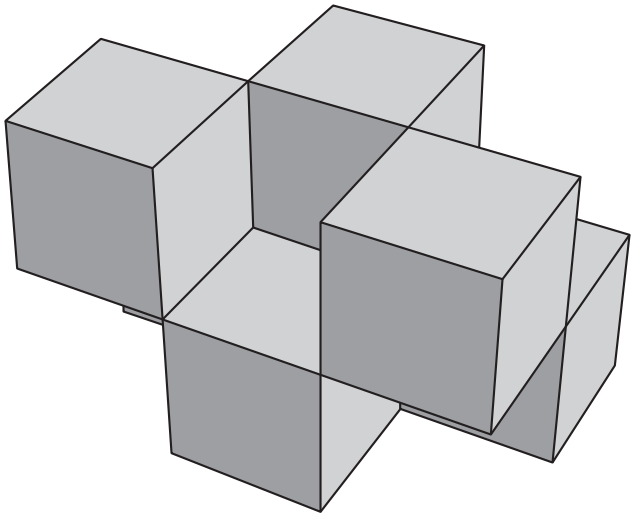

Fig. 1. "Even" cubes in checker board order.

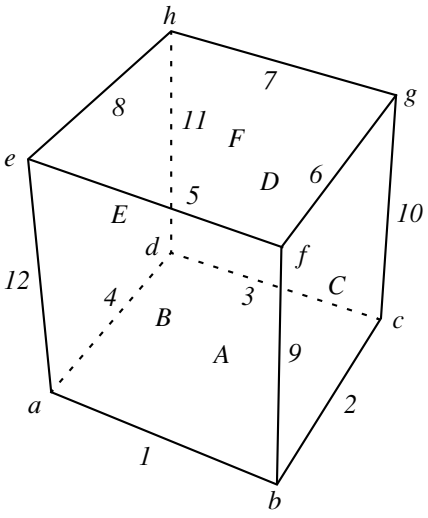

Fig. 2. An elementary even cube.

One can write the traces of products of four D-functions over plaquettes in various ways. To be systematic, we adhere to the following rule. Link variables in the plaquette are taken in the counterclockwise order, as viewed from the center of the even cube to which the given plaquette belongs. If the link goes in the positive

direction of the x-, y-, z-axes, we assign the U variable to it; otherwise, we assign the $U^\dagger$ variable to it.

With these rules, the six plaquettes of the elementary cube shown in Fig. 2 bring in the following six traces of the D-function products:

$$\begin{aligned}\text{Cube} \;=\; & [D^{J_A}_{i_a i_b}(U_1) D^{J_A}_{i_b i_c}(U_2) D^{J_A}_{i_c i_d}(U_3^\dagger) D^{J_A}_{i_d i_a}(U_4^\dagger)] \\ & \times [D^{J_B}_{j_b j_a}(U_1^\dagger) D^{J_B}_{j_a j_e}(U_{12}) D^{J_B}_{j_e j_f}(U_5) D^{J_B}_{j_f j_b}(U_9^\dagger)] \\ & \times [D^{J_C}_{k_c k_b}(U_2^\dagger) D^{J_C}_{k_b k_f}(U_9) D^{J_C}_{k_f k_g}(U_6) D^{J_C}_{k_g k_c}(U_{10}^\dagger)] \\ & \times [D^{J_D}_{l_d l_c}(U_3) D^{J_D}_{l_c l_g}(U_{10}) D^{J_D}_{l_g l_h}(U_7^\dagger) D^{J_D}_{l_h l_d}(U_{11}^\dagger)] \\ & \times [D^{J_E}_{m_e m_a}(U_{12}^\dagger) D^{J_E}_{m_a m_d}(U_4) D^{J_E}_{m_d m_h}(U_{11}) D^{J_E}_{m_h m_e}(U_8^\dagger)] \\ & \times [D^{J_F}_{n_f n_e}(U_5^\dagger) D^{J_F}_{n_e n_h}(U_8) D^{J_F}_{n_h n_g}(U_7) D^{J_F}_{n_g n_f}(U_6^\dagger)].\end{aligned} \tag{15}$$

Each link variable $U_{1\text{–}12}$ appears in this product twice: once as U and once as $U^\dagger$. For $D(U^\dagger)$, we use Eq. (A.5) to write it in terms of $D(U)$. We can then apply decomposition rule (A.10) to express pairs of D-functions in terms of one D-function and two 3jm symbols. The new D-functions correspond to the links and carry the angular momenta denoted by j. The 3jm symbols have "magnetic" indices that are contracted when all the indices related to a given corner of the cube are assembled together. Although straightforward, this exercise is rather lengthy, and we relegate it to Appendix B. As a result, we rewrite (15) as

$$\begin{aligned}\text{Cube} \;=\; & \sum_{j_1 \ldots j_{12}} (2j_1+1)\ldots(2j_{12}+1) \\ & \times D^{j_1}_{o_a o_b}(U_1) D^{j_2}_{-p_c,-p_b}(U_2^\dagger) D^{j_3}_{-q_c,-q_d}(U_3^\dagger) \\ & \times D^{j_4}_{r_a r_d}(U_4) D^{j_5}_{-s_f,-s_e}(U_5^\dagger) D^{j_6}_{t_f t_g}(U_6) \\ & \times D^{j_7}_{u_h u_g}(U_7) D^{j_s}_{-v_h,-v_e}(U_8^\dagger) D^{j_9}_{-w_f,-w_b}(U_9^\dagger) \\ & \times D^{j_{10}}_{x_c x_g}(U_{10}) D^{j_{11}}_{-y_h,-y_d}(U_{11}^\dagger) D^{j_{12}}_{z_a z_e}(U_{12}) \\ & \times \begin{pmatrix} j_1 & j_4 & j_{12} \\ o_a & r_a & z_a \end{pmatrix} \begin{pmatrix} j_1 & j_9 & j_2 \\ -o_b & w_b & p_b \end{pmatrix} \\ & \times \begin{pmatrix} j_2 & j_3 & j_{10} \\ p_c & q_c & -x_c \end{pmatrix} \begin{pmatrix} j_4 & j_3 & j_{11} \\ -r_d & q_d & y_d \end{pmatrix} \\ & \times \begin{pmatrix} j_{12} & j_8 & j_5 \\ -z_e & v_e & s_e \end{pmatrix} \begin{pmatrix} j_6 & j_5 & j_9 \\ -t_f & s_f & w_f \end{pmatrix} \\ & \times \begin{pmatrix} j_6 & j_{10} & j_7 \\ t_g & x_g & u_g \end{pmatrix} \begin{pmatrix} j_7 & j_{11} & j_8 \\ -u_h & y_h & v_h \end{pmatrix} \\ & \times \begin{Bmatrix} j_7 & j_{11} & j_8 \\ J_E & J_F & J_D \end{Bmatrix} \begin{Bmatrix} j_6 & j_{10} & j_7 \\ J_D & J_F & J_C \end{Bmatrix} \\ & \times \begin{Bmatrix} j_6 & j_5 & j_9 \\ J_B & J_C & J_F \end{Bmatrix} \begin{Bmatrix} j_{12} & j_8 & j_5 \\ J_F & J_B & J_E \end{Bmatrix} \\ & \times \begin{Bmatrix} j_4 & j_3 & j_{11} \\ J_D & J_E & J_A \end{Bmatrix} \begin{Bmatrix} j_2 & j_3 & j_{10} \\ J_D & J_C & J_A \end{Bmatrix} \\ & \times \begin{Bmatrix} j_1 & j_9 & j_2 \\ J_C & J_A & J_B \end{Bmatrix} \begin{Bmatrix} j_1 & j_4 & j_{12} \\ J_E & J_B & J_A \end{Bmatrix},\end{aligned} \tag{16}$$

where $j_{1\text{–}12}$ are the angular momenta attached to the links of the cube, $(\ldots)$ are the 3jm symbols, and $\{\ldots\}$ are the 6j symbols. We see that a 6j symbol is attached to each corner of the even cube; its arguments are three plaquette momenta J and three link momenta j intersecting at a given corner. The 3jm symbols involve only the link variables j.

We have, thus, rewritten all twelve pairs of D^J-functions entering a cube as a product of single D^j-functions, where j's are the new momenta associated with the links.

This procedure must be applied to all even cubes of the lattice. As the result, one has only two D^j-functions of the same link variable U for all links of the lattice, which makes it straightforward to integrate over the link variables using Eq. (13).

It is convenient to simultaneously integrate over six links entering one lattice site, because this leads to the full contraction over all the "magnetic" numbers. The derivation is again straightforward but lengthy: the details are given in Appendix C. The result is that the 3jm factors in Eq. (16) are contracted with analogous 3jm symbols arising from neighboring even cubes, which gives 6j symbols attached to every lattice site and that are composed of the six link momenta j intersecting at a given lattice site. In the notation of Fig. 3, the result for the a and b vertices has the form

$$\begin{aligned}\text{"}a\text{"} \;&=\; \begin{Bmatrix} j_1 & j_4 & j_{12} \\ j_{15} & j_{14} & j_{13} \end{Bmatrix}, \\ \text{"}b\text{"} \;&=\; \begin{Bmatrix} j_1 & j_9 & j_2 \\ j_{17} & j_{18} & j_{16} \end{Bmatrix},\end{aligned} \tag{17}$$

and the expressions for the other vertices are similar. A sign factor $(-1)^{2j}$ must be attributed to every link of the lattice. As shown in Appendix C, it is actually equivalent to the sign factor $(-1)^{2J}$ attributed to every lattice plaquette.

3. LATTICE PARTITION FUNCTION AS A PRODUCT OF 6*j* SYMBOLS

We now summarize the recipe derived in the previous section. One first divides all 3-cubes into two classes: even and odd. They form a 3-dimensional checker board shown in Fig. 1. All even cubes are characterized by their plaquette momenta J. The edges of even cubes have link momenta j; each link is shared by two even cubes.

To each of the eight corners of an even cube, one assigns a 6j symbol of the type

$$\left\{ \begin{array}{ccc} j_1 & j_2 & j_3 \\ J_A & J_B & J_C \end{array} \right\}, \tag{18}$$

where J's are the plaquette and j's are the link momenta intersecting at a given corner of the cube. The rule is that link 1 is perpendicular to plaquette A, link 2 is perpendicular to plaquette B, and link 3 is perpendicular to plaquette C. Four triades—$(j_1 J_B J_C)$, $(j_2 J_A J_C)$, $(j_3 J_A J_B)$, and $(j_1 j_2 j_3)$—satisfy triangle inequalities.

To each lattice site, one assigns a 6j symbol of the type

$$\left\{ \begin{array}{ccc} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{array} \right\}, \tag{19}$$

where j's are the six link momenta entering the chosen lattice site. The rule is that link 4 is a continuation of link 1 lying in the same direction, link 5 is a continuation of link 2, and link 6 is a continuation of link 3. Four triades—$(j_1 j_2 j_3)$, $(j_1 j_5 j_6)$, $(j_2 j_4 j_6)$, and $(j_3 j_4 j_5)$—satisfy triangle inequalities.

Actually, each lattice site has five 6j symbols associated with it: four originate from the corners of the even cubes adjacent to the site and are of type (18) and one is of type (19).

The lattice partition function (1) or (12) can be identically rewritten as a product of the 6j symbols described above. Independent summations over all possible plaquette momenta J and all possible link momenta j are understood. We write the partition function in a symbolic form as

$$\mathcal{Z} = \left[\frac{2}{\beta} I_1(\beta)\right]^{\text{number of plaquettes}} \times \sum_{J_P,\, j_l} \prod_{\text{plaquettes}} (2J_P + 1) T_{J_P}(\beta) (-1)^{2J_P} \prod_{\text{links}} (2j_l + 1) \tag{20}$$

$$\times \prod_{\text{even cubes corners}} \left\{ \begin{array}{ccc} j & j & j \\ J & J & J \end{array} \right\} \prod_{\text{lattice sites}} \left\{ \begin{array}{ccc} j & j & j \\ j & j & j \end{array} \right\}.$$

The plaquette weights $T_J(\beta)$ are given by Eq. (11). Apart from the sign factor, the essentially identical expression was given in [3, 4].[1] The sign factor is equal to ± 1 if the total number of half-integer plaquettes J's is even (odd). Since plaquettes with half-integer momenta form closed surfaces, it may seem that the sign factor can be omitted. This is not so, however, in the general case involving vacuum averages of operators; therefore, it is preferable to keep the sign factor. It is also important for getting a smooth continuum limit, see Section 10.

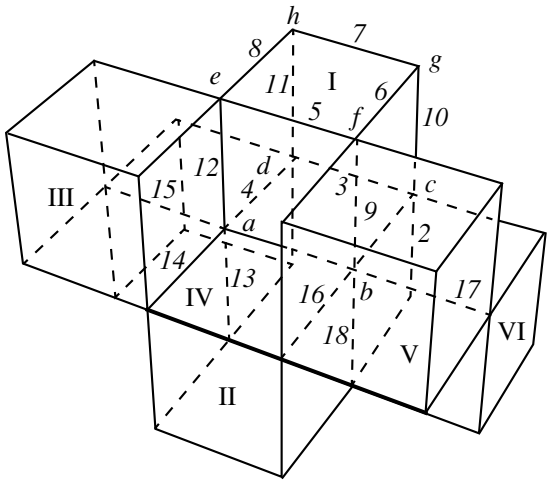

Fig. 3. Several cubes combine to produce 6j symbols composed of link momenta j. These are the same cubes as in Fig. 1.

4. A SIMPLE EXAMPLE: THE $d = 2$ YANG–MILLS THEORY

In a simple case of the exactly soluble 2-dimensional SU(2) theory, every link is shared by only two plaquettes. Therefore, the link integration is of the type given by Eq. (13): it requires that all plaquettes on the lattice have identical momenta J. The partition function thus becomes a single sum over the common J,

$$\mathcal{Z} = \left[\frac{2}{\beta} I_1(\beta)\right]^{\text{number of plaquettes}} \times \sum_J [T_J(\beta)]^{\text{number of plaquettes}} \tag{21}$$

(the number of plaquettes being equal to V/a^2), where V is the full lattice volume (full area in this case) and a is the lattice spacing.

A slightly less trivial exercise is to compute the average of the Wilson loop. Let the Wilson loop be in the representation j_s. This means that $D^{j_s}(U)$ must be inserted for each link along the loop. The result is given by integrals of two D-functions outside and inside the loop; integrals of three D-functions, for the links along the loop. The first integral says that all the plaquettes outside the loop are equal to a common J. The second

[1] We are grateful to P. Pobylitsa, who has independently derived Eq. (20).

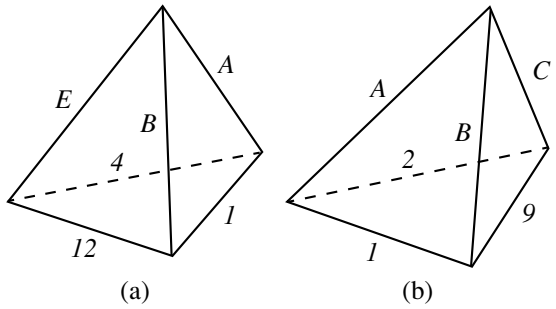

Fig. 4. Tetrahedra corresponding to the 6*j* symbols sitting at vertices (a) and (b).

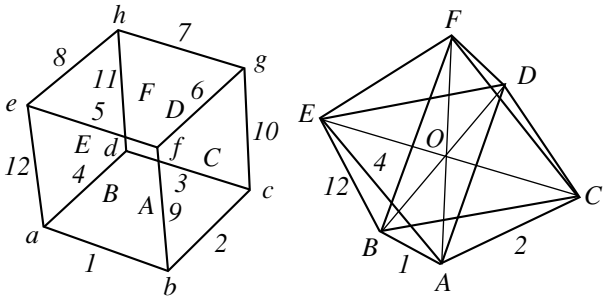

Fig. 5. Octahedron dual to the even cube.

integral says that all the plaquettes inside the loop are equal to a common J'. The integrals along the loop require that J, J', and j_s satisfy the triangle inequality. Thus, we have the average of the Wilson loop of area S given by

$$\langle W_{j_s}(S)\rangle = \frac{\sum_J [T_J(B)]^{V/a^2} \sum_{J'=|J-j_s|}^{J+j_s} \left[\frac{T_{J'}(\beta)}{T_J(\beta)}\right]^{S/a^2}}{\sum_J [T_J(\beta)]^{V/a^2}}. \quad (22)$$

This is an exact expression for the lattice Wilson loop; however, we wish to explore its continuum limit. In taking the continuum limit, we have $V/a^2 \longrightarrow \infty$ and $S/a^2 \longrightarrow \infty$ but $S \ll V$; we also have $\beta \longrightarrow \infty$ and $a \longrightarrow 0$, but $\beta a^2 = 4/g_2^2$ is fixed, where g_2^2 is the physical coupling constant with the dimension of mass squared, see Eq. (4).

In taking the $V/a^2 \longrightarrow \infty$ limit first, we see that only the $J = 0$ term contributes to the sum, with $T_0(\beta) \equiv 1$; consequently, all momenta inside the loop are equal to the source momentum: $J' = j_s$. Taking the large-β asymptotic form of $T_J(\beta)$ into account, Eq. (11), we obtain

$$\langle W_{j_s}(S)\rangle = [T_{j_s}(\beta)]^{S/a^2} = \exp\left[-\frac{g_2^2}{2} j_s(j_s+1)S\right], \quad (23)$$

which is, of course, the well-known area behavior of the Wilson loop with the string tension proportional to the Casimir eigenvalue.

5. THE DUAL LATTICE: TETRAHEDRA AND OCTAHEDRA

We now turn to the construction of the dual lattice. Each 6*j* symbol of the exact partition function (20) encodes four triangle inequalities between the plaquette J's and the link j's. It is therefore natural to represent each 6*j* symbol by a tetrahedron such that the lengths of its six edges are equal to the six momenta of the 6*j* symbol. Four faces of the tetrahedron form four triangles, and the triangle inequalities for the momenta are therefore automatically satisfied.

We first consider the eight 6*j* symbols corresponding to the eight corners of an even cube. These eight 6*j* symbols are given explicitly in Eq. (16) with the notation shown in Fig. 2. We represent all of them by tetrahedra having edges of the appropriate lengths. For example, the tetrahedra corresponding to the corners *a* and *b* are shown in Fig. 4. We denote the plaquette momenta J_A, … by their Latin labels *A*, *B*, etc., and the link momenta j_1, j_2, … by their numerical indices 1, 2, etc. We notice immediately that the two tetrahedra have a pair of equal faces; in this case, it is the triangle (*A*, *B*, 1). Therefore, we can glue the two tetrahedra together such that, this triangle becomes their common face. The gluing can be done in two ways. To be systematic, we always glue tetrahedra such that their volumes do not overlap.

Other tetrahedra are glued together in the same way. Being glued together, the eight tetrahedra of the cube form an octahedron shown in Fig. 5. Its center point *O* is connected with six lines to the vertices denoted as *A*–*F*; the lengths of these lines are equal to the corresponding plaquette momenta J_{A-F}. The lengths of the external twelve edges of the octahedron are equal to the link momenta j_{1-12}. The eight faces of the octahedron correspond to the eight vertices of the original even cube. One can say that the octahedron is dual to the cube: the faces become vertices and vice versa; the edges remain edges.

It is clear that for generic J's and j's, the octahedron cannot be placed into a flat 3-dimensional space. Indeed, we have 6 + 12 = 18 given momenta (i.e., fixed lengths) but only 7 points defining the octahedron, including the center one. In three dimensions, this gives 21 degrees of freedom, from which we must subtract 3 + 3 to allow for rigid translations and rotations. Therefore, we are left with only 15 degrees of freedom instead of the required 18. (In four dimensions, the arithmetic would match: $7 \times 4 - 4 - 6 = 18$.)

Each even cube of the original lattice has twelve neighboring even cubes sharing edges with the first one and with themselves. If we represent the neighboring even cubes by their own dual octahedra, the octahedra also share common edges. This network of octahedra

does not cover the space because there are holes between them. However, we have not yet used the 6*j* symbols (19) made solely from the link momenta *j*'s. If we represent these 6*j* symbols by tetrahedra, their triangle faces will coincide with the faces of the octahedra corresponding to the even cubes adjacent to the site. For example, if we consider the 6*j* symbols corresponding to site *a* (see Fig. 3 and Eq. (17)),

$$\left\{ \begin{matrix} j_1 & j_4 & j_{12} \\ j_{15} & j_{14} & j_{13} \end{matrix} \right\},$$

it has a common triangle face (j_1, j_4, j_{12}) with the octahedron shown in Fig. 5. The other faces of this tetrahedron match the octahedra dual to cubes II, III, and IV (see Fig. 3).

The octahedra corresponding to the cubes supplemented by the tetrahedra corresponding to the lattice sites cover the space without holes and, therefore, serve as a simplical triangulation (see Fig. 6).

An equivalent view on the dual lattice was proposed in [3]. One can connect the centers of neighboring cubes (both even and odd) and assign the plaquette momenta *J* to these lines. The link momenta *j* are then assigned to the diagonal lines connecting only neighboring even sites of that dual lattice (see Fig. 7).

The dual lattice can be understood in two senses. On the one hand, one can build a regular cubic dual lattice with additional face diagonals, as shown in Figs. 6 and 7, and assign *J*'s and *j*'s to its edges. On the other hand, since the variables living on the links of the dual lattice are positive numbers, one can build a lattice with the lengths of the edges equal to the appropriate angular momenta. We always use the dual lattice in this second sense.

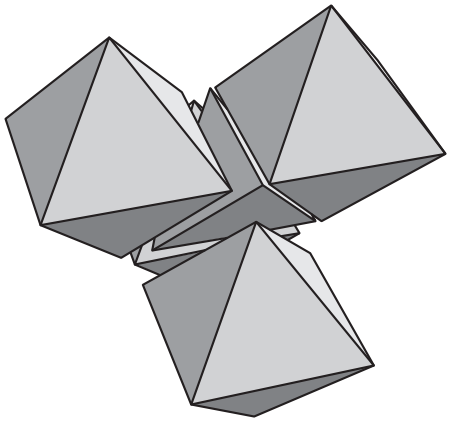

Fig. 6. A tetrahedron corresponding to the lattice site fits precisely into a hole between four octahedra corresponding to the four corners of the even cubes adjacent to the site (shown in motion).

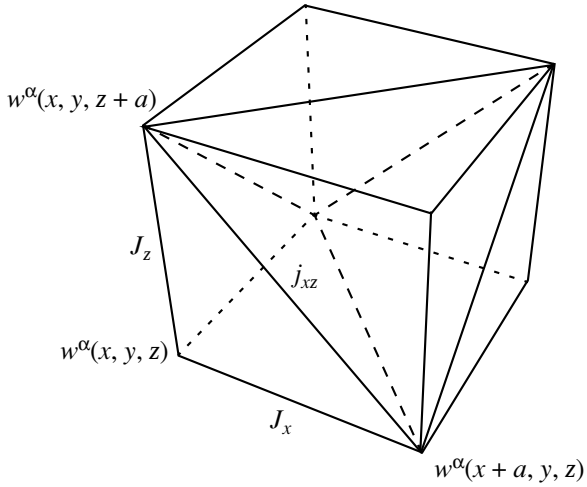

Fig. 7. Another view on the dual lattice.

6. THE DUAL LATTICE COORDINATES AS NEW VARIABLES

In the previous section, we encountered a situation where an octahedron that is dual to a cube does not fit into a 3-dimensional flat space: at least four dimensions were necessary. As one enlarges the triangulation complex, more dimensions are needed to match the number of the degrees of freedom. In the limiting case of an infinite lattice, one needs 6 flat dimensions. This number of dimensions follows from the number of the degrees of freedom involved: at each lattice site, there are three plaquette momenta *J* and three link momenta *j*, and there is a one-to-one correspondence between the lattice sites and cubes.

Therefore, the dual lattice (understood in the second sense, see above) spans a 3-dimensional manifold that can be embedded into a 6-dimensional flat space. We note that this is the maximum number of flat dimensions needed to embed a general 3-dimensional Riemannian manifold; it can be counted from the number of components of the metric tensor, which is 6 in three dimensions. Only very special configurations of *J*'s and *j*'s can be embedded into a lower dimensional flat space.

We are primarily interested in the continuum limit of the lattice theory, that is, in the case where *a* is small and β is large. This implies that large angular momenta $J \sim \sqrt{\beta}$ are involved, and we can pass from the summation over *J*'s and *j*'s to the integration over these variables in the partition function (20). We replace

$$\sum_{J = 0, 1/2, 1, \ldots} (2J + 1)\ldots \longrightarrow 2\int_0^\infty dJ^2 \ldots \qquad (24)$$

and for the summation over link momenta *j* as well.

The next step is to assign a 6-dimensional Lorentz scalar field $w^\alpha(x)$, $\alpha = 1, \ldots, 6$ to the centers of all cubes of the original lattice (see Fig. 7). We call them the coordinates of the dual lattice. They are scalars because

the cubes are scalars in three dimensions. The argument of the six-component scalar field is the coordinate of the center of the cube in question; however, we consider $w^\alpha(x)$ as continuous functions. Since six functions depend on only three coordinates, there are three relations between $w^\alpha(x)$ at any point; these relations define a curved 3-dimensional manifold whose triangulation is given by the set of J's and j's.

We next define 6-dimensional angular momenta as the differences of $w^\alpha(x)$ taken at the centers of neighbor cubes,

$$J_x^\alpha\left(x+\frac{a}{2},y,z\right) = w^\alpha(x+a,y,z)-w^\alpha(x,y,z)$$
$$= a\partial_x w^a + \frac{a^2}{2}\partial_x^2 w^\alpha + \ldots, \tag{25}$$
$$j_{xz}^\alpha\left(x+\frac{a}{2},y,z+\frac{a}{2}\right) = w^\alpha(x+a,y,z)-w^\alpha(x,y,z+a)$$
$$= a(\partial_x-\partial_z)w^\alpha + O(a^2),$$

and so on. By construction, the lengths of these 6-vectors are the lengths of the edges of the dual lattice.

The six functions $w^\alpha(x)$ can be called external coordinates of the manifold; they induce the metric tensor of the manifold given by

$$g_{ij}(x) = \partial_i w^\alpha \partial_j w^\alpha. \tag{26}$$

As is standard in differential geometry, one can define the Christoffel symbol

$$\Gamma_{i,jk}(x) = \frac{1}{2}(\partial_j g_{ik}+\partial_k g_{ij}-\partial_i g_{jk})$$
$$= \partial_i w^\alpha \partial_j \partial_k w^\alpha \equiv (w_i\cdot w_{jk}) \tag{27}$$

and the Riemann tensor

$$R_{ijkl}(x) = \frac{1}{2}(\partial_j\partial_k g_{il}+\partial_i\partial_l g_{jk}-\partial_j\partial_l g_{ik}-\partial_i\partial_k g_{jl})$$
$$+\Gamma_{m,jk}\Gamma^m_{il}-\Gamma_{m,jl}\Gamma^m_{ik} \tag{28}$$
$$= [(w_{ik}\cdot w_{jl})-g^{pq}(w_p\cdot w_{ik})(w_q\cdot w_{il})]-[k\longleftrightarrow l].$$

The contravariant tensor is inverse to the covariant one,

$$g^{ij}g_{jk} = \delta^i_k, \tag{29}$$

and can be used to raise indices and to make contractions. The determinant of the metric tensor is

$$g = \det g_{ij}$$
$$= \frac{1}{3!}\epsilon^{ijk}\epsilon^{lmn}(w_i\cdot w_l)(w_j\cdot w_m)(w_k\cdot w_n), \tag{30}$$

and the contravariant metric tensor is

$$g^{ij} = \frac{1}{2g}\epsilon^{ikl}\epsilon^{jmn}(w_k\cdot w_m)(w_l\cdot w_n). \tag{31}$$

In three dimensions, there is a useful identity for the antisymmetrized product of two contravariant tensors:

$$g^{ik}g^{jl}-g^{il}g^{jk} = \epsilon^{jim}\epsilon^{kln}g_{mn}/g. \tag{32}$$

The scalar curvature is obtained as the full contraction:

$$R = g^{ik}g^{jl}R_{ijkl} = \frac{1}{2}(g^{ik}g^{jl}-g^{il}g^{jk})R_{ijkl}$$
$$= \frac{1}{2g^2}\epsilon^{ijk}\epsilon^{i'j'k'}(w_k\cdot w_{k'})[2g(w_{ii'}\cdot w_{jj'}) \tag{33}$$
$$-\epsilon^{plm}\epsilon^{ql'm'}(w_p\cdot w_{ii'})(w_q\cdot w_{jj'})(w_l\cdot w_{l'})(w_m\cdot w_{m'})].$$

Recalling that w^α is a 6-dimensional vector, we can rewrite the scalar curvature in another form:

$$R = \frac{1}{72g^2}\epsilon_{\alpha\beta\gamma\delta\varepsilon\eta}\epsilon_{\alpha'\beta'\gamma'\delta'\varepsilon\eta}\epsilon^{ijk}\epsilon^{i'j'k'}\epsilon^{pll'}\epsilon^{qmm'}$$
$$\times w_i^\alpha w_{i'}^{\alpha'} w_j^\beta w_{j'}^{\beta'} w_k^\gamma w_{k'}^{\gamma'} w_{lm}^\delta w_{l'm'}^{\delta'} w_p^\zeta w_q^\zeta. \tag{34}$$

This form makes it clear that the scalar curvature vanishes if w^α has only three nonzero components, which corresponds to a flat 3-dimensional manifold.

Finally, we consider the Jacobian for the change of integration variables from the set of the lengths of the tetrahedra edges (J_i^2 and j_i^2 specified at all lattice sites) to the external coordinates w^α. In the continuum limit, this Jacobian is quite simple. It is given by the determinant of a 6×6 matrix composed of the second derivatives:

$$\prod_x dJ_i^2(x)dj_i^2(x) = \prod_x dw^\alpha(x)\mathrm{Jac}(w),$$
$$\mathrm{Jac}(w) = \det w^\alpha_{ij}. \tag{35}$$

Since $w^\alpha_{ij} = w^\alpha_{ji}$, there are six independent second derivatives. The Jacobian is zero in the degenerate case where the triangulation by tetrahedra can be embedded in less than 6 dimensions.

7. CONTINUUM DUALITY TRANSFORMATION AND THE BIANCHI IDENTITY

At this point, it is instructive to compare the duality transformation on the lattice with that in the continuum theory. The continuum partition function (3) can be written with the help of an additional Gaussian integration over the "dual field strength" J^a_{ij} as

$$\mathscr{Z} = \int DJ^a_{ij}DA^a_i$$
$$\times\exp\int d^3x\left[-\frac{g_3^2}{4}J^2_{ij}+\frac{i}{2}J^a_{ij}(\partial_i A^a_j-\partial_j A^a_i+\epsilon^{abc}A^b_iA^c_j)\right]. \tag{36}$$

This representation is usually referred to as the first-order formalism.

In the Abelian case, where the A_i commutator term is absent, the integration over A_i results in the δ-function of the Bianchi identity:

$$\partial_i J_{ij} = 0, \text{ or } \epsilon_{ijk}\partial_i J_k = 0, \qquad J_k = \frac{1}{2}\epsilon_{ijk}J_{ij}. \tag{37}$$

Because of this identity, one can parametrize $J_k = \partial_k w$ and obtain the partition function

$$\mathcal{Z}_{\text{Abel}} = \int Dw \exp\int d^3x\left[-\frac{g_3^2}{2}(\partial_k w)^2\right]. \tag{38}$$

It represents a theory of a free massless scalar field *w*. This is in accordance with a 3D Abelian theory containing only one physical (transverse) polarization. It is easy to check that gauge-invariant correlation functions of field strengths coincide with those computed in the original formulation.

In the non-Abelian case, the integration over A_i^a is more complicated and there is no simple Bianchi identity for

$$J_k^a = (1/2)\epsilon_{ijk}J_{ij}^a.$$

However, one can formally perform the Gaussian integration over A_i^a [5] with the result

$$\mathcal{Z} = \int DJ_i^a \det{}^{1/2}(\mathcal{J}^{-1}) \times \exp\int d^3x\left[-\frac{g_3^2}{2}(J_i^a)^2 - \frac{i}{2}(\epsilon_{ijm}\partial_j J_m^a)(\mathcal{J}^{-1})_{ik}^{ab}(\epsilon_{kln}\partial_l J_n^b)\right], \tag{39}$$

where $\mathcal{J}^{-1}$ is the inverse matrix,

$$(\mathcal{J}^{-1})_{ik}^{ab}\epsilon^{bcd}\epsilon_{klm}J_m^d = \delta^{ac}\delta_{il}, \qquad \det(\mathcal{J}^{-1}) = (\det J_k^a)^{-3}. \tag{40}$$

We note that the second term in the exponent is purely imaginary; the full partition function is real because for each configuration $J_i^a(x)$, there exists a configuration involving $-J_i^a(x)$, which adds a complex conjugate expression.

We now turn to the discretized version of the dual theory. As explained above, we need 6 flat dimensions to embed the dual lattice, and we have introduced 6-dimensional momenta J^α [see Eq. (25)]. These momenta obviously satisfy, e.g., the identity (see Fig. 7 for the notation)

$$\begin{aligned} &J_z^\alpha\left(x, y, z+\frac{a}{2}\right) - J_x^\alpha\left(x+\frac{a}{2}, y, z\right) \\ &= w^\alpha(x, y, z+a) - w^\alpha(x+a, y, z) \\ &= J_z^\alpha\left(x+a, y, z+\frac{a}{2}\right) - J_x^\alpha\left(x+\frac{a}{2}, y, z+a\right) \end{aligned} \tag{41}$$

as well as other components. This is nothing but a discretized version of the Bianchi identity:

$$\epsilon_{ijk}\partial_i J_k^\alpha = 0, \quad \alpha = 1, \ldots, 6. \tag{42}$$

Therefore, we recover the simple (flat) form of the Bianchi identity for the dual field strength in 6 dimensions. One can say that the complicated (nonlinear) form of the usual non-Abelian Bianchi identity is the result of projecting the flat Bianchi identity onto the curved color space.

8. THE WILSON LOOP

In this section, we present the Wilson loop in the representation j_s,

$$W_{j_s} = \frac{1}{2j_s+1}\text{Tr Pexp}\, i\oint dx_i A_i^a T^a, \tag{43}$$

in terms of dual variables.

In terms of the original lattice, the Wilson loop corresponds to adding a product of the $D^{j_s}(U)$ functions to all links along the loop, with a chain contraction of the "magnetic" indices. Because of these insertions, the links contained in the loop correspond to the integration over three *D*-functions instead of two (as for all other links). As the result, we obtain additional 3*jm* symbols along the loop that combine into new 9*j* symbols assigned to all lattice sites (see Appendix D). For example, the 9*j* symbols assigned to vertices *a* and *b* are (see Fig. 3 for notations)

$$\text{“}a\text{”} = \begin{Bmatrix} j_4 & j_1 & j_{12} \\ j'_{15} & \mathbf{j_s} & j_{15} \\ j_{13} & j'_1 & j_{14} \end{Bmatrix}, \qquad \text{“}b\text{”} = \begin{Bmatrix} j_2 & j_1 & j_9 \\ j'_{17} & \mathbf{j_s} & j_{17} \\ j_{18} & j'_1 & j_{16} \end{Bmatrix}. \tag{44}$$

The accompanying sign factors are given in Appendix D. Six triades of the 9*j* symbols corresponding to all its rows and columns satisfy the triangle inequalities.

Unlike the 6*j* symbol, the 9*j* symbol cannot be represented by a geometrical figure with the edges equal to the 9*j* symbol entries. In addition, the link momenta

along the loop split into pairs: j_1 and j_1', j_{15} and j_{15}', j_{17} and j_{17}', and so on. The "primed" and "nonprimed" angular momenta satisfy triangle inequalities, with the source j_s being the third edge of the triangles. If j_s is an integer, there always exists a contribution with $j_1' = j_1$ (and so on). If j_s is a half-integer, one necessarily has $j_1' \neq j_1$.

Thus, there appears to be a fundamental difference between Wilson loops in integer and half-integer representations. For integer representations, one can proceed as in the vacuum case and parametrize the dual lattice sites by the coordinates $w^\alpha(x)$ related to angular momenta through Eq. (25). In the half-integer case, one cannot uniquely parametrize the dual lattice by the coordinates $w^\alpha(x)$. In the presence of a Wilson loop in a half-integer representation, the dual space w^α is not simply connected: there is an infinitely thin cylindrical "hole" in the dual space along the loop.

9. ASYMPTOTIC FORM OF THE 6*j* SYMBOLS

In the continuum limit as $\beta \longrightarrow \infty$ and $J, j \longrightarrow \infty$, one can replace the 6*j* symbols by their asymptotic forms. The asymptotic form was ingeniously guessed in a seminal paper by Ponzano and Regge [6] and later explicitly derived and improved by Schulten and Gordon [7]. The results of these works can be summarized as follows.

First of all, one draws a tetrahedron with the edges equal to $j_n + 1/2$, where j_n are the six momenta of a given 6*j* symbol. It should be stressed that although four momenta triades satisfy triangle inequalities, this is not necessarily true for the same triades shifted by 1/2. In that case, the 6*j* symbol is said to be "classically forbidden" and is exponentially suppressed at large j_n. If j_n lie in the "classically allowed" region, the asymptotic form is given by the Ponzano–Regge formula

$$\left\{ \begin{array}{ccc} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{array} \right\} = \frac{1}{\sqrt{12\pi V(j)}} \times \cos\left[\sum_n \left(j_n + \frac{1}{2}\right)\theta_n + \frac{\pi}{4}\right], \tag{45}$$

where $V(j)$ is the 3-dimensional volume of the tetrahedron and θ_n is the dihedral angle in the tetrahedron corresponding to the edge $j_n + 1/2$. Since we are interested in the large-j_n limit, we systematically neglect the shifts by 1/2. The tetrahedron volume can be found from the Cayley formula

$$V(j)^2 = \frac{1}{288}\begin{vmatrix} 0 & j_4^2 & j_5^2 & j_6^2 & 1 \\ j_4^2 & 0 & j_3^2 & j_2^2 & 1 \\ j_5^2 & j_3^2 & 0 & j_1^2 & 1 \\ j_6^2 & j_2^2 & j_1^2 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}. \tag{46}$$

The dihedral angle corresponding, for example, to the edge j_1 can be found from

$$\cos\theta_1 = \frac{1}{16}[j_1^4 + j_1^2(2j_4^2 - j_5^2 - j_6^2 - j_2^2 - j_3^2) + (j_2^2 - j_3^2)(j_6^2 - j_5^2)][S(j_1, j_2, j_3)S(j_1, j_5, j_6)]^{-1}, \tag{47}$$

where

$$S(j_1, j_2, j_3) = \frac{1}{4}[(j_1 + j_2 + j_3)(j_2 + j_3 - j_1) \times (j_1 - j_2 + j_3)(j_1 + j_2 - j_3)]^{1/2} \tag{48}$$

is the area of the triangle built on the edges $j_{1,2,3}$. The dihedral angles are defined such that $0 \le \theta \le \pi$. Because the 6-dimensional angular momenta j^α defined in Section 6 are such that their lengths are the edges of the tetrahedra, we can find the dihedral angles from simpler formulas involving scalar products of momenta in the 6-dimensional space. For example, Eq. (47) can be rewritten as

$$\cos\theta_1 = \frac{(j_1 \cdot j_2)(j_1 \cdot j_6) - j_1^2(j_2 \cdot j_6)}{\sqrt{j_1^2 j_2^2 - (j_1 \cdot j_2)^2}\sqrt{j_1^2 j_6^2 - (j_1 \cdot j_6)^2}}. \tag{49}$$

We note that the angle is defined to be equal to π (not 0!) when the two vectors—j_2 and j_6—coincide; it is zero when they point in the opposite directions. We use this formula in what follows.

10. THE ANGLE DEFECT

The Yang–Mills partition function (20) is a product of many 6*j* symbols, for each of which we use the asymptotic form (45) in approaching the continuum limit. Each cosine can be written as a half-sum of exponentials of an imaginary argument. Therefore, we must consider the sum of products of many imaginary exponents:

$$\prod_n^N \cos\Omega_n = \frac{1}{2^N}\sum_{\{\epsilon_n = \pm 1\}} \exp\left(i\sum_n \epsilon_n \Omega_n\right), \tag{50}$$

where Ω_n denotes the cosine argument in Eq. (45) for the *n*th 6*j* symbol and the sum runs over all signs $\epsilon_n = \pm 1$.

The expression in the exponent of Eq. (50) can be rearranged as follows. We first pick one of the edges of

the dual lattice, whose length is a link j_l or a plaquette J_P, and combine all dihedral angles θ_n related to this edge as coming from the nth tetrahedron. We then sum over all edges of the dual lattice. Therefore, we can write

$$\sum_n \epsilon_n \Omega_n = \sum_P J_P \left(\sum_{n=1}^{4} \epsilon_n \theta_n(J_P) \right) + \sum_l j_l \left(\sum_{n=1}^{6} \epsilon_n \theta_n(j_l) \right), \quad \epsilon_n = \pm 1. \tag{51}$$

As can be seen, e.g., from Fig. 7, each plaquette J enters four tetrahedra; therefore, the corresponding sum over n in Eq. (51) goes from 1 to 4. Each link j enters six tetrahedra, and in this case the sum is over six dihedral angles $\theta_n(j)$ with the appropriate signs ϵ_n.

We consider the contribution to Eq. (50) where all signs $\epsilon_n = +1$, and we assume for a moment that the dual lattice spans a 3-dimensional Euclidean manifold. The sum of the dihedral angles about an edge is then equal to $4\pi - 2\pi = 2\pi$ when summing over four tetrahedra and equal to $6\pi - 2\pi = 4\pi$ when summing over six tetrahedra. In the first case, we obtain

$$\exp(2\pi i J) = (-1)^{2J};$$

in the second case, we obtain

$$\exp(4\pi i j) = (-1)^{4j} = +1.$$

We note that the sign factor $(-1)^{2J}$ compensates exactly the same factor in the partition function (20). We conclude that if the configuration of the momenta is "flat," there exists a contribution to the sum (50) that does not oscillate with varying J's and j's. In fact, there are exactly two such contributions corresponding to taking all signs $\epsilon_n = \pm 1$ simultaneously. Contributions of any other choice of the signs oscillate fast at large J's and j's, and thus die out in the continuum limit.

A generic configuration of momenta cannot be embedded into a flat 3-dimensional space. Therefore, the sum of dihedral angles about the respective edges J and j generally differs from 2π and 4π. These differences are sometimes called the angle deficiencies or angle defects (we use the second term). We denote them as

$$\Theta(J) = \sum_{n=1}^{4} \theta_n(J) - 2\pi, \tag{52}$$

$$\Theta(j) = \sum_{n=1}^{6} \theta_n(j) - 4\pi. \tag{53}$$

Our task is to identify those contributions to Eq. (50) that survive the continuum limit in the general case where the dual lattice is a curved 3-dimensional manifold. To be more precise, we must consider the sum of all momenta on the lattice times their angle defects,

$$\exp i \left[\sum_P J_P \Theta(J_P) + \sum_l j_l \Theta(j_l) \right], \tag{54}$$

and to find the contribution of the order a^3 to this exponent, where a is the lattice spacing. The $O(a^3)$ order is needed to compensate for the $1/a^3$ factor arising in passing from the summation over the lattice sites to the integration over the 3-dimensional space.

In the continuum limit, we assume that the momenta are given by the gradients of a 6-component function $w^\alpha(x)$ having the meaning of 6-dimensional coordinates of the dual lattice sites (see Eq. (25)). If we restrict ourselves to the first terms in the gradient expansion in Eq. (25), the momenta are expressed only through three vectors: $\partial_x w^\alpha$, $\partial_y w^\alpha$, and $\partial_z w^\alpha$. These vectors define a flat 3-dimensional space; therefore, the angle defects Θ are zero in the first-derivative approximation. To obtain a nonzero angle defect, it is necessary to expand the momenta in Eq. (25) up to the second derivatives of w^α. In what follows, we see that this is also sufficient in three dimensions.

Since the angle defects Θ vanish if j's are taken in the first approximation of the gradient expansion, the expansion of Θ's starts from linear terms in the lattice spacing a. In accordance with Eq. (25), the expansion of the momenta also starts from linear terms in a. Therefore, one can expect that the expansion of the exponent in Eq. (54) starts from $O(a^2)$ terms. If that were so, the configuration would be too "ultraviolet" and would not survive the continuum limit. Fortunately, an exact cancellation of all $O(a^2)$ terms appears to occur in the sum over several neighboring edges of the dual lattice, and the exponent in Eq. (54) then proves to be finite in the continuum limit.

We next embark on the rather tedious enterprise of calculating the angle defects about six plaquette J's in a cube (each entering four tetrahedra) and about twelve link j's that are the edges of that cube (each involved in six tetrahedra, see Section 5). Unfortunately, this seems to be the smallest elementary group that is repeated through the lattice. This means that we must compute $6 \times 4 + 12 \times 6 = 96$ dihedral angles, expressing them through the first and second gradients of the 6-component function w^α using Eqs. (25) and (49). This formidable calculation has been performed by heavily exploiting *Mathematica*. The intermediate results are very lengthy and we do not present them here. However, the final result is beautiful. From a

direct calculation, we obtain

$$\exp i\left[\sum_P J_P\Theta(J_P) + \sum_l j_l\Theta(j_l)\right] \tag{55}$$
$$= \exp i \sum_{\text{points } x} a^3 \frac{1}{2}\sqrt{g(w)}R(w) = \exp\frac{i}{2}\int d^3x\sqrt{g(w)}R(w),$$

where g is the determinant of the induced metric tensor given by Eq. (30) and R is the corresponding scalar curvature given by Eq. (33). Actually, we obtain the expression for the left-hand side of Eq. (55) in the form of Eq. (33) (written in components, 384 terms!), in which we recognize the scalar curvature.

In fact, this result is a concrete realization of a more general theory developed many years ago by Regge [6, 8]. In these papers, it was shown that the left-hand side of (55) must be equal to its right-hand side for any simplicial triangulation, provided it has a smooth continuum limit. However, no relation of the scalar curvature R to any concrete triangulation was given. We feel that it is the first time that this ingenious relation is derived explicitly for a concrete triangulation, and the continuum limit is shown to exist.

11. THE FULL PARTITION FUNCTION

Having dealt with the 6j symbols of partition function (20), we now turn to the weight factors $T_J(\beta)$. For large β and J, it follows from Eq. (11) that

$$\prod_{\text{plaquettes}} T_J(\beta) = \exp\left[-\sum_{\text{plaquettes}}\frac{2J^2}{\beta}\right] = \exp\left[-\int\frac{d^3x}{a^3}\frac{2J_i^2}{\beta}\right] \tag{56}$$
$$= \exp\left[-\int d^3x\frac{g_3^2}{2}\partial_i w^a\partial_i w^\alpha + O(a^2)\right],$$

where relation (5) between β and the physical coupling constant g_3^2 has been used together with the gradient expansion for the angular momenta in Eq. (25). Combining Eqs. (55) and (56) and using

$$\partial_i w^\alpha\partial_i w^\alpha = g_{ii}(w),$$

we finally obtain the Yang–Mills partition function

$$\mathcal{Z} = \int Dw^\alpha(x)\mathrm{Jac}(w)g(w)^{-5/4} \tag{57}$$
$$\times\exp\int d^3x\left[-\frac{g_3^2}{2}g_{ii} + \frac{i}{2}\sqrt{g}R\right].$$

The second term is the Einstein–Hilbert action with a purely imaginary Newton constant; it is invariant under global 6-dimensional rotations of the external coordinates $w^\alpha(x)$ and, more importantly, under local 3-dimensional diffeomorphisms

$$w^\alpha(x) \longrightarrow w^\alpha(x'(x)).$$

The first term in Eq. (57) can be viewed as a "matter" source,

$$-\frac{g_3^2}{2}\int d^3x g_{ii} = -\frac{g_3^2}{2}\int d^3x\sqrt{g}T^{ij}g_{ij}, \tag{58}$$

with the stress-energy tensor $T^{ij}\sqrt{g} = \delta^{ij}$.

violating the invariance under diffeomorphisms. Since it is homogeneous in space, it can be called the "ether."

The functional measure in Eq. (57) arises from two sources. One factor is the Jacobian for the change of variables from the tetrahedra edges J's and j's to w^α [see Eq. (35)]. The other factor arises from the tetrahedra volumes in the asymptotic form of the 6j symbols (Eq. (45)). In the continuum limit, the tetrahedron volume can be written as $V(j) \sim \sqrt{g}$, and there are 5 tetrahedra per lattice site (see Section 3).

Once the partition function is written in covariant terms, we can forget the origin of the external coordinates w^α (as the dual lattice coordinates) and consider the metric tensor components g_{ij} as independent dynamical variables that are integrated over in Eq. (57). The Jacobian for this change of variables can easily be worked out: in fact, it is the inverse of Jac(w) introduced in Eq. (35). We thus obtain the integration measure for partition function (57) as

$$\int Dg_{ij}g^{-5/4}, \text{ instead of } \int Dg_{ij}g^{-2}, \tag{59}$$

which would be the invariant measure in three dimensions. We give an independent check of the power –5/4 in the next section. It any case, it is a local counterterm not affecting the physics.

We stress that the partition function written in terms of the metric tensor does not contain explicit color degrees of freedom. Implicitly, however, the theory does contain three gluons at short distances. This follows from a simple dimensional analysis of Eq. (57). The dimension of the first term in (57) is $g_3^2\partial^2 w^2$ (we are counting the number of derivatives and the overall power of w); the dimension of the second term is $\partial^3 w^1$. At short distances, where quantum fluctuations of $w^\alpha(x)$ vary rapidly, the second term dominates the first one. On the other hand, the second term is a fast-oscillating functional at nonzero R. Therefore, the leading contribution to the functional integral arises from zero-curvature fluctuations of w^α, that is, from the 3-dimensional w^α. Being inserted into the first term, the three components of w^α describe three massless scalar fields. These fields correspond to three gluons of SU(2) with one physical (transverse) polarization (which is similar to Eq. (38) for free electrodynamics). This is the correct result for the non-Abelian theory at short distances in three dimensions.

At large distances or at low field momenta, on the contrary, the first term is dominant because it has less derivatives. It describes six (instead of three) massless

scalar degrees of freedom. This is the correct number of gauge-invariant degrees of freedom in the SU(2) theory. However, the theory remains strongly nonlinear, and it is not clear so far whether massless modes survive in the physical spectrum.

12. QUANTUM GRAVITY FROM THE FIRST-ORDER CONTINUUM FORMALISM

In this section, we give another derivation of partition function (57) directly in the continuum theory starting from the first-order formalism (see Section 7). We show that the two terms in the exponent of Eq. (36) are in fact in a one-to-one correspondence with the two terms in Eq. (57) and that the integration measure coincides with that in Eq. (59).

Actually, in the previous section we saw that the first terms in Eqs. (36) and (57) are equal to

$$\begin{aligned} S_1 &= -\frac{g_3^2}{2}\int d^3x (J_i^a)^2 \\ &= -\frac{g_3^2}{2}\int d^3x \partial_i w^\alpha \partial_i w^\alpha = -\frac{g_3^2}{2}\int d^3x g_{ii}. \end{aligned} \tag{60}$$

We now derive a less trivial relation for the second terms,

$$\begin{aligned} S_2 &= \frac{i}{2}\int d^3x \epsilon^{ijk} J_i^a (\partial_j A_k^a - \partial_k A_j^a + \epsilon_{abc} A_j^b A_k^c) \\ &= \frac{i}{2}\int d^3x \sqrt{g} R. \end{aligned} \tag{61}$$

This derivation is done in two steps. We first show, following Witten [9], that the left-hand side of (61) can be represented as a certain Chern–Simons term. Second, we show that it is formally equal to the Einstein–Hilbert action. A subtle question about the integration measure is discussed at the end of this section.

The left-hand side of (61) is obviously invariant under ordinary gauge transformations

$$\begin{aligned} \delta A_i^a &= -\partial_i \delta^{ab}\omega^b + \epsilon_{abc}\omega^b A_i^c = -D_i^{ab}(A)\omega^b, \\ \delta J_i^a &= \epsilon_{abc}\omega^b J_i^c, \end{aligned} \tag{62}$$

where

$$D_i^{ab}(A) = \partial_i \delta^{ab} + \epsilon_{abc} A_i^c$$

is the covariant derivative. Less evidently, it is also invariant under the local transformation

$$\delta J_i^a = -\partial_i \rho^a + \epsilon_{abc}\rho^b A_i^c, \quad \delta A_i^a = 0. \tag{63}$$

Indeed, after the integration by parts, the variation of the action becomes

$$\begin{aligned} \delta S_2 &= \frac{i}{2}\int d^3x \rho^a \epsilon_{ijk} D_i^{ab}(A) F_{jk}^b, \\ F_{jk}^b &= \partial_j A_k^b - \partial_k A_j^b + \epsilon_{bcd} A_j^c A_k^d. \end{aligned} \tag{64}$$

This variation vanishes owing to the Bianchi identity $\epsilon_{ijk} D_i^{ab} F_{jk}^b = 0$.

The two transformations form a 6-parameter gauged Poincaré group called ISO(3). We introduce three "momentum" generators P_i and three "angular momentum" generators L_i satisfying the Poincaré algebra

$$\begin{aligned} [P_a, P_b] = 0, \quad [L_a, L_b] &= i\epsilon_{abc} L_c, \\ [L_a, P_b] &= i\epsilon_{abc} P_c. \end{aligned} \tag{65}$$

We next introduce a 6-component vector field $\hat{B}_i$,

$$\begin{aligned} \hat{B}_i &= J_i^a P_a + A_i^a L_\alpha \equiv B_i^\alpha T^\alpha, \\ T^\alpha &= \begin{cases} P_a, & \alpha = a = 1, 2, 3, \\ L_a, & \alpha = 3 + a = 4, 5, 6. \end{cases} \end{aligned} \tag{66}$$

Its gauge transformation has the standard form

$$\begin{aligned} \hat{B}_i &\longrightarrow S^{-1}\hat{B}_i S + iS^{-1}\partial_i S, \\ S &= \exp[i\rho^a P_a + i\omega^a L_a]. \end{aligned} \tag{67}$$

Using Poincaré algebra (65), it is easy to verify that its infinitesimal form coincides with Eqs. (62) and (63).

Because the left-hand side of Eq. (61) is invariant under these 6-parameter transformations, it can be rewritten in an explicitly ISO(3)-invariant form. For this purpose, we note that the invariant tensor of this group is

$$M_{\alpha\beta} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{68}$$

where each "1" is a unit 3×3 matrix. This matrix defines the scalar product $B^\alpha M_{\alpha\beta} C^\beta$ that is invariant under global (x-independent) transformations (67). Using this invariant tensor, we build a local gauge-invariant action having the form of the Chern–Simons term:

$$S_2 = \frac{i}{2}\int d^3x \epsilon^{ijk} M_{\alpha\beta} B_i^\alpha \left(\partial_j B_k^\beta + \frac{1}{3} F_{\gamma\delta}^\beta B_j^\gamma B_k^\delta\right), \tag{69}$$

where $F_{\beta\gamma}^\alpha = -F_{\gamma\beta}^\alpha$ are the ISO(3) structure constants. Explicitly,

$$\begin{aligned} F_{bc}^\alpha = 0, \quad F_{3+b,c}^a &= \epsilon_{abc}, \\ F_{3+b,c}^{3+a} = F_{3+b,3+c}^a = 0, \quad F_{3+b,3+c}^{3+a} &= \epsilon_{abc}. \end{aligned} \tag{70}$$

Using definition (66), it is easy to verify that Eq. (69) coincides with the left-hand side of (61); however, it is explicitly invariant under the 6-parameter gauge transformation (67).

Equation (69) has the form of the Chern–Simons term in the Yang–Mills theory. Although our derivation above is for the gauge group SU(2), it can be easily generalized to any Lie group: it suffices to replace the SU(2) structure constants ϵ_{abc} with the structure constants f_{abc} of the gauge group under consideration. We note in passing that in four dimensions, the mixed $iJ^a_{\mu\nu}F^a_{\mu\nu}(A)$ term of the first-order formalism also possesses an additional local symmetry. To unveil it, it is sufficient to replace the scalar parameter ρ^a in the transformation (63) by a 4-vector parameter ρ^a_μ: the invariance again follows from the Bianchi identity, this time in four dimensions.

The second step in the derivation is more standard. We introduce the dreibein e^a_i, $a = 1, 2, 3$ satisfying the condition $e^a_i e^{bi} = \delta^{ab}$ such that the metric tensor is represented as $g_{ij} = e^a_i e^a_j$ and the spin connection

$$\omega^{ab}_i = \frac{1}{2}e^{ak}(\partial_i e^b_k - \partial_k e^b_i) - \frac{1}{2}e^{bk}(\partial_i e^a_k - \partial_k e^a_i) - \frac{1}{2}e^{ak}e^{bl}e^c_i(\partial_k e^c_l - \partial_l e^c_k). \quad (71)$$

We then identically rewrite $\sqrt{g}R$ as

$$\sqrt{g}R = \frac{1}{2}\epsilon^{ijk}e^a_i(\partial_j\omega^a_k - \partial_k\omega^a_j + \epsilon_{abc}\omega^b_j\omega^c_k), \quad (72)$$

where

$$\omega_{ai} = \omega^a_i = \frac{1}{2}\epsilon_{abc}\omega^{bc}_i.$$

If we now identify the dreibein with the dual field strength $e^a_i = J^a_i$ and the connection with the Yang–Mills potential $\omega^a_i = A^a_i$, Eq. (72) takes exactly the form of the left-hand side of (61). This parallel was first noticed in [10].

Since the pure gravity action can be rewritten as the Chern–Simons term in Eq. (69), it is actually a topological field theory [9] with no real propagating particles. It is the "ether" term that violates the invariance under diffeomorphisms and restores the propagation of gluons, as it should be in the Yang–Mills theory (see the end of the previous section).

Finally, we remark that the integration measure (59) could be anticipated from the first-order formalism as well. Indeed, integrating over A^a_i in Eq. (39), we obtain Eq. (40), with the integration measure over the dreibein being

$$(\det J)^{-3/2} \sim g^{-3/4}.$$

The Jacobian for the change of variables from the dreibein to the metric tensor is

$$de^a_i \sim dg_{ij}g^{-1/2}.$$

Adding the powers, we obtain

$$-\frac{3}{4} - \frac{1}{2} = -\frac{5}{4},$$

as in Eq. (59).

13. CONCLUSIONS AND AN OUTLOOK

We have studied the dual transformation of the SU(2) Yang–Mills theory in 3 dimensions, from both the continuum and lattice points of view.

On the lattice, one can introduce dual variables that are the angular momenta of the plaquettes (J's) and of the links (j's). The partition function can be rewritten as a product of 6j symbols made of those angular momenta. A Wilson loop is described by replacing the 6j symbols along the loop with a product of 9j symbols. One can construct a dual lattice made of tetrahedra whose edges have the lengths equal to J's and j's; the tetrahedra span a curved 3D manifold that can be embedded into a flat 6D space.

In the continuum limit, the angular momenta are large and we have introduced continuum 6D Euclidean external coordinates $w^\alpha(x)$ to describe the curved dual space. The Bianchi condition for the Yang–Mills field strength has been shown to be trivially soluble in six flat dimensions.

At large angular momenta, one can use the asymptotic form of the 6j symbols given by Ponzano and Regge. Using a specific simplicial triangulation of the dual space (as dictated by the original lattice), we have shown that the product of 6j symbols does have a smooth continuum limit that appears to be the Einstein–Hilbert action, with the metric tensor g_{ij} and the scalar curvature R being expressed through the flat external coordinates $w^\alpha(x)$. Although this result cannot be considered as particularly new (it is the cornerstone of the Regge's simplicial gravity), it is to the best of our knowledge the first time that this result has been explicitly derived from a concrete triangulation of the curved space and the continuum limit is shown to exist. We have also found the integration measure for the continuum limit.

The continuum Yang–Mills partition function can be rewritten as a quantum gravity theory but with an "ether" term violating the invariance with respect to the general coordinate transformations (diffeomorphisms). This term, however, revives gluons at short distances, in

contrast to the topological pure gravity theory, where no particles propagate.

The presentation of the Yang–Mills theory in the quantum gravity form (57) is explicitly color gauge–invariant because the metric tensor of the dual space is color-neutral. Thus, we have formulated the Yang–Mills theory solely in terms of colorless "glueball" degrees of freedom.[2] It turns out to be an interacting theory of six massless scalar fields. Nevertheless, it correctly reproduces the propagation of gluons at small distances. At the moment, it is not clear to us how to proceed best in order to reveal its large-distance behavior. We now discuss several possibilities.

One possibility is to exploit the fact that the pure quantum gravity theory is topological and, therefore, essentially a free theory. One can try to make a perturbative expansion in g_3^2 about it.

Another possibility is to use the Chern–Simons term (69), obtained from integrating over heavy fermions, in this case belonging to some ISO(3) representation. The subsequent integration over the bosonic fields A_i and J_i is trivial because there is no kinetic energy term for these fields: the result would be a local four-fermion theory with infinitely heavy fermions; it might be soluble, at least, in the large-N_c limit.

Probably the most promising possibility is to pursue the analogy with quantum gravity and corresponding methods. One can average Eq. (57) over 3D diffeomorphisms: the second term is invariant, while the first term is not. Integrating the first term over diffeomorphisms produces a diffeomorphism-invariant effective action containing growing powers of the curvature. The effective action may lead to a nonzero vacuum expectation value of the scalar curvature, thereby yielding a mass gap for the diffeomorphism–non-invariant correlation functions, for example, the correlation functions of $F_{\mu\nu}^2$.

There are several other tasks for the future lying on the surface. First, it would be interesting to generalize the present approach to color groups other than SU(2). In view of the sad fact that the theory of the "6*j* symbols" for higher Lie groups is not sufficiently developed, it will probably be difficult to make a straightforward generalization of the lattice formulation. A more promising approach would be to start from the first-order formalism, in particular, because the wide local symmetry revealed in Section 12 can be directly generalized to any Lie group. Second, it would be interesting to make a transformation similar to the one considered in this paper in $d = 4$. The lattice 6*j* symbols have been known for a while in this case [4] [for the SU(2) color]; however, it again seems that the first-order formalism is more promising due to the additional gauge symmetry noticed in Section 12.

[2] A somewhat similar line was developed in [11] for the 3 + 1 dimensional Yang–Mills theory in the Hamiltonian approach; see also [12].

ACKNOWLEDGMENTS

We are grateful to Pavel Pobylitsa for many fruitful discussions. D.I. Diakonov acknowledges the very useful conversation with Ben Mottelson. V.Yu. Petrov is grateful to NORDITA for the hospitality extended to him in Copenhagen, in particular, for the partial support by a Nordic Project grant. The work was supported in part by the Russian Foundation for Basic Research (project no. 97-27-15L).

APPENDIX A

D-FUNCTIONS, 3*jm*, 6*j*, AND 9*j* SYMBOLS

The Wigner *D*-functions are eigenfunctions of the square of the angular momentum operator (for example, written in terms of three Euler angles α, β, and γ):

$$\mathbf{J}^2 D^J_{mn}(\alpha, \beta, \gamma) = J(J+1) D^J_{mn}(\alpha, \beta, \gamma),$$
$$J = 0, \frac{1}{2}, 1, \frac{3}{2}\ldots, \quad -J \le m, n \le J. \tag{A.1}$$

These functions can be referred to as the eigenfunctions of a spherical top; they are $(2J+1)^2$-fold degenerate. The "magnetic" quantum numbers m and n have the meaning of the projections of the angular momentum of the spherical top on the third axis in the "body-fixed" and "lab" frames. One can parametrize a 2×2 unitary matrix by Euler angles as

$$U = \exp(i\alpha\tau^3/2)\exp(i\beta\tau^2/2)\exp(i\gamma\tau^3/2). \tag{A.2}$$

It is convenient to use the unitary matrix U as a formal argument of the *D*-functions. Their main properties are as follows.

(1) Multiplication law:

$$D^J_{kl}(U_1U_2) = D^J_{km}(U_1)D^J_{ml}(U_2) \tag{A.3}$$

(summation over repeated indices understood);

(2) unitarity:

$$D^J_{kl}(U^\dagger) = (D^J_{lk}(U))^* \tag{A.4}$$

("*" denotes complex conjugation);

(3) phase condition:

$$(D^J_{lk}(U))^* = (-1)^{l-k} D^J_{-l,-k}(U),$$
$$D^J_{kl}(1) = \delta^{(2J+1)}_{kl}; \tag{A.5}$$

(4) orthogonality and normalization:

$$\int dU D^{J_1}_{kl}(U^\dagger) D^{J_2}_{mn}(U) = \frac{1}{2J_1+1}\delta_{J_1J_2}\delta_{kn}\delta_{lm}, \tag{A.6}$$

with the integration taken over the Haar measure,

$$\int dU\ldots = \int d(SU)\ldots = \int d(US)\ldots,$$
$$\int dU = 1; \tag{A.7}$$

(5) completeness (with the δ-function understood in the Haar measure sense):

$$\delta(U, V) = \sum_J (2J+1) D^J_{kl}(U^\dagger) D^J_{lk}(V); \qquad \text{(A.8)}$$

(6) matrix element:

$$\int dU D^{J_1}_{a_1 b_1}(U) D^{J_2}_{a_2 b_2}(U) D^{J_3}_{a_3 b_3}(U) = \begin{pmatrix} J_1 & J_2 & J_3 \\ a_1 & a_2 & a_3 \end{pmatrix} \begin{pmatrix} J_1 & J_2 & J_3 \\ b_1 & b_2 & b_3 \end{pmatrix}, \qquad \text{(A.9)}$$

where (…) denote 3*jm* symbols;

(7) decomposition of a direct product of irreducible representations,

$$D^{J_1}_{a_1 b_1}(U) D^{J_2}_{a_2 b_2}(U) = \sum_J (2J+1) \begin{pmatrix} J & J_1 & J_2 \\ -c & a_1 & a_2 \end{pmatrix} \times \begin{pmatrix} J & J_1 & J_2 \\ -d & b_1 & b_2 \end{pmatrix} (-1)^{d-c} D^J_{cd}(U). \qquad \text{(A.10)}$$

The last two factors can be replaced by $D^J_{-d,-c}(U^\dagger)$ using Eq. (A.5).

The 3*jm* symbols are symmetric under cyclic permutations of the columns. An interchange of two columns gives a sign factor:

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ k & l & m \end{pmatrix} = (-1)^{j_1+j_2+j_3} \begin{pmatrix} j_2 & j_1 & j_3 \\ l & k & m \end{pmatrix}, \text{ etc.} \qquad \text{(A.11)}$$

If one changes the signs of all the "magnetic" quantum numbers or projections, the 3*jm* symbol also acquires a sign factor:

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ k & l & m \end{pmatrix} = (-1)^{j_1+j_2+j_3} \begin{pmatrix} j_1 & j_2 & j_3 \\ -k & -l & -m \end{pmatrix}. \qquad \text{(A.12)}$$

A "practical" definition of the 6*j* symbol {…} is via the contraction over projections in three 3*jm* symbols:

$$\sum_{klm} (-1)^{j_4-k+j_5-l+j_6-m} \begin{pmatrix} j_5 & j_1 & j_6 \\ l & p & -m \end{pmatrix} \begin{pmatrix} j_6 & j_2 & j_4 \\ m & q & -k \end{pmatrix} \times \begin{pmatrix} j_4 & j_3 & j_5 \\ k & r & -l \end{pmatrix} = \begin{pmatrix} j_1 & j_2 & j_3 \\ -p & -q & -r \end{pmatrix} \begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix}. \qquad \text{(A.13)}$$

The summation over the projections *k*, *l*, and *m* is such that $p = m - l$, $q = k - m$, and $r = l - k$ are kept fixed.

Another definition of the 6*j* symbol is via the full contraction of projections in four 3*jm* symbols:

$$\sum_{klmnop} (-1)^{j_4+n+j_5+o+j_6+p} \begin{pmatrix} j_1 & j_2 & j_3 \\ k & l & m \end{pmatrix} \begin{pmatrix} j_1 & j_5 & j_6 \\ k & o & -p \end{pmatrix} \times \begin{pmatrix} j_4 & j_2 & j_6 \\ -n & l & p \end{pmatrix} \begin{pmatrix} j_4 & j_5 & j_3 \\ n & -o & m \end{pmatrix} = \begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix}. \qquad \text{(A.14)}$$

Since the three *j*'s of any 3*jm* symbol satisfy the triangle inequalities, e.g.,

$$|j_1 - j_2| \le j_3 \le j_1 + j_2, \text{ etc.},$$

the following four triades of the 6j symbols must satisfy the triangle inequalities: $(j_1 j_2 j_3)$, $(j_1 j_5 j_6)$, $(j_2 j_4 j_6)$, and $(j_3 j_4 j_5)$; otherwise, the 6*j* symbol is zero.

The 6*j* symbols are symmetric under the permutation of any two columns and under a simultaneous interchange of the upper and lower arguments in any two columns, e.g.,

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix} = \begin{Bmatrix} j_1 & j_3 & j_2 \\ j_4 & j_6 & j_5 \end{Bmatrix} = \begin{Bmatrix} j_4 & j_2 & j_6 \\ j_1 & j_5 & j_3 \end{Bmatrix}, \text{ etc.} \qquad \text{(A.15)}$$

The full contraction of six 3*jm* symbols yields a 9*j* symbol,

$$\sum \begin{pmatrix} j_1 & j_2 & j_3 \\ k & l & m \end{pmatrix} \begin{pmatrix} j_4 & j_5 & j_6 \\ n & o & p \end{pmatrix} \begin{pmatrix} j_7 & j_8 & j_9 \\ q & r & s \end{pmatrix} \begin{pmatrix} j_1 & j_4 & j_7 \\ k & n & q \end{pmatrix} \times \begin{pmatrix} j_2 & j_5 & j_8 \\ l & o & r \end{pmatrix} \begin{pmatrix} j_3 & j_6 & j_9 \\ m & p & s \end{pmatrix} = \begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \\ j_7 & j_8 & j_9 \end{Bmatrix}. \qquad \text{(A.16)}$$

The 9*j* symbol is symmetric under the transposition and under even permutations of rows and columns; under odd permutations, it acquires the sign factor $(-1)^{j_1+\ldots+j_9}$. As follows from the definition, six momenta triades corresponding to the rows and columns of the 9*j* symbol satisfy triangle inequalities.

A convenient reference book on *D*-functions, 3*jm*, and 6*j* and 9*j* symbols is [1], from which we have borrowed the definitions.

APPENDIX B

6*j* SYMBOLS IN AN "EVEN" CUBE

In this Appendix, we decompose the D^J-functions of two plaquettes into a sum of single D^j-functions labelled by the link angular momenta *j*. We then assemble the arising 3*jm* symbols into 6*j* symbols attached to

the corners of the even cubes. The notation is given in Fig. 2.

We find it convenient (although not necessary) to write the decomposition for the pairs containing $U_{1,4,12,6,7,10}$ (these links are at the lower left and upper right corners of the cube) in terms of $D(U)$; the rest, in terms of $D(U^\dagger)$.

Using Eq. (A.10) of Appendix A, we obtain

$$D^{J_A}_{i_a i_b}(U_1)D^{J_B}_{j_b j_a}(U_1^\dagger) = (-1)^{j_a - j_b}\sum_{j_1}(2j_1+1)$$

$$\times\begin{pmatrix} j_1 & J_A & J_B \\ -o_a & i_a & -j_a \end{pmatrix}\begin{pmatrix} j_1 & J_A & J_B \\ -o_b & i_b & -j_b \end{pmatrix}(-1)^{o_b - o_a}D^{j_1}_{o_a o_b}(U_1),$$

$$D^{J_A}_{i_b i_c}(U_2)D^{J_C}_{k_b k_c}(U_2^\dagger) = (-1)^{k_b - k_c}\sum_{j_2}(2j_2+1)$$

$$\times\begin{pmatrix} j_2 & J_A & J_C \\ -p_b & i_b & -k_b \end{pmatrix}\begin{pmatrix} j_2 & J_A & J_C \\ -p_c & i_c & -k_c \end{pmatrix}D^{j_2}_{-p_c,-p_b}(U_2^\dagger),$$

$$D^{J_D}_{l_d l_c}(U_3)D^{J_A}_{i_c i_d}(U_3^\dagger) = (-1)^{i_d - i_c}\sum_{j_3}(2j_3+1)$$

$$\times\begin{pmatrix} j_3 & J_D & J_A \\ -q_d & l_d & -i_d \end{pmatrix}\begin{pmatrix} j_3 & J_D & J_A \\ -q_c & l_c & -i_c \end{pmatrix}D^{j_3}_{-q_c,-q_d}(U_3^\dagger),$$

$$D^{J_E}_{m_a m_d}(U_4)D^{J_A}_{i_d i_a}(U_4^\dagger) = (-1)^{i_a - i_d}\sum_{j_4}(2j_4+1)$$

$$\times\begin{pmatrix} j_4 & J_E & J_A \\ -r_a & m_a & -i_a \end{pmatrix}\begin{pmatrix} j_4 & J_E & J_A \\ -r_d & m_d & -i_d \end{pmatrix}(-1)^{r_d - r_a}D^{j_4}_{r_a r_d}(U_4),$$

$$D^{J_B}_{j_e j_f}(U_5)D^{J_F}_{n_f n_e}(U_5^\dagger) = (-1)^{n_e - n_f}\sum_{j_5}(2j_5+1)$$

$$\times\begin{pmatrix} j_5 & J_B & J_F \\ -s_e & j_e & -n_e \end{pmatrix}\begin{pmatrix} j_5 & J_B & J_F \\ -s_f & j_f & -n_f \end{pmatrix}D^{j_5}_{-s_f,-s_e}(U_5^\dagger),$$

$$D^{J_C}_{k_f k_g}(U_6)D^{J_F}_{n_g h_f}(U_6^\dagger) = (-1)^{n_f - n_g}\sum_{j_6}(2j_6+1)$$

$$\times\begin{pmatrix} j_6 & J_C & J_F \\ -t_f & k_f & -n_f \end{pmatrix}\begin{pmatrix} j_6 & J_C & J_F \\ -t_g & k_g & -n_g \end{pmatrix}(-1)^{t_g - t_f}D^{j_6}_{t_f t_g}(U_6), \tag{B.1}$$

$$D^{J_F}_{n_h n_g}(U_7)D^{J_D}_{l_g l_h}(U_7^\dagger) = (-1)^{l_h - l_g}\sum_{j_7}(2j_7+1)$$

$$\times\begin{pmatrix} j_7 & J_F & J_D \\ -u_h & n_h & -l_h \end{pmatrix}\begin{pmatrix} j_7 & J_F & J_D \\ -u_g & n_g & -l_g \end{pmatrix}(-1)^{u_g - u_h}D^{j_7}_{u_h u_g}(U_7),$$

$$D^{J_F}_{n_e n_h}(U_8)D^{J_E}_{m_h m_e}(U_8^\dagger) = (-1)^{m_e - m_h}\sum_{j_8}(2j_8+1)$$

$$\times\begin{pmatrix} j_8 & J_F & J_E \\ -\nu_e & n_e & -m_e \end{pmatrix}\begin{pmatrix} j_8 & J_F & J_E \\ -\nu_h & n_h & -m_h \end{pmatrix}D^{j_8}_{-\nu_h,-\nu_e}(U_8^\dagger),$$

$$D^{J_C}_{k_b k_f}(U_9)D^{J_B}_{j_f j_b}(U_9^\dagger) = (-1)^{j_b - j_f}\sum_{j_9}(2j_9+1)$$

$$\times\begin{pmatrix} j_9 & J_C & J_B \\ -w_b & k_b & -j_b \end{pmatrix}\begin{pmatrix} j_9 & J_C & J_B \\ -w_f & k_f & -j_f \end{pmatrix}D^{j_9}_{-w_f,-w_b}(U_9^\dagger),$$

$$D^{J_D}_{l_c l_g}(U_{10})D^{J_C}_{k_g k_c}(U_{10}^\dagger) = (-1)^{k_c - k_g}\sum_{j_{10}}(2j_{10}+1)$$

$$\times\begin{pmatrix} j_{10} & J_D & J_C \\ -x_c & l_c & -k_c \end{pmatrix}\begin{pmatrix} j_{10} & J_D & J_C \\ -x_g & l_g & -k_g \end{pmatrix}(-1)^{x_g - x_c}D^{j_{10}}_{x_c x_g}(U_{10}),$$

$$D^{J_E}_{m_d m_h}(U_{11})D^{J_D}_{l_h l_d}(U_{11}^\dagger) = (-1)^{l_d - l_h}\sum_{j_{11}}(2j_{11}+1)$$

$$\times\begin{pmatrix} j_{11} & J_E & J_D \\ -y_d & m_d & -l_d \end{pmatrix}\begin{pmatrix} j_{11} & J_E & J_D \\ -y_h & m_h & -l_h \end{pmatrix}D^{j_{11}}_{-y_h,-y_d}(U_{11}),$$

$$D^{J_B}_{j_a j_e}(U_{12})D^{J_E}_{m_e m_a}(U_{12}^\dagger) = (-1)^{m_a - m_e}\sum_{j_{12}}(2j_{12}+1)$$

$$\times\begin{pmatrix} j_{12} & J_B & J_E \\ -z_a & j_a & -m_a \end{pmatrix}\begin{pmatrix} j_{12} & J_B & J_E \\ -z_e & j_e & -m_e \end{pmatrix}(-1)^{z_e - z_a}D^{j_{12}}_{z_a z_e}(U_{12}).$$

We now combine the 3*jm* symbols related to the same vertices (they are marked by the appropriate indices of the projections *a*, *b*, *c*, *d*, *e*, *f*, *g*, *h*): three 3*jm* symbols for each vertex together with the appropriate sign factors. The three 3*jm* symbols per vertex combine into 6*j* symbols, one for each vertex of the cube.

Vertex *a*. Related to vertex *a* are the factors

$$\sum_{i_a, j_a, m_a}(-1)^{i_a + j_a + m_a - o_a - r_a - z_a}\begin{pmatrix} j_1 & J_A & J_B \\ -o_a & i_a & -j_a \end{pmatrix}$$

$$\times\begin{pmatrix} j_4 & J_E & J_A \\ -r_a & m_a & -i_a \end{pmatrix}\begin{pmatrix} j_{12} & J_B & J_E \\ -z_a & j_a & -m_a \end{pmatrix}$$

(we use $o_a + r_a + z_a = 0$, make cyclic permutations in all the 3*jm* symbols, and change the summation indices as $i, j, m \longrightarrow -i, -j, -m$)

$$= \sum_{i_a, j_a, m_a} (-1)^{-i_a - j_a - m_a} \begin{pmatrix} J_B & j_1 & J_A \\ j_a & -o_a & -i_a \end{pmatrix} \times \begin{pmatrix} J_A & j_4 & J_E \\ i_a & -r_a & -m_a \end{pmatrix} \begin{pmatrix} J_E & j_{12} & J_B \\ m_a & -z_a & -j_a \end{pmatrix} \quad \text{(B.2)}$$

$$= (-1)^{-J_A - J_B - J_E} \begin{pmatrix} j_1 & j_4 & j_{12} \\ o_a & r_a & z_a \end{pmatrix} \begin{Bmatrix} j_1 & j_4 & j_{12} \\ J_E & J_B & J_A \end{Bmatrix}.$$

In the last transformation, we used the definition of the 6*j* symbol, Eq. (A.13).

Vertex *b*. Related to vertex *b* are the factors

$$\sum_{i_b, j_b, k_b} (-1)^{o_b + k_b}\Big|_{o_b = i_b - j_b} \begin{pmatrix} j_1 & J_A & J_B \\ -o_b & i_b & -j_b \end{pmatrix} \times \begin{pmatrix} j_2 & J_A & J_C \\ p_b & i_b & -k_b \end{pmatrix} \begin{pmatrix} j_9 & J_C & J_B \\ -w_b & k_b & -j_b \end{pmatrix}$$

(we interchange the first two columns in the first 3*jm* symbol and change the signs of all its projections, which does not change the sign of the 3*jm* symbols; we then make cyclic permutations of the last two 3*jm* symbols and change the summation indices as $i, j, k \longrightarrow -i, -j, -k$)

$$= \sum_{i, j, k} (-1)^{-i + j - k} [\text{insert } 1 = (-1)^{2J_B - 2j}] \times \begin{pmatrix} J_A & j_1 & J_B \\ i & o_b & -j \end{pmatrix} \begin{pmatrix} J_B & j_4 & J_C \\ j & -w_b & -k \end{pmatrix} \begin{pmatrix} J_C & j_9 & J_A \\ k & -p_b & -i \end{pmatrix} \quad \text{(B.3)}$$

$$= (-1)^{J_B - J_A - J_C} \begin{pmatrix} j_1 & j_9 & j_2 \\ -o_b & w_b & p_b \end{pmatrix} \begin{Bmatrix} j_1 & j_9 & j_2 \\ J_C & J_A & J_B \end{Bmatrix}.$$

In each case, we combine the three 3*jm* symbols and the sign factors such that they suit the definition of the 6*j* symbol given in Appendix A, Eq. (A.13).

An important property of the sign factors is as follows: if j_1, J_A, and J_B enter the same 3*jm* symbol, there is the equality

$$(-1)^{\pm 2j_1 \pm 2J_A \pm 2J_B} = 1, \quad \text{(B.4)}$$

where all signs can occur. This is because there are either zero or two half-integer momenta among the three momenta. Another important property is that if the momentum *J* enters a certain 3*jm* symbol and m is its projection, then

$$(-1)^{2J \pm 2m} = +1.$$

This is because *J* and *m* are either integer or half-integer, but simultaneously.

In what follows, we list the expressions for other vertices of the cube without a detailed derivation (which is quite similar to the derivations given above).

Vertex *c*.

$$(-1)^{J_A + J_D - J_C} \begin{pmatrix} j_2 & j_3 & j_{10} \\ p_c & q_c & -x_c \end{pmatrix} \begin{Bmatrix} j_2 & j_3 & j_{10} \\ J_D & J_C & J_A \end{Bmatrix}. \quad \text{(B.5)}$$

Vertex *d*.

$$(-1)^{J_A - J_D - J_E} \begin{pmatrix} j_4 & j_3 & j_{11} \\ -r_d & q_d & y_d \end{pmatrix} \begin{Bmatrix} j_4 & j_3 & j_{11} \\ J_D & J_E & J_A \end{Bmatrix}. \quad \text{(B.6)}$$

Vertex *e*.

$$(-1)^{J_E - J_B - J_F} \begin{pmatrix} j_{12} & j_8 & j_5 \\ -z_e & v_e & s_e \end{pmatrix} \begin{Bmatrix} j_{12} & j_8 & j_5 \\ J_F & J_B & J_E \end{Bmatrix}. \quad \text{(B.7)}$$

Vertex *f*.

$$(-1)^{J_B + J_C - J_F} \begin{pmatrix} j_6 & j_5 & j_9 \\ -t_f & s_f & w_f \end{pmatrix} \begin{Bmatrix} j_6 & j_5 & j_9 \\ J_B & J_C & J_F \end{Bmatrix}. \quad \text{(B.8)}$$

Vertex *g*.

$$(-1)^{J_C + J_D + J_F} \begin{pmatrix} j_6 & j_{10} & j_7 \\ t_g & x_g & u_g \end{pmatrix} \begin{Bmatrix} j_6 & j_{10} & j_7 \\ J_D & J_F & J_C \end{Bmatrix}. \quad \text{(B.9)}$$

Vertex *h*.

$$(-1)^{J_E + J_F - J_D} \begin{pmatrix} j_7 & j_{11} & j_8 \\ -u_h & y_h & v_h \end{pmatrix} \begin{Bmatrix} j_7 & j_{11} & j_8 \\ J_E & J_F & J_D \end{Bmatrix}. \quad \text{(B.10)}$$

Combining all these factors, we obtain Eq. (16) corresponding to the cube.

APPENDIX C

6*j* SYMBOLS AT THE LATTICE SITES

In this Appendix, we show how the integration over the link variables combines the 3*jm* factors in Eq. (16) into 6*j* symbols composed of the link momenta *j*, one for each site of the lattice. The notation is given in Fig. 3.

We consider the integration over the link variables $U_{1,4,12,13,14,15}$ entering the vertex *a* shown in Fig. 3.

This vertex is an intersection of four even cubes denoted as I, II, III, and IV in Fig. 3. Link 1 is common to cubes I and II, link 4 is common to I and IV, and so on.

The analytical expression for cube I is given in Eq. (16). The factors relevant to vertex a are

$$D^{j_1}_{o_a o_b}(U_1) D^{j_4}_{r_a r_d}(U_4) D^{j_{12}}_{z_a z_e}(U_{12}) \begin{Bmatrix} j_1 & j_4 & j_{12} \\ o_a & r_a & z_a \end{Bmatrix}. \quad \text{(C.1)}$$

It is not necessary to directly compute the corresponding expressions for cubes II–IV. It is sufficient to draw a correspondence between the links and the sites of other cubes with those of cube I. For example, link 1, as seen from the viewpoint of cube II, is analogous to link 7 of cube I; vertex *a* seen from the viewpoint of cube II is analogous to vertex *h* of cube I, and vertex *b* is analogous to vertex *g*. In Table 1, we give the list of the "analogs" of links in cubes II–IV to those of cube I.

Using the correspondence given in Table 1, we can immediately read off from Eq. (16) the expressions relevant to vertex *a*, arising from cubes II–IV.

From cube II:

$$D^{j_1'}_{u_a u_b}(U_1) D^{j_{13}}_{-y_a, -y_\beta}(U^\dagger_{13}) \times D^{j_{14}}_{-v_a, -v_a}(U^\dagger_{14}) \begin{Bmatrix} j_1' & j_{13} & j_{14} \\ -u_a & y_a & v_a \end{Bmatrix}. \quad \text{(C.2)}$$

From cube III:

$$D^{j_{12}'}_{x_a x_e}(U_{12}) D^{j_{14}'}_{-p_a, -p_\alpha}(U^\dagger_{14}) \times D^{j_{15}}_{-q_a, -q_\epsilon}(U^\dagger_{15}) \begin{Bmatrix} j_{12}' & j_{14}' & j_{15} \\ -x_a & p_a & q_a \end{Bmatrix}. \quad \text{(C.3)}$$

From cube IV:

$$D^{j_4'}_{u_a u_b}(U_4) D^{j_{15}'}_{-s_a, -s_\epsilon}(U^\dagger_{15}) \times D^{j_{13}'}_{-w_a, -w_\beta}(U^\dagger_{13}) \begin{Bmatrix} j_4' & j_{14}' & j_{13}' \\ -t_a & s_a & w_a \end{Bmatrix}. \quad \text{(C.4)}$$

Integrating over $U_{1,4,12,13,14,15}$, we obtain

$$\int dU_1 D^{j_1}_{o_a o_b}(U_1) D^{j_1'}_{u_a u_b}(U_1) = \frac{\delta_{j_1 j_1'}}{2j_1+1}(-1)^{u_b - u_a} \delta_{o_a, -u_a} \delta_{o_b, -u_b}, \quad \text{(C.5)}$$

$$\int dU_4 D^{j_4}_{r_a r_d}(U_4) D^{j_4'}_{t_a t_d}(U_4) = \frac{\delta_{j_4 j_4'}}{2j_4+1}(-1)^{t_d - t_a} \delta_{r_a, -t_a} \delta_{r_d, -t_d}, \quad \text{(C.6)}$$

$$\int dU_{12} D^{j_{12}}_{z_a z_e}(U_{12}) D^{j_{12}'}_{x_a x_e}(U_{12}) = \frac{\delta_{j_{12} j_{12}'}}{2j_{12}+1}(-1)^{x_e - x_a} \delta_{z_a, -x_a} \delta_{z_e, -x_e}, \quad \text{(C.7)}$$

$$\int dU_{13} D^{j_{13}}_{-y_a, -y_\beta}(U^\dagger_{13}) D^{j_{13}'}_{-w_a, -w_\beta}(U^\dagger_{13}) = \frac{\delta_{j_{13} j_{13}'}}{2j_{13}+1}(-1)^{w_a - w_\beta} \delta_{w_a, -y_a} \delta_{w_\beta, -y_\beta}, \quad \text{(C.8)}$$

$$\int dU_{14} D^{j_{14}}_{-v_a, -v_\alpha}(U^\dagger_{14}) D^{j_{14}'}_{-p_a, -p_\alpha}(U^\dagger_{14}) = \frac{\delta_{j_{14} j_{14}'}}{2j_{14}+1}(-1)^{p_a - p_\alpha} \delta_{p_a, -v_a} \delta_{p_\alpha, -v_\alpha}, \quad \text{(C.9)}$$

$$\int dU_{15} D^{j_{15}}_{-q_a, -q_\epsilon}(U^\dagger_{15}) D^{j_{15}'}_{-s_a, -s_\epsilon}(U^\dagger_{15}) = \frac{\delta_{j_{15} j_{15}'}}{2j_{15}+1}(-1)^{s_a - s_\epsilon} \delta_{s_a, -q_a} \delta_{s_\epsilon, -q_\epsilon}. \quad \text{(C.10)}$$

The four 3*jm* symbols in Eqs. (C.1)–(C.4) are now fully contracted over all indices. This results in a 6*j* symbol, in accordance with Eq. (A.14). Indeed, for vertex *a*, we have

$$\text{"}a\text{"} = \sum_{o r q v y z} (-1)^{o+r+z-q-v-y} \begin{Bmatrix} j_1 & j_4 & j_{12} \\ o & r & z \end{Bmatrix} \times \begin{Bmatrix} j_1 & j_{13} & j_{14} \\ o & y & v \end{Bmatrix} \begin{Bmatrix} j_{12} & j_{14} & j_{15} \\ z & -v & q \end{Bmatrix} \begin{Bmatrix} j_4 & j_{15} & j_{13} \\ r & -q & -y \end{Bmatrix} \quad \text{(C.11)}$$

(we note that $o + r + z = 0$; we change the summation variable as $y \longrightarrow -y$ and interchange the last two columns in the second 3*jm* symbol and the first two columns in the last two 3*jm* symbols, which gives the sign factors $(-1)^{j_1 + j_{13} + j_{14}}$, $(-1)^{j_{12} + j_{14} + j_5}$, and $(-1)^{j_4 + j_{13} + j_{15}}$;

Table 1

II	I
1	7
13	11
14	8
a	*h*

III	I
12	10
14	2
15	3
a	*c*

IV	I
4	6
13	9
15	5
a	*f*

Table 2

II	I
1	7
16	6
18	10
b	*g*

V	I
9	11
16	4
17	3
b	*d*

VI	I
2	8
17	5
18	12
b	*e*

we then insert two unities in the form $1 = (-1)^{2v-2j_{14}}$ and $1 = (-1)^{2q-2j_{15}}$)

$$= (-1)^{j_1+j_4+j_{12}+j_{13}-j_{14}-j_{15}} \times \sum_{orqvyz} \begin{pmatrix} j_1 & j_4 & j_{12} \\ o & r & z \end{pmatrix} \begin{pmatrix} j_1 & j_{14} & j_{13} \\ o & v & -y \end{pmatrix} \times \begin{pmatrix} j_{15} & j_4 & j_{13} \\ -q & r & y \end{pmatrix} \begin{pmatrix} j_{15} & j_{14} & j_{12} \\ q & -v & z \end{pmatrix} \quad \text{(C.12)}$$

$$= (-1)^{j_1+j_4+j_{12}+j_{13}-j_{14}-j_{15}} \begin{Bmatrix} j_1 & j_4 & j_{12} \\ j_{15} & j_{14} & j_{13} \end{Bmatrix}$$

$$= (-1)^{j_1+j_4-j_{12}+j_{13}+j_{14}+j_{15}} \begin{Bmatrix} j_1 & j_4 & j_{12} \\ j_{15} & j_{14} & j_{13} \end{Bmatrix}, \quad \text{(C.13)}$$

where we used equation $(-1)^{j_{12}-j_{14}-j_{15}} = (-1)^{-j_{12}+j_{14}+j_{15}}$ (see Eq. (B.4)), which is valid because j_{12}, j_{14}, and j_{15} come from the same 3*jm* symbol. This is the final result for vertex *a*: the six angular momenta assigned to the six links entering this vertex combine to produce a 6*j* symbol.

One can treat vertex *b* similarly, see Fig. 3. The links labelled by 1, 2, 9, 16, 17, and 18 enter this vertex; they are pairwise shared by cubes I, II, V, and VI. The correspondence between the links viewed from the viewpoint of cubes II, V, and VI with those of cube I is given in Table 2.

Performing the same steps as in deriving the 6*j* symbol for vertex *a*, we arrive at the following result for the vertex *b*:

$$\text{“}b\text{”} = (-1)^{j_1+j_2+j_9+j_{16}+j_{17}-j_{18}} \begin{Bmatrix} j_1 & j_9 & j_2 \\ j_{17} & j_{18} & j_{16} \end{Bmatrix}. \quad \text{(C.14)}$$

We note that vertex *a* is of the “even” and vertex *b* is of the “odd” type: all other vertices of the lattice can be considered as either “even” or “odd.” Therefore, Eqs. (C.13) and (C.14) give the full result. Combining them, we find that the sign factor

$$(-1)^{2j} = (-1)^{-2j} \quad \text{(C.15)}$$

must be attributed to all links of the lattice.

We now prove that this sign factor is equivalent (in the vacuum!) to the sign factor

$$(-1)^{2J} = (-1)^{-2J} \quad \text{(C.16)}$$

attributed to all plaquettes of the lattice. We recall that all links are shared by two even cubes whose faces carry plaquette values *J*. We first attribute all links to only one (out of the two possible) cube according to some rule. Many such rules can be suggested, the only requirement being that each link is attributed to one and only one even cube. An example is given by the following construction: we choose edges 12, 5, 9, 2, and 7 (see Fig. 2) as “belonging” to the cube shown in that figure. The remaining six edges then “belong” to one of the neighboring even cubes. For example, edge 1 is counted as “belonging” to cube II (see Fig. 3). Indeed, from the cube II point of view, this edge has type 7, and so forth. It can be seen that in this scheme, every link of the full lattice “belongs” to one and only one even cube.

We have, therefore, the sign factor

$$(-1)^{2j_{12}+2j_5+2j_9+2j_2+2j_7} \quad \text{(C.17)}$$

attributed to cube I. Next, we recall that, e.g., j_{12} enters the 3*jm* symbol together with the plaquette angular momenta J_B and J_E (see Eq. (B.2)). Using Eq. (B.4) appropriate to this case, we can replace $(-1)^{2j_{12}} = (-1)^{2J_B+2J_E}$. Similarly $(-1)^{2j_5} = (-1)^{2J_B+2J_F}$ and so on. As a result, we see that sign factor (C.17) is equal to

$$(-1)^{2J_A+2J_B+2J_C+2J_D+2J_E+2J_F}. \quad \text{(C.18)}$$

This procedure can be repeated for all even cubes of the lattice. This proves the above statement that the product of all link sign factors (C.15) can be replaced by the product of all plaquette sign factors (C.14). We stress that this proof is valid only for the vacuum, i.e., for the partition function itself but, generally speaking, not for the averages of operators.

APPENDIX D

9*j* SYMBOLS FROM THE WILSON LOOP

Let the Wilson loop in the representation j_s go through links …, 15, 1, 17, …, see Fig. 3 for notation. This means that one has to integrate three *D*-functions of the link variables $U_{15,1,\ldots}$, instead of two, as was the

case in Eqs. (C.5) and (C.10), with the other integrations remaining unchanged. We now have

$$\int dU_1 D^{j_1}_{o_a o_b}(U_1) D^{j'_1}_{u_a u_b}(U_1) D^{j_s}_{m_a m_b}(U_1) = \begin{pmatrix} j_1 & j'_1 & j_s \\ o_a & u_a & m_a \end{pmatrix} \begin{pmatrix} j_1 & j'_1 & j_s \\ o_b & u_b & m_b \end{pmatrix}, \tag{D.1}$$

$$\int dU_{15} D^{j_{15}}_{-q_a, -q_\epsilon}(U^\dagger_{15}) D^{j'_{15}}_{-s_a, -s_\epsilon}(U^\dagger_{15}) D^{j_s}_{m_b m_\epsilon}(U_{15}) = (-1)^{m_\epsilon - m_a} \begin{pmatrix} j_{15} & j'_{15} & j_s \\ q_\epsilon & s_\epsilon & m_\epsilon \end{pmatrix} \begin{pmatrix} j_{15} & j'_{15} & j_s \\ q_a & s_a & m_a \end{pmatrix}. \tag{D.2}$$

Using the other 3*jm* symbols related to vertex *a* [see Eqs. (C.1)–(C.4)] and the Kronecker symbols from Eqs. (C.6)–(C.9), we obtain the vertex *a* in the form

$$\text{“}a\text{”} = \sum (-1)^{r+z+w+p-m} \begin{pmatrix} j_1 & j_4 & j_{12} \\ o & r & z \end{pmatrix} \times \begin{pmatrix} j'_1 & j_{13} & j_{14} \\ -u & -w & -p \end{pmatrix} \begin{pmatrix} j_{14} & j_{15} & j_{12} \\ p & q & z \end{pmatrix} \begin{pmatrix} j_4 & j'_{15} & j_{13} \\ r & s & w \end{pmatrix} \times \begin{pmatrix} j_1 & j'_1 & j_s \\ o & u & m \end{pmatrix} \begin{pmatrix} j_{15} & j'_{15} & j_s \\ q & s & m \end{pmatrix} \tag{D.3}$$

(we note that $r + z = -o$, $w + p = -u$, and $o + u + m = 0$; hence, the sign factor is +1; we change the signs of all projections in the second 3*jm* symbol and permute the columns in the other 3*jm* symbols to match the definition of the 9*j* symbols in Eq. (A.16))

$$= (-1)^{j'_1 - j_4 + j_{14} + j_{15} + j_s} \begin{Bmatrix} j_4 & j_1 & j_{12} \\ j'_{15} & j_s & j_{15} \\ j_{13} & j'_1 & j_{14} \end{Bmatrix}. \tag{D.4}$$

To obtain the final sign factor, we have used the relation $(-1)^{\pm 2j_1 \pm 2j_2 \pm 2j_3} = +1$ valid for any $j_{1,2,3}$ originating from the same 3*jm* symbol.

Proceeding in the same way, we obtain the vertex *b*,

$$\text{“}b\text{”} = \sum (-1)^{-\nu - y - t - x + m} \begin{pmatrix} j_1 & j_9 & j_2 \\ -o & -y & -\nu \end{pmatrix} \times \begin{pmatrix} j_{16} & j_{18} & j'_1 \\ t & x & u \end{pmatrix} \begin{pmatrix} j_{16} & j_{17} & j_9 \\ t & q & y \end{pmatrix} \begin{pmatrix} j_{18} & j_2 & j'_{17} \\ x & \nu & s \end{pmatrix} \times \begin{pmatrix} j_1 & j'_1 & j_s \\ o & u & m \end{pmatrix} \begin{pmatrix} j_{17} & j'_{17} & j_s \\ q & s & m \end{pmatrix} = (-1)^{j'_1 - j_2 + j_{16} + j_{17} + j_s} \begin{Bmatrix} j_2 & j_1 & j_9 \\ j'_{17} & j_s & j_{17} \\ j_{18} & j'_1 & j_{16} \end{Bmatrix}. \tag{D.5}$$

REFERENCES

1. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Scientific, Singarore, 1988).
2. J.-M. Drouffe and J.-B. Zuber, Phys. Rep. **102**, 1 (1983).
3. R. Anishetty, S. Cheluvaraja, H. S. Sharatchandra, and M. Mathur, Phys. Lett. B **314**, 387 (1993).
4. I. G. Halliday and P. Suranyi, Phys. Lett. B **350**, 189 (1995).
5. M. B. Halpern, Phys. Rev. D **16**, 1798 (1977).
6. G. Ponzano and T. Regge, in *Spectroscopic and Group Theoretical Methods in Physics*, Ed. by F. Bloch (North-Holland, Amsterdam, 1968).
7. K. Schulten and R. G. Gordon, J. Math. Phys. **16**, 1971 (1975).
8. T. Regge, Nuovo Cimento **19**, 558 (1961).
9. E. Witten, Nucl. Phys. B **311**, 46 (1988).
10. F. A. Lunev, Phys. Lett. B **295**, 99 (1992).
11. P. E. Haagensen and K. Johnson, Nucl. Phys. B **439**, 597 (1995); hep-th/9408164; P. E. Haagensen, K. Johnson, and C. S. Lam, Nucl. Phys. B **477**, 273 (1996); hep-th/9511226; R. Schiappa, Nucl. Phys. B **517**, 462 (1998); hep-th/9704206.
12. F. A. Lunev, J. Math. Phys. **37**, 5351 (1996); hep-th/9503133.

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 894–898.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1036–1040.

ATOMS, SPECTRA, RADIATION

Quantum Teleportation of an Einstein–Podolsy–Rosen Pair Using an Entangled Three-Particle State

V. N. Gorbachev* and A. I. Trubilko**

St. Petersburg Institute of Moscow State University of Printing, St. Petersburg, 190000 Russia
**e-mail: vn@vg3025.spb.edu*
***e-mail: tai@at3024.spb.edu*
Received October 13, 1999

Abstract—Quantum teleportation of an Einstein–Podolsky–Rosen pair using maximally entangled triplets to two receivers is studied. The projection basis for combined three-particle measurements, from the results of which the unknown state can be reconstructed, is found. The basis contains states where only two of the three particles are maximally entangled. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Quantum teleportation was proposed in [1]. It enables the sender *A* (conventionally named Alice) to transmit an unknown state to the user *B* (Bob) located in a different location of space. Teleportation of a state of a two-level quantum system or a qubit (quantum bit) requires an Einstein–Podolsky–Rosen (EPR) pair and a classical communication channel, along which the result of combined measurements performed in the Bell basis is transmitted. Various variants of this process, which employ two-particle entangled states, are examined in [2]. The case of quantum teleportation of a polarized photon and single-mode coherent light has been demonstrated experimentally in [3, 4]. In [4] light from an optical parametric amplifier was used as the source of the EPR pair, but the physical nature of the particles can be arbitrary. Thus, in [5] an ensemble of EPR correlated atoms was obtained. Such two-particle entangled states can be used in schemes for interspatial teleportation [6], where a quantum state is transferred between particles of a different physical nature [7].

The quantum teleportation procedure can be regarded as a computational process [8], corresponding to a network consisting of rotation type operations, logical elements c-NOT (controlled NOT), and other components. Such schemes are interesting because the components can be implemented using physical systems of various nature, for example, optical, such as a half-transmitting mirror or a polarized divider [9].

In the present paper we consider A variant of quantum teleportation of an entangled pair to two receivers *B* and *C* (Clair) using a three-particle entangled state of the type GHZ (Greenberger–Horne–Zeilinger). We note that the GHZ triplet has been realized experimentally in [10].

The main problem arising here concerns the form of the three-particle projection basis required for the combined measurements. The basis found, in contrast to the variants of quantum teleportation of one qubit, is not maximally entangled. It contains a state with two maximally entangled particles. Two receivers can reconstruct the unknown wave function of the EPR pair from the results of combined measurements in this basis, but neither receiver can do so separately. As shown in [11], in such a scheme with a GHZ triplet, where one qubit participates instead of an EPR pair, the unknown single-particle state can be reconstructed by only one receiver. The results obtained were extended to the case of the quantum teleportation of an *N*-particle entangled state of the EPR type. The basic features of the quantum teleportation process for one particle are presented in Section 2. The types of initial states of the entangled pair and the triplet are determined in Section 3, and the projection operators for combined measurements for them are constructed in Section 4. In Section 5 a protocol and a quantum teleportation scheme for an EPR pair are presented and the case of an *N*-particle entangled state of the EPR type is studied.

2. TELEPORTATION

According to [1], the teleportation of an unknown state of a quantum system by a sender *A* to a receiver *B* located at a different point in space involves the following basic aspects.

Let *A* possess a physical system, for example, a two-level or qubit system, in the unknown state

$$|\psi_1\rangle = \alpha|0\rangle + \beta|1\rangle, \quad (1)$$

where $|\alpha|^2 + |\beta|^2 = 1$. An EPR pair in the maximally entangled state $|\Psi_{23}\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$ is distributed between *A* and *B*, so that qubit 2 is located at *A* and qubit 3 is located at *B*. First, *A* performs combined

measurements on qubits 1 and 2 in the Bell basis, consisting of the for projecors $\prod_k = |\pi_k\rangle\langle\pi_k|$, $k = 1, \ldots, 4$, $\sum_k \prod_\kappa = 1$, $|\pi_1\rangle = |\Phi_{12}^+\rangle$, $|\pi_2\rangle = |\Phi_{12}^-\rangle$, $|\pi_3\rangle = |\Psi_{12}^+\rangle$, $|\pi_4\rangle = |\Psi_{12}^-\rangle$, where the Bell states represent the maximally entangled two-particle states

$$|\Phi^\pm\rangle = \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle), \quad (2)$$

$$|\Psi^\pm\rangle = \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle). \quad (3)$$

As result of combined measurements on the particles 1 and 2, the density matrix of the entire system $\rho = |\psi_1\rangle\langle\psi_1| \otimes |\Psi_{23}\rangle\langle\Psi_{23}|$, determined in the three-particle Hilbert space $H_1 \otimes H_2 \otimes H_3$, projects onto one of the four Bell states. The main circumstance ensuring the success of the procedure is that the probability of the kth outcome $\Pr(k) = \mathrm{Sp}\{\prod_k \rho \prod_k^\dagger\} = 1/4$ does not depend on α and β, and the reduced density matrix $\rho_3(k) = \mathrm{Sp}_{12}\{\prod_k \rho \prod_k^\dagger\}$ of qubit 3 is related with the unknown matrix of qubit one by the unitary transformation U_k:

$$\rho_3(k) = U_k \tilde{\rho}_1 U_k^\dagger, \quad (4)$$

where $\tilde{\rho}_1$ is the density matrix in H_3, corresponding to $\rho_1 = |\psi_1\rangle\langle\psi_1|$, and U_k is a set of unitary operators consisting of the Pauli matrices $U_1 = \sigma_x$, $U_2 = -i\sigma_y$, $U_3 = 1$, $U_4 = \sigma_z$.

According to the protocol of [1], the sender A sends the kth outcome of the measurement to the receiver B, who performs on his qubit 3 from the EPR pair one of four operations U_k, according to the message received. Ultimately, a qubit in the state $|\psi_1\rangle$ arises at B and teleportation has been successfully completed.

3. RESOURCES

Teleportation of an EPR pair requires a maximally entangled triplet of particles. On the basis of this assertion we shall examine the possible types of initial states.

The entangled pair can be described by a wave function of the form

$$|\Psi_{12}\rangle = \alpha|00\rangle + \beta|11\rangle, \quad (5)$$

where $|\alpha|^2 + |\beta|^2 = 1$, or

$$|\Psi_{\mathrm{EPR}}\rangle = \alpha|01\rangle + \beta|10\rangle, \quad (6)$$

which corresponds to the case of an EPR pair. Eight states with three maximally entangled particles can be indicated, for example,

$$\begin{aligned} &(|000\rangle \pm |111\rangle)/\sqrt{2}, \quad (|001\rangle \pm |110\rangle)/\sqrt{2}, \\ &(|010\rangle \pm |101\rangle)/\sqrt{2}, \quad (|100\rangle \pm |011\rangle)/\sqrt{2}. \end{aligned} \quad (7)$$

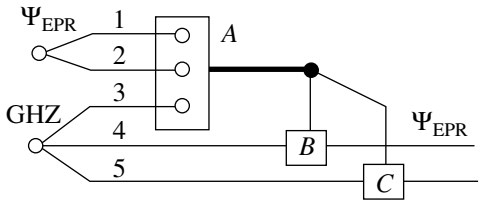

Fig. 1. Network for teleportation of an EPR pair using a GHZ triplet.

Without loss of generality we choose from the set of possible initial states presented above the state (6) and a GHZ triplet:

$$|\Psi_{\mathrm{GHZ}}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle). \quad (8)$$

Then the initial wave function of the entire system is a product

$$|\Psi\rangle = |\Psi_{\mathrm{EPR}}\rangle \otimes |\Psi_{\mathrm{GHZ}}\rangle. \quad (9)$$

Figure 1 displays a scheme for teleportation of an EPR pair, formed by particles 1 and 2, using an entangled triplet of particles 3, 4, and 5. The GHZ triplet is distributed between the sender A and the receivers B and C, to whom A sends the results of combined measurements on particles 1, 2, and 3. Eight projection operators $\prod_k$, forming a complete basis in which the wave function $|\Psi\rangle$ can be expanded, are required in order to perform combined measurements. The choice of such a basis is the main ingredient for solving the problem.

4. PROJECTION BASIS

It could be inferred that the set of projection operators $\prod_k$ will consist of the maximally entangled states (7). We denote this basis by $\pi_{(123)}$. However, a direct calculation shows that this is not so. The problem is that the projections of the initial wave function $|\Psi\rangle$ on the four vectors $(|010\rangle \pm |101\rangle)/\sqrt{2}$ and $(|100\rangle \pm |011\rangle)/\sqrt{2}$ from $\pi_{(123)}$ are 0. It is impossible to reconstruct the states of an EPR pair from these measurements. Consequently, here, in contrast to teleportation of one particle, the maximally entangled basis does not solve the problem.

In the present case a basis with two maximally entangled particles 1, 3 or 1, 2 is suitable. However, the presence of an entangled pair is only a necessary condition. To study the possible realizations of the operators $\prod_k$ we introduce a classification where one of the indicators will be the number of maximally entangled particles (two or three). Since all complete sets are related with one another by a unitary transformation, one can be chosen as the initial set. Let the initial basis be π_{123}:

$$|\pi_{123}\rangle = |ijk\rangle, \quad i, j, k = 0, 1, \quad (10)$$

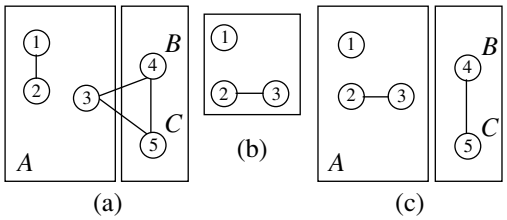

Fig. 2. Structure of the wave functions: (a) initial state with entangled particles 1, 2 from an EPR pair and 3, 4, 5 from a GHZ triplet; (b) projection basis $\pi_{1(23)}(4)$; (c) state after a measurement.

where each of the eight components describes a state with three independent particles. Any component of a different basis can be represented as a linear combination of $s \le 8$ vectors from π_{123}. In what follows, we shall assume the number s to be the same within a single basis and we shall use it for classification. This gives a basis $\pi_{1(23)}(s)$ in which particles 2 and 3 are maximally entangled. For $s = 2$ it has the form

$$|\pi_{1(23)}(2)\rangle = \{|i\rangle|\Phi^{\pm}_{23}\rangle;\ |i\rangle|\Phi^{\pm}_{23}\rangle\}, \quad i = 0, 1, \qquad (11)$$

where each of eight vectors, for example, $|0\rangle|\Phi^{\pm}_{23}\rangle = (|000\rangle \pm |011\rangle)/\sqrt{2}$, is represented by two elements from π_{123}. For the case $s = 4$

$$|\pi_{1(23)}(4)\rangle = \{|\pi^{\pm}_1\rangle|\Phi^{\pm}_{23}\rangle;\ |\pi^{\pm}_1\rangle|\Phi^{\pm}_{23}\rangle\}, \qquad (12)$$

where a pair of vectors forming a single-particle basis has the form

$$|\pi^{\pm}_1\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm \exp(i\varphi)|1\rangle). \qquad (13)$$

Any of the sets presented above is complete and orthonormal, but $\pi_{1(23)}(2)$ does not solve the problem. The reason is that the probabilities of the outcomes of measurements in this basis depend on the parameters of the teleportation wave function $|\Psi_{\text{EPR}}\rangle$: $\Pr(k) = |\alpha|^2/4$, $|\beta|^2/4$. Consequently, the unitary transformation relating the reduced density matrix of particles 3 and 4 with the state of the EPR pair will depend on α and β, i.e., on the unknown state. We note that for the basis $\pi_{(12)3}(4)$ their arises a situation, just as for $\pi_{(123)}(2)$, where half of the expansion coefficients are zero.

Two bases with $s = 4$ are suitable for teleportation of an EPR pair: $\pi_{1(23)}(4)$ or $\pi_{(13)2}(4)$, where the particles 2, 3 or 1, 3 are entangled. The structure of the initial state, the projection basis $\pi_{1(23)}(4)$, and the wave function of the entire system after the measurements are shown in Fig. 2.

5. TELEPORTATION OF AN EPR PAIR

The expansion of the wave function of the initial state in the basis $\pi_{1(23)}$, determined according to Eq. (12), where we set the phase $\varphi = 0$, has the form

$$\begin{aligned}|\Psi\rangle = {} & |\pi^{+}_1\rangle|\Phi^{+}_{23}\rangle|1\rangle + |\pi^{+}_1\rangle|\Phi^{-}_{23}\rangle|2\rangle \\ & + |\pi^{-}_1\rangle|\Phi^{+}_{23}\rangle|3\rangle + |\pi^{-}_1\rangle|\Phi^{-}_{23}\rangle|4\rangle \\ & + |\pi^{+}_1\rangle|\Phi^{+}_{23}\rangle|5\rangle + |\pi^{+}_1\rangle|\Phi^{-}_{23}\rangle|6\rangle \\ & + |\pi^{-}_1\rangle|\Phi^{+}_{23}\rangle|7\rangle + |\pi^{-}_1\rangle|\Phi^{-}_{23}\rangle|8\rangle,\end{aligned} \qquad (14)$$

where

$$\begin{aligned}|1, 2\rangle &= \beta|00\rangle \pm \alpha|11\rangle, \\ |3, 4\rangle &= -(\beta|00\rangle \mp \alpha|11\rangle), \\ |5, 6\rangle &= \beta|11\rangle \pm \alpha|00\rangle, \\ |7, 8\rangle &= -(\beta|11\rangle \mp \alpha|00\rangle).\end{aligned} \qquad (15)$$

Equations (15) mean that for measurements in the chosen basis the reduced density matrix $\rho_{45}(k) = \text{Sp}_{123}\{\Pi_k|\Psi\rangle\langle\Psi|\Pi_k\}$ of particles 4 and 5 will be related with the density matrix of the EPR pair by a unitary transformation

$$\rho_{45}(k) = U_k\tilde{\rho}_{EPR}U^{\dagger}_k, \quad k = 1, \ldots, 8, \qquad (16)$$

where $\tilde{\rho}_{\text{EPR}} = |\tilde{\Psi}_{\text{EPR}}\rangle\langle\tilde{\Psi}_{\text{EPR}}|$, $\tilde{\Psi}_{\text{EPR}}$ is a wave function from the Hilbert space $H_4 \otimes H_5$, corresponding to Ψ_{EPR}. The following expressions are valid for the unitary operator from Eq. (16): $U_1 = \sigma_{x4} \otimes I_5$, $U_2 = -U_3 = i\sigma_{y4} \otimes I_5$, $U_4 = -U_1$, $U_5 = I_4 \otimes \sigma_{x5}$, $U_6 = -U_7 = I_4 \otimes (-i\sigma_{y5})$, $U_8 = -U_5$. Here the Pauli operators $\sigma_{\gamma j}$ ($\gamma = x, y, z$) and the identity operators I_j act on the variables of the particle $j = 4, 5$.

In this scheme the teleportation of an EPR pair is accomplished according to the following protocol.

1. The sender A performs eight measurements in the basis $\pi_{1(23)}(4)$ on particles 1, 2, and 3, whose outcomes she transmits to B and C.

2. For the outcomes $k = 1$–4 the receivers B uses the unitary transformations σ_x, $i\sigma_y$, $-i\sigma_y$, $-\sigma_x$ on his particle. Ultimately, an EPR pair in the state Ψ_{EPR} arises at B and C.

3. For the outcomes $k = 5, 6$ the receiver C applies the unitary operations σ_x, $-i\sigma_y$, $i\sigma_y$, $-\sigma_x$ to her particle in order to reconstruct the state of the EPR pair.

In the protocol presented above, in half the cases only one receiver applies a unitary operation to her particle; the other receiver at the same time "operates" with the identity operator on her particle. This variant is not unique. The problem is that the initial states can be reconstructed by different methods. For example, the vector $|2\rangle$ from Eq. (15) can be obtained in two ways:

$$\beta|00\rangle - \alpha|11\rangle = i\sigma_{y4} \otimes I_5|\tilde{\Psi}_{\mathrm{EPR}}\rangle = \sigma_{x4} \otimes \sigma_{z5}|\tilde{\Psi}_{\mathrm{EPR}}\rangle. \quad (17)$$

The expression (17) means that for the outcome of a measurement, where $k = 2$, both receivers, *B* and *C*, must simultaneously operate on their particles, just as in the protocol presented, applying the operations σ_{x4} and σ_{z5} instead of $i\sigma_{y4}$ and the identity operation. However, these differences do not change the final results. The general feature here is the presence of two receivers, neither of which can solve the problem separately.

Figure 3 displays a network that simulates the teleportation of an EPR pair. It was constructed similarly to the case of one particle [8, 9, 12] and contains the c-NOT operation and the Hadamard transformation *H*. At the output of the EPR block particles 1 and 2 are in the entangled state Ψ_{EPR} after the c-NOT operation C_{12}, acting on qubits 1 and 2. The operation C_{12} does not change the state of qubit 1 hand to flips qubit 2 (the target) only when 1 is in a state corresponding to the logical 1. In the GHZ a maximally entangled GHZ triplet is produced by means of the Hadamard transformation *H*, operating qubit 3, ($(H|0\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, $H|1\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$), and two c-NOT operations C_{34} and C_{35}. At the output of the entire scheme the qubits 4 and 5 are in the entangled state Ψ_{EPR}, which does not depend on the states of the other qubits. We note that here, together with Ψ_{EPR}, a pair with the wave function (5) can be used.

The procedure examined above can be extended to the case of the teleportation of an *N*-particle entangled state of the EPR type

$$|\Psi_N\rangle = \alpha|0\rangle^N + \beta|1\rangle^N \quad (18)$$

using $N + 1$ qubits in a maximally entangled state of the GHZ type

$$|\Psi_{(N+1)}\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{N+1} + |1\rangle^{N+1}), \quad (19)$$

where $|i\rangle^N = |i\rangle \otimes \ldots |i\rangle$, $i = 0, 1$. This scheme involves $2N + 1$ particles, the sender *A*, and *N* receivers, between which $N + 1$ particles from Eqs. (19) are distributed. The initial wave function, which is determined in the Hilbert space $H_1 \otimes \ldots H_{2N+1}$, is a product $|\Psi\rangle = |\Psi_N\rangle \otimes |\Psi_{(N+1)}\rangle$.

To perform combined measurements on $N + 1$ particles 1, 2, …, *N*, $N + 1$ it is necessary to have 2^{N+1} projection operators, forming a complete set with a maximally entangled pair *M*, $N + 1$, $M = 1, \ldots, N$. For $M = N$ such a basis can be represented in the form

$$\pi_{1,\ldots,N-1(N,N+1)}(s) = \{|\pi_{1,\ldots,N-1}\rangle \otimes |\Phi^{\pm}_{N,N+1}\rangle; |\pi_{1,\ldots,N-1}\rangle \otimes |\Phi^{\pm}_{N,N+1}\rangle\}, \quad (20)$$

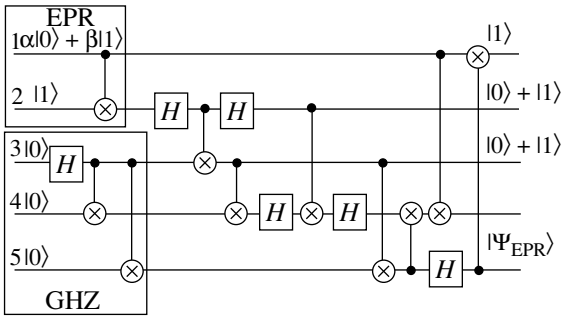

Fig. 3. Network for teleportation of an EPR pair.

where the particles of *N*, $N + 1$ are entangled, and $|\pi_{1,\ldots,N-1}\rangle$ are vectors from $H_1 \otimes \ldots H_{N-1}$. As noted above, the presence of an entangled pair is only a necessary condition for a projection basis. The value of the parameter *s*, whose value and the form of the vectors $\pi_{1,\ldots,N-1}$ will be determined below, expanding the wave function of the entire system in the basis (20), will be a sufficient condition. The expansion has the form

$$|\Psi\rangle = \{P_{N-1}\alpha|0\rangle^N \pm Q_{N-1}\beta|1\rangle^N\} \times |\pi_{1,\ldots,N-1}\rangle|\Phi^{\pm}_{N,N+1}\rangle + \{P_{N-1}\alpha|1\rangle^N \pm Q_{N-1}\beta|0\rangle^N\} \times |\pi_{1,\ldots,N-1}\rangle|\Phi^{\pm}_{N,N+1}\rangle, \quad (21)$$

where

$$P_{N-1} = \langle\pi_{1,\ldots,N-1}|0\rangle^{N-1}, \quad Q_{N-1} = \langle\pi_{1,\ldots,N-1}|1\rangle^{N-1}. \quad (22)$$

Teleportation requires that

$$P_{N-1} \neq 0, \quad Q_{N-1} \neq 0. \quad (23)$$

This means that all coefficients in the expansion must contain a linear combination of two terms $\alpha|i\rangle^N$ and $\beta|j\rangle^N$, $i \neq j = 0, 1$, which can be reconstructed from Eq. (18) by a unitary transformation that does not depend on α and β, i.e., on the parameters of the unknown state. The set of vectors

$$|\pi_{1,\ldots,N-1}\rangle = \{|\pi_1^{\pm}\rangle^{N-1}\}, \quad (24)$$

where $\pi_1^{\pm}$ is the single-particle basis determined according to Eq. (13), satisfies the conditions (23). This can be shown by noting that the set (24) consists of 2^{N-1} elements, each of which always contains a pair of terms $|i\rangle^{N-1}$, $i = 0, 1$.

For the basis found from Eqs. (20) and (24), the parameter *s* will have the value $s = 2^N$. We note that the case $s < 2^N$, where bases containing more than a pair of entangled particles (two pairs or a triplet) arise, does not solve the problem. The following assertion can serve as the final result. Teleportation of an *N*-particle entangled state of the EPR type, using maximally

entangled $N + 1$ particles for combined measurements, requires a projection basis containing one maximally entangled pair. Each element of the basis must consist of 2^N vectors corresponding to the states of $N + 1$ independent particles.

ACKNOWLEDGMENTS

We thank the Delzell Foundation for partial support of this work.

REFERENCES

1. C. H. Bennett, G. Brasard, C. Crepeau, *et al.*, Phys. Rev. Lett. **70**, 1895 (1993).
2. L. Vaidman, Phys. Rev. A **49**, 1473 (1994); S. N. Molotkov, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 248 (1998) [JETP Lett. **68**, 263 (1998)]; P. van Loock, S. L. Braunstein, and H. J. Kimble, quant-ph/9902030.
3. D. Bouwmeester, J.-W. Pan, M. Mattle, *et al.*, Nature **390**, 575 (1997); D. Boschi, S. Branca, F. De Martini, *et al.*, Phys. Rev. Lett. **80**, 1121 (1998).
4. A. Furusawa, J. L. Sorensen, S. L. Braunstein, *et al.*, Science **282**, 706 (1998).
5. A. Kusmich, K. Molmer, and E. S. Polzik, Phys. Rev. Lett. **79**, 4782 (1997); J. L. Sorensen, J. Hald, and E. S. Polzik, Phys. Rev. Lett. **80**, 3487 (1998).
6. C. S. Maierle, D. A. Lidar, and R. A. Harris, quant-ph/9807020.
7. A. S. Parkins and H. J. Kimble, quant-ph/9904054.
8. G. Brassard, quant-ph/9605035.
9. C. A. Adami and N. J. Cerf, quant-ph/9806048.
10. D. Bouwmeester, J. Pan, M. Daniell, *et al.*, quant-ph/9810035; R. J. Nelson, D. G. Cory, and S. Lloyd, quant-ph/9905028.
11. A. Karlson and M. Bourennane, Phys. Rev. A **58**, 4394 (1998).
12. S. L. Braunstein, Phys. Rev. A **53**, 1900 (1996).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 899–904.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1041–1047.

ATOMS, SPECTRA, RADIATION

Tunneling Ionization of Atoms with Excitation of the Core

B. A. Zon
Voronezh State University, Voronezh, 394693 Russia
e-mail: zon@niif.vsu.ru
Received November 4, 1999

Abstract—A general formula is obtained for the probability of tunneling ionization of an atom accompanied by excitation of the core. This formula is a generalization of the Carlson formula for the probability of a single-photon two-electron transition in atoms. The limiting case of this formula, just as that of the Carlson formula, is the well-known random-perturbations approximation. Numerical results are presented for Zn, Sr, and Cd atoms. For these atoms the contribution of the excited states of singly charged ions to the probability of the formation of doubly charged ions is a nonmonotonic function of the laser radiation intensity. Analysis of the tunneling ionization of molecules shows that with overwhelming probability an ion is formed in the ground vibrational state, while for the standard photoionization the distribution over vibrational states is determined by the Franck–Condon factors. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The formation of multiply charged ions of atoms in a strong laser field has been actively studied in the last few years both theoretically and experimentally [1, 2]. Nonetheless, the effect of the excited states of atoms and ions on the ionization probability has still not been adequately studied. The excitation of the atomic core in multiphoton ionization of the Sr atom was observed in [3, 4]. In [5] the concept of shaking ionization of a "core" electron when an "optical" electron is detached from the atom was used to describe the formation of He^{2+} ions in the tunneling regime by 614 nm laser radiation. Experimental data obtained in [6] for single-photon dielectronic ionization of the He atom were used to estimate the probability of the process.

In the present paper the role of electronic excitations of the core in the formation of a multiply charged atomic ion as a result of the tunneling detachment of several electrons from an atom is studied. Generally speaking, tunneling ionization cannot be treated as a "fast" process, and other approaches differing from the theory of instantaneous perturbations must be used. The corresponding general formula is derived in Section 3 using quantum-defect methods. However, it is interesting that the instantaneous-perturbations approximation follows from it as a limiting case with a clear physical content.

For molecules the excited states of the core can appear as vibrationally excited levels of the molecular ion formed. It is important to know the occupation probability of these levels in order to understand the subsequent behavior of the ion. For the standard photoeffect the distribution over the vibrational levels of the ion is described by the Franck–Condon factors [7], if, of course, the photon energy is much greater than the ionization potential. For the tunneling effect, however, this distribution has nothing in common with the Franck–Condon factors and is described using the indicated general formula, which is derived for atoms but which is also applicable for describing the excitation of the core in the tunneling ionization of molecules. We note that [8] is devoted to the question of the change in the Franck–Condon factors in a strong field, when the tunneling ionization regime is possible.

2. QUALITATIVE DESCRIPTION OF THE ROLE OF THE EXCITED STATES OF AN ATOM IN THE FORMATION OF MULTIPLY CHARGED IONS

It is clear from physical considerations that the probability of tunneling ionization of a singly charged atomic ion from an excited states $A^{+}*$ resulting in the formation of a doubly charged ion A^{2+} is greater than the probability of ionization of this singly charged ion from the ground state A^{+}:

$$W(A^{+}* \longrightarrow A^{2+} + e) > W(A^{+} \longrightarrow A^{2+} + e). \qquad (1)$$

However, the probability of ionization of a neutral atom A with excitation of a singly charged ion is less than the ionization probability of a neutral atom, when the singly charged ion remains in the ground state:

$$W(A \longrightarrow A^{+}* + e) < W(A \longrightarrow A^{+} + e). \qquad (2)$$

Indeed, when the core is excited the energy of the tunneling electron in the initial state decreases, and the electron drops even farther below the barrier, decreasing the tunneling probability.

1063-7761/00/9105-0899$20.00

Thus, if A^{2+} ions are produced as a result of two-cascade reactions

$$A \begin{array}{c} \nearrow A^{+} + e \\ \searrow A^{+*} + e \end{array} \begin{array}{c} \searrow \\ \nearrow \end{array} A^{2+} + 2e, \qquad (3)$$

then it follows from Eqs. (1) and (2) that the probability of the first cascade in the upper reaction is greater than the probability of the first cascade in the lower reaction, and the reverse is true for the probabilities of the second cascades in these reactions. Thus, from general considerations it cannot be concluded which of the two reactions (3) is most likely to occur—concrete calculations are required.

The probability of the reaction (3) proceeding via the upper cascade can be calculated using the standard formulas of the theory of tunneling ionization [9–11], known in the modern literature as the Ammosov–Delone–Krainov (ADK) theory. In the next section a general formula, from which the probability of the process of shaking excitation follows as a limiting case, is developed for calculating the probability of the reactions (3) proceeding through the lower cascade. The same limiting case also follows from the Carlson formula [12], describing the standard single-photon ionization of an atom with excitation of the core or the formation of a doubly charged ion. Since the form of the Carlson formula and the formula obtained in the present paper is very similar to that of the formulas from the instantaneous-perturbations theory, in Section 4 a detailed discussion is given of the problem, differing somewhat from that given in textbooks and a number of review works. Although this question is in a certain sense methodological, its discussion here is entirely relevant, since the tunneling process is usually considered to be adiabatic rather than fast. Consequently, the assumption that tunneling in certain limiting cases should be regarded as a fast process merits a separate investigation. In Section 4 it is shown that, specifically, the concept of relative rapidity and slowness, which are used in the theory of instantaneous perturbations, are not always precisely defined and must be replaced by energy relations, which are identical to the conventional concepts of the rapidity and slowness in cases where these relations can be obtained, for example, from semiclassical considerations.

The general results obtained are illustrated in the following sections by numerical examples for a number of atoms with two electrons in the outer shell and for the H_2 molecule. The atomic system of units is used throughout.

3. BASIC FORMULAS

To calculate the ionization probability of an atom we shall employ the algorithm proposed for this purpose by Keldysh [13], since in this approach the tunneling effect can be described on the basis of the *S* matrix formalism. For simplicity, we shall confine our attention to atoms with two *s* electrons in the outer shell.

The exact amplitude of a quantum transition is determined, as is well-known [14], by the expression

$$M_{if} = \langle \Psi_f(\mathbf{r}_1, \mathbf{r}_2, t)|V(\mathbf{r}_1, \mathbf{r}_2, t)|\Phi_i(\mathbf{r}_1, \mathbf{r}_2, t)\rangle. \qquad (4)$$

Here

$$V(\mathbf{r}_1, \mathbf{r}_2, t) = -(\mathbf{r}_1 + \mathbf{r}_2)\cdot\mathbf{F}\cos(\omega t) \qquad (5)$$

is the interaction of a two-electron atom an external field in the dipole approximation, ω and $\mathbf{F}$ are the frequency and amplitude of the electric field of the laser wave, Φ_i is the wave function of a free atom in the initial state before the field is switched on, and Ψ_f is the exact solution of the Schrödinger equation for two electrons, corresponding to the final state of the system, in the field of an atomic core and an ac external field. Of course, the explicit expression for Ψ_f is unknown, but to solve our problem the fact that for a periodic external field the time-dependence of this function is of a quasienergy character is sufficient.

Since in the ground state of a two-electron atom the electron spins are oppositely directed, the coordinate part of the wave function is symmetric. Assuming in the two-electron wave function to e factorable, we write

$$\begin{aligned} \Phi_i(\mathbf{r}_1, \mathbf{r}_2, t) &= \phi_0(\mathbf{r}_1)\phi_0(\mathbf{r}_2)\exp(-iE_0 t), \\ \Psi_f(\mathbf{r}_1, \mathbf{r}_2, t) &= \frac{1}{\sqrt{2}} \\ \times\left[\psi_j(\mathbf{r}_1)\sum_n g_n(\mathbf{r}_2)\exp(-in\omega t) + \{1 &\longleftrightarrow 2\}\right] \\ &\times \exp[-i(\epsilon_j + \varepsilon)t]. \end{aligned} \qquad (6)$$

Here ϕ_0 are the wave functions of the electrons in the initial state and $E_0 = -(U_1 + U_2)$, is their total energy in this state, $U_{1,2}$ are the first and second ionization potentials of the atom, ψ_j are is the wave function of the *j*th state of the atomic core corresponding to energy ϵ_j. If the excitation energy of the atomic residue Δ_j, $0 \le \Delta_j < U_2$, is introduced, then $\epsilon_j = -U_2 + \Delta_j$. Generally speaking, the functions ψ_j and ϕ_0 are not orthogonal. In Eq. (6) g_n is the *n*th quasienergy harmonic of a free electron, ε is the desired quasienergy of this electron determined in the standard manner: $\varepsilon \longrightarrow -U_1$ for adiabatically slowly switching on or off of the ac field, if the atomic core is assumed to be "frozen." We neglect the effect of an external field on the state of an ion, i.e., the quasienergy structure of the function ψ_j.

Substituting Eqs. (5) and (6) into Eq. (4) and performing the time integration we find

$$M_{if} = \sqrt{2}\langle\psi_j|\phi_0\rangle m_{if}, \qquad (7)$$

$$m_{if} = -\pi\sum_n \langle g_n|\mathbf{r}\cdot\mathbf{F}|\phi_0\rangle\delta[\epsilon_j + \varepsilon + (n+1)\omega - E_0]. \qquad (8)$$

From Eq. (8) follows the value of the quasienergy:

$$\varepsilon \equiv \varepsilon_j = E_0 - \epsilon_j = -U_1 - \Delta_j. \quad (9)$$

The quantity m_{if} is the amplitude for the ionization of an atom in the Keldysh model. As is well-known, in this model the Coulomb interaction of the exiting electron with the atomic core can be taken into account only in the tunneling limit [10, 11]. We recall that the tunneling effect in a laser field is possible if the electron can tunnel through the potential barrier in a time equal to the half-period of the field. This condition is formulated in the form of the smallness of the Keldysh parameter:

$$\gamma \equiv \frac{Z\omega}{F\nu} \ll 1, \quad (10)$$

where Z is the charge of the ion core and ν is the effective principal quantum number of the tunneling electron. According to (9),

$$\nu \equiv \nu_j = \frac{Z}{\sqrt{2(U_1 + \Delta_j)}}. \quad (11)$$

Thus the probability of tunneling ionization of an atom with excitation of the core in a state j is

$$W_j = 2|\langle\psi_j|\phi_0\rangle|^2 w_t(\nu_j, \mathbf{F}). \quad (12)$$

Here w_t is the probability of the tunneling effect, which depends only on the effective principal quantum number of the electron and the electric field amplitude of the wave. The coefficient 2 in Eq. (12) appeared from Eq. (7) and is associated with the equivalence of the electrons and the possibility that any electron can tunnel. A formula similar to Eq. (12) for two-electron single-photon ionization of an atom and for single-electron ionization with excitation of the core was obtained by Carlson in [12] (a detailed discussion of this formula is given in [15]). However, the case of single-photon ionization is simpler, since the energy absorbed by the atom is known and it is known how the energy is distributed between the electrons. In the tunneling effect, however, the question of the amount of energy absorbed by the atom is not so trivial. In the next section it is shown that the energy relations in this problem can be taken as the basic relations.

4. SHAKING APPROXIMATION

The presence of the overlap integral of the core wave functions in Eq. (12) is reminiscent of a formula of the theory of sudden perturbations [16–18]. It is well-known [15] that the appearance of such an overlap integral is due to fact that the interaction of the subsystem only in the initial state is taken into account.[1] If this interaction is completely neglected, the wave functions ψ_j and ϕ_0 become mutually orthogonal. On the other hand, the possibility of a limitation due to the fact that the interaction is taken into account only in the initial state in the theory of sudden-perturbations is justified by the shortness of the interaction time in the final state, where one subsystem "flies away" rapidly. In the theory of tunneling the neglect of the interaction in the final state is justified by the large spatial separation of the tunneling and core electrons [9, 16].

At the same time, Eq. (12), just as the Carlson formula, contains a fundamental difference from the formulas of the theory of sudden perturbations: the ionization probability depends on the state in which the core is formed. It follows from Eq. (11) that the tunneling probability no longer depends on the final state of the core, when the first ionization potential of the atom is much greater than the excitation energy of the core:

$$\Delta_j \ll U_1. \quad (13)$$

When the condition (13) is satisfied Eq. (12) is identical to the result of the theory of sudden perturbations, and this circumstance requires a special discussion.

It is easy to see that the idea of fast and slow subsystems, which are ordinarily used in the theory of sudden perturbations, is actually contained in inequalities similar to Eq. (13), which is a particular case of them. Indeed, if the change in the Hamiltonian of the slow subsystem as result of any quantum processes in the fast subsystem occurred in a time τ, then according to the uncertainty relation the energy of the fast subsystem is $E \sim \hbar/\tau$. If the characteristic time of the motion of the slow subsystem is T, then it's energy is $\varepsilon \sim \hbar/T$, and the condition $\tau \ll T$ is analogous to the condition (13): $\varepsilon \ll E$.

The difficulties of a priori substantiation of the applicability of the theory of sudden perturbations to a specific physical problem are due precisely to estimation of the times T and τ because in quantum mechanics there is no definite concept of a "quantum transition time." For example, the decay time of a quantum system cannot, of course, be taken for the time τ: according to a witty remark by the authors of [18], the lifetime of a β-active nucleus can be hundreds of years, but the shaking theory describes very well the ionization of the atom during β decay.

These difficulties are most acute in applications of the theory of sudden perturbations to equivalent atomic electrons belonging to the same shell. An example of such a problem is examined in the present paper, and a large number of similar problems is presented in [18]. Their solutions obtained in [18] by the distorted-wave method also have their own limiting cases of the formula from the shaking theory. Analysis shows that in all cases the shaking approximation is applicable when relations of the type (13) are satisfied.

Thus, we arrive at the conclusion that purely quantum energy relations are more suitable for establishing the applicability of the shaking approximation then considerations based on the notions of relative rapidity

[1] A less important approximation is due to the possibility of factorizing the wave functions of the subsystems in the initial state. The accuracy of this approximation for the two-electron photoeffect in He is analyzed in [19].

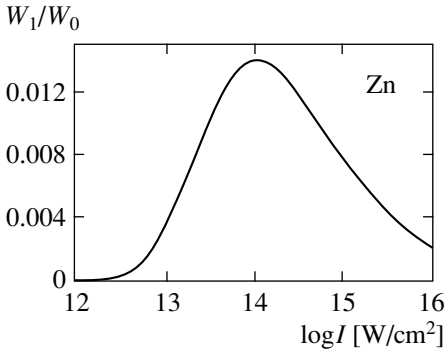

Fig. 1. Laser radiation intensity dependence of the ratio of the probability of cascade two-electron ionization of the Zn atom via an excited state of Zn^+ to the same ionization probability via the ground state of Zn^+.

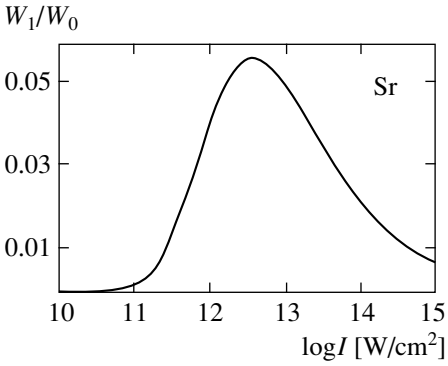

Fig. 2. Same as in Fig. 1, but for the Sr atom.

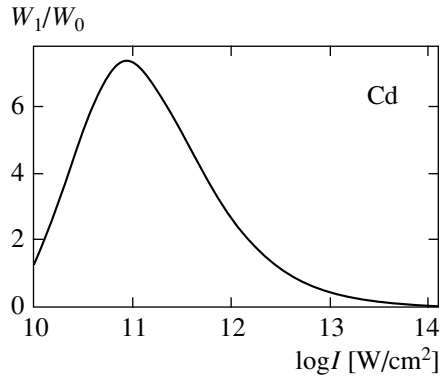

Fig. 3. Same as in Fig. 1 but for the Cd atom.

and slowness. Specifically, the subsystem where a quantum transition occurs with a change in energy greater than the change in the energy of another subsystem must be taken as the "fast" subsystem from the standpoint of the theory of sudden perturbations and, correspondingly, the second subsystem must be taken as "slow." It is assumed here that as result of quantum transitions the subsystems are spatially separated, so that the interaction between them in the final state can be neglect.

The tunneling effect studied above is an example of the constructiveness of this definition, aside from the examples presented in [18]. Another example concerns the already mentioned classical Migdal–Feinberg problem of the ionization of an atom during β decay, to solve which the approximation of sudden perturbations was first proposed. The Migdal–Feinberg theory correctly describes the experimental data, though the velocity of the β particle is only slightly greater than the velocity of the electron in the K orbit of the heavy atom, differing, as is well-known, from the velocity of light by a factor of $Z/137$. It is the high energy of the βparticle, as compared with the ionization potential of a K electron and not its velocity that makes it possible to use the approximation of sudden perturbations [20].

A definition of fast and slow motions, which is actually identical to the one presented above, under the conditions of applicability of the semiclassical correspondence principle was used in the well-known monograph [21] to study nonadiabatic processes in the Landau–Zener system (see Eqs. (13.46) and (13.47) in [21] and the accompanying discussion). On the basis of the correspondence principle the characteristic time of the motion in a slow subsystem is determined in [21] has $\hbar/\Delta E$, where ΔE is the energy of a quantum transition. The fact that different subfields of quantum mechanics such as the theory of sudden perturbations and the theory of adiabatic collisions work with identical definitions of rapidity and slowness is a very positive sign.

5. NUMERICAL EXAMPLES

5.1. Two-Electron Atoms

In this section we present the result of a calculation of the reactions (3) for neutral Zn, Sr, and Cd atoms with outer-shell configurations $4s^2$, $5s^2$ and $5s^2$, respectively. In Eq. (11) we must set $Z = 1$. The $5s$ level for Zn^+ and the $6s$ level for Sr^+ and Cd^+ where taken as the excited levels of the singly charged ions. The probability of the tunneling effect for ions is determined by the effective principal quantum number

$$\nu_j^{(+)} = \frac{Z}{\sqrt{2(U_2 - \Delta_j)}}, \quad Z = 2. \tag{14}$$

The ion wave functions presented in [22] were used to calculate the overlapped integrals on the basis of Eq. (12).

The ratios of the probabilities of the formation of doubly charged ions by tunneling ionization via the indicated excited levels of singly charged ions (W_1, upper reaction (3)) and via the ground states of these ions [W_0, lower reaction (3)] are presented in Figs. 1–3. As one can see, the role of the excitation of the core increases appreciably from light to heavier atoms. This means that the decrease in the probability of the first cascade in the reaction (3) for light atoms is not com-

pensated by an increase in the probability of the second cascade (3). For heavy atoms the situation is reversed. However, it should be noted that for the Cd atom the range where $W_1 > W_0$ corresponds to intensities that are too low: the probability of the tunneling effect here is negligibly small, so that this example is purely illustrative.

Analysis shows that the above-noted feature in the behavior of the probabilities $W_{0,1}$ is due to the smaller value of the excitation energy Δ_j in heavier atoms. For this reason, in atoms with a different configuration of the outer shell, differing from the configuration s^2 considered here, the behavior noted can change. For the He atom, which was studied in [5], the high excitation energy of the He^+ ions makes it possible to neglect completely the influence of the excited states of the ion on the tunneling ionization of a neutral atom.

It is also interesting that the ratio W_1/W_0 is described by curves with maxima. It follows from the ADK theory that the extremal value of the intensity of a laser field is determined in this case by the formula

$$F_{\text{extr}} = \frac{\nu_1^{-3} + (2/\nu_1^{(+)})^3 - \nu_0^{-3} - (2/\nu_0^{(+)})^3}{3(\nu_1 + \nu_1^{(+)} - \nu_0 - \nu_0^{(+)})}. \qquad (15)$$

Here ν_j and $\nu_j^{(+)}$ are determined by the formulas (11) and (14) and $\Delta_0 = 0$, while Δ_1 is equal to the energy of the excited level under study. For all three atoms shown in Figs. 1–3 the numerator and denominator in Eq. (15) are positive, so that F_{extr} corresponds to a maximum. For atomic transitions in which the numerator and denominator in Eq. (15) are negative, the value of F_{extr} will correspond to a minimum. However, if the right-hand side in Eq. (15) is negative, then the dependence of W_1/W_0 on the intensity of the field is a monotonic function that increases with the intensity if the numerator in Eq. (15) is negative or decreases with increasing intensity if the denominator in Eq. (15) is negative.

To complete the picture of the tunneling formation of doubly charged ions the simultaneous detachment (in a half-period of the field) of two electrons from an atom

$$A \longrightarrow A^{2+} + 2e \qquad (16)$$

should be calculated. The corresponding formulas are presented in [23]. However, since the reaction (16) is a single-cascade reaction, its dependence and that of the reaction (3) on the duration of the laser pulse are different. Consequently, the relative role of these reactions is not universal but rather it depends on the experimental conditions.

Unfortunately, it is impossible to compare the results obtained to the data obtained in [3, 4] for the Sr atom, since an inequality opposite to (10) was satisfied under the experimental conditions of [3, 4].

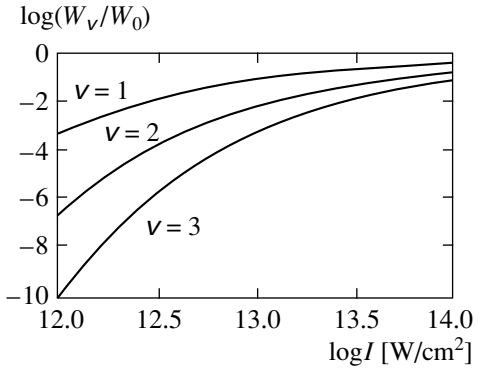

Fig. 4. Laser radiation intensity dependence of the relative excitation probabilities of the vibrational states $v = 1, 2, 3$ of the H^{2+} ion as result of tunneling ionization of the H_2 molecule.

5.2. H_2 Molecule

The formulas obtained in Section 3 are applicable not only to atoms but also to any quantum systems for which the ADK theory makes it possible to estimate the probability of a tunneling effect. Specifically, molecules are such systems. An ADK theory for electronically excited molecules has been developed in [24]. For molecules in the ground electronic state the differences from the simplest formulas of the ADK theory are greater [25]. However, even in this case it can be assumed that the ADK formulas describe qualitatively correctly the probability of the tunneling effect. This is even more convincing, if one is interested not in the absolute magnitude of the probability but rather the relative probabilities corresponding to different excited states of the molecular core.

If only the vibrational excitations of the molecular core are taken into account, Eqs. (11) and (12) can be used, setting $\nu \equiv \nu_v$, $\Delta_j \longrightarrow \Delta_v$, where v is the vibrational quantum number of the core and Δ_v is the energy of the state with this vibrational number, measured from the energy of the state with $v = 0$.

Figure 4 shows the probabilities calculated in this manner for the occupation of the states of the ion H^{2+} with quantum number v with respect to the occupation of the ground state $v = 0$, occurring as result of tunneling ionization of the H_2 molecule. The harmonic approximation with frequency 2297 cm^{-1} was used to calculate thee vibrational spectrum of H^{2+}.

As one can see, the probabilities are increasing functions of the intensity, differing very strongly in absolute magnitude. It is obvious that the distribution function of the vibrational ionic states in this case has nothing in common with the distribution determined by the Franck–Condon factors. For heavier molecules the vibrational frequency is smaller, and consequently the occupation probabilities for the ground and excited vibrational states will not differ as greatly. Nonetheless, even for heavy molecules the dependence of the tunnel-

ing probability on the energy of the ion formed plays a more important role than the difference of the Franck–Condon factors from 1.

ACKNOWLEDGMENTS

I am deeply grateful to C.P. Goreslavskiĭ, D.F. Zaretskiĭ, and V.P. Kraĭnov, whose critical remarks led to the inclusion of Section 4. I am also deeply grateful to N.B. Delone for his interest in this work and for helpful remarks. This work was supported by the Competitive Center for Fundamental Natural Science at St. Petersburg University (project 97-5.1-13).

REFERENCES

1. L. F. DiMauro and P. Agostini, Adv. At. Mol. Phys. **35**, 79 (1995).
2. N. B. Delone and V. P. Kraĭnov, Usp. Fiz. Nauk **168**, 531 (1998) [Phys. Usp. **41**, 469 (1998)].
3. P. Agostini and G. Petite, J. Phys. B **18**, L281 (1985).
4. P. Agostini and G. Petite, Phys. Rev. A **32**, 3800 (1985).
5. D. N. Fittinhoff, P. R. Bolton, B. Chang, and K. C. Kulander, Phys. Rev. Lett. **69**, 2642 (1992).
6. R. Wehlitz, F. Heiser, O. Hemmers, *et al.*, Phys. Rev. Lett. **67**, 3764 (1991).
7. L. A. Kuznetsova, N. E. Kuz'menko, Yu. A. Kuzyakov, and Yu. A. Plastinin, *Probabilities of Optical Transition of Diatomic Molecules* (Nauka, Moscow, 1980).
8. M. E. Sukharev and V. P. Kraĭnov, Zh. Éksp. Teor. Fiz. **110**, 832 (1996) [JETP **83**, 457 (1996)].
9. B. M. Smirnov and M. I. Chibisov, Zh. Éksp. Teor. Fiz. **49**, 841 (1965) [Sov. Phys. JETP **22**, 585 (1965)].
10. A. M. Perelomov, V. S. Popov, and M. V. Terent'ev, Zh. Éksp. Teor. Fiz. **50**, 1393 (1966) [Sov. Phys. JETP **23**, 924 (1966)].
11. M. V. Ammosov, N. B. Delone, and V. P. Kraĭnov, Zh. Éksp. Teor. Fiz. **91**, 2008 (1986) [Sov. Phys. JETP **64**, 1191 (1986)].
12. T. A. Carlson, Phys. Rev. **156**, 142 (1967).
13. L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1945 (1964) [Sov. Phys. JETP **20**, 1307 (1964)].
14. M. L. Goldberger and K. M. Watson, *Collision Theory* (Wiley, New York, 1964; Mir, Moscow, 1967).
15. M. Ya. Amus'ya, *The Photoelectric Effect in Atoms* (Nauka, Moscow, 1987), Chap. 8.
16. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1974; Pergamon, New York, 1977).
17. A. M. Dykhne and G. L. Yudin, Usp. Fiz. Nauk **125**, 377 (1978) [Sov. Phys. Usp. **21**, 549 (1978)].
18. V. I. Matveev and É. S. Parilis, Usp. Fiz. Nauk **138**, 573 (1982) [Sov. Phys. Usp. **25**, 881 (1982)].
19. M. I. Chibisov, Opt. Spektrosk. **38**, 236 (1975) [Opt. Spectrosc. **38**, 130 (1975)].
20. E. L. Feĭnberg, Yad. Fiz. **1**, 612 (1965) [Sov. J. Nucl. Phys. **1**, 438 (1965)].
21. N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Clarendon Press, Oxford, 1965; Mir, Moscow, 1969).
22. A. A. Radtsig and B. M. Smirnov, *Reference Data on Atoms, Molecules, and Ions* (Atomizdat, Moscow, 1980; Springer-Verlag, Berlin, 1985).
23. B. A. Zon, Zh. Éksp. Teor. Fiz. **116**, 410 (1999) [JETP **89**, 219 (1999)].
24. B. A. Zon, Zh. Éksp. Teor. Fiz. **112**, 115 (1997) [JETP **85**, 61 (1997)].
25. A. Talebpour, S. Larochelle, and S. L. Chin, J. Phys. B **31**, L49 (1998).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 905–920.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1048–1065.

ATOMS, SPECTRA, RADIATION

Simple Quantum Systems as a Source of Coherent Information

B. A. Grishanin* and V. N. Zadkov

International Laser Center, Moscow State University, Moscow, 119899 Russia

**e-mail: grishan@comsim1.phys.msu.su*

Received January 11, 2000

Abstract—A set of very important simple quantum systems is analyzed from the standpoint of the amount of coherent information that is accessible when information channels corresponding to the systems are used. It is shown that for simple quantum models the coherent information can be calculated and used for estimating the potential possibilities of the corresponding quantum channel as a source of physical information in experiments associated with the effects of the coherence of quantum states. The following physical models are studied: a two-level atom in a laser radiation field, an aggregate of two two-level subsystems in a multilevel atom (hydrogen), a system of two two-level atoms in the process of joint quantum-deterministic evolution and under the action of transformations of quantum measurement and quantum duplication, as well as one and two two-level atoms in the process of emission. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Finding a completely quantum analog of Shannon's quantitative measure of information [1] that would satisfy the corresponding quantum coding theorem, i.e., guarantee transmission along a quantum channel with a fixed information capacity irrespective of the physical nature of the channel, has for a long time remained a central unsolved problem of quantum information theory. The solution of this problem is given in [2, 3] using the concept of coherent information

$$I_c = S_{\text{out}} - S_{\text{e}}, \tag{1}$$

where S_{out} describes the quantum entropy of the output variables of the channel and S_{e} is the exchange entropy, taken from a reservoir. If the measure I_c is positive, then expressed in qubits it gives the logarithm of the dimension of the Hilbert space, all states of which can be transmitted with probability $p = 1$ in the limit $N \longrightarrow \infty$ for long ergodic ensembles. In the opposite case, when the exchange entropy is greater than the output entropy and, correspondingly, the noise introduced by the channel completely nullifies the input information, we take $I_c = 0$.

There is every reason to expect that in application to physics coherent information will play a much larger role than Shannon's information. While in classical physics the information capacity of channels, arising in the process of a physical measurement, ordinarily can also be estimated without special calculations, at least in order of magnitude, this is far from being the case in the quantum situation. Analysis of the potentially accessible quantum information in the formulation of experiments in the newest directions of physics, associated with quantum computations, problems of quantum communication and quantum cryptography [4, 5], where the measure of coherent information of the physical channel used determines the potential information content of the data obtained, is especially important. However, in order to apply the concept of coherent information to physical systems the corresponding channel in the form of a superoperator transformation $\mathscr{C}$ must be specified for each system considered and the required quantum calculations, which, as a rule, are quite nontrivial, must be performed. It is shown in the present paper that this can be done, at least, for the most important simple quantum systems studied. The analysis is performed for systems of various physical nature, including channels with qualitatively different nature of the input and output of the type of atom in the electromagnetic field of the vacuum. The classification of the types of quantum channels considered, coupling two quantum systems, is given in Fig. 1.[1] The types of two-moment channels studied, where the information is transmitted from a state at an earlier moment $t = 0$ to a state at a later moment $t > 0$, must be supplemented by the corresponding single-moment analogs, in which information at the output concerning the state of the input at the same moment in time is considered. The first class is most closely associated with the problems of quantum communication and quantum measure-

[1] The specific limitations associated with the causality principle and due to the spatial localization of the systems 1 and 2 are important only for the channels $1 \longrightarrow 2$ and $1 \longrightarrow (1 + 2)$. The analysis performed below of a system of two atoms interacting via a radiation field requires that the relativistic retardation of the signal be taken into account in order to give a correct description of the dynamics at short times.

1063-7761/00/9105-0905$20.00

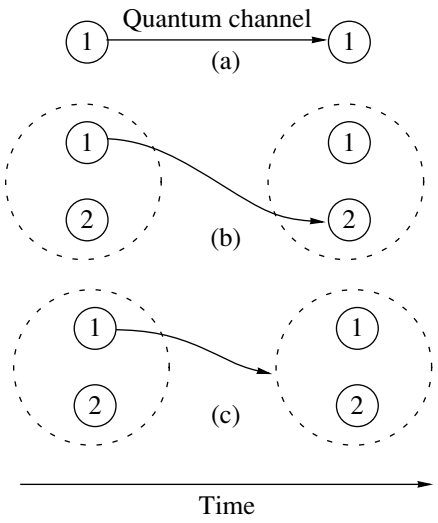

Fig. 1. Classification of possible quantum channels coupling two quantum systems: (a) 1 ⟶ 1—information is transferred from the initial state of a system to its final state; (b) 1 ⟶ 2—information is transferred from the subsystem 1 of the system (1 + 2) to the subsystem 2; (c) 1 ⟶ (1 + 2)—information is transferred from the subsystem 1 of the system (1 + 2) to the entire system (1 + 2).

ments, and the last class is associated with modern approaches to problems of quantum computations and quantum teleportation.

This paper is organized as follows. A description of the physical content of coherent information and the corresponding basic relations is given in Section 2. Section 3 is devoted to a description of the basic definitions and the technique of superoperator representations. The set of physical systems and the corresponding quantum channels is discussed in the next sections according to the classification presented in Fig. 1. The exchange of coherent information between the quantum states of a two-level atom (TLA) in a resonant laser field in two different moments in time (Fig. 1a) is discussed in Section 4. The same type of channel (1 ⟶ 1) is analyzed in Section 5 for a multilevel system, consisting of two systems of sublevels, for the example of the hydrogen atom. Section 6 examines the exchange of coherent information between two different quantum systems. It includes exchange between (1) two TLA coupled by a unitary transformation (Fig. 1b), (2) two TLA coupled via the procedure of quantum measurement (Fig. 1b), (3) an arbitrary system and its duplicated formed as a result of the quantum duplication procedure (Fig. 1c), (4) TLA and the field of the electromagnetic vacuum (Fig. 1b), and (5) two TLA coupled via a photon field of the vacuum (Fig. 1b). The basic results of this work are summarized in the conclusions.

2. QUALITATIVE MEANING OF COHERENT INFORMATION AND ITS RELATION WITH SHANNON'S INFORMATION

The classical measure of Shannon's information with error-free transmission of all possible values of a quantity x, which assumes M values, is given by $I = \log_2 M$, which for the given choice of the base of the logarithm it is conventionally assigned a "bit" as the unit of measurement. If the transmitted values x have different probabilities and are described by the probability distribution $P(x)$, then the definition presented is applicable not directly to x but to an ergodic sequence x_k ($k = 1, \ldots, n$) of statistically independent copies of x with the probability distribution $P(x_1) \cdot \ldots \cdot P(x_n)$. In this case, asymptotically for $n \longrightarrow \infty$, the set of sequences of nonzero probability consists of $M_n = 2^{nS(P)}$ approximately equally probable values, and one symbol corresponds to information $(\log_2 M_n)/n = S(p)$, where $S(P) = -\sum P(x)\log_2 P(x)$ is the Boltzmann entropy. This result, which, specifically, plays a fundamental role in statistical physics, gives the basis for assigning the value $I = S(P)$ to the information obtained with error-free transmission of all values of x with probability distribution $P(x)$. If errors are possible in transmission, then such a nontrivial information transmission channel is described by a conditional probability distribution $P(y|x)$ of the values of the output variable y for a fixed value of the input variable x. In this case, for long ergodic sequences the specific error-free transmitted information is described by the reciprocal Shannon information:

$$\begin{aligned} I &= S(P_x) + S(P_y) - S(P_{xy}) \\ &= S(P_y) - \sum_x S[P(y|x)]P(x). \end{aligned} \tag{2}$$

Here P_x, P_y, and P_{xy} are, respectively, the probability distributions for the input x, output y, and the pair (x, y). The first relation in Eq. (2) indicates the symmetric (reciprocal) character of Shannon's information with respect to input and output. The second relation gives the information as the difference of the entropy of the output variable y and the average value of the entropy introduced by the channel into the value of y for the transmission of a given a symbol x. The meaning of the latter relation is most transparent for a channel in which the transmitted values x are represented in transmission by nonoverlapping subsets M_x of the values of the quantity $y \in \cup M_x$, i.e., the distortions reduce to scatter of the output variable y in the regions M_x. The transmitted information is described, in this case, as the difference of the total entropy of y and the average entropy of the subsets M_x.

The initial definition of the coherent information is the relation $I_c = \log_2 \dim H$, where H is the Hilbert

space of the states of the input quantum system, all states of which are transmitted without distortions. The natural term for the unit of quantum information is the term "qubit," corresponding to a two-level quantum system with dimension $\dim H = 2$, that is used in the theory of quantum computations. The fundamentally new element of the theory is the quantum character of the transmitted information, which is described by an arbitrary coherent superposition of the basic elements. If the statistical distribution of the input states is described by the density matrix $\hat{\rho}_{in}$, then on the basis of considerations similar to those described above, with error-free transmission of quantum states $\psi \in H$ the measure of quantum information is the von Neuman entropy $I_c = S(\hat{\rho}_{in})$ where

$$S(\hat{\rho}) = -\mathrm{Tr}\hat{\rho}\log_2\hat{\rho}$$

is the direct operator generalization of the expression for Boltzmann's classical entropy. The simplest channel implementing error-free transmission of information is, for example, the dynamical quantum evolution of a closed system considered at two moments in time, $t = 0$ and $t \geq 0$.

For a quantum channel with distortions the input state is represented as a linear transformation of the input state $\hat{\rho}_{out} = \mathscr{C}\hat{\rho}_{in}$. The superoperator $\mathscr{C}$ of the channel is analogous to the conditional probability distribution $P(y|x)$, considered above, of a classical channel. The quantum generalization of the Shannon definition (2) is constructed on the basis of the second relation, in which the first term—the quantum entropy of the output—has a unique quantum generalization in the form of the corresponding von Neuman entropy. The second term, describing the entropy introduced by the channel—the so-called exchange entropy S_e—should give in the quantum case with error-free transmission, i.e., for the identity superoperator $\mathscr{C} = \mathscr{I}$, a zero quantity, and for a pure state at the input (analog of the classical deterministic state) it should be identical to the entropy at the output, which in this case is determined only by the entropy introduced by the channel. These requirements can be satisfied by considering instead of the input quantum system its expansion $H \otimes H'$, where the variables H' do not interact with the channel variables, but rather the state $\hat{\rho}_P$ in the aggregate system is pure and such that after averaging it gives the initial state $\hat{\rho}_{in}$ [2]. This procedure of replacing the initial quantum system is called "purification" of the mixed quantum state. The corresponding transformation, performed by the channel on the composite quantum system, has the form $\mathscr{C} \otimes \mathscr{I}$, where $\mathscr{I}$ corresponds to constancy of the variables of the additional system, and the resulting exchange entropy is identified with the entropy of the transformed composite system. The specific form of the purified state in $H \otimes H$, i.e., with the choice $H' = H$, is explicitly contained in the formula, obtained in [3], whence follows

$$\hat{\rho}_P = \sum_{ij}\sqrt{p_i p_j}|i\rangle\langle j| \otimes |i^*\rangle\langle j^*|, \tag{3}$$

where p_i, $|i\rangle$, and $\langle j|$ are the eigenvalues and the right/left eigenvectors of the density matrix $\hat{\rho}_{in}$, and $|i^*\rangle$ and $\langle j^*|$ denote the complex-conjugate vectors. The purified state is combined, therefore, from the input system and its "mirror image."[2] The corresponding exchange entropy has the form

$$S_e = S(\hat{\rho}_\alpha), \tag{4}$$

where

$$\hat{\rho}_\alpha = (\mathscr{C} \otimes \mathscr{I})\hat{\rho}_P. \tag{5}$$

The transformation $\mathscr{C}$ in the information channel, in general can describe the transfer of information to the output system with a different Hilbert space of states $H_{out} \neq H$.

For physical applications it is important to give an adequate physical interpretation of the density matrix (5) introduced in [3] and the density matrix, determined here, of the purified state (3), which initially appear from the above-described mathematical considerations. The expression (3) describes the combined state of the system consisting of the input and the mirror image, from which the quantum-mechanical state of the system input–output appears after transmission along the channel. In the classical theory the conditional probability $P(y|x)$ of the output with fixed input and, simultaneously, averaging with the distribution $P(x)$ over the states of the input corresponds to the state (5). The conditional distribution is represented by the superoperator $\mathscr{C}$, and averaging over the input is represented in the structure of the wave function

$$\Psi_P = \sum\sqrt{p_i}|i\rangle|i^*\rangle,$$

corresponding to the purified state (3). This two-particle state is entangled, i.e., it does not reduce to a statistical mixture of density matrices of the type $|\psi_i\rangle|\varphi_i\rangle\langle\varphi_i|\langle\psi_i|$, corresponding to pure states in the form of direct products $|\psi_i\rangle|\varphi_i\rangle$ of single-particle states. Its purely quantum fluctuations reproduce the fluctuations of a mixed nature, which are described by the density matrix $\hat{\rho}_{in}$, determined in the first space in the direct product $H \otimes H$. Therefore the density matrix (5) describes the state of the input–output system, where actually the input is replaced by the mirror-conjugate representation (see footnote 2). It determines the

[2] Compared with [3], here the complex conjugate, necessary for invariance of representation under study relative to rotations in subspaces corresponding to degenerate eigenvalues of the density matrix, is added. For real matrices $\hat{\rho}_{in}$ with a nondegenerate spectrum, this refinement is not essential.

exchange entropy in the channel and, on the basis of its physical meaning, is qualitatively different from the standard one-time density matrix, since the corresponding nonzero entropy appears only as a result of the transformation of the input state accompanying transmission along the channel. In the absence of distortions in the channel, in contrast to the standard two-particle density matrix, it always corresponds to a pure state and zero entropy.

3. BASIC DEFINITIONS AND THE SUPEROPERATOR REPRESENTATION TECHNIQUE

For purposes of the present paper, it is especially effective to use a combination of the technique of symbolic and matrix representation of superoperators [6]. The most general representation of a superoperator transformation is introduced by the symbolic expression

$$\mathcal{C} = \sum \hat{s}_{kl}\langle e_k| \odot |e_l\rangle, \tag{6}$$

where the substitution symbol $\odot$ must be replaced by an operator of the transformed physical quantity or the density matrix, while e_k describe an arbitrary vector basis in Hilbert space H where the transformed operator is defined. To describe physically realizable transformations of the density matrix $\hat{\rho}$, the operators $\hat{s}_{kl}$ must satisfy the positivity condition[3] of the block operator $\hat{S} = (\hat{s}_{kl})$ and the orthonormality condition

$$\mathrm{Tr}\hat{s}_{kl} = \delta_{kl}, \tag{7}$$

which ensures the required normalization for all normalized operators $\hat{\rho}$ with $\mathrm{Tr}\hat{\rho} = 1$.

Using the symbolic representation (6), it is possible to obtain the corresponding expression for the product of the superoperators $\mathcal{C}_1$ and $\mathcal{C}_2$, whence it is possible to give a symbolic representation of the superoperator algebra. For the case $\hat{s}_{kl} = |k\rangle\langle l|$ we obtain a representation of the identity superoperator $\mathcal{I}$, and for $\hat{s}_{kl} = |k\rangle\langle k|\delta_{kl}$ we obtain the representation of the quantum reduction superoperator

$$\mathcal{R} = \sum |k\rangle\langle k| \odot |k\rangle\langle k|.$$

The case $\hat{s}_{kl} = \delta_{kl}$ describes the superoperator of taking the trace $\mathrm{Tr}\odot$, which is a linear functional in the space of density matrices. The correspondence between the matrix form $S = (S_{mn})$ of the representation of the superoperator $\mathcal{C}$ in the orthonormal basis $\hat{e}_k$ and the representation (6) is given by the relation

$$\hat{s}_{kl} = \mathcal{C}(|k\rangle\langle l|) = \sum_{mn} S_{mn}\langle l|\hat{e}_n|k\rangle\hat{e}_m, \tag{8}$$

whose validity can be easily checked after substituting into the expression (6) and comparing with the standard definition of the matrix elements by means of the relation

$$\mathcal{C}\hat{e}_n = \sum_m S_{mn}\hat{e}_m.$$

The exchange entropy in the expression (1) for coherent information is determined by the relation (4), where the combined density matrix $\hat{\rho}_\alpha$ of the input–output variables is described in accordance with [3] and Eq. (5) by the relation

$$\hat{\rho}_\alpha = \sum_{ij} \mathcal{C}(|\rho_i\rangle\langle\rho_j|) \otimes |\rho_i^*\rangle\langle\rho_j^*|. \tag{9}$$

Here $|\rho_i\rangle = \hat{\rho}_{\mathrm{in}}^{1/4}|i\rangle$ are the transformed eigenvectors of the input density matrix

$$\hat{\rho}_{\mathrm{in}} = \sum p_i|i\rangle\langle i|,$$

$|\rho_i^*\rangle$ are the complex conjugates of $|\rho_i\rangle$, and $\mathcal{C}$ is the input–output transformation superoperator, so that $\hat{\rho}_{\mathrm{out}} = \mathcal{C}\hat{\rho}_{\mathrm{in}}$ describes the density matrix of the output variables. Using the superoperator representation in the form (6) and the above-defined eigenvectors $|i\rangle$, the density matrix (9) becomes

$$\hat{\rho}_\alpha = \sum_{ij} (p_i p_j)^{1/4}\hat{s}_{ij} \otimes |\rho_i^*\rangle\langle\rho_j^*|, \tag{10}$$

where the operators $\hat{s}_{ij}$ are the states of the output variables. Both the input and output partial density matrices can be represented as traces over the corresponding additional system: $\hat{\rho}_{\mathrm{out}} = \mathrm{Tr}_{\mathrm{in}}\hat{\rho}_\alpha$, $\hat{\rho}_{\mathrm{in}}^* = \mathrm{Tr}_{\mathrm{out}}\hat{\rho}_\alpha$.

To describe exchange of coherent information between two quantum systems via the quantum channels, shown in Figs. 1b, 1c ($1 \longrightarrow 2$ or $1 \longrightarrow (1+2)$) the initial combined density matrix must be given in the form of a direct product $\hat{\rho}_{1+2} = \hat{\rho}_{\mathrm{in}} \otimes \hat{\rho}_2$, where $\hat{\rho}_{in} = \hat{\rho}_1$ and $\hat{\rho}_2$ describes the initial partial density matrices, where the first one describes the input and the second describes the output channel. For a channel of the type $1 \longrightarrow 2$ the output are states of the second system, which contain information about the initial state of the first system, if a certain transformation over both systems is satisfied.

The temporal dynamics of the composite system $(1+2)$ is described by the superoperator $\mathcal{C}_{1+2}$, and the corresponding superoperator transformation of the channel $\hat{\rho}_{\mathrm{out}} = \mathcal{C}\hat{\rho}_{\mathrm{in}}$ can be written as

$$\mathcal{C} = \mathrm{Tr}_1\mathcal{C}_{1+2}(\odot \otimes \hat{\rho}_2),$$

where the trace is calculated over the final states of the first system. In terms of the representation (6) for the

[3] The operators $\hat{s}_{kl} \otimes \hat{1}$ must be introduced in order to check positivity completely [7].

composite system this transformation can be described as

$$\mathscr{C} = \sum_{k\kappa l\lambda}\sum_{n}\langle n|\hat{s}_{k\kappa,\,l\lambda}|n\rangle\langle\kappa|\hat{\rho}_2|\lambda\rangle\langle k|\odot|l\rangle, \qquad (11)$$

where the multiplicative basis $|k\rangle|\kappa\rangle$ is used, and the indices k and κ correspond to the first and second quantum systems. The operator coefficients $\hat{s}_{kl}$ in Eq. (6) now assume the form

$$\hat{s}_{kl} = \sum_{\kappa\lambda}\sum_{n}\langle n|\hat{s}_{k\kappa,\,l\lambda}|n\rangle\langle\kappa|\hat{\rho}_2|\lambda\rangle. \qquad (12)$$

Here $\mathscr{C}$ depends on the form of the combined dynamical transformation $\mathscr{C}_{1+2}$ and on the initial state $\hat{\rho}_2$ of the second system, and it maps the initial states of the first system into the final states of the second system.

Ordinarily, it is much easier to calculate the one-time amount of information, since the input–output density matrix is simply a one-time density matrix of the corresponding variables, which is calculated directly from the dynamical equations. For one system, the corresponding channel is described by the single superoperator $\mathscr{J}$ and the corresponding calculations are trivial: for the combined input–output density matrix (9) we obtain the pure state

$$\hat{\rho}_\alpha = \sum_i |\rho_i\rangle|\rho_i^*\rangle\sum_j\langle\rho_i^*|\langle\rho_j|,$$

and the corresponding exchange entropy $S_e = 0$ and coherent information $I_c = S_{\text{out}} = S_{\text{in}}$. For two systems, where the input–output density matrix is a combined density matrix $\hat{\rho}_{1+2}$, the corresponding coherent information in the system 2 about the system 1 at the time t is

$$I_c(t) = S[\hat{\rho}_2(t)] - S[\hat{\rho}_{1+2}(t)].$$

When the dynamics is described by a unitary transformation and the initial state of the second system is pure, all eigenstates $|i\rangle$ of the first system transform into the corresponding set of orthogonal states $\Psi_i(t)$ of the composite system (1 + 2), so that the combined entropy remains unchanged, and the coherent information becomes

$$I_c(t) = S[\hat{\rho}_2(t)] - S[\hat{\rho}_1(0)].$$

If the initial state of the first system is also pure, then we obtain simply $I_c(t) = S[\hat{\rho}_2(t)]$. For a TLA this gives $I_c = 1$ qubit, if the maximally entangled state is attained in a system of two qubits.

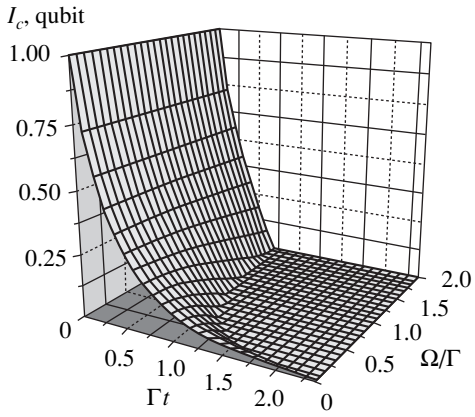

Fig. 2. Coherent information transferred from the initial state of a TLA at $t = 0$ state at the moment $t > 0$ as a function of the dimensionless time Γt and the Rabi frequency Ω/Γ.

4. TWO-LEVEL ATOM IN A RESONANT LASER FIELD

We shall consider the exchange of coherent information between the states of a TLA in a resonant laser field at two different times (Fig. 1a).

An example of a channel of this type was examined in [3], where only pure dephasing in the absence of an external field was studied. In the presence of a field and other relaxation mechanisms, the calculation of coherent information on the basis of the Markov approximation can be performed in the most general form by calculating the combined density matrix (9) using the technique of matrix representation of dynamical superoperators. One question of interest is the form of the dependence of the coherent information on the applied resonant field.

An external field changes the characteristic decay rates of the initial state of a TLA, which are described by the real parts of the eigenvalues λ_k of the dynamic Liouville operator $\mathscr{L} = \mathscr{L}_r + \mathscr{L}_E$, where the Liouville operators $\mathscr{L}_r$ and $\mathscr{L}_E$ describe relaxation and interaction with an external field. Here we confine our attention to relaxation represented only by pure dephasing in combination with the action of a laser field. The corresponding Liouville matrix in the operator basis $\hat{e}_k = \{\hat{I}, \hat{\sigma}_3, \hat{\sigma}_1, \hat{\sigma}_2\}$ has the form [8]

$$L = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Omega \\ 0 & 0 & -\Gamma & 0 \\ 0 & -\Omega & 0 & -\Gamma \end{pmatrix}, \qquad (13)$$

where Γ describes the rate of decay of the phase in the absence of the field, Ω is the Rabi frequency, and $\hat{\sigma}_1$, $\hat{\sigma}_2$, and $\hat{\sigma}_3$ are the Pauli matrices. The eigenvalues of

the matrix (13) have the form

$$\lambda_k = \{0, -\Gamma, -(\Gamma + \sqrt{\Gamma^2 - 4\Omega^2})/2, -(\Gamma - \sqrt{\Gamma^2 - 4\Omega^2})/2\}.$$

The laser field changes these quantities compared with their unperturbed values 0, $-\Gamma$, $-\Gamma$, and 0.

It is of interest to determine whether or not such a change in the decay rates results in a decrease of the decay rate of the coherent information, though from intuitive considerations it can be inferred beforehand that an information gain is possible only in the case of another effect related with the laser field—decrease of the relaxation parameters of the relaxation superoperator $\mathcal{L}_r$ itself [8–11].

Calculating the matrix of the dynamical superoperator $\mathcal{C} = \exp(\mathcal{L}t)$ and using the corresponding representation (6), we obtain an analytical expression for the combined density matrix (9) and then [using Eqs. (4) and (1)] we calculate the coherent information retained in the TLA at the time t relative to its initial state. The latter is chosen in the form of the density matrix $\hat{\rho}_0 = \hat{I}/2$ with maximum entropy $S(\hat{\rho}_{in}) = 1$ qubit. The computational results are displayed in Fig. 2. They are described by a threshold-type time dependence, typical for coherent information limited by coherence loss processes. In addition, it is clearly seen that the coherent information does not increase, and it even decreases somewhat with increasing field intensity, as described by the corresponding Rabi frequency.

The results presented demonstrate also the singularity of the first time derivative of the coherent information at time $t = 0$, which is a characteristic feature of the initial stage of its decay. Indeed, initially the input–output density matrix (9) of a TLA has the form of a pure state $\hat{\rho}_\alpha = \Psi\Psi^+$ with the input–output wave function $\Psi = \sum \sqrt{p_i}\,|i\rangle|i\rangle$. Its eigenvalues λ_k and the probabilities of the corresponding eigenvalues are all zero, except the one corresponding to Ψ. As a result of the singularity of the entropy function $\sum(-\lambda_k)\log_2\lambda_k$ at $\lambda_k = 0$, the derivative of the corresponding exchange entropy also possesses a logarithmic singularity.

Another interesting feature of the coherent information is the form of its dependence on the initial (input) state $\hat{\rho}_{in}$. If it were possible, it would make sense to choose it in the form of the characteristic Liouville operator

$$\hat{\rho}_{in} = \sum_{l=1}^{4} |k_{min}\rangle_l \hat{e}_l,$$

where $|k_{min}\rangle$ is the eigenvector corresponding to the minimum eigenvalue $|\mathrm{Re}\lambda_k| > 0$ of the matrix L. However, the vector $|k_{min}\rangle$ equals $\{0, (\Gamma + \sqrt{\Gamma^2 - 4\Omega^2})/2\Omega, 0, 1\}$, i.e., it describes an element of the linear subspace of operators with zero trace, since the first component is zero. Therefore the decay of coherent information cannot be decreased by decreasing the rate of decay of the coherence of the atomic state in a laser field.

5. EXCHANGE OF COHERENT INFORMATION BETWEEN TWO OPEN SUBSYSTEMS OF A SINGLE SYSTEM

Let us consider the quantum channel of the type $1 \longrightarrow 1$ (Fig. 1a) between two open subsystems A and B of a single closed system $\{A, B\}$ with the Hilbert space of states $H_A + H_B$, where H_A and H_B are the Hilbert subspaces of the systems A and B, respectively, and the "+" sign is used to denote a linear union.

In classical information theory this situation corresponds to transfer of only the part $A \subset X$ of the values of the input random variable $x \in X$. The realization where the detector does not obtain any message also carries information and means that x belongs to the complement of A, $x \in \bar{A}$. This situation can be described by the corresponding transformation of the choice $\mathcal{C} = P_A + P_0(1 - P_A)$, where P_A is the projection operator from X onto the subset A, $P_A x = x$ for $x \in A$ and $P_A x = \varnothing$ (empty set) for $x \in \bar{A}$, while P_0 is the projection operator from X onto an independent single-point set X_0 and $P_0 x = X_0$. This transformation corresponds to the classical reduction channel, which results in information losses, only if $\bar{A}$ does not consist of only one point. If A is only point, then it is possible to obtain a potential limit of information equal to 1 bit, because $\bar{A}$ replaces the second state of the bit, so that actually no information is lost.

In quantum mechanics the corresponding reduction channel is described by an obvious generalization of the classical selection operator—the selection superoperator

$$\mathcal{C} = \hat{P}_A \odot \hat{P}_A + |0\rangle\langle 0|\mathrm{Tr}(1 - \hat{P}_A) \odot (1 - \hat{P}_A), \quad (14)$$

where the state $|0\rangle$ is the quantum analog of the classical one-point set, which does not depend on all the other states. This transformation is positive and preserves the normalization of the density matrix, describing adequately the exchange of coherent information between open subsystems of a single system. The last term in Eq. (14) expresses conservation of normalization, provided that all states outside the set of B states are included. In our case these states are all included in the form of the projector $|0\rangle\langle 0|$, which does not take into account their coherence. In contrast to the classical one-bit case, for a TLA they do not carry any coherent information because of the complete loss of coherence.

Considering the coherent information transmitted from one part A to the part B of a system, whose state depends on time, we are dealing with a superoperator of this channel of the form

$$\mathcal{C}_{AB} = \mathcal{C}_B\mathcal{C}_0(t)\mathcal{C}_A, \quad \mathcal{C}_0(t) = U(t) \odot U^{-1}(t) \quad (15)$$

with a unitary temporal resolution operator $U(t)$ and selection superoperators $\mathcal{C}_A$ and $\mathcal{C}_B$ of the subsystems A, B. Here the selection superoperator $\mathcal{C}_A$ is presented only for the possibility of determining the complete superoperator of the channel irrespective of the form of the input density matrix. However, if the input density matrix $\hat{\rho}_{\text{in}}$ is determined only in the corresponding subspace H_A of the complete space H, this superoperator can be dropped.

Let us assume that the dynamical evolution of the system is given by a set of eigenstates $|k\rangle$ and the corresponding Bohr frequencies ω_k. Then, representing the projectors in terms of the corresponding input $|\psi_l\rangle$ and output $|\varphi_m\rangle$ wave functions, we obtain from Eq. (15) the representation of the temporal evolution specified in the form

$$\mathcal{C}_{AB}(t) = \sum_{ll' \in A}\Bigg[\hat{s}_{ll'}(t) + |0\rangle\langle 0| \times \sum_{m \notin B}\langle\varphi_m|\psi_l(t)\rangle\langle\psi_{l'}(t)|\varphi_m\rangle\Bigg]\langle\psi_l| \odot |\psi_{l'}\rangle,$$

$$\hat{s}_{ll'}(t) = \sum_{mm' \in B}\langle\varphi_m|\psi_l(t)\rangle\langle\psi_{l'}(t)|\varphi_{m'}\rangle|\varphi_m\rangle\langle\varphi_{m'}|, \quad (16)$$

$$|\psi_l(t)\rangle = \sum_k \exp(-i\omega_k t)\langle k|\psi_l\rangle|k\rangle.$$

Let us consider the case of an orthogonal choice of subsets of input/output wave functions, which is of special interest. Then, if there is only one common state $|\phi\rangle$ in the sets $|\psi_l\rangle$, $|\varphi_m\rangle$ and $U(t_0) = 1$ for a time t_0, we obtain the expression

$$\mathcal{C}_{AB}(t_0) = |\phi\rangle\langle\phi| \odot |\phi\rangle\langle\phi| + |0\rangle\langle 0| \sum_{\varphi_m \neq \phi}\langle\varphi_m| \odot |\varphi_m\rangle,$$

which means that the input system reduces to a classical bit of information, associated with the states $|\phi\rangle$ and $|0\rangle$, and no coherent information is transmitted into the system B. Nonetheless, it appears in the process of temporal evolution, provided that the eigenstates $|k\rangle$ of the operator $U(t)$ are different from the input/output states $|\psi_l\rangle$, $|\varphi_m\rangle$. Therefore, the information capacity of the channel is due to the quantum entanglement of the input and output on account of the corresponding contribution to the Hamiltonian systems.

To illustrate the exchange of coherent information in the channel of the type described, we shall consider

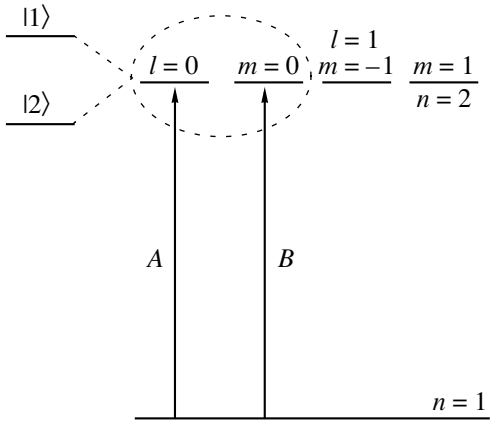

Fig. 3. Spinless model of the hydrogen atom. The information channel is formed from the input forbidden ($nlm \longrightarrow n'l'm'$) transition 100–200 and the output dipole-active transition 100–210.

the typical intraatom channel formed by two two-level systems constructed from two pairs of orthogonal states $A = \{|\psi_0\rangle, |\psi_1\rangle\}$ and $B = \{|\psi_0\rangle, |\psi_2\rangle\}$ of the same atom. As an example we shall use the spinless model of the hydrogen atom (Fig. 3): ψ_0 is the ground s state with $n = 1$, $\psi_{1,2}$ correspond to the s state with $l = 0$, $m = 0$ and the p state with $l = 1$, $m = 0$ of the first excited state $n = 2$.

In the absence of an external field the quantum channel does not transmit any coherent information, since the states $l = 0$, $m = 0$ and $l = 1$, $m = 0$ are not coupled. In the absence of an external electric field applied along the Z axis, the desired pair of four initially degenerate states with $n = 2$ splits as a result of the Stark effect and transforms into a pair of new eigenstates

$$|1\rangle = (|\psi_1\rangle + |\psi_2\rangle)/\sqrt{2}, \quad |2\rangle = (|\psi_1\rangle - |\psi_2\rangle)/\sqrt{2}$$

and the input state $l = 0$ oscillates with the frequency of the Stark shift:

$$|\psi_1(t)\rangle = \cos(\omega_s t)|\psi_1\rangle + \sin(\omega_s t)|\psi_2\rangle.$$

Therefore, on account of the applied electric field, the input states become entangled with the output states, which as a result contain coherent information about the input states.

In our model, Eq. (16) gives the operators $\hat{s}_{kl}$ in the form of a 3×3 matrices, where the third columns and rows correspond to a fictitious "vacuum" state $|0\rangle$:

$$\hat{s}_{11} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \hat{s}_{12} = \begin{pmatrix} 0 & \sin(\omega_s t) & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\hat{s}_{21} = \begin{pmatrix} 0 & 0 & 0 \\ \sin\omega_s t & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\hat{s}_{22} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sin^2(\omega_s t) & 0 \\ 0 & 0 & \cos^2(\omega_s t) \end{pmatrix}.$$

Zero values of the operators $\hat{s}_{12}$ and $\hat{s}_{21}$ correspond to the absence of coherent information at $t = 0$, i.e., the absence of entangled states. Choosing the input density matrix in the form $\hat{\rho}_{in} = \hat{I}/2$, we obtain the corresponding input–output density matrix:

$$\hat{\rho}_\alpha = \begin{pmatrix} \frac{1}{2} & 0 & 0 & \frac{x}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{x}{2} & 0 & 0 & \frac{x^2}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1-x^2}{2} \end{pmatrix},$$

where $x = \sin(\omega_s t)$ and the output density matrix $\hat{\rho}_{out}$ is diagonal with diagonal elements 1/2, $x^2/2$, and $(1 - x^2)/2$.

Calculating the nonzero eigenvalues $(1 \pm x^2)/2$ for $\hat{\rho}_\alpha$ and the entropies S_{out} and S_α, we obtain the coherent information

$$I_c = [(1 + x^2)\log_2(1 + x^2) - x^2\log_2(x^2)]/2.$$

This function is greater than zero everywhere with the exception of the point $x = 0$, where the coherent information is zero, and its maximum is one qubit with $x = \pm 1$, i.e., for the precession angle $\omega_s t = \pm\pi/2$. Hence it is evident that the coherent information about the states of the forbidden dipole transition is in principle accessible via the dipole transition using the Stark effect. Its average value in time is $\langle I_c \rangle = 0.46$ qubits.

The estimates made above indicate the potential possibility of observing experimentally the coherent effects due to the influence of the forbidden electronic transition on a dipole transition. Forbidden transitions were studied in [12, 13] as a potential source of information about the breakdown of spatial symmetry due to an interaction via weak neutral currents [14, 15]. If $I_c = 0$, then in principle only an incoherent effect of a forbidden transition via the population of the ground state n_0 is possible. In this case only one parameter—the population—can be measured, while the exact knowledge of the phase requires $I_c = 1$.

6. EXCHANGE OF COHERENT INFORMATION BETWEEN TWO CLOSED QUANTUM SYSTEMS

A variety of results associated with the exchange of coherent information between two atomic qubits, including analysis of the problem from the standpoint of the measure of quantum entanglements [16], analysis of the problem of eavesdropping [17], and a set of various experiments proposed in order to create a controllable entanglement in a system of two atoms [18, 19], has been published in the last few years. From the information standpoint coherent information exchanged in a system of two TLA coupled by a quantum channel depends on the specific form of the transformation realized by the quantum channel as well as on the initial states of the TLA. It is natural to take as the initial state the product of independent density matrices of the atoms: $\hat{\rho}_{1+2} = \hat{\rho}_{in} \otimes \hat{\rho}_2$.

In this section we give a systematic analysis of the processes leading to exchange of coherent information between two initially independent quantum systems. The following are included: (1) two unitarily coupled TLA (Section 6.1), (2) two TLA coupled by the quantum measurement procedure (Section 6.2), (3) an arbitrary system and its duplicated state (Section 6.3), (4) TLA and a photon field in free space (Section 6.4), and (5) two TLA interacting via the vacuum photon field (Section 6.5).

6.1. Two Unitarily Coupled Two-Level Atoms

We discuss first a noiseless deterministic quantum channel coupling two TLA (Fig. 1b). It can be described by a unitary two-atom transformation, given by the matrix elements $U_{ki,k'i'}$, $k, i, k', i' = 1, 2$. Then the superoperator of the transformation of channel $\mathcal{C}$, giving the mapping $\hat{\rho}_{in} \longrightarrow \hat{\rho}_{out} = \hat{\rho}_2'$, can be written in terms of the substitution symbol using the relation (6) with the operators

$$\hat{s}_{kl} = \sum_{\mu\nu} S_{kl,\mu\nu}|\mu\rangle\langle\nu|,$$

represented in accordance with Eqs. (8) and (12), by the matrix elements $\mathcal{C}$ in the form

$$S_{kl,\mu\nu} = \sum_{m\alpha\beta} \rho_{2\alpha\beta} U_{m\mu,k\alpha} U^*_{m\nu,l\beta}. \qquad (17)$$

The relation

$$\mathrm{Tr}\hat{s}_{kl} = \sum_{\mu} S_{kl,\mu\mu} = \delta_{kl}$$

holds and gives the correct normalization, and the positiveness of the block matrix

$$(\hat{s}_{kl}) = \begin{pmatrix} \hat{s}_{11} & \hat{s}_{12} \\ \hat{s}_{21} & \hat{s}_{22} \end{pmatrix}$$

guarantees positiveness of $\mathcal{C}$.

For a transformation of the form $U = U_1 \otimes U_2$, which does not lead to the creation of entangled states, Eqs. (6) and (17) give $\mathcal{C} = \hat{\rho}_2' \mathrm{Tr}\odot$, which signifies a transformation of the initial state $\hat{\rho}_1'$ of the first TLA into the final state, which is not entangled with the state $\hat{\rho}_2' = U_2\hat{\rho}_2 U_2^+$ of the second TLA.

Relation (17) can be simplified by considering pure states $\hat{\rho}_2'$, so that in combination with the possibility of choosing an arbitrary transformation U without entanglement it is useful, specifically, to single out especially the case of the state $\rho_{2\alpha\beta} = \delta_{\alpha\beta}\delta_{\alpha\alpha_0}$. Taking account of the linearity of the dependence of $S_{kl,\,\mu\nu}$ on $\hat{\rho}_2'$ and the convexity of the coherent information I_c as a function of $\mathcal{C}$ [20], the analysis of Eq. (17) can be reduced to analysis of the relation

$$S_{kl,\,\mu\nu} = \sum_m U_{m\mu,\,k\alpha_0} U^*_{m\nu,\,l\alpha_0}, \qquad (18)$$

which means that the quantum channel is described only by a unitary transformation U. Here the summation extends only over the states $|m\rangle$ of the first atom after the entangling transformation.

The coherent information transmitted, in the present system of two coupled TLA with

$$\hat{\rho}_{\mathrm{in}} = \hat{I}/2, \quad (\hat{\rho}_2)_{12} = \sqrt{(\hat{\rho}_2)_{11}[1-(\hat{\rho}_2)_{11}]}$$

is shown in Fig. 4. It is a convex function of $\hat{\rho}_2$ with a maximum at the boundary, $\rho_{11} = 0.1$. Just as in the case of one TLA, the coherent information preserves the typical threshold character of the dependence on the coupling angle, which describes the degree of coherent coupling of two TLA with respect to independent fluctuations of the second TLA.

6.2. Two Two-Level Atoms Coupled by the Quantum Measurement Procedure

We shall consider a special type of quantum channel coupling two TLA,[4] which can be described by a superoperator $\mathcal{C}$, defining the quantum measurement procedure. This procedure is related with a different approach to defining the quantum information [21], based on the so-called measured information.

Let us consider first a channel consisting of two identical two-level systems. In terms of the wave function the corresponding transformation of the complete measurement of the state of the first TLA has the form

$$\psi \otimes \varphi \longrightarrow \sum a_i|\phi_i\rangle|\phi_i\rangle, \quad a_i = \langle\phi_i|\psi\rangle. \qquad (19)$$

[4] In reality, the results of the present section are valid not only for two TLA but also for any quantum systems with finite dimension.

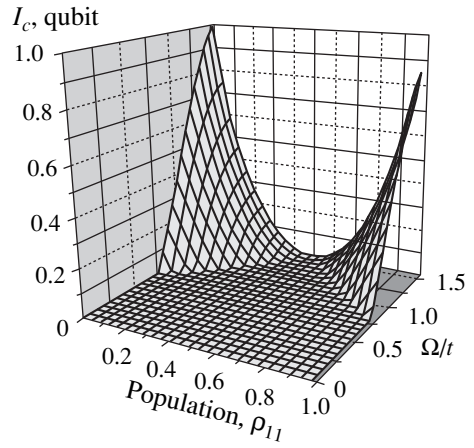

Fig. 4. Coherent information transmitted between two TLA coupled by a unitary transformation, as a function of the matrix element ρ_{11} of the diagonal initial density matrix of the second TLA and the procession angle $\varphi = \Omega t$.

This transformation describes the entanglement of certain basis states $|\phi_i\rangle$, which does not depend on the initial state φ of the second TLA. The latter is an indicator of the measuring setup, preserving completeness of information on the basis states in the initial state $\psi = \sum a_i|\phi_i\rangle$. Relation (19), given in the form of a single-valued transformation of the wave function, in reality is not a linear transformation with respect to φ and therefore cannot represent a deterministic transformation, not being unitary. The corresponding representation in terms of the two-atom density matrices has the form

$$\hat{\rho}_{12} \longrightarrow \sum_i \sum_j \langle\phi_i|\langle\phi_j|\hat{\rho}_{12}|\phi_j\rangle|\phi_i\rangle|\phi_i\rangle\langle\phi_i|\langle\phi_i|. \qquad (20)$$

It is linear with respect to $\hat{\rho}_{12}$ and satisfies the standard conditions of physical realizability [7, 20], i.e., completely positive and preserves normalization. The density matrix has the form $\sum p_i|\phi_i\rangle|\phi_i\rangle\langle\phi_i|\langle\phi_i|$, so that $S(\hat{\rho}_{12}) = S(\hat{\rho}_2)$, and in accordance with the relations of Section 3 the one-time coherent information is zero. This is due to the classical nature of the information, represented here only by the classical indices i.

The quantum transformation superoperator coupling two TLA in the case of a two-time channel can be obtained from the relation (6) with $\hat{s}_{kl} = |\phi_k\rangle\langle\phi_k|\delta_{kl}$, $\langle k| \longrightarrow \langle\phi_k|$ and $|k\rangle \longrightarrow |\phi_k\rangle$, which after taking the trace with respect to the first TLA and replacing $\hat{\rho}_{12}$ by the substitution symbol $\odot$ becomes

$$\mathcal{M} = \sum_k \hat{P}_k \mathrm{Tr}_1 \hat{E}_k \odot, \qquad (21)$$

where $\hat{P}_k = |\phi_k\rangle\langle\phi_k|$ are orthogonal projectors, representing the eigenstates of the "indicator" variable of the

second TLA, and $\hat{E}_k = |\phi_k\rangle\langle\phi_k|$ describes the orthogonal expansion of unity. It is constructed from the same operators, giving here the transformation of the quantum-classical reduction $\mathrm{Tr}_1 \hat{E}_k \odot = \langle\phi_k| \odot |\phi_k\rangle$, which gives the procedure for obtaining the classical information k from the first system. Applying the transformation (21) to $\hat{\rho}_{\mathrm{in}}$ and using Eq. (9), we obtain for the corresponding output and input–output density matrices

$$\hat{\rho}_{\mathrm{out}} = \sum_k \tilde{p}_k |\phi_k\rangle\langle\phi_k|,$$
$$\hat{\rho}_{\alpha} = \sum_k \tilde{p}_k |\pi_k\rangle|\phi_k\rangle\langle\phi_k|\langle\pi_k|, \tag{22}$$

where

$$\tilde{p}_k = \langle\phi_k|\hat{\rho}_{in}|\phi_k\rangle = \sum_i p_i |\langle\phi_k|i\rangle|^2$$

are the eigenvalues of the probabilities for the reduced density matrix, and

$$|\pi_k\rangle = \sum_i \sqrt{p_i/\tilde{p}_k}\langle\phi_k|i\rangle|i\rangle$$

are the normalized modified input states, coupled after the quantum measurement procedure with the output states $|\phi_k\rangle$. It should be noted especially that, as follows from Eqs. (22), there is no exchange of coherent information in the system, since the vectors $|\phi_k\rangle$ are orthogonal and the entropies of the density matrices (22) are obviously the same. Conversely, the measured information, introduced in [21], in this case is different from zero.

There is no difficulty in extending this result to the case of a channel of a more general form, where the output system has a structure that is different from the initial system and is described by a different Hilbert space. The latter difference is taken into account by replacing in the preceding relations the basal states $|\phi_i\rangle$ of the second system by a different orthogonal set $|\varphi_i\rangle = V|\phi_i\rangle$, where V is the isometric transformation from the Hilbert space of the states H_1 of the first system into a different Hilbert space H_2 of the second system. After simple, obvious transformations, we obtain the same final result—the absence of coherent information transmitted in such a channel. This result is characteristic for coherent information, in contrast to other information approaches (see, e.g., [21]).

It is of interest to discuss quantum-measurement transformations of a more general type, specifically, the procedure of indirect (generalized) measurement, first introduced in application to problems of the theory of optimal quantum detection and measurement in [22], and in a more general form of the nonorthogonal expansion of unity $\hat{\mathscr{E}}(d\lambda)$ in [23] ($\hat{\mathscr{E}}(d\lambda)$ is the equivalent of a positive-definite operator-valued measure (POVM), used in semiclassical variants of quantum information theory and quantum theory of optimal detection/measurement [5, 24, 25]). The corresponding transformation of the indirect measurement is obtained by averaging the transformation of the direct measurement, applied not directly to the system of interest but rather to its combination with an arbitrary independent system. In the general case the indirect measurement superoperator can be represented as

$$\mathscr{M} = \sum_q \hat{P}_q \mathrm{Tr}\hat{\mathscr{E}}_q \odot, \tag{23}$$

where $\hat{P}_q$ describe an arbitrary set of orthogonal projectors, and $\hat{\mathscr{E}}_q$ is a general nonorthogonal expansion of unity in the space H (POVM). We note that $\hat{\mathscr{E}}_q = |\varphi_q\rangle\langle\varphi_q|$ describes the case of a "pure" POVM, first introduced precisely in this form in the quantum theory of detection/measurement [22]. It corresponds to a complete measurement in $H \otimes H_a$ with the choice of the singular density matrix for the initial state of an auxiliary independent system $\rho^a_{bc} = \delta_{b0}\delta_{bc}$.

In Eq. (23) the classical index q represents the information exchange between the initial state and final state of the output. Since the number N_q of values of q can be greater than the dimension dim H, it can be inferred that a definite amount of coherent information might be attained as a result of this excess. The corresponding output and input–output density matrices have the form

$$\hat{\rho}_{out} = \sum_q \tilde{p}_q \hat{P}_q,$$
$$\hat{\rho}_{\alpha} = \sum_{qij} \sqrt{p_i p_j}\langle j|\hat{\mathscr{E}}_q|i\rangle \hat{P}_q \otimes |i\rangle\langle j|, \tag{24}$$

where $\tilde{p}_q = \mathrm{Tr}\hat{\mathscr{E}}_q \hat{\rho}_{in}$ describe the probabilities of states determined by indirect measurement. For the case of complete indirect measurement, based on the quantum analog [2] of the classical theorem of no increase in information in successive transformations of data and on the above-proved result concerning the complete direct measurement, it is not difficult to substantiate theoretically and confirm by numerical calculations that it is impossible to obtain coherent information. Therefore, to obtain as a result of a measurement procedure a nonzero amount of coherent information it is necessary to use incomplete ("soft") quantum measurements which require an independent, more detailed investigation.

6.3. Quantum Duplication Procedure

In counterpoint to the above-studied dequantizing-type measurement procedure, determined by the trans-

formation (20), which completely destroys coherent information, here we shall examine a transformation in a channel, shown in Fig. 1c, which preserves the coherent information:

$$\hat{\rho}_{12} \longrightarrow \hat{\rho}'_{12} = \sum_{ij} \langle\phi_i|\mathrm{Tr}_2\hat{\rho}_{12}|\phi_j\rangle|\phi_i\rangle\langle\phi_j|\langle\phi_j|.$$

It does not ignore the phase relations between the various ϕ_i because of the use of the off-diagonal matrix elements of the input density matrix $\hat{\rho}_1 = \hat{\rho}_{\mathrm{in}}$.

For the initial density matrix in the form of a product $\hat{\rho}_{\mathrm{in}} \otimes \hat{\rho}_2$, in terms of the transformation $\hat{\rho}_{\mathrm{in}} \longrightarrow \hat{\rho}'_{12}$ from H to $H \otimes H$ the corresponding superoperator has the form

$$\mathfrak{D} = \sum_{ij} |\phi_i\rangle|\phi_i\rangle\langle\phi_j|\langle\phi_i| \odot |\phi_j\rangle. \qquad (25)$$

This superoperator determines the transformation of a coherent measurement in counterpoint to an incoherent measurement, studied in [21]. The transformation of the coherent measurement converts $\hat{\rho}_{\mathrm{in}}$ into an $\hat{\rho}_2$-independent state

$$\hat{\rho}_{\mathrm{out}} = \hat{\rho}'_{12} = \sum_{ij} \langle\phi_i|\hat{\rho}_{in}|\phi_j\rangle|\phi_i\rangle\langle\phi_j|\langle\phi_j|, \qquad (26)$$

which results in duplication of the eigenstates ϕ_i of the input by the same states of the indicator variable

$$\hat{k} = \sum_k k|\phi_k\rangle\langle\phi_k|.$$

The pure input states transform into pure states of the composite (1 + 2) system by means of duplication of the indicator states:

$$\psi \longrightarrow \sum_i \langle\phi_i|\psi\rangle|\phi_i\rangle|\phi_i\rangle,$$

which repeats the mapping (19), which gives a multivalued description of the corresponding superoperator transformation. Of course, only the input states equal to the eigenstates ϕ_k of the chosen indicator variable are duplicated without distortion as a result of the incompatibility of the nonorthogonal states; this is the basic theorem of the impossibility of quantum cloning [26]. The entropy of the output state with density matrix (26), possessing the same matrix elements as $\hat{\rho}_{\mathrm{in}}$, obviously is identical to the entropy of the input state, $S_{\mathrm{out}} = S_{\mathrm{in}} = S[\hat{\rho}_{\mathrm{in}}]$, on account of the conservation of the coherence of all pure input states.

For combined input–output states the transformation (25) leads to the density matrix (9) in the space $H \otimes H \otimes H$ of the form

$$\hat{\rho}_\alpha = \sum_{kl} |\phi_k\rangle|\phi_k\rangle\langle\phi_l|\langle\phi_l| \otimes \sqrt{\tilde{p}_k\tilde{p}_l}|\chi_k\rangle\langle\chi_l|, \qquad (27)$$

where $\tilde{p}_k$ and $|\chi_k\rangle$ are the same as above; this makes it possible to construct a spectral expansion of the density matrix in the form

$$\hat{\rho}_{\mathrm{in}} = \sum_k \tilde{p}_k|\chi_k\rangle\langle\chi_k|.$$

Keeping in mind the fact that the first term of the tensor product in Eq. (27) is a set of transition projection operators $\hat{P}_{kl}$, $\hat{P}_{kl}\hat{P}_{mn} = \delta_{lm}\hat{P}_{kn}$, it is easy to prove the algebraic rule used for an arbitrary scalar function f:

$$f\left(\sum_{kl} \hat{P}_{kl} \otimes \hat{R}_{kl}\right) = \sum_{kl} \hat{P}_{kl} \otimes f(\hat{R})_{kl},$$

where $\hat{R} = (\hat{R}_{kl})$ is a block matrix, and

$$\mathrm{Tr}f\left(\sum_{kl} \hat{P}_{kl} \otimes \hat{R}_{kl}\right) = \mathrm{Tr}f(\hat{R}).$$

Here $\hat{R} = (\sqrt{\tilde{p}_k\tilde{p}_l}|\chi_k\rangle\langle\chi_l|)$, which equals simply $||\chi\rangle\rangle\langle\langle\chi||^+$ with $||\chi\rangle\rangle_{ki} = \sqrt{\tilde{p}_k}\chi_{ki}$, since this corresponds to a vector in the space $H \otimes H$. The eigenvalues λ_k of this matrix are {1, 0, 0, 0} with a single nonzero eigenvalue, corresponding to the vector $||\chi\rangle\rangle$.

A calculation of the exchange entropy gives $S_e = 0$ and therefore $I_c = S_{\mathrm{in}}$. This means that the quantum duplication procedure does not decrease the volume of coherent information in the channel $1 \longrightarrow (1 + 2)$ irrespective of whether or not the indicator $\hat{k}$ is compatible with the input density matrix, i.e., $[\hat{k}, \hat{\rho}_{\mathrm{in}}] = 0$.

If the channel considered is reduced to the channel shown in Fig. 1b and studied in the preceding section by taking the trace over the first or second system in Eq. (26), we obviously arrive at the measuring procedure examined in Section 6.2. As a result, we can conclude that coherent information is not associated with each system separately, i.e., it is strongly coupled with both systems. The specific nature of the quantum information, studied above, can be used, specifically, in algorithms for correcting quantum errors [27] or for producing stable entangled states [28].

6.4. Two Level Atom–Vacuum Field Channel

We now consider the interaction between a TLA and a vacuum electromagnetic field, i.e., the process of electromagnetic emission, as an information channel (Fig. 1b), which compared with a TLA in a given laser field (see Section 4) introduces a new object—the photon vacuum field—as the output.

For this purpose we shall employ a reduced model of the field based on the reduction of the Hilbert space in the Fock representation (Fig. 5). In a more general

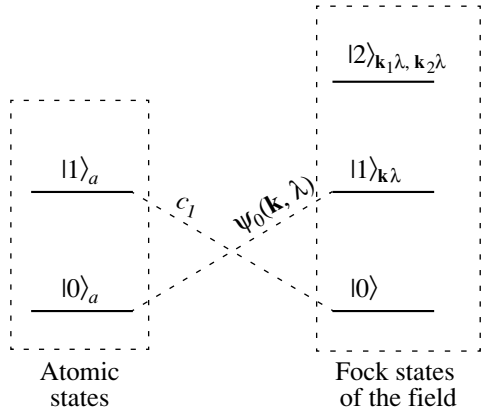

Fig. 5. Structure of the compound Hilbert space of the atom–field system. For the vacuum initial state of the field both atomic states and only two Fock states of the field ($|0\rangle$ and $|1\rangle$) are included in the dynamics of the atom–field composite system, which can be defined by only two states $|0\rangle_a|1\rangle_{\mathbf{k}\lambda}$ and $|1\rangle_a|0\rangle$, described by time-dependent functions $\psi_0(\mathbf{k}, \lambda)$ and c_1, respectively.

terminology, this problem corresponds to the problem of the dynamics of the interaction of a two-level system with a multimodal system of linear oscillators [29]. The solution of the latter problem on long-time scales, to which we shall confine our attention here, corresponds to the standard description of the emission of a single photon. Moreover, for purposes of information analysis of a system consisting of an atom and a field, there is no need to describe the coherent dependence of the wave function $\psi_0(\mathbf{k}, \lambda)$ of the photon on the photon wave vector (and polarization), since only the total probability of emission of the photon is important.

In the basis of states of the free atom and the Fock states of the free field for a vacuum initial state $\alpha_0 = 0$ we obtain from the relation (18)

$$S_{kl,\,\mu\nu} = \sum_m U_{m\mu,\,k0} U^*_{m\nu,\,l0},$$

where the Greek indices are used to denote states of the photon field, which in general correspond to the number of photons and their spatial coordinates or wave vectors. The calculation of this superoperator, performed on the basis of a unitary matrix of the temporal evolution of the atom–field system with matrix elements $U_{m\mu,\,k0}$, is illustrated in Tables 1 and 2.

Choosing $\psi_0(\mathbf{k}, \lambda)$ as the basis element of a single-photon subspace of states of the field reduces the matrix $S_{kl,\,\mu\nu}$ of the superoperator to a nonoperator transformation matrix, which in terms of the matrices $\hat{s}_{kl}$ has the form

$$\hat{s}_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

$$\hat{s}_{12} = \begin{pmatrix} 0 & (1-e^{-\gamma t})^{1/2} \\ 0 & 0 \end{pmatrix},$$

$$\hat{s}_{21} = \begin{pmatrix} 0 & 0 \\ (1-e^{-\gamma t})^{1/2} & 0 \end{pmatrix}, \qquad (28)$$

$$\hat{s}_{22} = \begin{pmatrix} e^{-\gamma t} & 0 \\ 0 & 1-e^{-\gamma t} \end{pmatrix},$$

where $|c_1|^2 = \exp(-\gamma t)$ describes the decay of the population of an initially completely populated excited state of the atom, and

$$\int\sum |\psi_0(\mathbf{k}, \lambda)|^2 d\mathbf{k} = 1 - \exp(-\gamma t)$$

describes the total probability of detecting a photon. It follows from Eqs. (28) that the structure of the photon is of no significance, and the transmitted information is determined only by the probability of emission of a photon by the time t. This reduction of the photon field (only the photon numbers $\mu, \nu = 0, 1$ are important) reduces it to an equivalent two-level system.

Applying the transformation (28) to the input density matrix

$$\hat{\rho}_{in} = \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{12} & 1-\rho_{11} \end{pmatrix},$$

where we have confined ourselves to the case of purely real off-diagonal matrix elements, we obtain the output density matrix

$$\hat{\rho}_{out} = \begin{pmatrix} \rho_{11} + \rho_{22}e^{-\gamma t} & \rho_{12}(1-e^{-\gamma t})^{1/2} \\ \rho_{12}(1-e^{-\gamma t})^{1/2} & \rho_{22}(1-e^{-\gamma t}) \end{pmatrix},$$

and the corresponding input–output density matrix, which for the case $\rho_{12} = 0$ has the form

$$\hat{\rho}_\alpha = \left(\begin{array}{cc|cc} \rho_{11} & 0 & 0 & [\rho_{11}\rho_{22}(1-e^{-\gamma t})]^{1/2} \\ 0 & \rho_{22}e^{-\gamma t} & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ [\rho_{11}\rho_{22}(1-e^{-\gamma t})]^{1/2} & 0 & 0 & \rho_{22}(1-e^{-\gamma t}) \end{array}\right).$$

For $t \longrightarrow \infty$ this expression gives a completely entangled state (in the sense of the absence of classical correlations, since it is a pure state) of the atom–photon system, leading to transfer to the photon of coherent fluctuations of the atomic state, which are equivalent to a mixed ensemble. The corresponding eigenvalues are

$$\lambda_\alpha = \{0, 0, 1 - \rho_{22}\exp(-\gamma t), \rho_{22}\exp(-\gamma t)\}.$$

The nonzero values describe the probabilities of atomic states at time t. The eigenvalues for the output density matrix (photon + vacuum) $\hat{\rho}_{out}$ are

$$\lambda_{out} = \{\rho_{22}[1 - \exp(-\gamma t)], 1 - \rho_{22}[1 - \exp(-\gamma t)]\}$$

and describe the probability of observing whether the photon is emitted or not. The set of quantities presented above determines the characteristic values of the probabilities of the combined input–output density matrix and the partial density matrices. The coherent information given by the corresponding entropy difference assumes the form

$$\begin{aligned} I_c &= x\rho_{22}\log_2(x\rho_{22}) \\ &- (1 - \rho_{22} + x\rho_{22})\log_2[1 - (1 - x)\rho_{22}] \\ &+ (1 - x\rho_{22})\log_2(1 - x\rho_{22}) \\ &- (1 - x)\rho_{22}\log_2(\rho_{22} - x\rho_{22}), \end{aligned} \tag{29}$$

where $x = \exp(-\gamma t)$. This formula is applicable for $I_c > 0$, while in the opposite case $I_c = 0$. The corresponding critical moment in time is determined by the relation $\exp(-\gamma t) = 1/2$, which corresponds to the probability $1 - \rho_{22}[1 - \exp(-\gamma t)]$ of the absence of a photon being equal to the probability $1 - \rho_{22}\exp(-\gamma t)$ of the bottom atomic level being occupied.

The results of the calculation of the coherent information are presented in Fig. 6 for two special cases: $\rho_{12} = 0$ (Fig. 6a) and $\rho_{11} = 1/2$, $0 \le \rho_{12} \le 1/2$ (Fig. 6b). It is easy to see from Fig. 6a that the coherent information is symmetric with respect to the symmetry point $\rho_{11} = 1/2$. A further increase of the population of the excited state $\rho_{22} = 1 - \rho_{11}$ and the corresponding level of photon emission does not increase the amount of coherent information. This is due to the decrease in the input entropy, which determines the potential maximum of coherent information. For the same reasons, the coherent information decreases if a nonzero coherent correction is made to the initial density matrix of the atom in the form of a state with maximum entropy and vanishes for a purely coherent initial state (Fig. 6b).

In accordance with Section 3 and taking account of the fact that the initial state of the field is pure, the one-time information is equal to the entropy difference only for the photon field represented by the density matrix $\hat{\rho}_{out}$ and the initial atomic state represented by $\hat{\rho}_{in}$. For a pure initial state of the atom in the form of an excited state $|2\rangle$ we obtain for all times the nonzero coherent information

$$I_c = -x\log_2 x - (1 - x)\log_2(1 - x),$$

which gives one qubit for the time when $x = 1/2$ and the population of the excited state coincides with the probability of emission of a photon.

Table 1. Unitary transformation atom–field–atom–field $U_{m\mu, k\alpha}$ for a vacuum initial state of the photon field; the indices m and k enumerate the atomic photons, μ and α are the photon numbers

$k\alpha$	$m\mu$			
	00	01	10	11
00	1	0	0	0
01	–	–	–	–
10	0	$\psi_0(\mathbf{k}, \lambda)$	c_1	0
11	–	–	–	–

Note: The second and fourth rows are the matrix elements that do not appear in the matrix elements $S_{kl, \mu\nu}$ which are computed (see Table 2).

Table 2. Atom–field superoperator $S_{kl, \mu\nu}$, setting the transformation $|k\rangle\langle l| \longrightarrow |\mu\rangle\langle\nu|$. The indices k and l enumerate the atomic photons, and μ, ν are the photon numbers

kl	$\mu\nu$			
	00	01	10	11
00	1	0	0	0
01	0	0	$\psi_0(\mathbf{k}, \lambda)$	0
10	0	$\psi_0^+(\mathbf{k}, \lambda)$	0	0
11	$\lvert c_1\rvert^2$	0	0	$\psi_0(\mathbf{k}, \lambda) \times \psi_0^+(\mathbf{k}', \lambda')$

6.5. Atom-to-Atom Transmission of Coherent Information Via a Free Vacuum Field

Let us consider a quantum channel of the type $1 \longrightarrow 2$ (Fig. 1b), where information is transferred from one atom to another through free space by emission of a photon, assuming that initially the second atom is the ground state. In addition, we introduce a limitation on the scale of the times considered, excluding from our analysis fast processes occurring on times of the order of and less than the period of atomic oscillations, i.e., ignoring the discrete nature of the electromagnetic signal that is associated with interatomic retardation [30–33]. In this approximation the problem under consideration is a Dicke problem [34], for which the solution in terms of two time-decaying symmetric and antisym-

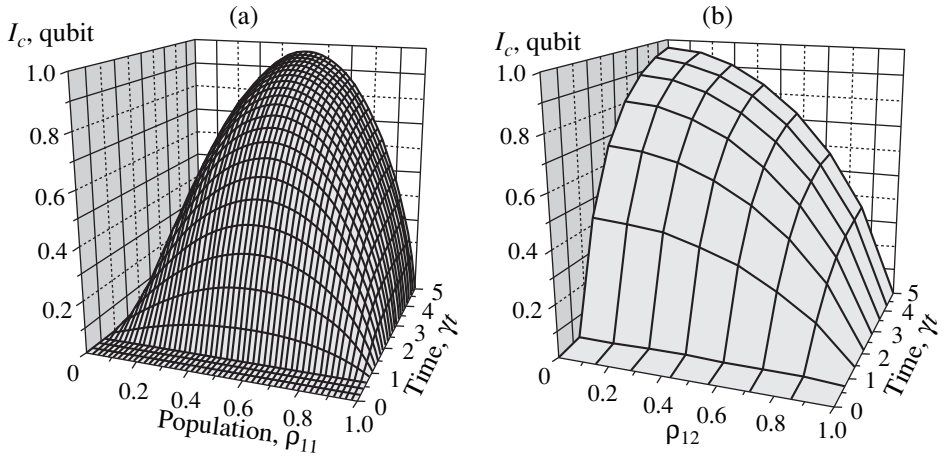

Fig. 6. Coherent information transmitted by the atom–field quantum channel as a function of time and the output density matrix of the atom: (a) the density matrix is diagonal with matrix element ρ_{11} of the ground state; (b) the density matrix is described by the sum of $\hat{I}/2$ and the real ("cosine"-type) coherent addition in the form of the off-diagonal term $\rho_{12}\hat{\sigma}_1$.

metric Dicke states $\|s\rangle\rangle = (|1\rangle|2\rangle + |2\rangle|1\rangle)/\sqrt{2}$, $\|a\rangle\rangle = (|1\rangle|2\rangle - |2\rangle|1\rangle)/\sqrt{2}$ and a stable vacuum state $\|0\rangle\rangle = |1\rangle|1\rangle$ in the following form is well known:

$$\begin{gathered} c_s(t) = c_s(0)\exp[-(\gamma_s/2 + i\Lambda)t], \\ c_a(t) = c_a(0)\exp[-(\gamma_a/2 + i\Lambda)t], \\ c_0(t) = c_0(0) \\ + [c_s(0)^2 + c_a(0)^2 - c_s(t)^2 - c_a(t)^2]^{1/2} e^{i\xi(t)}, \end{gathered} \tag{30}$$

where $c_0(t)$ is the complex amplitude of the vacuum component $|1\rangle|1\rangle$, including the incoherent correction due to the spontaneous radiative transitions from excited atomic states, $\xi(t)$ is the uniformly distributed phase of atomic oscillations, $\gamma_{s,a}$ and Λ are the decay rates and the frequency splitting (frequency shift), respectively, and $c_{s,a}$ are the complex amplitudes of the Dicke states.

In terms of multiplicative combinations of individual atomic states $|i\rangle|j\rangle$ for the corresponding amplitudes of the initial states $c_{12}(0) = 0$ and $c_{22}(0) = 0$ the dynamics of the system under study is described in accordance with the dynamics of Dicke states, determined by the relations (30), for the following equations:

$$\begin{gathered} c_{11}(t) = c_{11}(0) + f(t)e^{i\xi(t)}c_{21}(0), \\ c_{21}(t) = f_s(t)c_{21}(0), \quad c_{12}(t) = f_a(t)c_{12}(0), \\ c_{22}(t) = 0, \\ f(t) = \{1 - [\exp(-\gamma_s t) + \exp(-\gamma_a t)]/2\}^{1/2}, \\ f_s(t) = \{\exp[-(\gamma_s/2 + i\Lambda)t] \\ + \exp[-(\gamma_a/2 - i\Lambda)t]\}/2, \\ f_a(t) = \{\exp[-(\gamma_s/2 + i\Lambda)t] \\ - \exp[-(\gamma_a/2 - i\Lambda)t]\}/2. \end{gathered}$$

Using these expressions for the input operators of the form $c_{k1}(0)c_{l1}^*(0)|k\rangle\langle l|$ for the first atom and then averaging them over the final states of the first atom and fluctuations of the atomic field (the latter are represented here only by the variable $\xi(t)$), we obtain the symbolic representation of the superoperator of the transformation of the channel $\hat{\rho}^{(1)}(0) \longrightarrow \hat{\rho}^{(2)}(t) = \mathscr{C}(t)\hat{\rho}^{(1)}(0)$ and the corresponding operators $\hat{s}_{kl}$ in the form

$$\begin{gathered} \mathscr{C}(t) = |1\rangle\langle 1| \odot |1\rangle\langle 1| + [f(t)^2 + |f_s(t)|^2]|1\rangle \\ \times\langle 2| \odot |2\rangle\langle 1| + |f_a(t)|^2|2\rangle\langle 2| \odot |2\rangle\langle 2| \\ + f_a(t)|2\rangle\langle 2| \odot |1\rangle\langle 1| + f_a^*(t)|1\rangle\langle 1| \odot |2\rangle\langle 2|, \\ \hat{s}_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \hat{s}_{12} = \begin{pmatrix} 0 & f_a^*(t) \\ 0 & 0 \end{pmatrix}, \\ \hat{s}_{21} = \begin{pmatrix} 0 & 0 \\ f_a(t) & 0 \end{pmatrix}, \\ \hat{s}_{22} = \begin{pmatrix} f(t)^2 + |f_s(t)|^2 & 0 \\ 0 & |f_a(t)|^2 \end{pmatrix}. \end{gathered} \tag{31}$$

To make the problem more concrete we shall consider two identical atoms with parallel dipole moments, directed perpendicular to the vector connecting the atoms under study. Then only two dimensionless variables are important: γt, where γ describes the radiative

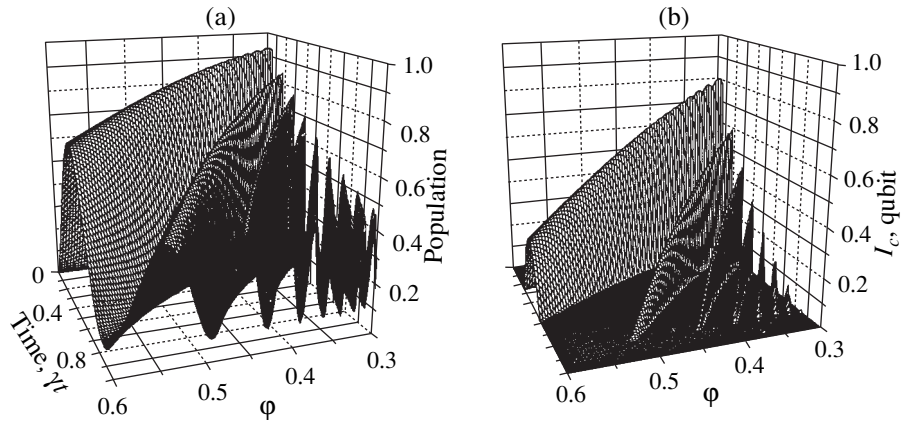

Fig. 7. (a) Population of the excited state of the second atom and (b) the coherent information in a system of two atoms interacting through free space as a function of the dimensionless time γt and the interatomic distance $\varphi = \omega_0 R/c$. The input density matrix corresponds to a state with maximum entropy $\hat{\rho}_{in} = \hat{I}/2$.

decay rate of an isolated atom, and a dimensionless interatomic distance $\varphi = k_0 R$, where R is the distance between the atoms and k_0 is the modulus of the wave vector corresponding to the frequency of the atom. The dimensionless two-atom radiative decay rates and the frequency shift due to the short-range dipole-dipole interaction are described by the corresponding relations [19, 28–33]:

$$\gamma_{s,a}/\gamma = 1 \pm g, \quad \Lambda/\gamma = (3/4)/\varphi^3,$$

where $g = (3/2)(\varphi^{-1}\sin\varphi + \varphi^{-2}\cos\varphi - \varphi^{-3}\sin\varphi)$.

The corresponding coherent information can be calculated as done in Section 6.4. Using the correspondence $\exp(-\gamma t) \longrightarrow f(t)^2 + |f_s(t)|^2$, the operators $\hat{s}_{kl}$ for the two cases are completely similar and the coherent information, once again, is described by the same relation (29) with $x = f(t)^2 + |f_s(t)|^2$. Nonetheless, in this case the dependence considered, as compared with the case of one atom (see Section 6.4), has specific qualitative features on account of the oscillatory character of the function $|f_{s,a}(t)|^2$ as a function of the interatomic distance φ.

If there were no oscillations due to the quasielectrostatic short-range dipole-dipole interaction, i.e., if one could set $\Lambda = 0$, then the coherent information would always be zero, since the threshold $x < 0.5$ cannot be reached. The parameter $1 - x$ corresponds to the population of the excited state of the second atom with the initial state $|2\rangle$ of the first atom, and for the optimal, from the standpoint of information, value of the population of the first atom ρ_{22}, equal to half its initial population, we obtain $1 - x \le 1/4$ and correspondingly $x \ge 3/4$. The oscillations in $|f_a(t)|^2$ lead to interference between two decaying Dicke components, so that the maximum of the population $n_2 = 1 - x$ also reaches larger values right up to $n_2 = 1$, and the corresponding coherent information becomes different from zero.

The functions $n_2(\varphi, \gamma t)$ and $I_c(\varphi, \gamma t)$, calculated using the relation (29), are shown in Fig. 7. They serve as a universal measure for a system of two atoms, being independent of their frequency or the magnitude of the dipole moment (the latter is valid only for a fixed geometry of the system, described above).

As one can see in Fig. 7a the population decays rapidly as a function of time because of the rapid decay of the short-lived Dicke component. The population and the coherent information undergoes strong oscillations (Fig. 7b) for small interatomic distances φ. As $\varphi \longrightarrow 0$ the population of the long-lived Dicke state remains substantial for unlimited long times, but no coherent information is associated with it because the other component decays completely.

7. CONCLUSIONS

It was shown in this work that the concept of coherent information can be used for obtaining the most general description of the interaction between two real quantum systems, including systems of qualitatively different physical nature, and for determining the role of quantum coherence in the composite system.

It was shown for a TLA in a resonant laser field that coherent information in the system does not increase with increasing intensity of the applied field, provided that the relaxation processes themselves are not suppressed.

The hydrogen atom was considered as an example of the information exchange between subsystems of a single system. It was shown that under the action of an applied electric field coherent information exchange occurs between forbidden and dipole-active atomic transitions as a result of the interaction due to the Stark effect.

It was shown for two unitarily coupled TLA that the maximum possible value of the coherent information $I_c = 1$ qubit is reached for maximum entanglement and

$I_c = 0$ for any type of measurement procedures studied in Section 6.2.

It was shown for information exchange between TLA and a free photon field in the process of emission of electromagnetic radiation that the coherent information reaches the threshold of nonzero values at the critical point of the decay exponential $\exp(-\gamma t) = 1/2$, where the probability of there being no emitted photon is equal to the population of the lower atomic state. At the maximum the coherent information can reach the value $I_c = 1$ qubit.

It was shown for information exchange between two atoms by means of the vacuum field, when the atoms are separated by a distance of the order of the wavelength, that the coherent information is nonzero only as a result of coherent oscillations between the Dicke states, which are due to short-range dipole-dipole quasielectrostatic interaction with spatial dependence $\propto 1/R^3$. In contradistinction to this, the semiclassical information extracted using the quantum detection procedure is associated with population correlations [28].

ACKNOWLEDGMENTS

This work was partially supported by the programs "Fundamental Metrology," "Physics of Quantum and Wave Phenomena," and "Nanotechnology" of the Russian Ministry of Science and Technology.

REFERENCES

1. R. G. Gallager, *Information Theory and Reliable Communication* (Wiley, New York, 1968; Sov. Radio, Moscow, 1974).
2. B. Schumacher and M. A. Nielsen, Phys. Rev. A **54**, 2629 (1996).
3. S. Lloyd, Phys. Rev. A **55**, 1613 (1997).
4. C. P. Williams and S. H. Clearwater, *Explorations in Quantum Computing* (Springer-Verlag, New York, 1998).
5. J. Preskill, in *Lecture Notes on Physics*, Vol. 229: *Quantum Information and Computation*, http://www.theory.caltech.edu/people/preskill/ph229/.
6. B. A. Grishanin, Kvantovaya Élektron. (Moscow) **9**, 827 (1979).
7. K. Kraus, *States, Effects, and Operations* (Springer-Verlag, Berlin, 1983).
8. B. A. Grishanin, Zh. Éksp. Teor. Fiz. **85**, 447 (1983) [Sov. Phys. JETP **58**, 262 (1983)].
9. É. G. Pestov and S. G. Rautian, Zh. Éksp. Teor. Fiz. **64**, 2032 (1973) [Sov. Phys. JETP **37**, 1025 (1973)].
10. V. S. Lisitsa and S. I. Yakovlenko, Zh. Éksp. Teor. Fiz. **68**, 479 (1975) [Sov. Phys. JETP **41**, 233 (1975)].
11. K. Burnett, J. Cooper, P. D. Kleiber, and A. Ben-Reuven, Phys. Rev. A **25**, 1345 (1982).
12. V. A. Alekseev, B. Ya. Zel'dovich, and I. I. Sobel'man, Usp. Fiz. Nauk **118**, 385 (1976) [Sov. Phys. Usp. **19**, 207 (1976)].
13. A. N. Moskalev, R. M. Ryndin, and I. B. Khriplovich, Usp. Fiz. Nauk **118**, 409 (1976) [Sov. Phys. Usp. **19**, 220 (1976)].
14. S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967).
15. A. Salam, in *Proceedings of the 8th Nobel Symposium*, Stokholm, 1968, p. 367.
16. S. Hill and W. K. Wooters, quant-ph/9703041 (1997).
17. C.-S. Niu and R. B. Griffits, quant-ph/9810008 (1999).
18. G. K. Brennen, I. H. Deutsch, and P. S. Jessen, quant-ph/9910031 (1999).
19. I. V. Bargatin, B. A. Grishanin, and V. N. Zadkov, submitted to Fortschr. Phys. (2000); quant-ph/9903056 (1999).
20. H. Barnum, M. A. Nielsen, and B. Schumacher, Phys. Rev. A **57**, 4153 (1998).
21. Y.-X. Chen, quant-ph/9906037 (1999).
22. C. W. Helstrom, J. W. S. Liu, and J. P. Gordon, Opt. Commun. **58**, 1578 (1970).
23. B. A. Grishanin, Izv. Akad. Nauk SSSR, Tekh. Kibern. **11** (5), 127 (1973).
24. C. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976; Mir, Moscow, 1979).
25. A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer, Dordrecht, 1993).
26. W. K. Wooters and W. H. Zurek, Nature **299**, 802 (1982).
27. P. W. Shor, Phys. Rev. A **52**, 2493 (1995).
28. B. A. Grishanin and V. N. Zadkov, Laser Phys. **8**, 1074 (1998); quant-ph/9906069 (1999).
29. C. Cohen-Tannoudji, J. Dupont-Roc, and G. Grynberg, *Atom-Photon Interactions* (Wiley, New York, 1992).
30. E. Fermi, Rev. Mod. Phys. **4**, 87 (1932).
31. J. Hamilton, Proc. R. Soc. London, Ser. A **62**, 12 (1949).
32. W. Heitler and S. T. Ma, Proc. R. Ir. Acad., Sect. A **52**, 109 (1949).
33. P. W. Milonni and P. L. Knight, Phys. Rev. A **10**, 1096 (1974).
34. R. H. Dicke, Phys. Rev. **93**, 9 (1954).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 921–937.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1066–1083.

ATOMS, SPECTRA, RADIATION

Bipolar Harmonics Method in the Semiclassical Theory of Sub-Doppler Cooling

A. V. Bezverbnyĭ
Nevel'skoĭ Far East State Marine Academy, Vladivostok, 690059 Russia
e-mail: alexb@fesma.ru
Received May 10, 2000

Abstract—The bipolar harmonics method is extended to the case of complex elliptic polarization vectors. The method is used to study, on the basis of the semiclassical theory, the multipole moments of the ground state of atoms under conditions of sub-Doppler cooling with a monochromatic light field possessing spatial gradients of the polarization. It is shown that for stationary atoms with an initial isotropic distribution over sublevels the multipole moments of rank κ decompose, in accordance with the parity κ of the rank, according to one of two minimal sets of bipolar harmonics with different symmetry under inversion. An expansion of the corrections, which are linear in the velocity, to the multipole moments with respect to the indicated minimal sets of bipolar harmonics is studied for a stationary state, and the expansion coefficients are analyzed. The orientation vector **J** of the atomic ensemble is studied on the basis of the proposed method for the dipole transition $1/2 \longrightarrow 1/2$, and the light-induced forces for a specific 2D configuration of the light field, including radiation friction forces and Lorentz-type forces, are analyzed. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Many problems in atomic physics are determined in the initial formulation by a certain set of vector quantities. Examples are problems of the collisions of atomic particles, photoionization of ions, optical pumping of atoms by a light field, and so on. Specifically, the polarization and spatial configuration of the light field are important for sub-Doppler cooling of atoms. The identification of the kinematic factors characterizing the transformation properties of the desired quantities with respect to spatial transformations and the dynamical factors is an important problem here. For two determining vectors,[1] $\mathbf{n}_1$ and $\mathbf{n}_2$, the method of bipolar harmonics

$$Y_j^{l,L}(\mathbf{n}_1, \mathbf{n}_2) = \{Y_l(\mathbf{n}_1) \otimes Y_L(\mathbf{n}_2)\}_j,$$

where $Y_l(\mathbf{n})$ are spherical functions of rank l, is effective. This method is based on the theorem [1, 2] on the expansion of bipolar harmonics of arbitrary rank $\{l, L\}$ with respect to a basis of "minimal" harmonics $\{\mathcal{Y}_j^{k,p}(\mathbf{n}_1, \mathbf{n}_2)\}$, $k = 0, \ldots, j$, $p = 0, 1$, where $\mathcal{Y}_j^{k,p}(\mathbf{n}_1, \mathbf{n}_2) = Y_j^{j-k,k+p}(\mathbf{n}_1, \mathbf{n}_2)$. We note that this basis consists of two independent sets, $\mathcal{Y}_j^{k,0}$ and $\mathcal{Y}_j^{k,1}$, whose harmonics have different parity under inversion. Such minimal sets can form natural bases for studying the multipole moments arising in the problem, where the expansion coefficients are the dynamical factors of the problem.

[1] Without loss of generality, these vectors can be taken as unit vectors.

The possibility of extending the spherical functions $Y_l(\mathbf{n})$ for a complex vector n, as noted in [3] and used in [4] for finding the stationary point of the radiation relaxation operator and in [5] to describe the "dark" states under conditions of coherent population trapping, also makes it possible to extend the method of bipolar harmonics to the case of the determining complex vectors $\mathbf{n}_1$ and $\mathbf{n}_2$. In this work such an extension is made in a study of the multipole moments of atoms in the semiclassical theory of sub-Doppler cooling. The processes of optical pumping of an atom play an important role in sub-Doppler cooling [6], where in general $\mathbf{e}(\mathbf{r})$ and $\mathbf{e}^*(\mathbf{r})$—the polarization vector of the general light field and its complex conjugate—are the determining vectors in the problem.

In Section 2 a reduced closed kinetic equation (11) for the multipole moments of the ground state of atoms, which is the initial equation for the subsequent analysis, is presented on the basis of the known approximations of the theory of sub-Doppler cooling.

The extension of bipolar harmonics and certain other quantities from the apparatus of the quantum theory of angular momentum [7] by means of analytic continuation to complex Euler angles is studied in Section 3. The Clebsch–Gordan theorem for the Wigner D functions of complex arguments and Clebsch–Gordan coefficients [8], which corresponds to the expansion of the known representations of the SU(2) group to the group of unimodular complex matrices SL(2, C) (see Appendix), plays a fundamental role here. In this interpretation the arbitrary elliptic polarization vector gives a certain direction characterized in some coordinate system by certain complex Euler angles. Correspond-

ingly, the analytic continuation to complex angles for the spherical functions $Y_l(\vartheta, \varphi)$ and bipolar harmonics $Y_j^{l,L}(\vartheta_1, \varphi_1; \vartheta_2, \varphi_2)$ as tensor products of spherical functions of different directions becomes justified. Many important algebraic properties of generalized functions remain in force.[2]

In Section 4 the structure of the multipole moments of the ground state of stationary atoms is studied using the method of bipolar harmonics, and in Section 5 the first corrections to them with respect to the velocity **v** of the atoms are studied. The vectors $\mathbf{n}_1 = \mathbf{e}/e$, $\mathbf{n}_2 = \mathbf{e}^*/e$, where $e = \sqrt{\mathbf{e}\cdot\mathbf{e}} = \sqrt{\cos 2\varepsilon} = \sqrt{l}$ corresponds to the degree of maximum linear polarization for a pure polarization state of the light field [9] (ε is the ellipticity angle), are chosen as the determining vectors. Obviously, the total number of bipolar harmonics in the $2j + 1$ basis permits expanding an arbitrary multipole moment of rank j with respect to it.

For optical pumping from an equilibrium (isotropic) state of atoms the multipole moments of the ground state are expanded only in terms of the minimal sets with $p = 0$ for even ranks j and sets with $p = 1$ for odd ranks j. For example, the orientation of the atoms ($j = 1$) is proportional to $\mathcal{Y}_1^{1,1}$, while only $\mathcal{Y}_2^{i,0}$ ($i = 0, 1, 2$) are present in the alignment tensor ($j = 2$). The dependence of the dynamical factors on the parameters of the light field is determined only by the scalar quantity $\chi = \mathbf{n}_1 \cdot \mathbf{n}_2 = 1/\cos 2\varepsilon = l^{-1}$, related with the ellipticity of the field.

The contributions linear in the velocity **v** to the multipole moments of atoms of ranks j are expanded in the general basis $\{\mathcal{Y}_j^{k,p}\}$. The dependence of the velocity is characterized by four scalars:[3] $\Lambda = ((\mathbf{v}\cdot\nabla)\mathbf{n}_1)\cdot\mathbf{n}_2$, $\Upsilon = ((\mathbf{v}\cdot\nabla)\mathbf{n}_1)\cdot[\mathbf{n}_1\times\mathbf{n}_2]$, and their complex conjugates. Together with χ they appear in the expansion coefficients which are the dynamical factors of the problem. The physical content of these scalars is due to the spatial gradients of polarization vectors of various nature. Thus, the real part Λ is proportional to the gradient of the ellipticity of the light field, and the imaginary part is proportional to the gradient of the rotation angle of the polarization ellipse in the initial plane, while Υ is proportional to the gradients of the rotation of the plane of the polarization ellipse itself in space. The invariant method of finding the dynamical factors, which is based on the property that the minimal sets $\{\mathcal{Y}_j^{k,0}; \mathcal{Y}_j^{k,1}\}$ are linearly independent separately, is examined.

In Section 6 the model of the simplest transition $1/2 \longrightarrow 1/2$ in a monochromatic light field, which has gradients of the polarization, is studied. In this case the polarization of the atomic medium is described only by the orientation vector **J**, which is a multipole moment of rank 1 ($j = 1$). A calculation of **J** up to linear corrections in the velocity **v** is performed on the basis of the method presented. Further, the light-induced forces, including radiation friction forces and Lorentz-type forces, are analyzed on the basis of the specific two-dimensional (2D) configuration of the light field.

[2] For example, the Clebsch–Gordan expansion and the addition theorem for spherical functions are satisfied.

[3] Υ is a pseudoscalar.

2. KINETIC EQUATION FOR THE MULTIPOLE MOMENTS OF ATOMS IN THE GROUND STATE

Let us consider the resonance interaction of an atomic medium with a weak light field, which in general is a superposition of coherent monochromatic plane waves with optical frequency ω:

$$\mathbf{E}(\mathbf{r}, t) = \frac{1}{2}[e^{-i\omega t}\mathcal{E}(\mathbf{r})\mathbf{e}(\mathbf{r}) + \text{c.c.}] = \frac{1}{2}\{e^{-i\omega t}\mathbf{E}(\mathbf{r}) + e^{i\omega t}\mathbf{E}^*(\mathbf{r})\}. \quad (1)$$

Here $\mathcal{E}(\mathbf{r}) = |\mathcal{E}|\exp(i\phi(\mathbf{r}))$ is the total complex amplitude of the light field taking account of the general spatial phase of the field ϕ, and $\mathbf{e}(\mathbf{r})$ is a unit polarization vector: $\mathbf{e}\cdot\mathbf{e}^* = 1$.

In a local cyclic basis, called in this paper the local natural basis, the unit vector $\mathbf{e}_0(\mathbf{r}) \sim \mathbf{e}\times\mathbf{e}^*$ is perpendicular to the plane of the polarization ellipse, and the unit polarization vector **e** and its complex conjugate **e*** have the form

$$\mathbf{e} = -\cos(\varepsilon - \pi/4)e^{i\varphi_0}\mathbf{e}_+ - \sin(\varepsilon - \pi/4)e^{-i\varphi_0}\mathbf{e}_- = \epsilon^+\mathbf{e}_+ + \epsilon^-\mathbf{e}_-, \quad (2)$$

$$\mathbf{e}^* = \cos(\varepsilon - \pi/4)e^{-i\varphi_0}\mathbf{e}_- + \sin(\varepsilon - \pi/4)e^{i\varphi_0}\mathbf{e}_+ = (\epsilon^+)^*\mathbf{e}_- - (\epsilon^-)^*\mathbf{e}_+, \quad (3)$$

where $-\pi/4 \le \varepsilon \le \pi/4$ is the ellipticity angle and φ_0 is the rotation angle of the polarization ellipse.

The light field (1) is resonant with the closed atomic dipole transition $j_g \longrightarrow j_e$ with frequency ω_0, where j_g and j_e are the total angular momenta for the ground and excited states of the atom. Let the atomic medium be described by the density operator in the Wigner representation, $\hat{\rho}(\mathbf{r}, \mathbf{p}, t)$. If the optical transition is closed, it is sufficient to confine attention to the following components of the density operator: $\hat{\rho}^{ee} = \hat{\rho}^e$—from the excited state; $\hat{\rho}^{eg}$, $\hat{\rho}^{ge}$—from optical coherences; and, $\hat{\rho}^{gg} = \hat{\rho}^g$—from the ground state. We shall consider as the atomic medium an ensemble of precooled (slow) atoms, so that the condition for the velocities of the atoms $\gamma \gg \mathbf{k}\cdot\mathbf{v}$ is satisfied (γ is the radiative relaxation constant for the excited state of an atom). On the other hand the temperature of the ensemble is still quite high

compared with the recoil energy due to the emission of one photon, $mv^2/2 \gg (\hbar k)^2/2m$, so that the recoil parameter $\hbar k/p \ll 1$ is the small parameter. We shall neglect all relaxation processes (for example, those due to interatomic collisions) except for radiative relaxation processes. This model is typical for studying the initial states of sub-Doppler cooling of atoms in light fields, when an analysis can be performed staying within the semiclassical approximation.

It is well known [10] how a closed approximation for the density operator of only the ground state $\hat{\rho}$ can be derived in the zeroth approximation in the recoil parameter in a weak light field, where the rate of optical pumping of the excited state is low compared with γ, from the main equation for $\hat{\rho}^g$. In accordance with the standard reduction method [11, 12] the optical coherences in the adiabatic approximation and in the linear approximation in the field intensity can be represented in the form

$$\hat{\rho}^{eg} = (\hat{\rho}^{ge})^\dagger \approx \frac{\Omega}{\delta + i\gamma/2}\hat{V}\hat{\rho}^g, \tag{4}$$

and the operator for the density of the excited state $\hat{\rho}^e$ in the linear approximation in the field intensity for the times considered below, which exceed the characteristic time γ^{-1} for establishing a stationary regime for $\hat{\rho}^e$, can be represented as

$$\hat{\rho}^e \approx S\hat{V}\hat{\rho}^g\hat{V}^\dagger. \tag{5}$$

Here

$$S = \frac{|\Omega|^2}{\gamma^2/4 + \delta^2} \tag{6}$$

is the saturation parameter (smallness parameter, $S \ll 1$), $\Omega = -\mathcal{E}d/\hbar$ is the Rabi frequency, d is the reduced matrix element of the dipole moment for a given optical transition, $\delta = \omega - \omega_0$ is the detuning of the light-field frequency from resonance, $\hat{V} = \hat{V}^{eg}$ and $\hat{V}^\dagger = \hat{V}^{ge}$ are the projections of the lowering and raising operators of the reduced dipole moments [11, 12] on the polarization direction. The explicit form of these operators is determined by the representation chosen. Thus, in the *Jm* representation of the eigenvectors of the angular momentum ($|j_e, \mu_e\rangle$ for the excited and $|j_g, \mu_g\rangle$ for the ground states of the atom) the operator $\hat{V}$ [11] has the matrix elements

$$V_{\mu_e\mu_g} = \sum_{q=\pm1} (-1)^{j_e-\mu_e}\begin{pmatrix} j_e & 1 & j_g \\ -\mu_e & q & \mu_g \end{pmatrix}\epsilon^q, \tag{7}$$

where the Wigner 3*jm* symbols [7] are used and the coefficients ϵ^q were determined in Eqs. (2) and (3).

In the approximations (4) and (5) and in zeroth order in the recoil parameter the closed equation for the density operator of the ground state, $\hat{\rho}^g$, has the form [11, 12]

$$\frac{d}{dt}\hat{\rho}^g = \hat{\gamma}\{S\hat{V}\hat{\rho}^g\hat{V}^\dagger\} - \gamma S((\Delta)^*\hat{V}^\dagger\hat{V}\hat{\rho}^g + \Delta\hat{\rho}^g\hat{V}^\dagger\hat{V}), \tag{8}$$

where $\Delta = 1/2 - i\delta/\gamma$. This equation describes the dynamics of the ground state of the atomic ensemble as a result of optical pumping processes. The first term (arrival operator) on the right-hand side describes the arrival of atoms in the ground state as a result of spontaneous radiative decay of the excited state; the remaining terms describe outgoing processes and light-induced shifts of the energy levels of the ground state. The explicit form of the arrival operator in the *Jm* representation [11] is

$$(\hat{\gamma}\{\hat{\rho}^e\})_{\mu_g\mu_g'} = \gamma(2j_e+1) \times \sum_{\substack{q'=0;\pm1 \\ \mu_e,\mu_e'}} (-1)^{j_e-\mu_e}\begin{pmatrix} j_e & 1 & j_g \\ -\mu_e & q' & \mu_g \end{pmatrix} \times \rho^e_{\mu_e\mu_e'}(-1)^{j_e-\mu_e'}\begin{pmatrix} j_e & 1 & j_g \\ -\mu_e' & q' & \mu_g' \end{pmatrix}. \tag{9}$$

Stationary solutions were obtained in the stationary-atoms approximation for all types of closed transitions $j_g \longrightarrow j_e$ ($j \longrightarrow j$ for half-integer j in [13], $j \longrightarrow j+1$ in [3]), including for arbitrary saturations S, while for transitions with coherent population trapping ($j \longrightarrow j$ for integer j and $j \longrightarrow j-1$) they were obtained in [14]. Specifically, it has been shown that the solutions $\hat{\rho}^g$ in the low-saturation limit, $S \ll 1$, do not depend on the saturation parameter S and, which is surprising, on the detuning δ of the light field, while their tensor structure is determined only by the irreducible tensor products of the vectors **e** and **e*** with the corresponding rank.

We note one final circumstance: evidently, the tensor structure of the solutions of Eq. (8) will be determined by the vectors **e** and **e*** in general also (for a nonstationary regime and taking account of the motions of the atoms). To this end we switch to the representation of irreducible tensors, otherwise called the $\kappa\xi$ representation. In accordance with the Wigner–Eckart theorem, the multipole moments of the atomic ensemble are related with the matrix elements $\rho_{\mu\mu'}$ of the density operator (in the *Jm* representation):

$$\hat{\rho}_{\kappa\xi} = \sum_{-j\le\mu;\,\mu'\le j} \Pi(\kappa)(-1)^{j-\mu}\begin{pmatrix} j & \kappa & j \\ -\mu & \xi & \mu' \end{pmatrix}\rho_{\mu\mu'}, \tag{10}$$

where

$$\Pi(x, y, \ldots) = \sqrt{(2x+1)(2y+1)\ldots}$$

The components $\hat{\rho}_{\kappa\xi}$ ($-\kappa \le \xi \le \kappa$) are projections of the multipole moment of rank κ of the ground ($j = j_g$) state of the atom and transform with respect to the corresponding irreducible representations of the group SU(2) [7]. The physical content of these quantities is studied, e.g., in [15].

In the $\kappa\xi$ representation Eq. (8) has the form [16]

$$\frac{d\hat{\rho}^g_\kappa}{dt} = \gamma S \sum_{\kappa_1, \kappa_2} \mathscr{F}_\kappa^{\kappa_1\kappa_2}(\delta, j_g, j_e)\{\hat{\rho}_{\kappa_1} \otimes \hat{\rho}^g_{\kappa_2}\}_\kappa, \qquad (11)$$

where the matrix $\mathscr{F}$ is expressed in terms of the 6j and 9j symbols:

$$\mathscr{F}_\kappa^{\kappa_1\kappa_2}(\delta, j_g, j_e) = \Pi(\kappa_1, \kappa_2)(-1)^{j_e - j_g + \kappa}$$

$$\times\left((-1)^{2j_e}(2j_e+1)\begin{Bmatrix} j_g & j_g & \kappa \\ j_e & j_e & 1 \end{Bmatrix}\begin{Bmatrix} 1 & 1 & \kappa_1 \\ j_g & j_g & \kappa_2 \\ j_e & j_e & \kappa \end{Bmatrix}\right.$$

$$\left. - (\Delta^* + (-1)^{\kappa+\kappa_1+\kappa_2}\Delta)\begin{Bmatrix} \kappa & \kappa_1 & \kappa_2 \\ j_g & j_g & j_g \end{Bmatrix}\begin{Bmatrix} 1 & \kappa_1 & 1 \\ j_g & j_e & j_g \end{Bmatrix}\right).$$

The summation in Eq. (11) extends over possible values of the multipole moments of the photon density matrix ($\kappa_1 = \{0; 1; 2\}$) and the atomic density matrix ($\max(0, \kappa - \kappa_1) \le \kappa_2 \le \min(2j_g, \kappa + \kappa_1)$); $\{\hat{\rho}_{\kappa_1} \otimes \hat{\rho}^g_{\kappa_2}\}_\kappa$ is the tensor product of the irreducible tensors of the photon $\hat{\rho}$ and atomic density matrices. The multipole moments of the photon density matrix can be represented as a tensor product of the vectors (2) and (3) appearing in the problem:

$$\hat{\rho}_x = \{\mathbf{e} \otimes \mathbf{e}^*\}_x. \qquad (12)$$

The equation (11) is the starting equation for the subsequent analysis.

3. ELLIPTICAL POLARIZATION VECTORS AND THE SL(2, *C*) GROUP

An arbitrary irreducible tensor of rank κ transforms according to the corresponding representation of the SU(2) group. Thus, on switching to a different coordinate system the transformation is performed using the Wigner D operator (rotation matrices):

$$\hat{\rho}_{\kappa\xi} = \sum_{\xi'} D^\kappa_{\xi'\xi}(U)\hat{\rho}_{\kappa\xi'}, \qquad (13)$$

where the explicit form of the matrix elements $D^\kappa_{\xi'\xi}(U)$ is determined by the specific parameterization of the matrices U of the SU(2) group. Ordinarily, the Euler angles α, β, and γ are used for these purposes [7].

Remaining in the $\kappa\xi$ representation, the concept of "rotation matrix" can be extended to the larger group SL(2, *C*) (see Appendix). For example, complex quantitities can be used as the Euler angles, and many important algebraic properties from the apparatus of irreducible tensors remain in force. This also permits studying in a natural manner arbitrary elliptic polarization vectors as vectors (directions) in a complex three-dimensional space, which are uniquely given by certain complex angles. For example, we shall employ Eq. (A.9) from the Appendix and represent the initial vectors **e** and **e*** (2) and (3) in a natural local basis in terms of the expanded spherical functions of rank 1:

$$\mathbf{e} = \mathbf{e}(\vartheta_1, \varphi_1) = \sqrt{\frac{4\pi}{3}}e\tilde{Y}_1\left(\frac{\pi}{2}, \varphi_0 + \tilde{\varphi}\right), \qquad (14)$$

$$\mathbf{e}^* = \mathbf{e}^*(\vartheta_2, \varphi_2) = \sqrt{\frac{4\pi}{3}}e\tilde{Y}_1\left(\frac{\pi}{2}, \varphi_0 - \tilde{\varphi}\right), \qquad (15)$$

where the length of the elliptic polarization vector,

$$e = \sqrt{\mathbf{e}\cdot\mathbf{e}} = \sqrt{\mathbf{e}^*\cdot\mathbf{e}^*} = \sqrt{\cos(2\varepsilon)} = \sqrt{l}, \qquad (16)$$

corresponds to the degree of maximum linear polarization, and

$$\exp(i\tilde{\varphi}) = \sqrt{\tan(\varepsilon + \pi/4)}$$

is the imaginary azimuthal angle (minus φ_0). We underscore that the extension of the Euler angles to complex values and the subsequent transition to generalized spherical coordinates $\mathbf{z} = \{z, \tilde{\vartheta}, \tilde{\varphi}\}$ for vectors in a three-dimensional complex space necessarily lead to a Euclidean scalar product of complex vectors,

$$\mathbf{z}_1 \cdot \mathbf{z}_2 = \sum_i (\mathbf{z}_1)_i(\mathbf{z}_2)_i,$$

and therefore length of the vector (16), which is important for the following exposition.

We shall employ the notation $\tilde{Y}_\kappa(\mathbf{z}) = \tilde{Y}_\kappa(\vartheta, \varphi)$ [3] for an arbitrary unit vector $\mathbf{z} = \{1, \tilde{\vartheta}, \tilde{\varphi}\}$. Specifically, in a local natural basis

$$\tilde{Y}_\kappa(\mathbf{n}_1) = \tilde{Y}_\kappa\left(\frac{\pi}{2}, \varphi_0 + \tilde{\varphi}\right), \quad \tilde{Y}_\kappa(\mathbf{n}_2) = \tilde{Y}_\kappa\left(\frac{\pi}{2}, \varphi_0 - \tilde{\varphi}\right).$$

The vectors

$$\mathbf{n}_1 = \frac{\mathbf{e}}{e}, \quad \mathbf{n}_2 = \frac{\mathbf{e}^*}{e} \qquad (17)$$

everywhere below denote the directions of elliptical polarization of the light field. We underscore an impor-

tant relation between the components of these functions:

$$(\tilde{Y}_{\kappa,\xi}(\mathbf{n}_1))^* = (-1)^{\xi}\tilde{Y}_{\kappa,-\xi}(\mathbf{n}_2). \quad (18)$$

It should be noted that the case of circular polarizations ($\mathbf{e} = \mathbf{e}_{\pm}$, $\mathbf{e}^* = -\mathbf{e}_{\mp}$) is a limiting case, since because $e = e^* \longrightarrow 0$ it corresponds to the limit $\tilde{\varphi} \longrightarrow \pm i\infty$.

The representations of the photon density matrix in terms of the expanded spherical functions and the generalized bipolar harmonics are presented below:

$$\hat{\rho}_x = \frac{4\pi}{3}\cos(2\varepsilon)\{\tilde{Y}_1(\mathbf{n}_1) \otimes \tilde{Y}_1(\mathbf{n}_2)\}_x, \quad (19)$$

$$\hat{\rho}_x = \frac{4\pi}{3}\cos(2\varepsilon)\tilde{Y}_x^{1,1}(\mathbf{n}_1, \mathbf{n}_2). \quad (20)$$

4. STRUCTURE OF THE MULTIPOLE MOMENTS OF STATIONARY ATOMS

We shall now examine the structure of the multipole moments of stationary atoms ($\mathbf{v} = 0$), representing the solution of Eq. (11) as an asymptotic series in time for $\hat{\rho}_\kappa^g(t)$ starting from the initial conditions $\hat{\rho}_\kappa^g(0) = \delta_{\kappa,0}/\Pi(j_g)$ (isotropic distribution with respect to the Zeeman sublevels).

The stationary-atoms approximation here means that $v/\gamma S \ll \lambda$, i.e., the displacement of atoms in the characteristic optical pumping time $(\gamma S)^{-1}$ is much smaller than the wavelength of the light. Then

$$\frac{d\hat{\rho}^g}{dt} \approx \frac{\partial\hat{\rho}^g}{\partial t},$$

and the iteration series has the form

$$\hat{\rho}_\kappa^g(t) = \sum_{n=0}^{\infty} R_{n,\kappa}\frac{t^n}{n!}, \quad R_{0,\kappa} = \frac{\delta_{\kappa,0}}{\Pi(j_g)}, \quad (21)$$

$$R_{n,\kappa} = \gamma S\sum_{\kappa_1,\kappa_2}\mathcal{F}_\kappa^{\kappa_1\kappa_2}(\delta, j_g, j_e)\{\hat{\rho}_{\kappa_1} \otimes R_{(n-1),\kappa_2}\}_\kappa. \quad (22)$$

Thus, $R_{1,\kappa} \sim \tilde{Y}_x^{1,1}(\mathbf{n}_1, \mathbf{n}_2)$ in accordance with Eq. (20). It is easy to show that the tensor structure of all other coefficients $R_{n,\kappa}$ is determined only by the generalized bipolar harmonics $\tilde{Y}_\kappa^{l,L}(\mathbf{n}_1, \mathbf{n}_2)$ with even $l + L$, if the following relation is used:

$$\{Y_{\kappa_1}^{1,1}(\mathbf{n}, \mathbf{n}') \otimes Y_{\kappa_2}^{l,L}(\mathbf{n}, \mathbf{n}')\}_\kappa = \frac{3\Pi(\kappa_1, \kappa_2)}{4\pi}\sum_{p,q=\pm1} G_{p,q}^{l,L}Y_\kappa^{l+p,L+q}(\mathbf{n}, \mathbf{n}'), \quad (23)$$

where

$$G_{p,q}^{l,L} = (-1)^{(p-q)/2}\sqrt{\left(l+\frac{p+1}{2}\right)\left(L+\frac{q+1}{2}\right)} \times \begin{Bmatrix} 1 & 1 & \kappa_1 \\ l & L & \kappa_2 \\ l+p & L+q & \kappa \end{Bmatrix},$$

which follows directly from the formulas for the change in the coupling for four commuting irreducible tensors and the Clebsch–Gordan theorem for spherical functions [1] and is true for an arbitrary pair of unit vectors $\mathbf{n}$ and $\mathbf{n}'$. This relation remains in force also for the generalized bipolar harmonics $\tilde{Y}_\kappa^{l,L}(\mathbf{n}_1, \mathbf{n}_2)$. The parameters p and q in Eq. (23) assume only the values ±1, and therefore the quantities $l + L$ and $l + L + p + q$ have the same parity.

Thus, in the general case the multipole moments of the ground state have the form

$$\hat{\rho}_\kappa^g(t) = \sum_{l+L=2k}^{\infty}\tilde{a}_\kappa^{l,L}(t)\tilde{Y}_\kappa^{l,L}(\mathbf{n}_1, \mathbf{n}_2), \quad (24)$$

where the expansion coefficients $\tilde{a}_\kappa^{l,L}(t)$ depend on δ, j_g, j_e, and S but not on the polarization parameters of the light field. The existence of a stationary solution $t \longrightarrow \infty$ for Eq. (24) is due to relaxation processes in the system, and this solution also has the form (24) with $\tilde{a}_\kappa^{l,L} = \text{const}$. The universality of the expansion (24) breaks down only for optical transitions $j \longrightarrow j - 1$, for which, as is well known [14], the stationary solution depends on the choice of initial conditions. It can thereby be asserted that with the exception of the transitions $j \longrightarrow j - 1$ the general form of the stationary solutions,

$$\hat{\rho}_\kappa^g = \sum_{l+L=2k}^{\infty}\tilde{a}_\kappa^{l,L}\tilde{Y}_\kappa^{l,L}(\mathbf{n}_1, \mathbf{n}_2), \quad (25)$$

is universal for the description of the optical orientation processes for stationary atoms irrespective of the choice of initial conditions.

This infinite series can be represented as a finite sum, if the reduction relation (A.11) is used. We introduce the parameter $\tilde{\kappa} = 2[(\kappa + 1)/2]$, equal to κ for even κ and $\kappa + 1$ for odd κ ($[a]$ is the integer part of the number a). Then, for Eqs. (11) the application of the reduction formula (A.11) necessarily leads to expansions in the bipolar harmonics of the form

$$\hat{\rho}_\kappa^g = \sum_{l=\tilde{\kappa}-\kappa}^{\kappa} a_\kappa^l(\chi)\tilde{Y}_\kappa^{l,\tilde{\kappa}-l}(\mathbf{n}_1, \mathbf{n}_2), \quad (26)$$

where the expansion coefficients a_κ^l $(0 \le \kappa \le 2j_g)$ are now functions of the cosine of a complex angle

$$\chi = \mathbf{n}_1 \cdot \mathbf{n}_2 = \frac{\mathbf{e} \cdot \mathbf{e}^*}{\mathbf{e} \cdot \mathbf{e}} = \frac{1}{\cos(2\varepsilon)} = l^{-1}, \qquad (27)$$

inversely proportional to the degree l of the maximum linear polarization of the field [9].

We underscore that the expansion (26) has an invariant form, where the kinematic part of the problem (the transformation properties of the solution with respect to the transformations of the coordinate system, as well as inversion) is determined by the generalized bipolar harmonics, and the dynamical part is described by the expansion coefficients a_κ^l. Specifically, it follows from the fact that the sum $l + L$ is even that $\hat{\rho}_\kappa^g$ are invariant under inversion $\{\mathbf{e} \longrightarrow -\mathbf{e};\ \mathbf{e}^* \longrightarrow -\mathbf{e}^*\}$, i.e., the multipole moments of even rank are true tensors and those of odd rank are pseudotensors. The quantity χ remains invariant in any case.

Thus, the problem of finding the stationary solution of initial Eq. (11) and the problem of finding the dynamical solution with an isotropic initial distribution reduces to the problem of finding the dynamical factors $a_\kappa^l(\chi)$ $(a_\kappa^l(\chi, t))$. We shall now determine the number of these parameters as a function of the value of the total angular momentum j_g of the ground state.

In accordance with the condition for choosing the general phase for the components of irreducible tensors [7], the relation $\hat{\rho}_{\kappa,\xi}^* = (-1)^\xi \hat{\rho}_{\kappa,-\xi}$ holds. We shall use Eq. (18). Then the expansion coefficients (26) are related with one another by the condition $(a_\kappa^l)^* = a_\kappa^{\tilde{\kappa}-l}$. We separate the real and imaginary components of these coefficients:

$$m_\kappa^l = \mathrm{Re}\, a_\kappa^l = \frac{a_\kappa^l + a_\kappa^{\tilde{\kappa}-l}}{2},$$
$$\mathfrak{m}_\kappa^l = \mathrm{Im}\, a_\kappa^l = \frac{a_\kappa^l - a_\kappa^{\tilde{\kappa}-l}}{2i}. \qquad (28)$$

It is obvious that of the entire set $\{m_\kappa^l, \mathfrak{m}_\kappa^l\}$ the coefficients m_κ^l with values $l \le \tilde{\kappa}/2$ and $\mathfrak{m}_\kappa^l$ with $l \le \tilde{\kappa}/2 - 1$ are sufficient. For fixed κ their total number is $2(2\kappa - \tilde{\kappa} + 1)$. Next, we perform a summation over all ranks up to the maximum value $\kappa_{\max} = 2j_g$ and take into account the normalization condition $\rho_0^g = 1/\Pi(j_g)$. Finally, the total number of independent coefficients in the expansion (26) is

$$N_{\mathrm{gen}} = N_g - 1,$$

where

$$N_g = (j_g + 1)(2j_g + 1) - \frac{\tilde{\kappa}_{\max}}{2}.$$

We underscore that N_{gen} also can be found by analyzing the structure of initial Eq. (11) in a local natural basis, in which the photon density matrix has the following nonzero components: $\{\rho_{00};\ \rho_{10};\ \rho_{20};\ \rho_{2\pm2}\}$. It is obvious that in this basis the system of Eqs. (11) separates for even and odd projections of the atomic density matrix, where the number of independent elements is determined by the number of components $\rho_{\kappa,\xi}$ with $\kappa > 0$ and with even projections ξ, which coincides with N_{gen}.

Thus, for a transition with $j_g = 3/2$ there are seven such coefficients. For example, the following can be chosen: $\{m_1^1, m_2^0, m_2^1, m_3^1, m_3^2\}$ and $\{\mathfrak{m}_2^0, \mathfrak{m}_3^1\}$. The coefficients in the first group do not depend on the sign of the detuning δ, while the coefficients in the second group show a dispersion dependence on δ. This property remains in force for arbitrary values of j_g.

The stationary solutions for the ground state of an atomic medium were obtained in [4, 13, 14], including for arbitrary saturations S. These solutions possess a number of unusual properties. In the first place, the number of independent dynamical parameters a_κ^l in them is much smaller than N_{gen}, since all $\mathfrak{m}_\kappa^l = 0$. In the second place, the coefficients m_κ^l do not depend on the detuning δ. We also note that for transitions with coherent population trapping, which have "dark" states, the part m_κ^l vanishes.

The coefficients a_κ^l are invariants and, correspondingly, they can be found by a method that is not related with the choice of any distinguished coordinate system. For example, let us consider the stationary state. We substitute Eq. (26) into Eq. (11). Then the equations for a_κ^l become

$$\sum_{\kappa_1,\kappa_2} \Pi(\kappa_1, \kappa_2) \mathcal{F}_\kappa^{\kappa_1,\kappa_2} \sum_{l=\tilde{\kappa}_2-\kappa_2}^{\kappa_2} a_{\kappa_2}^l \times \sum_{p,q=\pm1} G_{p,q}^{l,\tilde{\kappa}_2-l} \tilde{Y}_\kappa^{l+p,\tilde{\kappa}_2-l+q}(\mathbf{n}_1, \mathbf{n}_2) = 0. \qquad (29)$$

We shall use the property of "linear independence" [1] of the minimum set of bipolar harmonics $\mathcal{Y}_\kappa^{l,0}(\mathbf{n}, \mathbf{n}') = Y_\kappa^{\kappa-l,l}(\mathbf{n}, \mathbf{n}')$ with $0 \le l \le \kappa$: relations of the form

$$\sum_l C_\kappa^l (\mathbf{n} \cdot \mathbf{n}') \mathcal{Y}_\kappa^{l,0}(\mathbf{n}, \mathbf{n}') = 0 \qquad (30)$$

are possible only if $C^l_\kappa = 0$. Evidently, this relation also holds for generalized bipolar harmonics. This is easy to see by switching to a distinguished coordinate system with quantization axis in the direction of one of the vectors **n** or **n**'. Then Eqs. (29) for even ranks κ lead to the form, similar to Eq. (30) taking account of the substitution $\mathbf{n} \cdot \mathbf{n}' \longrightarrow \chi$, if the reduction relation (A.11) is used. The system of equations (30) which ultimately arises is equivalent to the linear equations of the form

$$C^l_\kappa(\chi) = \sum_{l', \kappa'} X^{l', \kappa'}_{l, \kappa} a^{l'}_{\kappa'} = 0 \qquad (31)$$

with coefficients

$$X^{l', \kappa'}_{l, \kappa} = c_0(l', \kappa', l, \kappa) + c_1(l', \kappa', l, \kappa)\chi + c_2(l', \kappa', l, \kappa)\chi^2 \ldots,$$

where c_0, c_1, … depend only on the detuning δ. For fixed κ the number of these equations is κ + 1.

A similar method is also applicable for odd ranks κ with the clarification that here the linearly independent set is formed by harmonics of the form

$$\mathcal{Y}^{l, 1}_\kappa(\mathbf{n}, \mathbf{n}') = Y^{\kappa + 1 - l, l}_\kappa(\mathbf{n}, \mathbf{n}') \quad 1 \le l \le \kappa.$$

Thus, Eqs. (29) reduce to a system $N_{gen} = N_g - 1$ of linear inhomogeneous equations of the form (31) with nonuniformity due to the contributions from the total population of the ground state with $a^0_0 = 4\pi/\Pi(j_g)$ in accordance with the normalization condition.

The calculation, performed by the indicated method, of the dynamical factors a^l_κ for the simplest transitions $j_g < 3$ agrees with the known solutions [4, 13] for the transitions $j \longrightarrow j$ (*j* are half integer values) and $j \longrightarrow j + 1$ in the limit of small saturations $S \ll 1$. We now present some results:

$$j_g = 1/2 \longrightarrow j_e = 1/2, \quad j_g = 1/2 \longrightarrow j_e = 3/2$$

$$a^1_1 = \frac{-4\pi}{3\chi},$$

$$j_g = 1 \longrightarrow j_e = 1$$

$$a^1_1 = a^1_2 = \frac{-4\pi}{3\chi}, \quad a^0_2 = a^2_2 = 0, \qquad (32)$$

$$j_g = 1 \longrightarrow j_e = 2$$

$$a^1_1 = a^1_2 = \frac{100\pi\chi}{3(8 - 25\chi^2)},$$

$$a^0_2 = a^2_2 = -\sqrt{\frac{10}{3}}\frac{8\pi}{8 - 25\chi^2}.$$

Taichenachev *et al.* [17] used a different method to obtain general results for the dynamical factors for arbitrary j_g and the indicated types of optical transitions.

5. STRUCTURE OF THE CORRECTIONS LINEAR IN THE VELOCITY TO THE MULTIPOLE MOMENTS

Taking account of the motion of the atoms results in qualitatively new corrections for the multipole moments of the ground state of an atomic ensemble. It is well known [6] that these corrections describe the retardation in the optical ordering of atoms over the Zeeman sublevels, which arises as a result of motion in light fields with spatial gradients of the polarization (ε(**r**), φ_0(**r**)). Ordinarily, the characteristic scale of these gradients is of the order of the wavelength of the light λ, and the characteristic time for establishing a stationary distribution over the Zeeman sublevels of the ground state is of the order of $t_{\text{pump}} \sim (\gamma S)^{-1}$ in accordance with Eq. (11). Then the degree of retardation can be estimated as

$$\eta_{\text{ret}} = \frac{v t_{\text{pump}}}{\lambda} \sim \frac{k v}{\gamma S}.$$

We shall consider an atomic ensemble, for which the condition $\eta_{\text{ret}} < 1$, is satisfied. This is also the condition for the velocity groups of atoms subjected to sub-Doppler cooling. Correspondingly, to find the contributions that are linear in the velocity we shall confine ourselves to corrections which are of first-order in η_{ret}. We note the importance of these contributions: in most models of sub-Doppler cooling these contributions can be used to determine the radiation friction forces acting on atoms in light fields. In this connection, it should be noted that the condition $\eta_{\text{ret}} < 1$ breaks down at the nodes of the light field, where $\mathcal{E}(\mathbf{r}) = 0$. In these regions the dependence of the multipole moments of the atoms on the velocity **v** must be taken into account exactly [18].

The initial equation for finding the corrections linear in **v** in accordance with Eq. (11) has the form

$$\frac{\mathbf{v} \cdot \nabla}{\gamma S}\hat{\rho}^g_\kappa(0) = \sum_{\kappa_1, \kappa_2} \mathcal{F}^{\kappa_1 \kappa_2}_\kappa(\delta, j_g, j_e)\{\hat{\rho}_{\kappa_1} \otimes \hat{\rho}^g_{\kappa_2}(1)\}_\kappa, \qquad (33)$$

where $\hat{\rho}^g_\kappa(0)$ are the known multipole moments of the stationary atoms and $\hat{\rho}^g_\kappa(1)$ are the corrections which are first in the retardation parameter η_{ret}.

Applying the above-described method of temporal asymptotic series to Eq. (11), we can determine the ten-

sor structure of these collections using the previously considered minimal sets of bipolar harmonics:

$$\rho^g_\kappa(1) = \sum_{l=\delta_0}^{\kappa} (\mathbf{v}\cdot\nabla\chi) f^l_\kappa \tilde{\mathcal{Y}}^{\kappa+\delta_0-l,\delta_0}_\kappa + \sum_{\kappa'=\kappa-1}^{\kappa+1}\sum_{l=(\delta_1)'}^{\kappa'} \{(s^l_{\kappa,\kappa'}(\mathbf{v}\cdot\nabla)\mathbf{n}_1 + t^l_{\kappa,\kappa'}(\mathbf{v}\cdot\nabla)\mathbf{n}_2) \otimes \tilde{\mathcal{Y}}^{\kappa'+(\delta_1)'-l,\delta_1'}_{\kappa'}\}_\kappa, \quad (34)$$

where f^l_κ, $s^l_{\kappa,\kappa'}$, and $t^l_{\kappa,\kappa'}$ are coefficients which depend on δ, ξ, and S and the type of transition, $\delta_0 = 0$ for even κ and $\delta_0 = 1$ for odd κ, and $\delta_1 = 1$ for even κ and $\delta_1 = 0$ for odd κ. The tripolar harmonics [7] appearing in Eq. (34) in the double sum reduce to bipolar harmonics, if a local natural basis $\{\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3 = [\mathbf{n}_1\times\mathbf{n}_2]\}$ can be used as a basis for expanding an arbitrary vector. This basis is not orthonormalized: thus, $(\mathbf{n}_1\cdot\mathbf{n}_2) = \chi \neq 0$ in the general case. The use of this basis for a linear polarization field requires an additional analysis because this basis reduces to the single vector $\mathbf{e}$. For example, in n-dimensional ($n = 2, 3$) configurations of a light field with polarization gradients the regions with linear polarization ordinarily form a hypersurface of dimension $n-1$, and here the passage to the limit $\chi \longrightarrow 1$ ordinarily does not result in multivaluedness and divergence.[4] The field configuration $\sigma_+ - \sigma_-$ [6], where the polarization is linear everywhere, is a special case with $n = 1$. Here a different basis must be used. For example, the set of vectors $\{\mathbf{e}, (\mathbf{v}\cdot\nabla)\mathbf{e}, [\mathbf{e}\times(\mathbf{v}\cdot\nabla)\mathbf{e}]\}$ can be used.

We shall perform an expansion, in the basis $\{\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3\}$, of the newly arising vectors in the problem $(\mathbf{v}\cdot\nabla)\mathbf{n}_1$ and $(\mathbf{v}\cdot\nabla)\mathbf{n}_2$, different from the previously determined $\mathbf{n}_1$ and $\mathbf{n}_2$ (17). For example,

$$(\mathbf{v}\cdot\nabla)\mathbf{n}_1 = \nu_1\mathbf{n}_1 + \nu_2\mathbf{n}_2 + \nu_3\mathbf{n}_3, \quad (35)$$

where the expansion coordinates have the form

$$\nu_2 = \frac{(\mathbf{v}\cdot\nabla)\mathbf{n}_1\cdot\mathbf{n}_2}{1-\chi^2}, \quad \nu_1 = -\nu_2\cdot\chi, \quad \nu_3 = \frac{(\mathbf{v}\cdot\nabla)\mathbf{n}_1\cdot\mathbf{n}_3}{1-\chi^2}. \quad (36)$$

A similar expansion is possible for $(\mathbf{v}\cdot\nabla)\mathbf{n}_2$.

Using an expansion of the type (35), it can be shown that the corrections linear in $\mathbf{v}$ to the multipole moments of the ground state have the structure

$$\rho^g_\kappa(1) = (\gamma S)^{-1}\left[\sum_{l=\delta_0}^{\kappa}(\Lambda\mathbb{A}^l_\kappa + \Lambda^*\mathbb{B}^l_\kappa)\tilde{\mathcal{Y}}^{\kappa+\delta_0-l,\delta_0}_\kappa + \sum_{l=\delta_1}^{\kappa}(\Upsilon\cdot\mathbb{C}^l_\kappa + \Upsilon^*\cdot\mathbb{D}^l_\kappa)\tilde{\mathcal{Y}}^{\kappa+\delta_1-l,\delta_1}_\kappa\right], \quad (37)$$

[4] In this connection, see the result (52) below for the optical transition $1/2 \longrightarrow 1/2$.

where $\mathbb{A}^l_\kappa$, $\mathbb{B}^l_\kappa$, $\mathbb{C}^l_\kappa$, and $\mathbb{D}^l_\kappa$ depend on χ, δ, and the type of transition and are dynamical parameters similar to the previously studied coefficients a^l_κ in Eq. (26), while the kinematic part is determined by the minimum sets of generalized bipolar harmonics $\{\tilde{\mathcal{Y}}^{k,p}_\kappa\}$ of the corresponding rank κ.

The scalar functions, formed from the products of new $((\mathbf{v}\cdot\nabla)\mathbf{n}_1, (\mathbf{v}\cdot\nabla)\mathbf{n}_2)$ and previous $(\mathbf{n}_1, \mathbf{n}_2)$ vectors

$$\Lambda = ((\mathbf{v}\cdot\nabla)\mathbf{n}_1)\cdot\mathbf{n}_2, \quad \Lambda^* = ((\mathbf{v}\cdot\nabla)\mathbf{n}_2)\cdot\mathbf{n}_1, \quad \Upsilon = ((\mathbf{v}\cdot\nabla)\mathbf{n}_1)\cdot\mathbf{n}_3, \quad \Upsilon^* = -((\mathbf{v}\cdot\nabla)\mathbf{n}_2)\cdot\mathbf{n}_3, \quad (38)$$

appear linearly in Eq. (37) in accordance with the linear approximation being considered. Under inversion Λ and Λ^* are true scalars, while Υ and Υ^* are pseudoscalars, so that the multipole moments (37) of even rank are true tensors, while those of odd rank are pseudotensors. In this connection, we note that for multipole moments of stationary atoms the only scalar formed from the defining vectors is the cosine of the complex angle $\chi = (\mathbf{n}_1\cdot\mathbf{n}_2)$ between the directions of the elliptical polarization and its complex conjugate. In terms of the geometric content of the new scalars, it can be shown that the quantity $\mathrm{Re}\Lambda$ is proportional to the gradient of the ellipticity ε, while $\mathrm{Im}\Lambda$ is proportional to the gradient of the rotation angle $\varphi_{0,\perp}$ of the local polarization ellipse relative to the initial axis $\mathbf{n}_3(\mathbf{r})$, perpendicular to the plane of the initial polarization ellipse. We shall call such a rotation the first-kind rotation of the polarization ellipse. The quantity $\mathrm{Re}\Upsilon$ is proportional to the gradient of the rotation of one of the principal axes ($\boldsymbol{\epsilon}_1 \sim \mathrm{Re}\mathbf{n}_1$) of the polarization ellipse relative to the $\boldsymbol{\epsilon}_2 \sim \mathrm{Im}\mathbf{n}_1$-axis, perpendicular to $\boldsymbol{\epsilon}_1$ and $\mathbf{n}_3$. Correspondingly, the quantity $\mathrm{Im}\Upsilon$ is proportional to the gradient of the rotation angle of another principal axis $\boldsymbol{\epsilon}_2$ of the local polarization ellipse of the field relative to the axis $\boldsymbol{\epsilon}_1$. We shall call this type of rotation of the plane of the polarization ellipse, polarization ellipse rotations of the second kind. Evidently, they do not occur in models of a one-dimensional sub-Doppler cooling: $\Upsilon = 0$.

The condition for choosing the general phase $(\hat{\rho}^g_{\kappa,\xi}(1))^* = (-1)^\xi\hat{\rho}^g_{\kappa,-\xi}(1)$ leads to the following relations for these quantities:

$$(\mathbb{A}^l_\kappa)^* = \mathbb{B}^{\kappa+\delta_0-l}_\kappa, \quad (\mathbb{C}^l_\kappa)^* = \mathbb{D}^{\kappa+\delta_1-l}_\kappa. \quad (39)$$

A substantial difference between the kinematic structure of the correction (37) and the expansion for stationary atoms (26) is the presence of bipolar harmonics $\tilde{\mathcal{Y}}^{k,\delta_1}_\kappa$ from the minimal sets, in addition to the previously examined minimal sets $\tilde{\mathcal{Y}}^{k,\delta_0}_\kappa$ and possessing a different parity with respect to inversion. Thus, for rank $\kappa = 1$ the previously considered set reduces to the

only vector $\mathbf{n}_3 = \mathbf{n}_1 \times \mathbf{n}_2 \sim \tilde{\mathcal{Y}}_1^{1,1} \sim \mathbf{e}_0$, while the new set consists of the vectors $\{\tilde{\mathcal{Y}}_1^{0,0}, \tilde{\mathcal{Y}}_1^{1,0}\} \sim \{\mathbf{n}_1; \mathbf{n}_2\} \sim \{\mathbf{e}_+; \mathbf{e}_-\}$. The corrections linear in the velocity to the multipole moments characterized by the sets $\tilde{\mathcal{Y}}_\kappa^{k,\delta_1}$ are qualitatively different from those considered previously. For example, for the simplest case $j_g = 1/2$ in the *Jm* representation it is easy to see that the component $\rho^g_{1,\,\mathrm{pop}}(1) \sim \tilde{\mathcal{Y}}_1^{1,1}$ gives a correction linear in the velocity to the difference of the populations of the Zeeman sublevels in a local natural basis, while the remaining part $\rho^g_{1,\,coh}(1) = \beta_1\tilde{\mathcal{Y}}_1^{0,0} + \beta_2\tilde{\mathcal{Y}}_1^{1,0}$ ($\beta_{1,2}$ are the expansion coefficients) describes the appearance of coherences between the sublevels $\mu_1 = -1/2$ and $\mu_2 = 1/2$ in this basis. In the general case of arbitrary values of j_g the contributions to $\tilde{\mathcal{Y}}_\kappa^{k,\delta_0}$ describe the populations of the sublevels and the coherences between the sublevels with the same parity ($\mu - \mu' = \pm 2, \pm 4\ldots$), while the contributions with $\tilde{\mathcal{Y}}_\kappa^{k,\delta_1}$ describe the coherences of the sublevels of different parity ($\mu - \mu' = \pm 1, \pm 3\ldots$) and, as follows from Eq. (37), are due to a rotation of the polarization plane in space. The total number of dynamical parameters[5] $\mathbb{A}_\kappa^l$ and $\mathbb{C}_\kappa^l$ can be easily determined by taking into account the fact that the total number of minimal harmonics $\{\tilde{\mathcal{Y}}_\kappa^{k,p}\}$ of a given rank j is $2j + 1$. Thus, the number of dynamical factors $N_{\mathrm{gen},1} = (2j_g + 1)^2 - 1$ is equal to the number of elements of the density matrix in the *Jm* representation minus the normalization condition $\rho_0^g(1) = 0$.

The invariant method for finding the coefficients $\mathbb{A}_\kappa^l$, $\mathbb{B}_\kappa^l$, $\mathbb{C}_\kappa^l$, and $\mathbb{D}_\kappa^l$ is similar to the method used previously to determine a_κ^l and is based on the previously considered property of the linear independence of minimal sets of bipolar harmonics. We take account of the fact that

$$(\mathbf{v}\cdot\nabla)Y_l(\mathbf{n}_1) = \sqrt{(2l+1)}\,l\{Y_{l-1}\otimes(\mathbf{v}\cdot\nabla)\mathbf{n}_1\}_\kappa,$$

$$(\mathbf{v}\cdot\nabla)\chi = \Lambda + \Lambda^*.$$

Then, the left-hand side of Eq. (33) can be reduced to a form with kinematic structure similar to that of Eq. (37). The coefficients of the parameters Λ, Λ^*, Υ, and Υ^* should vanish independently; this results in separation of the general system of equations into four parts. There are a number of reasons for the possibility of such a reduction. In the first place, as noted previously, the bipolar harmonics $\tilde{\mathcal{Y}}_\kappa^{k,\delta_0}$ and $\tilde{\mathcal{Y}}_\kappa^{k,\delta_1}$ for fixed κ form independent minimal sets, and consequently the general system of equations immediately separates into two parts: $\mathbb{A}_\kappa^l$ and $\mathbb{B}_\kappa^l$ determined from the first part, and the coefficients $\mathbb{C}_\kappa^l$ and $\mathbb{D}_\kappa^l$ are determined from the second part. In the second place further separation of the two systems of equations is due to the different geometric content of the scalars (39). For example, $\mathrm{Re}\,\Lambda$ and $\mathrm{Im}\,\Lambda$ are determined by the spatial gradients of the ellipticity and the rotation angle of the first kind of the ellipse, which are linearly independent vectors. This is easy to see for the two- and three-dimensional configurations of the light field. We shall choose the velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ to be the same in magnitude but different in direction. Then, in the general case these directions can be chosen so that $\Lambda_1/\Lambda_1^* \neq \Lambda_2/\Lambda_2^*$, which corresponds to the parameters $\mathrm{Re}\,\Lambda$ and $\mathrm{Im}\,\Lambda$ and, correspondingly, Λ and Λ^* being linearly independent. Therefore, the terms with Λ and Λ^* should vanish separately.

As a result of the reduction, the basic equation for finding $\mathbb{A}_\kappa^l$ becomes

$$\sum_{l=\delta_0}^{\kappa}\left[\left(\partial_\chi a_\kappa^l - \frac{l\chi a_\chi^l}{1-\chi^2}\right)\tilde{\mathcal{Y}}_\kappa^{\tilde\kappa-l,\delta_0} + \frac{a_\kappa^l\Pi^2(l)\sqrt{l(\tilde\kappa-l+1)}}{1-\chi^2}(-1)^\kappa \times\begin{Bmatrix} 1 & l-1 & l \\ \kappa & \tilde\kappa-l & \tilde\kappa-l+1\end{Bmatrix}\tilde{\mathcal{Y}}_\kappa^{\tilde\kappa-l+1,\delta_0}\right] = \frac{\gamma S}{\chi}\sum_{\kappa',\kappa''}\Pi(\kappa',\kappa'')\mathcal{F}_\kappa^{\kappa',\kappa''}\sum_{l''=(\delta_0)''}^{\kappa''}\mathbb{A}_{\kappa''}^{l''} \times\sum_{p,q=\pm1}G_{p,q}^{l'',\tilde\kappa''-l''}\tilde{Y}_\kappa^{l''+p,\tilde\kappa''-l''+q}, \tag{40}$$

where $\tilde\kappa = \kappa + \delta_0$ in accordance with the previously used notation. Then $\mathbb{B}_\kappa^l$ can be determined from Eq. (39).

The basic equation for finding $\mathbb{C}_\kappa^l$ has the form

$$i(-1)^\kappa\sum_{l=\delta_0}^{\kappa}\frac{a_\kappa^l\Pi(l,\tilde\kappa-l)\sqrt{l(l+1)}}{1-\chi^2}$$

[5] The dynamical parameters $\mathbb{B}_\kappa^l$ and $\mathbb{D}_\kappa^l$ are bound according to Eq. (39).

$$\times \sum_{h=\tilde{\kappa}-l\pm 1} \begin{Bmatrix} 1 & l & l \\ \kappa & \tilde{\kappa}-l & h \end{Bmatrix} C^{h,0}_{1,0;\,\tilde{\kappa}-l,0} \tilde{Y}^{l,h}_{\kappa}$$

$$= \frac{\gamma S}{\chi} \sum_{\kappa',\kappa'} \Pi(\kappa',\kappa') \mathcal{F}^{\kappa',\kappa'}_{\kappa} \sum_{l''=(\delta_1)''}^{\kappa''} \mathbb{C}^{l''}_{\kappa''} \tag{41}$$

$$\times \sum_{p,q=\pm 1} G^{l'',\kappa''+(\delta_1)''-l''}_{p,q} \tilde{Y}^{l''+p,\kappa''+(\delta_1)''-l''+q}_{\kappa},$$

and $\mathbb{D}^l_\kappa$ can be determined from Eq. (39).

We shall present the computational results for some of the simplest transitions ($\tilde{\delta} = \delta/\gamma$):

$$j_g = 1/2 \longrightarrow j_e = 1/2$$

$$\mathbb{A}^1_1 = \frac{3}{\chi^4 - \chi^2},$$

$$\mathbb{C}^0_1 = -\frac{3\sqrt{6}(6\tilde{\delta}(\chi^2-1) + i(1+3\chi^2))}{(\chi^2-1)(9\chi^2 - 1 + 36\tilde{\delta}^2(\chi^2-1))}, \tag{42}$$

$$\mathbb{C}^1_1 = \frac{12i\sqrt{6}\chi}{(\chi^2-1)(9\chi^2 - 1 + 36\tilde{\delta}^2(\chi^2-1))}.$$

$$j_g = 1 \longrightarrow j_e = 1$$

$$\mathbb{A}^1_1 = \frac{8(2\pi(1+4\tilde{\delta}^2) + \tilde{\delta}(3i\chi + 2\tilde{\delta}(\sqrt{30}+3\chi)))}{(1+4\tilde{\delta}^2)\chi^2(\chi^2-1)},$$

$$\mathbb{A}^0_2 = \frac{-12(10\tilde{\delta} + i\sqrt{30}\chi)}{5(-i+2\tilde{\delta})\chi(\chi^2-1)},$$

$$\mathbb{A}^1_2 = [4(4\pi(1+4\tilde{\delta}^2) + 6i\tilde{\delta}\chi + \sqrt{30}\tilde{\delta}^2$$

$$+ 4\tilde{\delta}^2(\sqrt{30} + 3\chi + \sqrt{30}\chi^2))][(1+4\tilde{\delta}^2)\chi^2(\chi^2-1)]^{-1}, \tag{43}$$

$$\mathbb{A}^2_2 = \frac{-24\tilde{\delta}}{(i+2\tilde{\delta})\chi(\chi^2-1)},$$

$$\mathbb{C}^0_1 = \frac{-8\sqrt{6}\pi}{(2\tilde{\delta}-i)(\chi^2-1)},$$

$$\mathbb{C}^1_1 = \frac{8\sqrt{6}\pi}{(2\tilde{\delta}-i)\chi(\chi^2-1)},$$

$$\mathbb{C}^1_2 = \frac{16\sqrt{3/5}\pi}{(2\tilde{\delta}-i)\chi(\chi^2-1)},$$

$$\mathbb{C}^2_2 = 0.$$

The remaining dynamical parameters are found from Eq. (39).

The parameters for the transition $j_g = 1/2 \longrightarrow j_e = 3/2$ differ from the parameters presented above for the transition $1/2 \longrightarrow 1/2$ only by the general numerical factor and the substitution $\tilde{\delta} \longrightarrow -\tilde{\delta}$. The parameters for more complicated transitions have a very complicated form and are not presented here, but we note an important property of these coefficients: the real part of $\mathbb{A}^l_\kappa$ is an even function of the detuning δ of the field, while the imaginary part of δ exhibits a dispersion dependence; for $\mathbb{A}^l_\kappa$, conversely, the real part exhibits a dispersion dependence on $\mathbb{C}^l_\kappa$, while the imaginary part is an even function of δ.

6. MODEL OF SUB-DOPPLER COOLING

6.1. Transition $1/2 \longrightarrow 1/2$ *in a Nonuniformly Polarized Light Field*

We shall consider the simplest type of resonance dipole optical transition $j_g = 1/2 \longrightarrow j_e = 1/2$ in a monochromatic light field (1). In this case the initial equations (8) and (11) can be put into a vector form by taking account of the fact that in the ground state only multipole moments of ranks $\kappa = 0$ and 1 are possible, and $\rho^g_0 = 1/\sqrt{2}$ because of the normalization condition. The multipole moment of the first rank $\rho^g_1 = \mathbf{J}(\mathbf{r})$ is a vector quantity—the optical orientation of the atomic ensemble [15]. Then the system of equations (11) reduces to a single equation for the optical orientation vector:

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\nabla\right)\mathbf{J} = -\frac{\gamma S}{9}\Big[\mathbf{J} + \frac{i}{\sqrt{2}}\mathbf{e}\times\mathbf{e}^*$$

$$+ \left(\frac{1}{2} - 3i\tilde{\delta}\right)(\mathbf{e}^*\cdot\mathbf{J})\mathbf{e} + \left(\frac{1}{2} + 3i\tilde{\delta}\right)(\mathbf{e}\cdot\mathbf{J})\mathbf{e}^*\Big], \tag{44}$$

where $\mathbf{e}$ and $\mathbf{e}^*$ are polarization vectors of the field (2) and its complex conjugate (3). The equation (44) is an equation for optical pumping of the ground state of an atomic ensemble assuming the approximations $S \ll 1$ and $\mathbf{k}\cdot\mathbf{v}/\gamma \ll 1$. Likewise, for sub-Doppler cooling the condition $\mathbf{k}\cdot\mathbf{v}/(\gamma S) \ll 1$ holds, and we shall seek the solution $\mathbf{J} \approx \mathbf{J}^{(0)} + \mathbf{J}^{(1)}$ to first-order in this small parameter inclusively for the stationary optical pumping.

The minimum set of bipolar harmonics for this case is equivalent to a local natural basis $\{\mathbf{e}, \mathbf{e}^*, \mathbf{e}_\perp = \{\mathbf{e}\otimes\mathbf{e}^*\}_1\}$. The zeroth-order solution, describing the orientation of stationary atoms,

$$\mathbf{J}^{(0)} = -\mathbf{e}_\perp, \tag{45}$$

corresponds to the previously presented solution (32).

The projections of the first correction $\mathbf{J}^{(1)}$ to the vectors of the local natural basis are

$$J_e = \mathbf{e}\cdot\mathbf{J}^{(1)} = -\frac{18}{\gamma S}\frac{(3-6i\tilde{\delta})\tilde{\Upsilon}-(1-6i\tilde{\delta})\cos 2\varepsilon\cdot\tilde{\Upsilon}^*}{9-\cos^2 2\varepsilon+36\tilde{\delta}^2\sin^2 2\varepsilon}, \quad (46)$$

$$J_{e^*} = \mathbf{e}^*\cdot\mathbf{J}^{(1)} = (J_e)^*, \quad (47)$$

$$J_\perp = \mathbf{e}_\perp\cdot\mathbf{J}^{(1)} = \frac{9}{4\gamma S}(\mathbf{v}\cdot\nabla)\sin^2(2\varepsilon), \quad (48)$$

where $\tilde{\Upsilon} = (\mathbf{v}\cdot\nabla)\underline{\mathbf{e}}\cdot\mathbf{e}_\perp = -i\sqrt{2}\,e^3\Upsilon$ is proportional to the previously determined quantity Υ (38), and the underbar denotes the quantity on which the differentiation operator ∇ acts.

The final expression for the first correction, equivalent to the expansion of the given multipole moment (orientation vector) in terms of the minimal set of bipolar harmonics in a local natural basis, has the form

$$\mathbf{J}^{(1)} = \frac{(J_e)^*-\cos 2\varepsilon J_e}{\sin^2 2\varepsilon}\mathbf{e} + \frac{J_e-\cos 2\varepsilon(J_e)^*}{\sin^2 2\varepsilon}\mathbf{e}^* + \frac{2J_\perp}{\sin^2 2\varepsilon}\mathbf{e}_\perp \quad (49)$$

and agrees with the previously presented result (42). It follows from the explicit form (49) that this solution must be examined additionally for linear polarization of the light field: $\sin 2\varepsilon \longrightarrow 0$. Taking account of the fact that in a local natural basis

$$\mathbf{e}_\perp = (\sqrt{2})^{-1}\sin(2\varepsilon)\mathbf{e}_0, \quad (50)$$

$$\mathbf{e}-\mathbf{e}^* = -\sqrt{2}\sin\varepsilon(\mathbf{e}_+e^{i\varphi_0}+\mathbf{e}_-e^{-i\varphi_0}) \quad (51)$$

($\mathbf{e}_0$ is a vector of unit length), it is easy to show that the asymptotic form of the correction (49) is

$$\lim_{\sin 2\varepsilon\to 0}\mathbf{J}^{(1)} = \frac{9\sqrt{2}}{\gamma S}(\mathbf{v}\nabla)\varepsilon(\mathbf{r})\cdot\mathbf{e}_0. \quad (52)$$

Thus, for the transition under consideration the correction linear in the velocity to the optical orientation exists and is finite for all admissible values of the polarization parameters of the light field.

Proceeding from the results (45) and (49), we can study various kinetic and optical characteristics of such an atomic ensemble. We shall confine our attention below to determining the resonance light-induced force $\mathbf{F}$ acting on the atoms in a light field. Following the results of [19], we shall employ for this the definition of this force in terms of the multipole moments of the ground state of an atomic ensemble:

$$\mathbf{F} = \frac{\hbar i/4}{\gamma/2+i\delta}(-1)^{j_g+j_e} \times\sum_{\kappa=0,1,2}\begin{Bmatrix}1 & 1 & \kappa\\ j_g & j_g & j_e\end{Bmatrix}\nabla(\hat{\rho}^g_\kappa\cdot\{\underline{\mathbf{E}}\otimes\mathbf{E}^*\}_\kappa)+\text{c.c.}, \quad (53)$$

where the operation $(\ldots\cdot\ldots)$ denotes a scalar product [7] of irreducible tensors of the same rank κ. The electric field vectors $\mathbf{E}$ and $\mathbf{E}^*$ were determined previously in Eq. (1). Then we have for the transition $1/2 \longrightarrow 1/2$

$$\mathbf{F} = \mathbf{F}^{(0)}+\mathbf{F}^{(1)} = \frac{\hbar i/12}{\gamma/2+i\delta} \times\nabla(\{\underline{\mathbf{E}}\otimes\mathbf{E}^*\}_1\cdot(\mathbf{J}^0+\mathbf{J}^1))+\text{c.c.} \quad (54)$$

6.2. Model of 2D Cooling

We shall now consider the manifestation of the force (54) in the simple 2D model of sub-Doppler cooling: the symmetric configuration of the light field is given by three coherent linearly polarized traveling waves:

$$\mathbf{E} = E_0\sum_{m=1}^{3}\mathbf{e}_m\exp(i\mathbf{k}_m\cdot\mathbf{r}) \quad (55)$$

where E_0 is the amplitude of each wave, and the wave vectors $\mathbf{k}_m$, $m = 1, \ldots, 3$ lie in the same plane making angles $2\pi/3$ with one another, so that in the distinguished Cartesian coordinate system $\{\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z\}$ (the $\mathbf{e}_z$-axis is perpendicular to the plane of the wave vectors) their explicit form is

$$\mathbf{k}_1 = k\mathbf{e}_x,\quad \mathbf{k}_2 = \frac{k}{2}(-\mathbf{e}_x+\sqrt{3}\mathbf{e}_y),\quad \mathbf{k}_3 = \frac{k}{2}(-\mathbf{e}_x-\sqrt{3}\mathbf{e}_y). \quad (56)$$

Let the linear polarization vectors $\mathbf{e}_m$, $m = 1, \ldots, 3$, make the same angles θ with the $\mathbf{e}_z$ axis. Then

$$\mathbf{e}_1 = \sin\theta\mathbf{e}_y+\cos\theta\mathbf{e}_z,\quad \mathbf{e}_2 = -\frac{1}{2}\sin\theta(\sqrt{3}\mathbf{e}_x+\mathbf{e}_y)+\cos\theta\mathbf{e}_z,\quad \mathbf{e}_3 = \frac{1}{2}\sin\theta(\sqrt{3}\mathbf{e}_x-\mathbf{e}_y)+\cos\theta\mathbf{e}_z. \quad (57)$$

For the configuration (55) the local parameters of the field can be expressed using the following invariants:

$$\mathbf{E}^*\cdot\mathbf{E} = E_0^2[3+c(ZZ^*-3)] = |\mathscr{E}|^2, \quad (58)$$

$$\mathbf{E}\cdot\mathbf{E} = E_0^2(Z^2-2(1-c)Z^*), \quad (59)$$

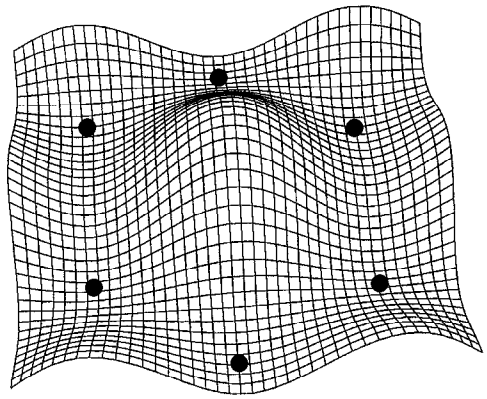

Fig. 1. Form of the potential $\Phi(\mathbf{r})$ for $\delta > 0$. The circles mark the regions of localization of the atoms.

where

$$Z = \sum_{m=1}^{3} \exp(i\mathbf{k}_m \cdot \mathbf{r}),$$

$$c = (\mathbf{e}_m \cdot \mathbf{e}_{m'})_{m \neq m'} = \cos^2\theta - \frac{1}{2}\sin^2\theta,$$

$\mathscr{E}(\mathbf{r})$ is the total amplitude of the field and c is the cosine of the angle between the polarization vectors. Therefore, the ellipticity $\varepsilon(\mathbf{r})$ of the total field, the total phase $\phi(\mathbf{r})$, and the polarization vector of the total field can be represented in the form:

$$\cos(2\varepsilon) = 1/\chi = \frac{\sqrt{(Z^2 - 2(1-c)Z^*)((Z^*)^2 - 2(1-c)Z)}}{3 + c(ZZ^* - 3)}, \quad (60)$$

$$\exp(4i\phi) = \frac{Z^2 - 2(1-c)Z^*}{(Z^*)^2 - 2(1-c)Z}, \quad (61)$$

$$\mathbf{e} = \frac{\mathbf{E}}{|\mathscr{E}|e^{i\phi}}. \quad (62)$$

For the model presented an analysis of the possible structures of two-dimensional atomic gratings, arising as a result of the force $\mathbf{F}^{(0)}$, is made in [12] for the transitions $j \longrightarrow j$ (j are half-integers). For simplicity, we shall confine our attention to the case $c = 0$, where the polarization vectors (57) are orthogonal to one another, and the intensity $\mathscr{E}^2 = 3E_0^2$ of the general field is uniform. This configuration of the light field was first considered in [20] for models of optical transitions with coherent population trapping. The uniformity of the intensity of the general light field simplifies substantially the analysis of the force $\mathbf{F}^{(1)}$, since here the light field has no nodes.

1. Let us consider the force $\mathbf{F}^{(0)}$. In general, it is determined by the gradients of the ellipticity and the total phase.[6] However, for large detunings $\tilde{\delta} \gg 1$ the ellipticity gradient makes the main contribution [12], and the force $\mathbf{F}^{(0)}$, to a high degree of accuracy, can be treated as a potential force: $\mathbf{F}^{(0)} = -\nabla\Phi$, where the potential

$$\Phi(\mathbf{r}) = \frac{\hbar S\delta}{6}\cos^2(2\varepsilon),$$

$$\cos^2(2\varepsilon) = \frac{(Z^2 - 2Z^*)((Z^*)^2 - 2Z)}{9} \quad (63)$$

$$= \frac{3 + 4\cos(3kx)\cos(\sqrt{3}ky) + 2\cos(2\sqrt{3}ky)}{9}$$

is presented in Fig. 1. The centers of localization for $\delta > 0$ in this case are found from the condition that the following invariant function is zero:

$$W = Z^2 - 2Z^* = \sum_{m=1}^{3} \exp(2i\mathbf{k}_m \cdot \mathbf{r}) = 0, \quad (64)$$

and at these points the field (5) is circularly polarized.

2. We now consider the force $\mathbf{F}^{(1)}$. We take account of the fact that for $c = 0$ the relation $\nabla(\underline{\mathbf{E}} \cdot \mathbf{E}^*) = 0$ holds. This force can be represented in the form

$$\mathbf{F}^{(1)} = \mathbf{F}_1^{(1)} + \mathbf{F}_2^{(1)}.$$

The component

$$\mathbf{F}_1^{(1)} = \mathbf{F}_{Sys} + \mathrm{F}_{\mathrm{scat}}, \quad (65)$$

$$\mathbf{F}_{Sys} = \frac{3\hbar\tilde{\delta}}{8}\mathbf{f}_1(\mathbf{f}_1 \cdot \mathbf{v}), \quad (66)$$

$$\mathrm{F}_{\mathrm{scat}} = \frac{\hbar}{192}\mathbf{f}_2 \cdot (\mathbf{f}_1 \cdot \mathbf{v}) = \hat{\mathbb{R}}_1\mathbf{v} + \mathbf{v} \times \mathbf{B}_{\mathrm{eff},1}, \quad (67)$$

where

$$\mathbf{f}_1 = \frac{\nabla\cos^2(2\varepsilon)}{|\sin(2\varepsilon)|},$$

$$\mathbf{f}_2 = i[W^*\nabla W - W\nabla W^*]$$

is due to the gradients of the ellipticity, the total phase, and the angle of rotation of the first kind of the polarization ellipse. For an atom moving toward the center of localization, it decreases to zero, just as $\mathbf{F}^{(0)}$. The components appearing in Eq. (65) correspond to the force contributions which are already known from the 1D models of sub-Doppler cooling.

Thus, the contribution $\mathbf{F}_{Sys} \sim \mathbf{f}_1$ has a dispersion dependence on the detuning and is associated only with

[6] The general configuration of the light field and for an arbitrary dipole transition the force $\mathbf{F}^{(0)}$ also depends on the gradients of the intensity and the angle of polarization-ellipse rotation of the first kind [10].

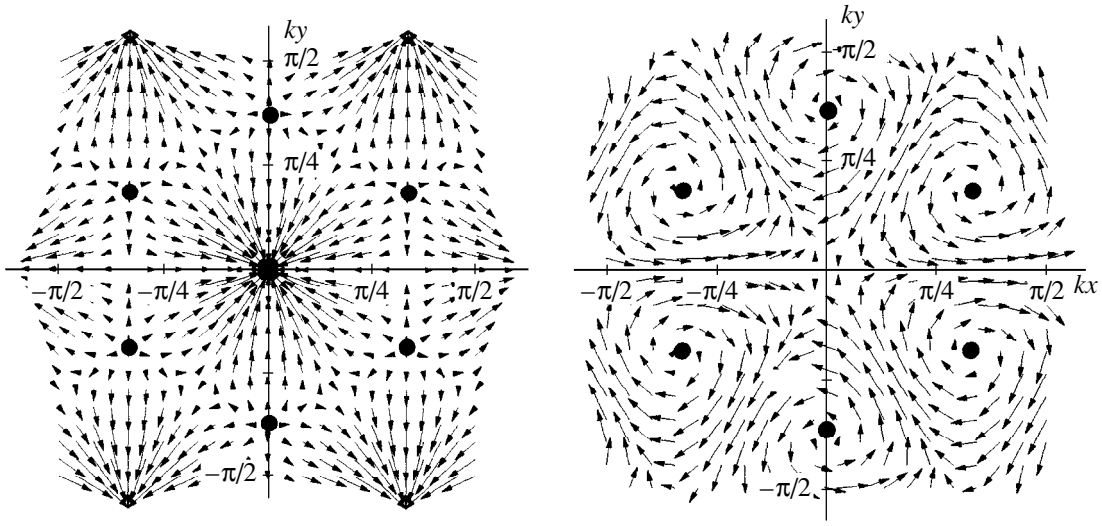

Fig. 2. The vector fields (left panel) $\mathbf{f}_1$ and (right panel) $\mathbf{f}_2$. The circles mark the regions of localization.

the spatial gradient of the ellipticity $\varepsilon(\mathbf{r})$, and the vector field $\mathbf{f}_1(\mathbf{r})$, represented in Fig. 2 (left panel), determines the direction of action of this force. Force $\mathbf{F}_{Sys}$ results in cooling of the atoms for $\delta > 0$ by the sisyphus mechanism [6], due to the redistribution of the photons scattering by the atoms between the light beams, forming the initial light field (55). We note that $\mathbf{F}_{Sys}$ predominates for large detunings $\tilde{\delta} \gg 1$.

The contribution $\mathbf{F}_{\text{scat}} \sim \mathbf{f}_2$ arises from the gradient of the ellipticity and the gradients of the angle of rotation of the first kind of the ellipse and the total phase of the light field and is due to scattering of photons of the initial light field (55) by atoms as a result of spontaneous scattering. The direction of the force is given by the vector field $\mathbf{f}_2(\mathbf{r})$ and is represented in Fig. 2 (right panel), but the magnitude of the force is proportional to the projection of the velocity of the particle on $\mathbf{f}_1$. The effect of radiation friction is characterized by the tensor

$$(\hat{\mathbb{R}}_1)_{i,j} \sim (\mathbf{f}_1)_i(\mathbf{f}_2)_j + (\mathbf{f}_2)_i(\mathbf{f}_1)_j,$$

and there is also a contribution to the "Lorentz force," where the corresponding effective field $\mathbf{B}_{\text{eff},1}(\mathbf{r}) \sim \mathbf{f}_2 \times \mathbf{f}_1$ is directed along the z axis. In contrast to $\mathbf{F}_{Sys}$, the force $\mathbf{F}_{\text{scat}}$ does not depend on the detuning for a given type of transition, but rather it is determined only by the spatial gradients of the light field. An example of a similar force in the one-dimensional model is presented in [21], where a possible cooling mechanism is also proposed for this case.

The component $\mathbf{F}_2^{(1)}$ arises from the gradients of the angles of rotation of the second kind of the polarization ellipse and its structure is more complicated. For the light-field configuration considered here it can be represented in the form

$$\mathbf{F}_2^{(1)} = \hat{\mathbb{R}}_2\mathbf{v} + \mathbf{v} \times \mathbf{B}_{\text{eff},2}, \tag{68}$$

$$\hat{\mathbb{R}}_2\mathbf{v} = 24\hbar\tilde{\delta}\cos(2\varepsilon)[\mathbf{f}_4(\mathbf{v}\cdot\mathbf{f}_4)\cos^2\varepsilon - \mathbf{f}_3(\mathbf{v}\cdot\mathbf{f}_3)\sin^2\varepsilon] + 2\hbar\cos(2\varepsilon)[\mathbf{f}_3(\mathbf{v}\cdot\mathbf{f}_4) + \mathbf{f}_4(\mathbf{v}\cdot\mathbf{f}_3)], \tag{69}$$

$$\mathbf{B}_{\text{eff},2} = \hbar\left[18\tilde{\delta}^2\sin^2(2\varepsilon) + \frac{1}{2}(3 + \cos^2 2\varepsilon)\right]\mathbf{f}_3 \times \mathbf{f}_4, \tag{70}$$

where the vector fields $\mathbf{f}_3$ and $\mathbf{f}_4$ are presented in Fig. 3 on the left and right sides, respectively, and are determined by the relations

$$\mathbf{f}_3 = \frac{1}{N}\mathrm{Re}\left(e^{-i\phi}\sum_m \mathbf{b}_m \exp(-2i\mathbf{k}_m\cdot\mathbf{r})\right),$$

$$\mathbf{f}_4 = \frac{1}{N}\mathrm{Im}\left(e^{-i\phi}\sum_m \mathbf{b}_m \exp(-2i\mathbf{k}_m\cdot\mathbf{r})\right),$$

where

$$N^2 = 54\sin^2(2\varepsilon)[9 - \cos^2(2\varepsilon) + 36\tilde{\delta}^2\sin^2(2\varepsilon)],$$

and the vectors

$$\mathbf{b}_1 = \mathbf{k}_2 - \mathbf{k}_3, \quad \mathbf{b}_2 = \mathbf{k}_3 - \mathbf{k}_1, \quad \mathbf{b}_3 = \mathbf{k}_1 - \mathbf{k}_2$$

are reciprocal-lattice vectors (e.g., $\{\mathbf{b}_1; \mathbf{b}_2\}$) and determine the spatial periodicity of the light-field configuration under study [22].

The contribution $\hat{\mathbb{R}}_2\mathbf{v}$ is a dissipative force, but the explicit anisotropy of the radiation friction forces fol-

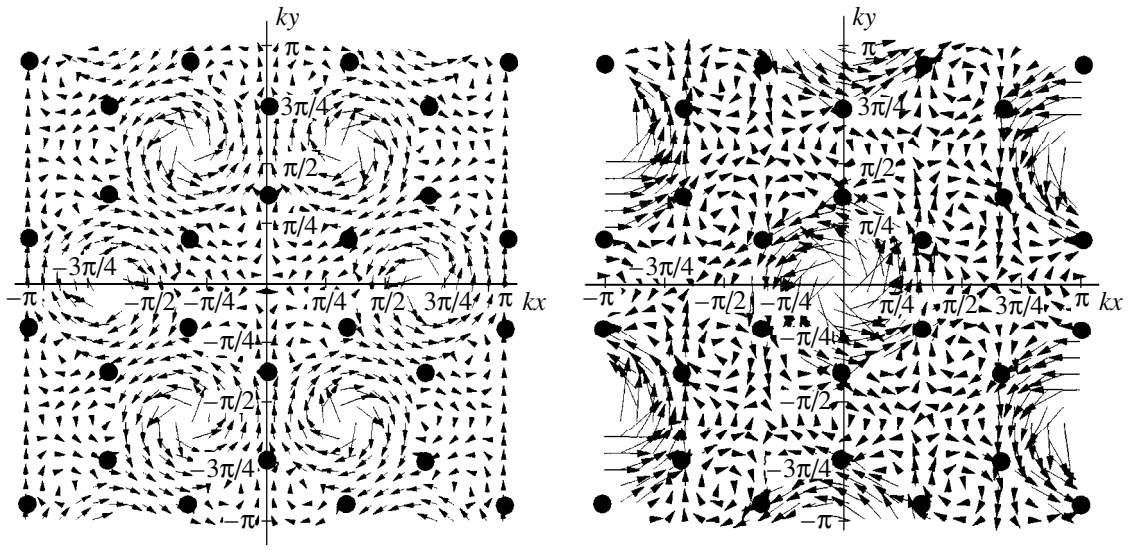

Fig. 3. The vector fields $\mathbf{f}_3$ (left side) and $\mathbf{f}_4$ (right side). The circles mark the regions of localization.

lows from its form (69). For example, for positive detunings $\delta > 0$ cooling occurs along the directions $\mathbf{f}_4$, while heating occurs along the directions $\mathbf{f}_3$. The dissipative force vanishes in the localization regions. It is important to note that for large detunings $\tilde{\delta} \gg 1$ this force decreases and, on the whole, becomes much less than the previously considered forces $\mathbf{F}_{Sys}$ and $\mathbf{F}_{scat}$. The symmetry of the vector fields $\mathbf{f}_1$ and $\mathbf{f}_2$ is determined by the reciprocal-lattice vectors $\{\mathbf{b}_3; \mathbf{b}_4\}$, so that the unit cell of the periodic structure is different from the unit cells of the previously considered fields $\mathbf{f}_1$ and $\mathbf{f}_2$. Apparently, the final spatial lattice of atoms also will be determined by the vectors $\{\mathbf{b}_1; \mathbf{b}_2\}$.

Another component contributes to the "Lorentz force," where the effective field (69) depends strongly on the detuning of the light field. Specifically, this component for large detunings predominates over the previously considered dissipative contribution, and in localization regions it does not vanish and is determined as

$$\mathbf{F}_2^{(1)}(\cos 2\varepsilon = 0) = \frac{\hbar}{108}\varepsilon_{mnp}\mathbf{b}_m(\mathbf{v}\cdot\mathbf{b}_n)\sin(2\mathbf{b}_p\cdot\mathbf{r}), \quad (71)$$

where summation over the indices is assumed, and ε_{mnp} is the Levi–Civita tensor.

The appearance of the force (69), just as the previously considered contributions (65), is due to retardation of the optical pumping of the ground state of the atoms in fields with polarization gradients. Corrections to the populations of the Zeeman sublevels $m_g = \{+1/2, -1/2\}$ in a local natural basis lead to the force (65), and these corrections arise even in one-dimensional configurations. The well-known interpretation of cooling mechanisms [6, 21] is also presented within this basis. In contrast to them, the force (69) is present only in the 2D and 3D models of the sub-Doppler cooling and is due to the appearance of coherences $\rho^g_{\pm 1/2;\,\mp 1/2}$ between the sublevels of the ground state in a local natural basis. For large detunings $\delta \gg \gamma$ this force is on the whole small compared to, e.g., $\mathbf{F}_{Sys}$. However, its influence on the formation of three-dimensional atomic gratings must be taken into account because of the different scales of the symmetry of the corresponding vector fields (Figs. 2 and 3).

7. CONCLUSIONS

The bipolar-harmonics method, studied in this paper and extended to the case of complex elliptic polarization vectors, is based on an expansion of the multipole moments which arise in physical problems in bases formed from the minimal sets of harmonics of any two chosen directions. This method is most effective in problems where there are two determining vectors, but it could also be helpful with a large number of vectors in the problem, as we showed in Section 5. The choice of determining vectors is not unique. For example, instead of the complex polarization vectors $\mathbf{e}$ and $\mathbf{e}^*$, the real vectors $\boldsymbol{\epsilon}_1 = \mathrm{Re}\,\mathbf{e}$ and $\boldsymbol{\epsilon}_2 = \mathrm{Im}\,\mathbf{e}$, giving the directions of the principal axes of the polarization ellipse, could be used. The advantage of the method is that the analysis is invariant: the basic dynamical factors are immediately identified; the number of such factors, for a number of reasons, can be much smaller than the initial dimension of the physical problem. On the other hand, the known structure of the irreducible tensors

makes it possible to analyze the characteristics of the medium (susceptibility tensor [23], kinetic coefficients) in their connection with the determining vectors of the problem. For example, in analyzing on the basis of the semiclassical approximation light-induced forces and diffusion coefficients, determining the kinetics of the cooling of atomic ensembles in light fields with 2D and 3D configurations, attention is focused on the dependence of these kinetic coefficients on the parameters of the light field: the spatial gradients of the intensity, the total phase, the ellipticity, and the rotation angle of the polarization ellipse. Depending on the type of gradient, light-induced forces are naturally classified, as one can see for the model considered in the last section. It can also be shown that the dynamical factors, which have the form of rational functions of the ellipticity parameter $\cos 2\varepsilon$ of the light field, will determine the specific nature of a particular atomic dipole transition.

ACKNOWLEDGMENTS

In conclusion, I thank A.M. Tumaĭkin, V.I. Yudin, and O.N. Prudnikov from Novosibirsk State University for their hospitality and fruitful discussions. I am especially grateful to A.V. Taichenachev for constructive remarks and his interest in this work, which made it possible to complete this work.

APPENDIX

Generalized Rotation Matrices

The parameterization of the matrices U of the SU(2) group using the Euler angles α, β, and γ [7] has the form

$$U = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} = \begin{pmatrix} \cos(\beta/2)e^{-i(\alpha+\gamma)/2} & \sin(\beta/2)e^{-i(\alpha-\gamma)/2} \\ -\sin(\beta/2)e^{i(\alpha-\gamma)/2} & \cos(\beta/2)e^{i(\alpha+\gamma)/2} \end{pmatrix}, \quad \text{(A.1)}$$

where the ranges of the parameters are $0 \le \beta \le \pi$, $0 \le \alpha \le 2\pi$, and $0 \le \gamma \le 2\pi$. In this parameterization the rotation matrix $D^{\kappa}_{\xi',\xi}(U)$ is represented in the form [8]

$$D^{\kappa}_{\xi',\xi}(\alpha,\beta,\gamma) = e^{-i(\xi'\alpha+\xi\gamma)} d^{\kappa}_{\xi'\xi}(\beta),$$

$$d^{\kappa}_{\xi'\xi}(\beta) = (-1)^{\xi}\sqrt{\frac{(\kappa-\xi)!(\kappa-\xi')!}{(\kappa+\xi)!(\kappa+\xi')!}}\cot^{\xi+\xi'}\frac{\beta}{2} \quad \text{(A.2)}$$

$$\times \sum_{j=\max(\xi,\xi')}^{\kappa} \frac{(\kappa+j)!(-1)^j}{(\kappa-j)!(j-\xi)!(j-\xi')!}\sin^{2j}\frac{\beta}{2}.$$

The direct representation $D^{\kappa}_{\xi',\xi}(U)$ in terms of the elements $u_{i,j}$ of the matrix U [8] is

$$D^{\kappa}_{\xi',\xi}(U) = \sqrt{(\kappa+\xi)!(\kappa-\xi)!(\kappa+\xi')!(\kappa-\xi')!} \times \sum_{j=\max(0,\xi-\xi')}^{\min(\kappa-\xi',\kappa+\xi)} (u_{11})^{\kappa+\xi-j}(u_{21})^{j}(u_{12})^{\xi'-\xi+j} \quad \text{(A.3)}$$

$$\times (u_{22})^{\kappa-\xi'-j}[(\kappa+\xi-j)!\,j!(\xi'-\xi+j)!(\kappa-\xi'-j)!]^{-1}.$$

The generalization of the representation (A.3) for arbitrary complex 2×2 matrices U possesses a multiplicativity property:

$$D^{\kappa}(U_1 \cdot U_2) = D^{\kappa}(U_1)\cdot D^{\kappa}(U_2). \quad \text{(A.4)}$$

If $\det U = 1$ ($U \in \mathrm{SL}(2, C)$)), then the generalized "rotation matrices" (A.3) will also satisfy the Clebsch–Gordan theorem [8]:

$$D^{\kappa_1}_{\xi_1',\xi_1}(U)D^{\kappa_2}_{\xi_2',\xi_2}(U) = \sum_{\kappa=|\kappa_1-\kappa_2|}^{\kappa_1+\kappa_2} C^{\kappa,\xi_1'+\xi_2'}_{\kappa_1,\xi_1';\kappa_2,\xi_2'} C^{\kappa,\xi_1+\xi_2}_{\kappa_1,\xi_1;\kappa_2,\xi_2} D^{\kappa}_{\xi_1'+\xi_2',\xi_1+\xi_2}(U). \quad \text{(A.5)}$$

These two properties make it possible to expand certain well-known algebraic relations from the apparatus of the quantum theory of angular momentum. For example, it is well known [8] that the matrices $U \in \mathrm{SL}(2, C)$ can be parameterized in accordance with Eq. (A.1), where now the angles α, β, and γ are complex, and their ranges are

$$0 \le \mathrm{Re}\,\beta \le \pi, \quad 0 \le \mathrm{Re}\,\alpha \le 2\pi, \quad 0 \le \mathrm{Re}\,\gamma < 2\pi.$$

The extension of the trigonometric functions of complex angles is done in the standard manner. The transformations for the Euler angles (7) from the superposition of two rotations $U = U_1 \cdot U_2$ remain valid even in the more general case of complex angles. The rotation matrices generalized in this manner are not unitary:

$$D^{\kappa}(U^{-1}) \ne (D^{\kappa}(U))^{\dagger}.$$

However, once again, the relation

$$D^{\kappa}(U^{-1}) = D^{\kappa}(-\gamma,-\beta,-\alpha) \quad \text{(A.6)}$$

and other similar symmetry properties are satisfied [7].

Expanded Spherical Functions

Expanded spherical functions (the term "expanded spherical functions" is sometimes used in the literature

to denote the Wigner D functions [24]) can be introduced into the analysis:

$$\begin{aligned}\tilde{Y}_{\kappa,\xi}(\beta,\alpha) &= \frac{\Pi(\kappa)}{\sqrt{4\pi}}(D^{\kappa})^{-1}_{0,\xi}(\alpha,\beta,\gamma)\\ &= \frac{\Pi(\kappa)}{\sqrt{4\pi}}D^{\kappa}_{0,\xi}(-\gamma,-\beta,-\alpha)\\ &= e^{i\xi\alpha}\sqrt{\frac{(2\kappa+1)(\kappa-\xi)!}{4\pi(\kappa+\xi)!}}P^{\xi}_{\kappa}(\cos\beta)\end{aligned} \tag{A.7}$$

by analogy with the definition of the spherical functions for the SU(2) group. Here $P^{\xi}_{\kappa}(\cos\beta)$ is an associated Legendre polynomial. The spherical functions defined in this manner possess the following properties:

(a) The Clebsch–Gordan expansion remains valid [7], which follows directly from Eqs. (A.5) and (A.7);

(b) these functions transform according to the law

$$\begin{aligned}&\hat{D}(\alpha,\beta,\gamma)\tilde{Y}_{\kappa,\xi'}(\vartheta',\varphi')\\ &= \sum_{\xi}\tilde{Y}_{\kappa,\xi}(\vartheta,\varphi)D^{\kappa}_{\xi,\xi'}(\alpha,\beta,\gamma),\end{aligned} \tag{A.8}$$

where ϑ, φ and ϑ', φ' are the complex spherical angles in the initial and the new coordinate systems; and,

(c) the symmetry properties

$$\begin{aligned}\tilde{Y}_{\kappa,\xi}(\beta,-\alpha) &= (-1)^{\xi}\tilde{Y}_{\kappa,-\xi}(\beta,\alpha),\\ \tilde{Y}_{\kappa,\xi}(-\beta,\alpha) &= (-1)^{\xi}\tilde{Y}_{\kappa,\xi}(\beta,\alpha).\end{aligned}$$

These functions can be represented as an irreducible tensor product of the corresponding complex vector z:

$$\begin{aligned}\tilde{Y}_{\kappa,\xi}(\vartheta,\varphi) &= \frac{1}{z^{l}}\sqrt{\frac{(2\kappa+1)!!}{4\pi\kappa!}}\\ &\times\{\ldots\{\{\mathbf{z}\otimes\mathbf{z}\}_2\otimes\mathbf{z}\}_3\ldots\otimes\mathbf{z}\}_{\kappa,\xi},\end{aligned} \tag{A.9}$$

if the parameterization vector in the form of generalized complex spherical coordinates (z, ϑ, φ) instead of the Cartesian complex coordinates (z_1, z_2, z_3) is used for this vector. The representation of the expanded spherical functions in the form (A.9) was used in [4] to analyze the tensor properties of the radiation relaxation operator and the stationary solution for the density matrix of atoms with optical orientation by elliptically polarized light.

Generalized Bipolar Harmonics

By analogy to the standard harmonics [7], we now introduce the generalized bipolar harmonics as functions of directions in a complex space, which are given by two arbitrary vectors, $\mathbf{z}$ and $\mathbf{z}'$, and are defined by the relation

$$\begin{aligned}\tilde{Y}^{l,L}_{j}(\vartheta,\varphi;\vartheta',\varphi') &= \tilde{Y}^{l,L}_{j}(\mathbf{z},\mathbf{z}')\\ &= \{\tilde{Y}_{l}(\mathbf{z})\otimes\tilde{Y}_{L}(\mathbf{z}')\}_{j}.\end{aligned} \tag{A.10}$$

Many algebraic properties of these functions are identical, because of Eqs. (A.4) and (A.5), to those for the standard bipolar harmonics [7], including properties such as the Clebsch–Gordan expansion and the transformation under a rotation of the coordinate system. In this connection, we note the reduction relation (1) for bipolar harmonics:

$$\begin{aligned}&A(0,0)\tilde{Y}^{l,L}_{j}(\mathbf{z},\mathbf{z}')\\ &= -2\Pi(l-1,L-1)(\mathbf{n}_z\cdot\mathbf{n}_{z'})\tilde{Y}^{l-1,L-1}_{j}\\ &+A(0,1)\tilde{Y}^{l,L-2}_{j}+A(1,0)\tilde{Y}^{l-2,L}_{j}-A(1,1)\tilde{Y}^{l-2,L-2}_{j},\\ &A(p,q) = \{[(p+q+(-1)^{p}l+(-1)^{q}L)^2-j^2]\\ &\times[(p+q+(-1)^{p}l+(-1)^{q}L)^2-(j+1)^2]\\ &\times[(2l+1-4p)(2l+1-4q)]^{-1}\}^{1/2},\end{aligned} \tag{A.11}$$

which makes it possible ultimately to represent $\tilde{Y}^{l,L}_{j}$ with arbitrary values $(l, L) \geq 0$ as a sum of bipolar harmonics with the order of the upper indices $(l', L') \geq 0$ satisfying the condition $2j \leq l' + L' \leq 2j + 1$. Here $\mathbf{n}_z = \mathbf{z}/z$.

REFERENCES

1. N. L. Manakov, S. I. Marmo, and A. V. Meremianin, J. Phys. B **29**, 2711 (1996).
2. N. L. Manakov, A. V. Meremianin, and A. Starace, Phys. Rev. A **61**, 022103 (2000).
3. N. L. Manakov, A. V. Meremianin, and A. F. Starace, Phys. Rev. A **57**, 3233 (1998).
4. G. Nienhuis, A. V. Taichenachev, A. M. Tumaikin, and V. I. Yudin, Europhys. Lett. **44**, 20 (1998).
5. A. V. Taichenachev, A. M. Tumaikin, and V. I. Yudin, Europhys. Lett. **45**, 301 (1999).
6. J. Dalibard and C. Cohen-Tannoudji, J. Opt. Soc. Am. B **6**, 2023 (1989).
7. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Scientific, Singapore, 1988).
8. N. Ya. Vilenkin, *Special Functions and the Theory of Group Representations* (Nauka, Moscow, 1965; American Mathematical Society, Providence, 1968).
9. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Nauka, Moscow, 1988; Pergamon, Oxford, 1975).
10. G. Nienhuis, P. van der Straten, and S.-Q. Shang, Phys. Rev. A **44**, 462 (1991).
11. A. V. Taĭchenachev, A. M. Tumaĭkin, and V. I. Yudin, Zh. Éksp. Teor. Fiz. **110**, 1727 (1996) [JETP **83**, 949 (1996)].
12. A. V. Bezverbnyi, G. Nienhuis, and A. M. Tumaikin, Opt. Commun. **148**, 151 (1998).

13. A. V. Taĭchenachev, A. M. Tumaĭkin, V. I. Yudin, and G. Nienhuis, Zh. Éksp. Teor. Fiz. **108**, 415 (1995) [JETP **81**, 224 (1995)].
14. V. S. Smirnov, A. M. Tumaĭkin, and V. I. Yudin, Zh. Éksp. Teor. Fiz. **96**, 1613 (1989) [Sov. Phys. JETP **69**, 913 (1989)].
15. K. Blum, *Density Matrix Theory and Its Applications* (Plenum, New York, 1981; Mir, Moscow, 1983).
16. A. M. Tumaĭkin, Doctoral Dissertation in Mathematical Physics (Novosibirsk State Univ., Novosibirsk, 1989).
17. A. V. Taĭchenachev, A. M. Tumaĭkin, and V. I. Yudin (2000) (in press).
18. V. Finkelstein, P. R. Berman, and J. Guo, Phys. Rev. A **45**, 1829 (1992).
19. P. Berman, G. Rogers, and B. Dubetsky, Phys. Rev. A **48**, 1506 (1993).
20. F. Mauri and E. Arimondo, Europhys. Lett. **8**, 171 (1991).
21. O. N. Prudnikov, A. V. Taĭchenachev, A. M. Tumaĭkin, and V. I. Yudin, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 439 (1999) [JETP Lett. **70**, 443 (1999)].
22. K. Petsas, A. Coates, and G. Grinberg, Phys. Rev. A **50**, 5173 (1994).
23. A. V. Bezverbnyĭ, V. S. Smirnov, and A. M. Tumaĭkin, Zh. Éksp. Teor. Fiz. **105**, 62 (1994) [JETP **78**, 33 (1994)].
24. I. M. Gel'fand, R. V. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Fizmatgiz, Moscow, 1958; Pergamon, Oxford, 1963).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 938–944.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1084–1091.

ATOMS, SPECTRA, RADIATION

Quantum-Dot Microlaser Operating on the Whispering Gallery Mode—A Source of Squeezed (Sub-Poissonian) Light

A. V. Kozlovsky* and A. N. Oraevsky**

Lebedev Physical Institute, Russian Academy of Sciences, Moscow, 117924 Russia
**e-mail: kozlovsky@neur.lpi.msk.su*
***e-mail: oraevsky@sci.lebedev.ru*

Received May 15, 2000

Abstract—The reduced density matrix method is used to calculate the quantum-statistical properties of the radiation of a quantum-dot laser operating on the whispering gallery mode of a dielectric microsphere. It is shown that under the conditions of strong coupling between the quantum dot and an electromagnetic field the radiation of such a laser can be in a nonclassical (sub-Poissonian) state. The laser scheme considered is characterized by an extremely low lasing threshold and a small number of saturation photons, as result of which lasing is possible with close to zero population inversion of the working levels, if $g \gg P \gg \gamma \gg \Gamma$, where g is the field–matter interaction constant, P is the pumping rate, γ is the loss rate of the resonator, and Γ is the spontaneous emission rate. The largest squeezing inside the resonator–microsphere (the Fano factor $F = 0.75$) obtains for $g \gg P \gg \gamma \gg \Gamma$, and the greatest squeezing in the fluctuation spectrum outside the resonator [$V(\omega = 0) \approx 0.25$] occurs for $g \sim P \sim \gamma \gg \Gamma$, and in this case a substantial deviation of the photon number statistics of the radiation leaving the resonator from the Poissonian statistics is observed. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Quantum dots are nanoscopic semiconductor quantum structures which possess a discrete spectrum of electronic states [1–3]. The single-atom microlaser is now an object of active theoretical and experimental investigations, which make it possible to study the fundamental quantum properties of the interaction of matter and light for the example of such an elementary light source.

As shown in [1,3], a dielectric microsphere with an atomic transition which is in resonance with the whispering gallery mode with its inherent high Q (up to 10^{10}) makes it possible to develop a microlaser with a record low threshold of excitation and a narrow spectral line.

In the present work we examine a four-level laser scheme based on a quantum dot with parallel pumping [3–5]. The quantum dynamics of the laser was calculated by finding an accurate numerical solution of the equations of motion for the reduced density operator in a basis of Fock states of the single-mode electromagnetic field of the whispering gallery mode of a microsphere interacting with the quantum dot. The conditions for generation of a squeezed (sub-Poissonian) state of light in the strong-coupling regime, which we found in a previous work [6] and which is characterized by the following relations between the field–matter interaction constant (g), the pumping rates (P_i), the loss rate of the resonator (γ), and the spontaneous emission rate (Γ), were determined: $P_i, \gamma \gg \Gamma$ and $P_i, \gamma \sim g$. It was established that under the optimal conditions 25% squeezing can be obtained inside the resonator and four-fold squeezing of the spectrum of the intensity fluctuations of the output radiation can be obtained at zero frequency. In a previous work [6] we showed that in the strong-coupling regime a two-level single-atom microlaser can produce a squeezed state of light with indicators of squeezing under the stationary conditions much smaller than in the four-level scheme with parallel pumping studied in the present work. At the same time, in the transitional regime squeezing in a two-level scheme can reach a factor of 10 at the output of the resonator and exceed the corresponding value for a four-level laser. The whispering gallery mode used as the resonator with high Q and a small effective mode volume [1] makes it possible to obtain the strong-coupling regime necessary for the generation of the squeezed state of light [6].

2. MODEL OF A QUANTUM DOT LASER

Since the size of a quantum dot is 10–30 nm, we can assume that only one discrete energy level for an electron and one level for a hole are present in it. As result of the Coulomb blockade effect [2, 3], no more than one electron and hole can be present simultaneously in each energy level. The excited state of a quantum dot (exciton) arises when both types of carriers are present simultaneously in the potential wells corresponding to each of them. An optical transition (λ = 980 nm for InAs/GaAs) occurs with the mutual annihilation of an

1063-7761/00/9105-0938$20.00

electron and a hole, as a result of which the quantum dot is transferred into the ground state (empty quantum dot). If only one carrier (electron or hole) enters the quantum dot by tunneling through the potential barrier, there arises a semiexcited [3] state of the quantum dot that does not interact with the electromagnetic field. A transition from a semiexcited state into an excited state is possible if a carrier with the opposite sign enters the quantum dot.

A quantum dot is coupled with the whispering gallery mode of the microsphere [1] placed nearby and is the active medium of the microlaser.

Following [5], we shall distinguish two semiexcited states [3] and study two levels $|C\rangle$ and $|D\rangle$ (through which the upper laser level $|A\rangle$ is pumped) arising when only one electron or only one hole is present in the quantum dot. In this excitation scheme there arises four-level laser model operating on a two-level quantum dot with parallel pumping (Fig. 1).

We shall neglect the process where the excited state $|A\rangle$ decays by means of the loss of one of the carriers and transitions into semiexcited states, and we shall also neglect the direct pumping arising when carriers with different signs have entered the quantum dot simultaneously.

The coupling constant between the quantum dot and the field of the whispering gallery mode in the dipole interaction approximation was taken to be real

$$g = d_{AB}\sqrt{\frac{2\pi\omega_{AB}}{\hbar V_{\text{eff}}}}, \tag{1}$$

where V_{eff} is the effective volume of the whispering gallery mode [1] interacting with the quantum dot, and d_{AB} is the matrix element of the dipole moment operator of the quantum dot characterizing the transition $|A\rangle \longrightarrow |B\rangle$ with the emission of one photon into the resonance whispering gallery mode of the microsphere.

We shall analyze the quantum stochastic dynamics of a four-level quantum-dot laser using a reduced density operator of a system consisting of the quantum dot and a single-mode field in a Fock basis:

$$\rho(t) = \sum_{x, y = \{A, B, C, D\}} \sum_{n, m = 0}^{\infty} \rho_{xy; nm}(t)|x\rangle|n\rangle\langle m|\langle y|. \tag{2}$$

We shall place the reference point of the energy of the quantum dot midway between the energies of the states $|A\rangle$ and $|B\rangle$, and we shall assume that the energies of the conditional semiexcited states $|C\rangle$ and $|D\rangle$ are also equally spaced from the energies of the laser states $|A\rangle$ and $|B\rangle$. Under these conditions, in the interaction representation and the Born–Markov approximation [6, 7] the reduced density operator of the atom–field

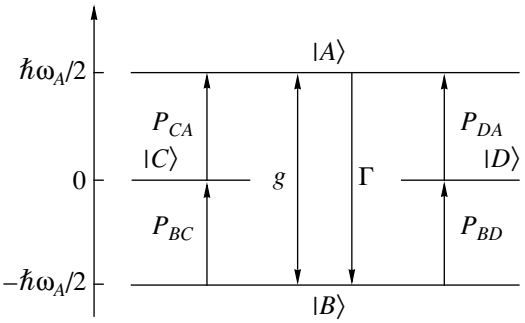

Fig. 1. Arrangement of the energy levels of a quantum dot. The states $|A\rangle$ and $|B\rangle$ are coupled with one another by the laser dipole transition with frequency ω_A and coupling constant g; Γ is the spontaneous emission rate: the states $|C\rangle$ and $|D\rangle$ correspond to the energy levels of a semiexcited quantum dot (see explanation in the text; see also [3, 5]); P_{xy} ($xy = BC, BD, CA, DA$) are the rates of pumping of the upper laser level through semiexcited states.

system (2), interacting with a reservoir, satisfies a Liouville equation of motion of the following form:

$$\begin{aligned}
\frac{\partial\rho}{\partial t} &= \frac{\partial}{\partial t}\mathrm{Sp}_R(\sigma_{R\oplus S}) = -i\Delta\left[\frac{\sigma_{AA}-\sigma_{BB}}{2}, \rho\right] \\
&- ig[(a^+\sigma_{AB} + \sigma_{BA}a), \rho] \\
&+ \frac{\gamma}{2}(n_T+1)(2a\rho a^+ - a^+a\rho - \rho a^+a) \\
&+ \frac{\gamma}{2}n_T(2a^+\rho a - aa^+\rho - \rho aa^+) \\
&+ \frac{\Gamma}{2}(N_T+1)(2\sigma_{AB}\rho\sigma_{BA} - \sigma_{BA}\sigma_{AB}\rho - \rho\sigma_{BA}\sigma_{AB}) \\
&+ \frac{\Gamma}{2}N_T(2\sigma_{BA}\rho\sigma_{AB} - \sigma_{AB}\sigma_{BA}\rho - \rho\sigma_{AB}\sigma_{AB}) \\
&+ \sum_{\{xy\} = CA, BC, BD, DA}\left\{\frac{P_{xy}}{2}(2\sigma_{xy}\rho\sigma_{yx} - \sigma_{yx}\sigma_{xy}\rho \right. \\
&\left. - \rho\sigma_{yx}\sigma_{xy}) + \frac{P_{yx}}{2}(2\sigma_{yx}\rho\sigma_{xy} - \sigma_{xy}\sigma_{yx}\rho - \rho\sigma_{xy}\sigma_{yx})\right\},
\end{aligned} \tag{3}$$

where $\Delta = \omega_{AB} - \omega_c$ is the detuning of the resonator frequency from the transition frequency, and the operators referring to the quantum dot have the following form: $\sigma_{xy} = |y\rangle\langle x|$, $x, y = A, B, C, D$. The quantities γ, Γ, and P_{xy} are, respectively, the rate of loss of the field at the mirrors, the spontaneous emission rate, and the incoherent pumping rate. The quantities P_{yx} are the rates of de-excitation of the excited and semiexcited states and will be neglected in the calculations below. The equilibrium average numbers of photons of the thermostat and excitation of the quantum dot at temperature T are denoted in Eq. (3) as n_T and N_T, respectively. Both

quantities are small in the optical frequency range of the field at moderate temperatures.

In the basis of states of the quantum dots the components of the reduced density operator of the system

$$\rho_{xy}(t) \equiv \langle x|\rho|y\rangle, \quad x, y = A, B, C, D,$$

which depend only on the electromagnetic-field variables, satisfy the following equations of motion obtained from Eqs. (3):

$$\begin{aligned}
\dot\rho_{AB} &= -\frac{\Gamma + P_{BD} + P_{BC}}{2}\rho_{AB} \\
&- i\Delta\rho_{AB} - ig(a\rho_{BB} - \rho_{AA}a) + L\rho_{AB}, \\
\dot\rho_{BA} &= -\frac{\Gamma + P_{BD} + P_{BC}}{2}\rho_{BA} \\
&+ i\Delta\rho_{BA} + ig(\rho_{BB}a^+ - a^+\rho_{AA}) + L\rho_{BA}, \\
\dot\rho_{CD} &= -\frac{P_{BD} + P_{BC} + P_{CA}}{2}\rho_{CB} \\
&- i\frac{\Delta}{2}\rho_{CB} + ig\rho_{CA}a + L\rho_{CB}, \\
\dot\rho_{BC} &= -\frac{P_{BD} + P_{BC} + P_{CA}}{2}\rho_{BC} \\
&- i\frac{\Delta}{2}\rho_{BC} - iga^+\rho_{AC} + L\rho_{BC}, \\
\dot\rho_{CA} &= -\frac{\Gamma + P_{CA}}{2}\rho_{CA} + i\frac{\Delta}{2}\rho_{CA} + ig\rho_{CB}a^+ + L\rho_{CA}, \\
\dot\rho_{AC} &= -\frac{\Gamma + P_{CA}}{2}\rho_{AC} - i\frac{\Delta}{2}\rho_{AC} - iga\rho_{BC} + L\rho_{AC}, \\
\dot\rho_{BD} &= -\frac{P_{DA} + P_{BD} + P_{BC}}{2}\rho_{BD} \\
&+ i\frac{\Delta}{2}\rho_{BD} - iga^+\rho_{AD} + L\rho_{BD}, \\
\dot\rho_{DB} &= -\frac{P_{DA} + P_{BD} + P_{BC}}{2}\rho_{DB} \\
&- i\frac{\Delta}{2}\rho_{DB} + ig\rho_{DA}a + L\rho_{DB}, \\
\dot\rho_{DA} &= -\frac{\Gamma + P_{DA}}{2}\rho_{DA} + i\frac{\Delta}{2}\rho_{DA} + ig\rho_{DB}a^+ + L\rho_{DA,} \\
\dot\rho_{AD} &= -\frac{\Gamma + P_{DA}}{2}\rho_{AD} - i\frac{\Delta}{2}\rho_{AD} - iga\rho_{BD} + L\rho_{AD}, \\
\dot\rho_{AA} &= -\Gamma\rho_{AA} + P_{CA}\rho_{CC} + P_{DA}\rho_{DD} \\
&- ig(a\rho_{BA} - \rho_{AB}a^+) + L\rho_{AA}, \\
\dot\rho_{BB} &= \Gamma\rho_{AA} - (P_{BD} + P_{BC})\rho_{BB} \\
&+ ig(\rho_{BA}a - a^+\rho_{AB}) + L\rho_{BB}, \\
\dot\rho_{CC} &= -P_{CA}\rho_{CC} + P_{BC}\rho_{BB} + L\rho_{CC}, \\
\dot\rho_{DD} &= P_{BD}\rho_{BB} - P_{DA}\rho_{DD} + L\rho_{DD}.
\end{aligned} \tag{4}$$

Here the operators representing the resonator losses caused by the interaction with the vacuum modes of the field have the form

$$L\rho_{xy} \equiv \frac{\gamma}{2}(2a\rho_{xy}a^+ - a^+a\rho_{xy} - \rho_{xy}a^+a). \tag{5}$$

It follows from the equations of motion (4) for the polarization operators coupling the nonlaser states with the laser states ρ_{xy} ($xy = CB, BC, AD, DA, CA, AC, BD, DB$) that if initially the values of these quantities are zero, then these variables remain zero fro all time. Consequently, in the calculations below the equations for the indicated polarizations of the quantum dot will be neglected, since the initial values will be taken as zero.

The equations for the matrix elements of the reduced density operator, which we solved numerically, can be obtained from Eqs. (4) in the Fock representation

$$\rho_{xy;\,nm}(t) \equiv \langle x|\langle n|\rho|m\rangle|y\rangle, \tag{6}$$

$$\begin{aligned}
xy = {} & AB, BA, CB, BC, CA, AC, BD, DB, \\
& DA, AD, AA, BB, CC, DD.
\end{aligned}$$

The system of equations with dimension $6 \times (n_{\max} + 1) \times (n_{\max} + 1)$, where $n_{\max}$ is the size of the basis of Fock states, was solved numerically by the fourth-order Runge–Kutta method. Initially, the field was in a pure vacuum state, and the quantum dot was in the lower state. Thus the density matrix of the quantum dot and the field, which did not interact with one another at the moment $t = 0$, is

$$\begin{aligned}
\rho(0) &= \rho_a \otimes \rho_f, \\
\rho_a &= |B\rangle\langle B|, \quad \rho_f = |0\rangle\langle 0|.
\end{aligned} \tag{7}$$

The statistical average values of the field and populations of the states of the quantum dot were determined as follows:

$$\langle n(t)\rangle = \sum_x \mathrm{Tr}(\rho_{xx}(t)) = \sum_{x = A, B, C, D}\ \sum_{n = 0}^{\infty} \rho_{xx;\,n,n}(t), \tag{8}$$

$$\langle P_x(t)\rangle = \mathrm{Tr}(\rho(t)\sigma_{xx}) = \sum_{n = 0}^{\infty} \rho_{xx;\,n,n}(t). \tag{9}$$

The variance (fluctuations inside the resonator) was found from the expression

$$\begin{aligned}\mathrm{var}(n(t)) &\equiv \langle(\Delta n(t))^2\rangle \\ &= \sum_{x=A,B,C,D}\sum_{n=0}^{\infty}(n-\langle n(t)\rangle)^2\rho_{xx;\,n,n}(t).\end{aligned} \tag{10}$$

The variances of the canonically conjugate quadratures of the field $X_+(t) = [a^+(t) + a(t)]/2$ and $X_-(t) = [a^+(t) - a(t)]/2i$ can be expressed in terms of the matrix elements of the density operator as

$$\begin{aligned}\langle(\Delta X_\pm)^2\rangle &= \frac{1}{4}\sum_{x=A,B,C,D}\left\{\sum_{n=0}^{\infty}(2n+1)\rho_{xx;\,n,n}(t)\right. \\ &\pm\sum_{n=2}^{\infty}\sqrt{n(n-1)}\rho_{xx;\,n,n-2}(t) \\ &\pm\sum_{n=0}^{\infty}\sqrt{(n+1)(n+2)}\rho_{xx;\,n,n+2}(t) \\ &\left.\mp\left[\sum_{n=0}^{\infty}\sqrt{n+1}\rho_{xx;\,n,n+1}(t)\pm\sum_{n=1}^{\infty}\sqrt{n}\rho_{xx;\,n,n-1}(t)\right]^2\right\}.\end{aligned} \tag{11}$$

It is assumed that inside the microsphere, which is the laser resonator, the electromagnetic field is in a state with discrete values of the frequencies (whispering gallery mode), while outside the resonator the field possesses a continuous spectrum. Consequently, the temporal fluctuations of the field inside the resonator are sources of fluctuations of the frequency spectrum of the radiation exiting through the surface of the microsphere. The field outside the spherical resonator can be represented as a sum of the laser field exiting the cavity and the noise field of the reservoir–thermostat, incident on the surface of the microsphere, i.e., $a^{\mathrm{out}}(t) = b^{\mathrm{in}}(t) + \sqrt{\gamma}\,a(t)$, where $b^{\mathrm{in}}(t)$ is the operator of the vacuum field incident on the surface of the microsphere [8–10]. The Heisenberg operator $a^{\mathrm{out}+}(t)a^{\mathrm{out}}(t)$ is the operator for the number of photons exiting through the surface of the microsphere per unit time. The quantity characterizing the statistics of the laser radiation outside the microsphere is the stationary spectrum of fluctuations of the form

$$\begin{aligned}V^{\mathrm{out}}(\omega) &= \lim_{t\to\infty}2\int_0^{\infty}d\tau\cos(\omega\tau) \\ &\times[\langle n^{\mathrm{out}}(t+\tau)n^{\mathrm{out}}(t)\rangle - \langle a^{\mathrm{out}+}(t+\tau)a^{\mathrm{out}}(t)\rangle^2].\end{aligned} \tag{12}$$

Since the two-time correlators under stationary conditions are even functions of t, the Fourier cosine transform is used in Eq. (12).

The commutation relations for the field operators, forming a continuous spectrum outside the cavity have the form [8–10]

$$[a^{\mathrm{out}}(t+\tau), a^{\mathrm{out}+}(t)] = \delta(\tau). \tag{13}$$

Using Eq. (13), we obtain the two-time correlation function of the photon number operators:

$$\begin{aligned}\langle n^{\mathrm{out}}(t+\tau)n^{\mathrm{out}}(t)\rangle &\equiv \langle a^+(t+\tau)a(t+\tau)a^+(t)a(t)\rangle^{\mathrm{out}} \\ &= \langle a^+(t+\tau)a(t)\rangle^{\mathrm{out}}\delta(\tau) \\ &+ \langle a^+(t)a^+(t+\tau)a(t+\tau)a(t)\rangle^{\mathrm{out}}.\end{aligned} \tag{14}$$

Thus the stationary spectrum (SS) of the fluctuations of the number of field photons at the cavity exit consists of the shot noise and the chronologically and normally ordered spectrum of the fluctuations:

$$V^{\mathrm{out}}(\omega) = \langle n^{\mathrm{out}}(t_{\mathrm{SS}})\rangle + :V^{\mathrm{out}}(\omega): . \tag{15}$$

In the case of one transmitting mirror the correlators of the field of the discrete mode of the radiation inside the microsphere are related with the correlators of the fields of the continuous spectrum outside the cavity, as shown in [8–10], as follows:

$$\langle a^+(t)a(t)\rangle^{\mathrm{out}} = \gamma\langle a^+(t)a(t)\rangle, \tag{16}$$

$$\begin{aligned}&\langle a^+(t)a^+(t+\tau)a(t+\tau)a(t)\rangle^{\mathrm{out}} \\ &= \gamma^2\langle a^+(t)a^+(t+\tau)a(t+\tau)a(t)\rangle.\end{aligned} \tag{17}$$

Substituting Eqs. (16) and (17) into Eq. (12), we find finally for the Fano spectral factor the formula

$$\begin{aligned}F(\omega) \equiv \frac{V^{\mathrm{out}}(\omega)}{\langle n\rangle^{\mathrm{out}}} &= 1 + \lim_{t\to\infty}\frac{2\gamma}{\langle a^+(t)a(t)\rangle} \\ &\times\int_0^{\infty}d\tau[\langle a^+(t)a^+(t+\tau)a(t+\tau)a(t)\rangle \\ &- \langle a^+(t+\tau)a(t+\tau)\rangle\langle a^+(t)a(t)\rangle]\cos(\omega\tau).\end{aligned} \tag{18}$$

The quantity $F(\omega) \geq 0$ assumes values less than 1 for the field outside the microsphere in a nonclassical state. For $F(\omega) = 0$ there is a shot noise level for all frequencies ω, i.e., the field is in a coherent state. The parameter characterizing the photon statistics outside the microsphere is the Mandel parameter

$$\begin{aligned}Q_{\mathrm{out}} \equiv \frac{\langle:\Delta n^{+\mathrm{out}}(t)\Delta n^{\mathrm{out}}(t):\rangle}{n^{\mathrm{out}}(t)} &= \frac{g}{\pi}\int_0^{\infty}[F(\tilde{\omega})-1]d\tilde{\omega}, \\ \tilde{\omega} &= \frac{\omega}{g}.\end{aligned} \tag{19}$$

The Fano factor is related with the Madnel parameter by the relation $F = Q + 1$; for the field in a nonclassical squeezed state the Mandel parameter is negative. We note that local squeezing in the spectrum (18) may not

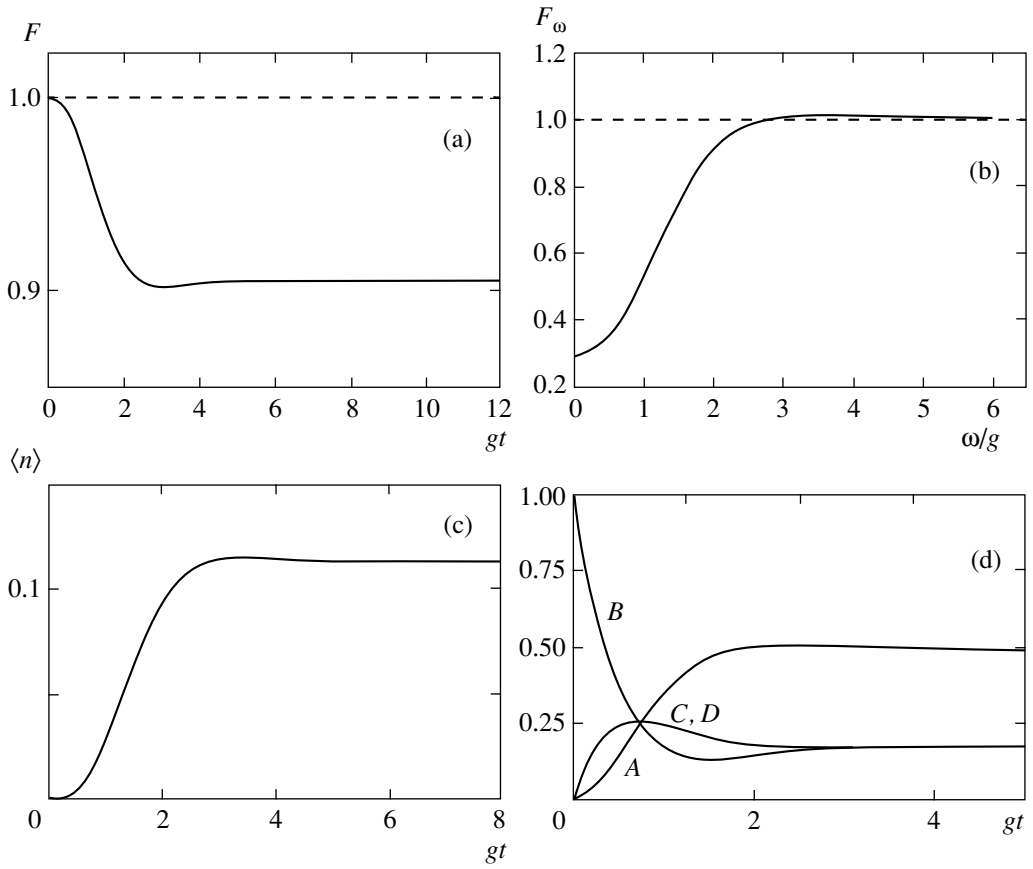

Fig. 2. (a) Fano factor for the radiation inside a quantum-dot laser resonator versus the dimensionless time for $P_{xy} = 3$, $\gamma = 1$, and $\Gamma = 0.001$ in units of g. (b) The spectrum of the fluctuations of the number of radiation photons outside the laser cavity. (c) Dependence of the average number of photons of radiation inside the quantum-dot laser resonator versus the dimensionless time for the same values of the parameters. (d) Populations of the states of a quantum dot versus the dimensionless time for the same values of the parameters.

lead to a frequency-integrated nonclassical sub-Poissonian photon-number distribution.

The two-time correlation functions appearing in Eq. (18) was found as follows. Under stationary conditions it is easy to obtain for the correlation functions of the field operators inside the resonator the following expression from the quantum regression theory (see, for example, [11]):

$$\langle a^+(t)a^+(t+\tau)a(t+\tau)a(t)\rangle_{SS} = \mathrm{Tr}(a^+a\tilde{\rho}(\tau)), \quad (20)$$

where the operator $\tilde{\rho}(\tau) \equiv \tilde{\rho}(t+\tau)$ satisfies the Liouville equation (3) with the initial condition ($\tau = 0$)

$$\tilde{\rho}_{n,m}(0) = \sqrt{(n+1)(m+1)}\rho_{n+1,m+1}(t_{SS}), \quad (21)$$

where $\rho_{n+1,m+1}(t_{SS})$ are the stationary values of the density matrix of the system.

3. STRONG COUPLING CONDITION AND GENERATION NONCLASSICAL LIGHT

In [6] we established that a single-atom two-level laser with incoherent pumping is capable of generating nonclassical radiation with sharply sub-Poissonian photon statistics. The necessary condition for generating such light is that the spontaneous emission rate must be small compared to the cavity pumping and loss rates. A necessary and sufficient condition for the radiation to be nonclassical is the so-called strong-coupling condition, when the field–atom coupling constant is of the order of the loss rate of the resonator and the pumping rate and much greater than the spontaneous emission rate, i.e., $g \sim P_{xy}$, $\gamma \gg \Gamma$. It should be noted that such a regime is different from the strong-coupling regime studied in, for example, [12] with $g \gg \gamma, \Gamma$, where the fluorescence spectrum possesses the characteristic Rabi doublet structure.

In this work, a strong-coupling regime similar to the one described in [6] is studied for a quantum-dot laser in a four-level scheme under conditions where the pump rates P_{xy} are all the same. Calculations using the relations (4)–(6) showed that, just as in the case of the two-level one-atom laser, a substantial degree of compression of the light generated is possible under stationary conditions as well as in a transitional regime. Figure 2a shows the dependence of the Fano factor $F = \langle(\Delta n)^2\rangle/\langle n\rangle$

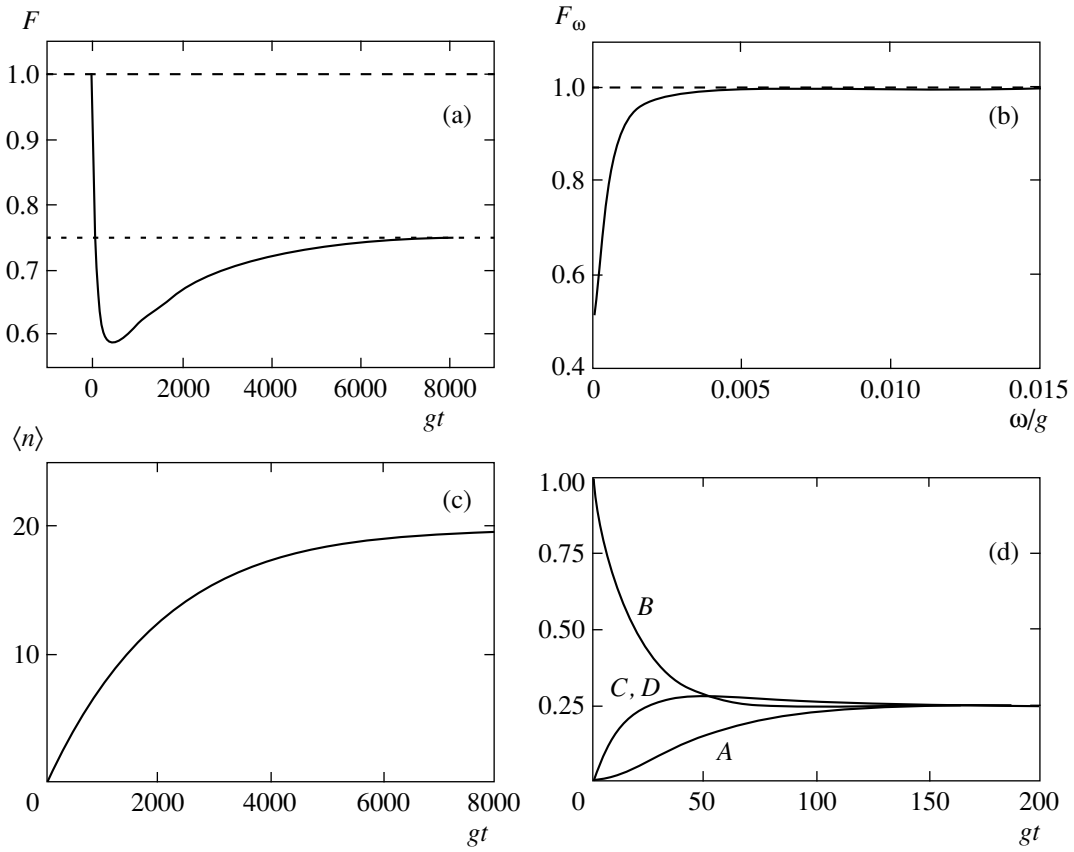

Fig. 3. Same as Fig. 2, for $P_{xy} = 0.02$, $\gamma = 0.005$, and $\Gamma = 0.0001$ in units of g.

inside the resonator on the dimensionless time for $P = 1$, $\gamma = 3$, and $\Gamma = 0.001$ in units of the coupling constant g, i.e., under strong-coupling conditions. As one can see from Fig. 2a, in this case sub-Poissonian photon statistics occurs in the transitional regime ($F_{\min} = 0.901$) and under stationary conditions $F_{SS} = 0.905$. It was found in [6] that a single-atom two-level microlaser with $\Gamma \ll \gamma \ll g \sim P_{xy}$ in the transitional regime produces a field with $F \ll 1$. The calculations performed in the present work show that for a quantum-dot microlaser transitional squeezing of the field intensity cannot exceed its stationary value by so much. Figures 2c and 2d show the quantum-mechanical averages of the number of photons and the populations as a function of time for the same values of the laser parameters. We note that in the process of establishing a stationary state, at some moment in time all populations of the states of the quantum dot become equal to one another, which is a characteristic feature of this four-level scheme and occurs for any values of the laser parameters. At the same time, as follows from Fig. 2d, under stationary conditions, for $g \sim P \sim \gamma \gg \Gamma$, there is a substantial population inversion of the laser levels of the quantum dot.

It is evident from Fig. 2b that substantial squeezing of the output radiation of the laser is present in the spectrum of the photon number fluctuations near zero frequency: $F(\omega = 0) = 0.289$. The Mandel parameter in this case is $Q_{out} = -0.29$, i.e., the distribution of outgoing radiation photons is sub-Poissonian.

The calculations show that the internal squeezing is minimal for strong-coupling conditions of the form $g \gg P \gg \gamma \gg \Gamma$. Varying the parameters P_{xy} and γ showed that the maximum squeezing attainable in this scheme inside the cavity is 25% (the Fano factor of $F \geq 0.75$). Figure 3 displays the computational results for close to optimal values of P and γ; here the squeezing inside the cavity $F = 0.76$ (see Fig. 3a). Squeezing in the outgoing radiation spectrum is substantial only for frequencies $\omega < \gamma$; the maximum squeezing obtains at zero frequency: $F(\omega = 0) = 0.504$ (see Fig. 3b). The Mandel parameter in this case $Q_{out} = -10^{-4}$, i.e., the state of the outgoing radiation is close to coherent. Thus, when the rate of resonator losses is small compared to be coupling constant g and the pumping rate, substantial squeezing of the fluctuations of the number of photons of intracavity radiation leads to a nearly Poissonian distribution of the photons outside the cavity.

The dynamics of the establishment of stationary lasing in the case $g \gg P \gg \gamma \gg \Gamma$ is of special interest. As one can see from Fig. 3d, the point where all level populations are the same is obtained even at the early stage of the evolution of the laser, when the average number

of photons in the cavity is small compared to the stationary value. In the stationary state the populations of all four levels become the same, i.e., generation occurs in the virtual absence of a population inversion for the working transition (the inversion does not exceed 10^{-5}).

Our calculations for $\Delta \neq 0$ showed that the detunings of the frequency of the resonator field from the transition frequency lead only to a decrease of the squeezing inside and outside the cavity.

As noted in [13–15], noise suppression is a characteristic property of multilevel lasers, including our quantum-dot laser. Under the conditions of the scheme with parallel pumping the residence of carriers in intermediate semiexcited states results in a time delay or a memory effect [14], a consequence of which is regularization of the pumping of the excited state, since a quantum dot in a semiexcited state no longer interacts directly with the field. An important factor making possible generation of a squeezed state is the equality of the rates of pumping performed synchronously from both semiexcited states.

4. CONCLUSIONS

The calculations performed in the present work have shown that an important condition for the generation of a squeezed state of light by a quantum-dot laser is that the coupling constant between the matter and the electromagnetic field must be large compared with the spontaneous emission rate. Using the whispering gallery mode of a dielectric microsphere, having a small effective mode volume [1], as a laser resonator makes it possible to increase the coupling constant to 10^9 Hz [1, 3]. The characteristic spontaneous emission rate of a quantum dot in empty space is of the order of 10^{10} Hz. However, the presence of a material body (dielectric microsphere) substantially changes the mode density of the electromagnetic field in the space surrounding the body and therefore the distribution of the zero fluctuations of the field and, consequently, the spontaneous emission rate [16–18]. Depending on the transition frequency of the quantum dot and the position of the dot relative to the surface of the microsphere, the spontaneous emission rate of the quantum dot can differ from the corresponding value in empty space [19] and therefore it can be substantially lower.

Another necessary condition for achieving substantial squeezing of intensity is the relation between the coupling constant and the radiation lost rate in the cavity $g \gg \gamma$. This condition can also be satisfied in the case of the whispering gallery mode because of its high Q [1]. On this basis the strong-coupling condition $g \gg P \gg \gamma \gg \Gamma$, permitting generation of nonclassical light by a quantum-dot laser, can be satisfied.

A characteristic feature of a quantum-dot laser operating on the whispering gallery mode is an extremely low lasing threshold, as a result of which it is possible to produce intense radiation with incoherent pumping under close to zero population inversion conditions.

REFERENCES

1. A. N. Oraevskiĭ, M. O. Scully, and V. L. Velichanskiĭ, Kvantovaya Élektron. (Moscow) **25** (3), 211 (1995).
2. A. Imamoglu and Y. Yamamoto, Phys. Rev. Lett. **72**, 210 (1994).
3. C. Wiele, F. Haake, C. Rocke, and A. Wixforth, Phys. Rev. A **58**, R2680 (1998).
4. M. Pelton and Y. Yamamoto, Phys. Rev. A **59**, 2418 (1999).
5. O. Benson and Y. Yamamoto, Phys. Rev. A **59**, 4756 (1999).
6. A. V. Kozlovskiĭ and A. N. Oraevskiĭ, Zh. Éksp. Teor. Fiz. **115**, 1210 (1999) [JETP **88**, 666 (1999)].
7. H. Haken, *Light* (Elsevier, Amsterdam, 1985), Vol. 1.
8. C. W. Gardiner and M. J. Collett, Phys. Rev. A **31**, 3761 (1985).
9. M. J. Collett and C. W. Gardiner, Phys. Rev. A **30**, 1386 (1984).
10. H. J. Carmichael, J. Opt. Soc. Am. B **4**, 1588 (1987).
11. M. Lax, *Fluctuations and Coherence Phenomena in Classical and Quantum Physics*, Ed. by M. Chretien, E. P. Gross, and S. Deser (Gordon and Breach, New York, 1968; Mir, Moscow, 1974).
12. H. J. Carmichael, R. J. Brecha, M. J. Raizen, and H. J. Kimble, Phys. Rev. A **40**, 5516 (1989).
13. H.-J. Briegel, G. M. Meyer, and B.-G. Englert, Phys. Rev. A **53**, 1143 (1996).
14. G. M. Meyer and H.-J. Briegel, Phys. Rev. A **58**, 3210 (1998).
15. Yu. M. Golubev, B.-G. Englert, M. O. Scully, *et al.*, Zh. Éksp. Teor. Fiz. **116**, 485 (1999) [JETP **89**, 258 (1999)].
16. E. M. Purcell, Phys. Rev. **69**, 681 (1946).
17. D. Kleppner, Phys. Rev. Lett. **47**, 233 (1981).
18. S. M. Barnett and R. Loudon, Phys. Rev. Lett. **77**, 2444 (1996).
19. *Optical Processes in Microcavities*, Ed. by R. K. Chang and A. J. Campillo (World Scientific, Singapore, 1996).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 945–951.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1092–1228.

ATOMS, SPECTRA, RADIATION

Calculation of the Probability of Optical Transitions in Strong Electric Fields

S. V. Bulyarskiĭ, N. S. Grushko, and A. V. Zhukov*
Ulyanovsk State University, Ulyanovsk, 432700 Russia
**e-mail: avg@ulsu.ru*
Received May 26, 2000

Abstract—An algorithm is proposed for calculating the spectrum of the cross section for photoionization of carriers on deep centers in electric fields on the basis of the form function of an optical transition. An experiment and calculations were performed for the complex V_{Ga}–S_{As} in GaAs. The proposed model is compared with theoretical works based on the single-coordinate approximation. It is concluded that the single-coordinate model is applicable for describing the field-dependence of the cross section for photoionization of an electron on a V_{Ga}–S_{As} center. Data on the influence of an external electric field on the change in the moments of the form function of the absorption band of the complex V_{Ga}–S_{As} in GaAs are obtained. It is concluded that an electric field influences the adiabatic potentials of the center investigated. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is known that an electric field increases the probability of optical transitions. This effect was predicted for intraband transitions by Franz and Keldysh independently [1] in 1958 and first observed experimentally by Vavilov and Britsin [2] in 1960. The theory was further elaborated in [3, 4] (the Franz–Keldysh effect in impurity absorption).

A more accurate calculation of the transition probability from deep impurities centers into the conduction band in the space charge region of a semiconductor in an electric Field was made in [5, 6] in 1972, and the results of this work were quickly confirmed experimentally in [7]. The Timashov formulas, presented in [6], are based on a single-coordinate model for describing deep centers in semiconductors. This model imposes quite stringent restrictions on the character of the oscillations in a system and must be verified in each individual case. Thus, the degeneracy of the electronic states of a crystal with an impurity center results in the breakdown of the adiabatic approximation and in vibrational mixing of the electronic levels. In this case the single-coordinate model may be in inapplicable for calculating field dependences.

It is shown in the present paper how to calculate the field dependences of the photoionization cross section in general case on the basis of the experimental form function of the optical transition. To confirm the calculation presented the inverse problem of finding the form function of an optical transition from the experimental spectrum of the photoionization cross section of electron in a deep level of the complex V_{Ga}–S_{As} in GaAs is solved.

2. CALCULATION OF THE PHOTOIONIZATION PROBABILITY FOR DEEP CENTERS IN STRONG ELECTRIC FIELDS

The basic characteristics of an electronic transition between two states of a system which are described by the wave functions $|1\rangle$ and $|2\rangle$ are determined by the transition probability, which according to [8, 9] can be represented in the form

$$W^{12}(h\nu) = \sum_{n,\,n'} \rho_{1n} \left|\left\langle 1_n \middle| \hat{M} \middle| 2_{n'} \right\rangle\right|^2 \delta(E_{2n'} - E_{1n} - h\nu), \quad (1)$$

where n and n' enumerate the vibrational states of the ground and excited electronic terms, and ρ_{1n} is the probability of finding an electron in a vibrational state with index n of the term 1, which taking account of the Boltzmann distribution has the form

$$\rho_{1n} = \exp\left(-\frac{E_{1n}}{kT}\right)\left[\sum_{n''} \exp\left(-\frac{E_{1n''}}{kT}\right)\right]^{-1},$$

where $\hat{M}$ is the electric or magnetic dipole moment operator of the system.

It is obvious that the value of the integral $W^{12}(h\nu)$ depends strongly on the form of the functions $|1\rangle$ and $|2\rangle$, so that for a given electronic transition $1 \longrightarrow 2$ it is determined by the character of the vibrational states being combined or in other words by the electron–phonon interaction. The latter determines the dependence of W^{12} on the photon energy. Knowing $W^{12}(h\nu)$, it is easy to find the light absorption coefficient $K^{12}(h\nu)$. We shall determine this coefficient, as usual, from the

relation $I = I_0\exp[-K^{12}(h\nu)l]$, where I and I_0 are, respectively, the intensity of the incident and transmitted light, and l is the thickness of the absorbing layer. Then [10]

$$K^{12}(h\nu) = \frac{4\pi^2 N\nu}{3\hbar c}W^{12}(h\nu), \tag{2}$$

where N is the number of absorbing centers per unit volume.

According to [8], Eq. (2) can be rewritten in the form

$$K^{12}(h\nu) \approx k\nu f(h\nu), \tag{3}$$

where $f(h\nu)$ is the form function of the optical transition (in our case with absorption of a photon), containing information about the electron–phonon interaction.

In practice it is often necessary to deal with level–band (conduction or valence) transitions. Having absorbed a photon, an electron moves from a deep level into any state of the band (for definiteness, we shall consider the conduction band). Consequently, even in the absence of an electron–phonon interaction the probability of a transition under the action of optical radiation depends on the electron density of states of the conduction band. In the parabolic approximation for the conduction band $E(k) = \hbar^2k^2/2m^*$ this dependence near the band edge has the well-known form [11]

$$\overline{W}(h\nu) \approx \frac{1}{h\nu}\sqrt{h\nu - E_0}. \tag{4}$$

Thus, above the ionization threshold E_0 the effective cross section is proportional to the electron density of states of the conduction band.

On this basis the transition probability from a deep impurity center into the conduction band is a convolution of Eq. (3) with the conduction-band states:

$$W_{\rm abs}(h\nu) = \int_{h\nu-\Delta}^{h\nu} \overline{W}(h\nu-\varepsilon)K^{12}(\varepsilon)d\varepsilon, \tag{5}$$

where Δ is the energy width of the allowed band (conduction band).

Then we obtain the following expression for the probability of an optical transition with absorption from a deep impurity center into the conduction band:

$$W_{\rm abs}(h\nu) = G_0\int_{h\nu-\Delta}^{h\nu} \sqrt{h\nu-\varepsilon}\,f(\varepsilon)d\varepsilon, \tag{6}$$

where G_0 it is a normalization constant.

The contribution of electric field F can be taken into account in the expression for the probability of a purely electronic transition. The latter was calculated theoretically by Vinogradov, using a three-dimensional δ-function for the potential of a deep center, neglecting the electron–phonon interaction [4]:

$$\overline{W}(h\nu) = A\frac{F}{h\nu\sqrt{2m^*E_0}}\Omega(z(h\nu, F)), \tag{7}$$

where

$$\Omega(z) = \frac{2}{3}\left\{z^2V^2(-z) + zV'^2(-z) - \frac{1}{2}V(-z)V'(-z)\right\},$$

$$z(h\nu, F) = \left(\frac{h\nu - E_F}{E_0}\right)\left(\frac{\sqrt{2m^*}E_0^{2/3}}{e\hbar F}\right)^{2/3},$$

$$E_F = E_0 + \left(\frac{e\hbar F}{32m^*E_0}\right)^2,$$

A is a normalization factor that depends on the number of impurity centers. Finally, we obtain

$$W_{\rm abs}(h\nu) = A'\int_{h\nu-\Delta}^{h\nu}\frac{E}{\sqrt{2m^*E_0}} \times \Omega(z(h\nu-\varepsilon, F))\,f(\varepsilon)d\varepsilon. \tag{8}$$

Therefore the problem of calculating the spectrum of the photoionization cross section in electric fields reduces to finding the form functions of the optical transition with absorption of a photon. The latter can be obtained [12] from the contour of the absorption band by dividing it by the photon energy or from the luminescence band, as shown in [9].

3. EXPERIMENTAL DETERMINATION OF PHOTOIONIZATION CROSS SECTIONS

Sulfur-doped GaAs was chosen as the material for checking the model experimentally. Group-VI impurities in GaAs occupy arsenic sites and become donors, forming shallow levels near the conduction-band bottom. In addition, it is well-known [13–15] that they form complexes consisting of gallium vacancies and a donor at an arsenic site (V_{Ga}–D_{As}). In [9] the photoluminescence spectrum of the complexes V_{Ga}–S_{As} in GaAs:S were investigated. Ni-GaAs contacts were produced by electrochemical deposition of nickel. The method for depositing nickel and an investigation of the structures obtained, which are described in [16], showed that the structures obtained are Schottky diodes.

The spectrum of the photoionization cross sections of deep centers was measured on the basis of a study of the charge transfer kinetics of deep levels under the action of illumination in the stage charge region (SCR) of a Schottky diode [17–19]—the so-called photocapacitance method [19]. A metal cryostat was used to perform the experiment at temperatures below room temperature. The sample was thermostated (the tem-

perature was maintained to within 1 K), and it was protected from the background light. An MDR-23 monochromator was used for excitation.

The analysis of the measurements was based on a simple kinetic equation, which in the absence of trapping of electrons and holes in the field of the space charge region is

$$\frac{dn_t}{dt} = -(Jq_n + e_n)n_t + (Jq_p + e_p)(N_t - n_t), \qquad (9)$$

where J is the photon flux in the SCR, $q_{n(p)}$ it is the photoionization cross section for electrons (holes), $e_{n(p)}$ is the rate of emission of electrons (holes) from the level, N_t is the density of complexes, and n_t is the density of electrons on the complexes. The emission rate includes a combination of all thermal field processes. Hence we obtain the time constant for the decrease of the capacitance with the light switched off: $\tau^{-1} = e_n + e_p + J(q_n + q_p)$.

Since the level lies closer to the valence band and the energy splittings from the bands exceeded $10kT$, it can be assumed that charge transfer on the level with the light switched off is completely determined by hole emission and the time constant of the process $\tau_{\text{off}}^{-1} = e_p$, and with the light switched it depends on the time constant $\tau_{\text{on}}^{-1} = e_p + Jq_n$. Then the photoionization cross section is

$$q_n = \frac{e_n^0}{J} = \frac{1}{\tau_{\text{on}}^{-1} - \tau_{\text{off}}^{-1}}\frac{1}{J}. \qquad (10)$$

The experiment was performed as follows. The time variation of the capacitance of the SCR of the sample with monochromatic illumination with photon energies from 1.27 to 1.46 eV switched off and on were measured for several fixed reverse-bias voltages applied to the sample.

The spectral curves of the optical emission rate and the photoionization cross section for electrons on a deep center, produced by the complex V_{Ga}–S_{As}, are presented in Fig. 1 for five different fixed fields in the SCR of the barriers investigated.

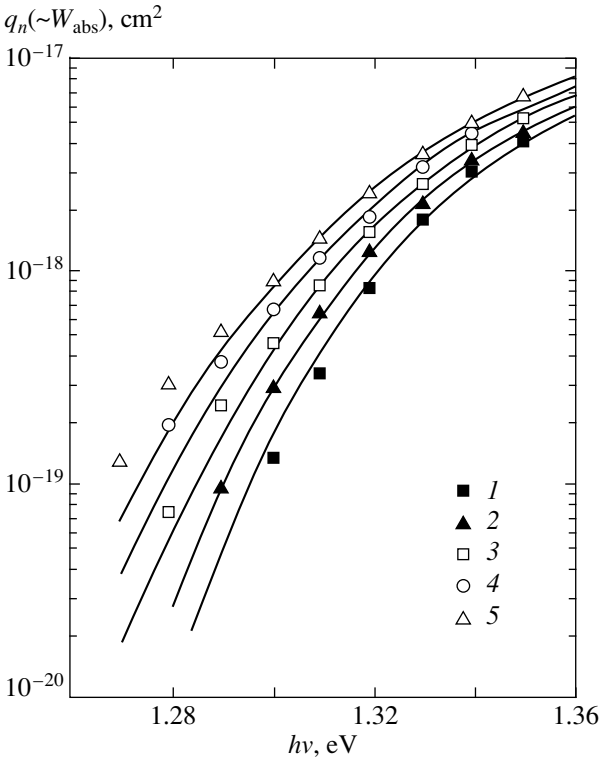

Fig. 1. Experimental photoionization cross sections for an electron on a deep level of the complex V_{Ga}–S_{As} (points) and the spectra computed using Eq. (8) (solid lines) for various fields in the SCR: (*1*) 8.3×10^4, (*2*) 8.8×10^4, (*3*) 9.5×10^4, (*4*) 1.05×10^5, and (*5*) 1.1×10^5 V/cm.

4. CALCULATION OF THE FORM FUNCTION FOR OPTICAL ABSORPTION

The quantity $W_{abs}(h\nu)$ on the left-hand side of Eq. (6) is simply the optical emission rate $e_n^0 = Jq_n$, where q_n is the photoionization cross section and J is the photon flux in the semiconductor. Therefore (taking account of the weak dependence of J on $h\nu$), the dependences which we measured for the photoionization cross section for electrons on deep centers of a V_{Ga}–S_{As} complex are described by the expression

$$q_n(h\nu) = \bar{G}_0 \int_{h\nu-\Delta}^{h\nu} \sqrt{h\nu - \varepsilon}\, f_{\text{abs}}(\varepsilon)\, d\varepsilon. \qquad (11)$$

To find $f(h\nu)$ it is necessary to solve Eq. (11), which is a Fredholm integral equation of the first kind. This equation can be solved for $h\nu$ using the Riemann–Liouville integral transformation:

$$f(h\nu) = \frac{4}{\bar{G}_0}\frac{d}{dh\nu}\left\{\int_{-\infty}^{h\nu} \frac{q_n'(\xi)\,d\xi}{\sqrt{h\nu - \xi}}\right\} = \frac{4}{\bar{G}_0} D^{1/2}\{q_n'(h\nu)\}, \qquad (12)$$

where $D^{1/2}$ is a derivative of degree 1/2 [20]. However, the problem of solving the equation is ill-posed because the solution is unstable. This means that even very small errors in $q_n(h\nu)$, which are definitely present and are due to errors in the measurements and calculations, can result in large errors in the solution. Such problems can be solve by regularization methods [21], which are based on the concept of a regularizing algorithm.

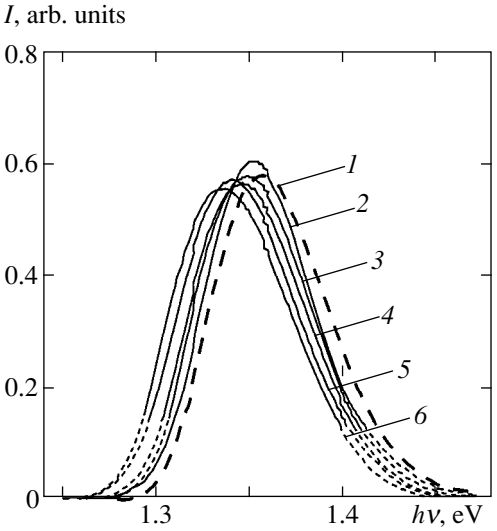

Fig. 2. Form function of the optical absorption band of the complex V_{Ga}–S_{As}, reconstructed from the emission spectrum (curve *1*). The form functions of optical transitions with absorption, which were calculated from the experimental spectrum of the photoionization cross sections, described by polynomials of degree eight, with Gaussian edges for various fields in the SCR: (*2*) 8.3×10^4, (*3*) 8.8×10^4, (*4*) 9.5×10^4, (*5*) 1.05×10^5, and (*6*) 1.1×10^5 V/cm.

The desired form functions were found by the zero-order Tikhonov regularization method, according to which the problem was reduced to solving a system of linear equations

$$\alpha f_{\text{abs}}(h\nu) + \int_a^b k(h\nu, s) f_{\text{abs}}(s) ds = w(h\nu), \quad a \le h\nu \le b, \tag{13}$$

where

$$k(h\nu, s) = \int_c^d K(t, h\nu) K(t, s) dt,$$

$$w(h\nu) = \int_c^d K(t, h\nu) W_{\text{abs}}(t) dt,$$

$$K(x, y) = \begin{cases} \sqrt{x - y}, & x > y \\ 0, & x < y, \end{cases}$$

a, c and b, d are, respectively, the lower and upper energy limits of the range of the measured spectrum of the photoionization rate. The adjustable parameter α for the spectrum varied from 10^{-4} to 3×10^{-3}.

The curves $f(h\nu)$ which we obtained in this manner were described by polynomials of degree eighth, normalized to 1 and corrected. The following was done for this: (1) sections were chosen on the short- and long-wavelength wings of the form functions that correspond to intensities from 0.15 to 0.30; (2) these sections were approximated by Gaussian functions and extrapolated to the wavelength range containing distortions caused by the inaccuracy of the Tikhonov method (Fig. 2).

The form functions [$f^F_{\text{abs}}(h\nu)$ below] presented in Fig. 2 can be used to calculate the field dependences of the photoionization cross sections using the formula (8).

To check the experimental form functions $f^F_{\text{abs}}(h\nu)$ it is helpful to compare them with the results of a luminescence investigation [9]. To this end, the form function of the optical absorption band of the complex V_{Ga}–S_{As} ($f^0_{\text{abs}}(h\nu)$ below), which is shown by the dashed line in Fig. 2, was reconstructed from the emission spectrum of the complex V_{Ga}–S_{As} [9] using the results of [12, 22, 23].

To compare $f^F_{\text{abs}}(h\nu)$ and $f^0_{\text{abs}}(h\nu)$ we shall calculate the first central moments of these functions. The central moment $\langle M_n \rangle$ of order n of the distribution function $f(\nu)$, calculated relative to the origin of coordinates, is determined by the formula

$$\langle M_n \rangle = \sum (-1)^i C_i^n \left(\frac{M_i}{M_0} \right)^i M_{n-i}, \tag{14}$$

where C_i^n is a binomial coefficients and $M_n = \int \varepsilon^n f(\varepsilon) d\varepsilon$ is the initial moments of order n [12]. The moments obtained are presented in the table, where it is evident that the first moments decreases as the electric field intensity increases. The second moment can be assumed to be constant. This indicates that the adiabatic potentials of the ground and excited states of the defect in an electric field draw together on the energy axis by the amount

$$\Delta E(F) = E_0(0) - E_0(F) = M_1(0) - M_1(F),$$

Table

Moment	$f^0_{\text{abs}}(h\nu)$	$f^F_{\text{abs}}(h\nu)$ with field intensity				
		8.3×10^4 V/cm	8.8×10^4 V/cm	9.5×10^4 V/cm	1.05×10^5 V/cm	1.1×10^5 V/cm
M_1	1.3658	1.3567	1.3534	1.3504	1.3449	1.3409
$\langle M_2 \rangle$	9.476×10^{-4}	9.434×10^{-4}	9.300×10^{-4}	9.187×10^{-4}	9.894×10^{-4}	9.549×10^{-4}

where $E_0(0)$ is the energy of a purely electronic transition in a zero electric field and $E_0(F)$ is the energy of a purely electronic transition in an electric field with intensity F. No changes in the curvature of the potentials occur. The first moment of the function $f_{\text{abs}}^{F}(h\nu)$ can be taken as the first moment of the form function $f_{\text{abs}}^{0}(h\nu)$ in a zero field. Figure 3 shows the deformations of the adiabatic potentials in an electric field which were constructed according to their first moments, clearly illustrating the conclusions drawn above.

5. CALCULATION OF THE FIELD DEPENDENCE OF THE CROSS SECTION FOR PHOTOIONIZATION OF A DEEP CENTER V_{Ga}–S_{As} IN GaAs ON THE BASIS OF THE FORM FUNCTION OF THE OPTICAL ABSORPTION BAND

The form function of the optical absorption band of the complex V_{Ga}–S_{As}, reconstructed from the initial band of the complexes using the parameters of the single-coordinate model, was used as $f(\varepsilon)$ to calculate the photoionization cross section from Eq. (8). The field dependences of the cross section calculated in this manner are displayed in Fig. 1. We note that the Vinogradov formula for the probability of a purely electronic transition (7) describes the field dependences of the transition probability only in a narrow range near the impurity absorption edge and below. Consequently, good agreement between the theoretical curves and experiments is obtained only in this range. The coefficient A' in Eq. (8) was chosen so as to obtain the best agreement between theory and experiment for $h\nu$ ranging from 1.30 to 1.33 eV with a 8.8×10^4 V/cm field. As one can see from Fig. 1, within the experimental error limits, which are associated with inaccuracies in measurements of the capacitance and calibration of the radiation source, the agreement between the computed and experimental data is satisfactory.

6. CALCULATION OF THE FIELD DEPENDENCE OF THE PHOTOIONIZATION PROBABILITY FOR A DEEP CENTER V_{Ga}–S_{As} IN GaAs IN THE SINGLE-COORDINATE APPROXIMATION

There are very few works on the exact theoretical calculation of the probability of optical transitions in electric fields. This is due to the difficulty of the calculations and the choice of a model. In [24] optical generation of charge carriers was studied, on the basis of the one-coordinate model, in a strong electric field with

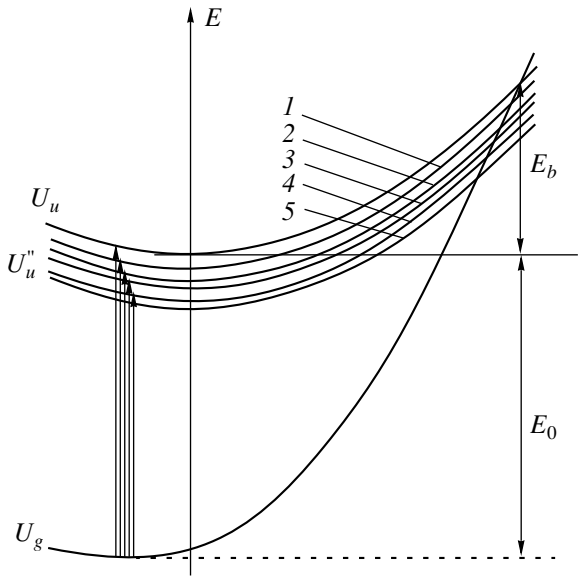

Fig. 3. Configuration-coordinate diagram of the complex V_{Ga}–S_{As}, illustrating the distortions of the adiabatic potentials in an electric field. Here U_g is the adiabatic potential of the ground state, U_u is the adiabatic potential of the excited state outside the field, U_u'' is the adiabatic potential of the excited state in a field with intensity F, and $F_5 > F_4 > F_3 > F_2 > F_1 > 0$ (curves *1–5*).

light absorption below the impurity absorption edge:

$$W(h\nu, F) = A\exp\left[\frac{\sigma^2}{kT}\left(\lambda_\omega - \frac{1}{2kT}\right) + \frac{h\nu - E_0}{kT}\right] \times \int_{-\infty}^{\sqrt{E_0/\hbar\Omega}} \left(\lambda_\omega - \frac{\hbar\Omega}{\sigma}\xi\right)^{-1} \times \exp\left[-\frac{\lambda_\omega^2\sigma^2}{2} + \lambda_\omega\hbar\Omega\xi - \frac{(\hbar\Omega)^2}{2\sigma^2}\xi^2\right]d\xi, \quad (15)$$

where

$$\lambda_\omega = \frac{(E_0 - S\hbar\omega - h\nu)kT + \sigma^2}{\sigma^2 kT},$$

$$\hbar\Omega = \frac{(eF\hbar)^{3/2}}{(2m^*)^{1/3}},$$

σ^2 is the second moment of the absorption form function, and A is a slowly varying function of the field and temperature. As $F \longrightarrow 0$, Eq. (15) passes into the Vinogradov expression [3]

$$W \approx \frac{1}{\sqrt{2\pi\sigma}}\exp\left\{\frac{h\nu - E_0}{kT} - \frac{(h\nu - E_0 - S\hbar\omega)^2}{2\sigma^2}\right\}.$$

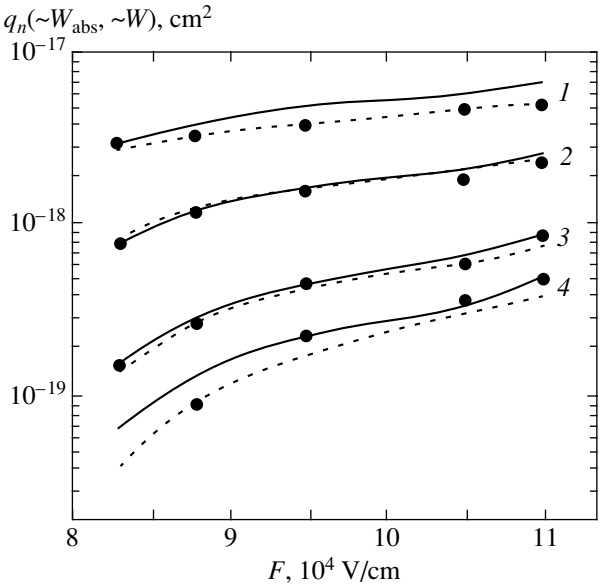

Fig. 4. Experimental dependences of the photoionization cross sections for an electron in a deep level of the complex V_{Ga}–S_{As} versus the field intensity (points); dependences calculated using Eq. (8) (broken lines) and the formula (15) (solid lines) for various incident photon energies: (*1*) 1.34, (*2*) 1.32, (*3*) 1.30, (*4*) 1.29 eV.

In [25] the electron–phonon interaction parameters in a one-coordinate model were calculated, using the formulas proposed in [12], from the moments of the form function of the optical radiative transition $f_{em}(h\nu)$. Thus, we have for the complex V_{Ga}–S_{As} the following values of the parameters of the single-coordinate model at temperature 100 K: E_0 = 1.3 eV, $\hbar\omega_u$ = 0.017 eV, $\hbar\omega_g$ = 0.025 eV, S = 3, where $\hbar\omega_u$ is the phonon energy describing the adiabatic potential of the excited state, $\hbar\omega_g$ is the phonon energy describing the adiabatic potential of the ground state, S is the Huang–Rice factor, and E_0 is the energy of a purely electronic transition from a local state near the conduction-band bottom to a deep level corresponding to the ground state of the center—this is essentially the activation energy of the level. The average between $\hbar\omega_u$ and $\hbar\omega$, specifically, $\hbar\omega$ = 0.21 eV, was used as $\hbar\omega$ in Eq. (15).

The computational results obtained with Eq. (15) are presented in Fig. 4. The coefficient A in the formula (15) was chosen to obtain the best agreement between theory and experiment with $h\nu$ ranging from 1.30 to 1.33 eV for a 8.8×10^4 V/cm field [just as in the calculation using Eq. (8)]. For comparison, the field dependences of the transition probability calculated using Eqs. (8) and (15) are presented in Fig. 4.

7. CONCLUSIONS

Thus, an algorithm based on experimental optical transition spectra has been developed to calculate the photoionization cross sections of deep centers in electric fields. For this, a quantum-mechanical calculation was performed of the optical ionization probability of deep centers in electric fields. The algorithm developed for calculating the photoionization cross sections of deep centers was compared with the calculations of the ionization probability of deep centers of the complex V_{Ga}–S_{As} performed on the basis of the single-coordinate model as well as with the experimental data. Good agreement was obtained between the experimental dependence and both theoretical dependences. However, preference was given to the scheme proposed for calculating the photoionization cross sections of deep centers on the basis of the experimental spectrum of optical transitions, because the latter was obtained without using any approximations, in contrast to methods based on the single-coordinate model, and is therefore more general. Calculations of the form function of the absorption band of the complex V_{Ga}–S_{As} in GaAs were performed. Data on the influence of an external electric field on the change in the moments of the form function of the absorption bands were obtained. It was concluded that an electric field affects the adiabatic potentials of the center studied.

REFERENCES

1. L. V. Keldysh, Zh. Éksp. Teor. Fiz. **34**, 1138 (1958) [Sov. Phys. JETP **7**, 788 (1958)].
2. V. S. Vavilov and K. T. Britsin, Fiz. Tverd. Tela (Leningrad) **2**, 1937 (1960) [Sov. Phys. Solid State **2**, 1746 (1960)].
3. V. S. Vinogradov, Fiz. Tverd. Tela (Leningrad) **12**, 3081 (1970) [Sov. Phys. Solid State **12**, 2493 (1970)].
4. V. S. Vinogradov, Fiz. Tverd. Tela (Leningrad) **13**, 3266 (1971) [Sov. Phys. Solid State **13**, 2745 (1971)].
5. S. F. Timashov, Fiz. Tverd. Tela (Leningrad) **14**, 2621 (1972) [Sov. Phys. Solid State **14**, 2267 (1973)].
6. S. F. Timashov, Fiz. Tverd. Tela (Leningrad) **14**, 171 (1972) [Sov. Phys. Solid State **14**, 136 (1972)].
7. S. V. Bulyarskiĭ, N. S. Grushko, and A. A. Gutkin, Fiz. Tekh. Poluprovodn. (Leningrad) **9**, 287 (1975) [Sov. Phys. Semicond. **9**, 187 (1975)].
8. Yu. E. Perlin and B. S. Tsukerblat, *The Effects of Electron-Phonon Interaction in Optical Spectra of Paramagnetic Impurity Ions* (Shtiintsa, Kishinev, 1976).
9. S. V. Bulyarskiĭ, N. S. Grushko, and A. V. Zhukov, Zh. Éksp. Teor. Fiz. **116**, 1027 (1999) [JETP **89**, 547 (1999)].
10. S. I. Pekar, Usp. Fiz. Nauk **50**, 197 (1953).
11. A. M. Stoneham, *Theory of Defects in Solids: the Electronic Structure of Defects in Insulators and Semiconductors* (Clarendon Press, Oxford, 1975; Mir, Moscow, 1978).
12. K. K. Rebane, A. P. Purga, and O. I. Sil'd, Tr. Inst. Fiz. Astron., Akad. Nauk Ést. SSR **14**, 31 (1961).
13. N. S. Averkiev, A. A. Gutkin, E. B. Osipov, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **25**, 50 (1991) [Sov. Phys. Semicond. **25**, 28 (1991)].

14. N. S. Averkiev, A. A. Gutkin, E. B. Osipov, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **25**, 58 (1991) [Sov. Phys. Semicond. **25**, 33 (1991)].
15. A. A. Gutkin, M. A. Reshchikov, and V. E. Sedov, Fiz. Tekh. Poluprovodn. (St. Petersburg) **31**, 1062 (1997) [Semiconductors **31**, 908 (1997)].
16. S. V. Bulyarskiĭ and A. V. Zhukov, Uch. Zap. Ul'yanovsk. Gos. Univ., Ser. Fiz. **2** (5), 98 (1998).
17. S. V. Bulyarskiĭ and N. S. Grushko, *Generation–Recombination Processes in Active Elements* (Mosk. Gos. Univ., Moscow, 1995).
18. S. Sah, A. Forbes, *et al.*, Solid-State Electron. **13**, 758 (1970).
19. N. S. Grushko and A. A. Gutkin, Fiz. Tekh. Poluprovodn. (Leningrad) **9**, 58 (1975) [Sov. Phys. Semicond. **9**, 37 (1975)].
20. *Tables of Integral Transforms (Bateman Manuscript Project)*, Ed. by A. Erdelyi (McGraw-Hill, New York, 1954; Nauka, Moscow, 1970), Vol. 2.
21. A. F. Verlan' and V. S. Sizikov, *Methods of Solution of Integral Equations with Computer Programs. Reference Manual* (Naukova Dumka, Kiev, 1978).
22. K. K. Rebane and O. I. Sil'd, Opt. Spektrosk. **9**, 521 (1960).
23. H. Cramer, *Mathematical Methods of Statistics* (Princeton Univ. Press, Princeton, 1946; Mir, Moscow, 1975).
24. S. F. Timashov, Doctoral Dissertation in Mathematical Physics (Moscow, 1975).
25. S. V. Bulyarskiĭ, N. S. Grushko, and A. V. Zhukov, Fiz. Tekh. Poluprovodn. (St. Petersburg) **34**, 41 (2000) [Semiconductors **34**, 40 (2000)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 952–956.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1099–1104.

ATOMS, SPECTRA, RADIATION

Zero-Noise Conversion of the Carrier Frequency in a Resonant Zero-Inversion Medium Consisting of Three-Level Atoms

Yu. M. Golubev* and G. R. Ershov

Institute of Physics, St. Petersburg State University, St. Petersburg, 198904 Russia
**e-mail: yugolubev@peterlink.ru*

Received June 14, 2000

Abstract—It is demonstrated that a three-level medium can be used to convert the carrier wave frequency without decreasing the signal/noise ratio. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The attractiveness of a three-level medium from the standpoint of quantum optics has been noted repeatedly. Lasers constructed using such objects are capable of emitting sub-Poissonian light because of their internal properties. It is sufficient to overcome a threshold, and the lasing arising will be automatically nonclassical. This is because the specific nonlinear properties a three-level medium negative feedback capable of stabilizing the electromagnetic field even below the quantum limits occurs [1]. Nonetheless, the prospects for practical applications of three-level lasers as sources of nonclassical light are at the present time small, since the quantum characteristics of this light are not very pronounced: shot noise can be suppressed in best case by only one-half.

At the same time there is another possibility of using the system for, specifically, the conversion of the frequency of an electromagnetic wave without substantially degrading the signal/noise ratio. One can imagine, for example, the following physical situation. An attempt is made to transmit and analyze information by appropriate modulation of laser radiation. In the usual physical situation one frequency of the laser carrier wave is more suitable for transfer of the radiation through space and the other is more suitable for photodetection. Thus, carrier frequency conversion is an important applied problem.

It will be shown below that, specifically, this problem can be solved in a three-level medium. Indeed, the arriving information-modulated wave can be used to excite this medium, as result of which lasing arises on an adjacent transition of the three-level medium, i.e., at a completely different but, as expected, also information-modulated frequency. Here we shall demonstrate that, in first place, the physical parameters can be adjusted so that the information modulation will be transmitted without distortion from the control channel into the generation channel of the converter and, in second place, if the information embedded in the initial beam in the form of amplitude modulation is to be analyzed, then information transfer from one carrier wave frequency to another can occur without a decrease of the signal/noise ratio.

2. INFORMATION TRANSFER BY LASER RADIATION

A. Information Modulation of Laser Radiation

Figure 1 shows schematically the experimental situation that we shall discuss. We have a single-mode laser source of Poissonian or sub-Poissonian light (*S*). One possible method of introducing information into the laser beam without destroying the quantum properties of the beam is Q-modulation of the resonator. This can be done, for example, by gluing a resonator mirror on a plate cut from a ceramic crystal; this will make it possible to change the perimeter of the resonator using suitable electric signals.

The information-modulated laser beam propagating in free space reaches the unit (*Tr*) in our experimental setup. The carrier frequency is changed in this unit. We shall assume that it consists of an optical high-Q resonator with a three-level medium.

The arriving modulated light excites the three-level medium in the resonator to the generation threshold, as result of which a laser field appears on an adjacent atomic transition. We assume that the new lasing will be information-modulated just as the initial lasing. Thus, the optical carrier frequency will be converted. As is well-known, this could be important, for example,

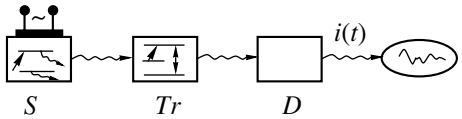

Fig. 1.

1063-7761/00/9105-0952$20.00

for photodetection, if the photodetector is insensitive to the initial lasing frequency.

Finally, the modulated radiation, which is now from the three-level laser (frequency converter), strikes the photodetector (*D*), whose electrical circuit is arranged in a manner so that it is possible to study the spectrum of the photocurrent.

In the present section we shall discuss only the first unit, specifically, the laser source of information-modulated Poissonian or sub-Poissonian light. The simplest models of two-level lasers make it possible to write the following equation for the number of laser photons $m_0(t)$ [2, 3]:

$$\dot{m}_0 = \left(-\kappa_0 + \frac{r_0\eta_2}{1+\eta m_0}\right)m_0. \tag{2.1}$$

Here, κ_0 is the spectral width of the laser mode, r_0 is the average rate of excitation of the upper laser level, η^{-1} is the number of photons which saturate the laser transition. The quantity η can be written in the explicit form

$$\eta = \eta_1 + \eta_2, \quad \eta_1 = \frac{2g_0^2}{\gamma_1^{(0)}\gamma_{12}^{(0)}}, \quad \eta_2 = \frac{2g_0^2}{\gamma_2^{(0)}\gamma_{12}^{(0)}}, \tag{2.2}$$

where $\gamma_2^{(0)}$ and $\gamma_1^{(0)}$ the relaxation rates, respectively, of the upper and lower laser levels (Fig. 2), $\gamma_{12}^{(0)}$ is the relaxation rates of the coherence on the laser transition, and g_0 is the dipole interaction constant for an atom interacting with the laser wave.

On the basis of this equation we shall assume that the resonator width κ_0 varies adiabatically in time as result of the corresponding electric oscillations of the piezoelectric ceramic:

$$\kappa_0 \longrightarrow \kappa_0(t) = \kappa_0 - \delta\kappa_0(t). \tag{2.3}$$

We shall assume that the degree of modulation is small $\kappa_0 \gg \delta\kappa_0(t)$, and then the number of photons in Eq. (2.1) can be written as

$$m_0(t) = m_0 + \delta m_0(t), \quad \delta m_0(t) \ll m_0. \tag{2.4}$$

Here, m_0 is the stationary solution in the absence of modulation and (which is important, as we shall see below, for our formulation of the problem) in the absence of the frequency converter *Tr*

$$\eta m_0 = \frac{r_0\eta_2}{\kappa_0} - 1. \tag{2.5}$$

We could proceed in the standard manner, linearizing Eq. (2.1) and writing a differential equation describing the modulation of the laser radiation. This is a linear, inhomogeneous, first-order equation and therefore it is easily solved. However, recalling that the problem for

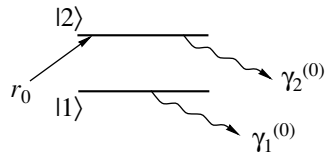

Fig. 2.

modulation is something in the adiabatic approximation, we can obtain

$$\frac{\eta\delta m_0(t)}{1+\eta m_0} = \frac{\delta\kappa_0(t)}{\kappa_0} \tag{2.6}$$

directly from the expression (2.5).

B. Limiting Possibilities for Observing Modulation: Signal/Noise Ratio with Photodetection of the Laser Radiation

The statistical theory in this case can be formulated precisely as done in [3], with no modulation, on the basis of the ideas developed in [2].

It is well-known that if there is no modulation, the photocurrent spectrum can be represented in the form

$$i_\omega^{(2)} = i_{\text{shot}}^{(2)}\left(1 + 2\xi_0\frac{\kappa_0\Gamma_0}{\Gamma_0^2+\omega^2}\right),$$
$$\xi_0 = \frac{1}{\eta m_0} - \frac{1}{2}\frac{\gamma_1^{(0)}}{\gamma_1^{(0)}+\gamma_2^{(0)}}. \tag{2.7}$$

Here, the first term (1 in parentheses) gives the shot-noise level, and the second term gives the level of the "excess" noise. For Poissonian light with saturation, $\eta m_0 \gg 1$, the Mandel parameter ξ_0 is 0 (the second term in ξ_0 is absent in this case) and therefore there is no excess noise. For sub-Poissonian light, likewise with situation, the second term in ξ_0 becomes the main term and then $\xi_0 = -1/2$, provided that the spontaneous decay of the upper laser level is sufficiently slow: $\gamma_1^{(0)} \gg \gamma_2^{(0)}$. It is evident that at close to zero frequencies the excess noise completely compensates the shot noise. The suppression of shot noise, which was predicted in [3], has been confirmed experimentally in [4].

If we now assume that the initial beam contains information-carrying low-contrast modulation, then an additional term due to this modulation will appear in the formula, and the complete spectrum at saturation will have the form

$$i_\omega^{(2)} = i_{\text{shot}}^{(2)}\left(1 + 2\xi_0\frac{\kappa_0^2}{\kappa_0^2+\omega^2} + \frac{\kappa_0}{n_0}\frac{|\delta n_0(\omega)|^2}{T}\right). \tag{2.8}$$

Here, *T* is the measurement time, which, in one hand, should be long enough so that explicit expression for the excess noise in Eq. (2.7) remains unchanged, while

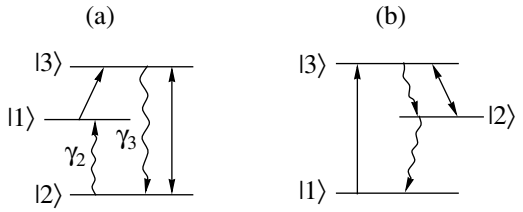

Fig. 3.

on the other hand it should not be so long that the spectral structure

$$\delta n_0(\omega) = \int_{-T/2}^{T/2} dt \delta n_0(t) \exp(i\omega t) \tag{2.9}$$

could fit into the spectral interval T^{-1}. The latter condition is in complete agreement with the initial adiabaticity condition.

The formula (2.8) can be viewed as the observed signal in our experiment on the detection of modulated laser radiation. In addition, for obvious reasons, the modulation is not observed in a pure form but rather against background noise, which is determined by Eq. (2.7).

It is natural to determine the signal/noise ratio SNR_ω as the ratio of the third term in Eq. (2.8) to the sum of the first two terms:

$$SNR_\omega^{\text{in}} = \frac{\kappa_0}{m_0 T} \frac{|\delta m_0(\omega)|^2}{1 + 2\xi_0}. \tag{2.10}$$

Here, we took into account the adiabaticity condition and assumed that the excess noise is spectrally uniform within the spectral width of the informative part of the signal and we have introduced the index in to underscore that the investigation is performed at the input of the frequency converter (but, in its absence).

To be able to measure and analyze the modulation it is natural to require that

$$SNR_\omega^{\text{in}} > 1. \tag{2.11}$$

Obviously, this condition for a Poissonian laser, $\xi = 0$, gives a limit on the degree of information modulation. At the same time, in a theoretical analysis for a sub-Poissonian laser, $\xi = -1/2$, modulation with arbitrary lack of contrast can, in principle, be noticeable.

3. CONVERSION OF THE CARRIER FREQUENCY ON THREE-LEVEL ATOMS

A. Semiclassical Theory

We shall now demonstrate that when a three-level laser is pumped by modulated laser light we can force the system to generate light that is modulated in precisely the same manner (but, naturally, at a different frequency) on an adjacent transition.

Let the carrier frequency of the initial signal entering from the source *S* (Fig. 1) into the input of the frequency converter *Tr* be in resonance with an atomic transition ($1 \longrightarrow 3$) (Fig. 3a) and excite the atomic medium, as result of which lasing arises on another mode frequency in resonance with the atomic transition ($2 \longrightarrow 3$). As one can see from Fig. 3a, for atoms the frequency can only increase. However, one can imagine a different structure which is mathematically complete equivalent to the first structure (Fig. 3b) but gives a decrease in frequency.

The smallest distortions of the modulation when the modulation is transferred from the pump channel to the lasing channel will occur when the pumping is weak:

$$\beta_0 n_0 \ll 1, \quad \beta_0 = 2g_d^2/\gamma_3^2. \tag{3.1}$$

Here, g_d is the dipole interaction constant between the pump wave and the transition (1–3) (the pump or control channel).

In addition, we require

$$\delta = \frac{\gamma_2}{\gamma_3} \ll 1. \tag{3.2}$$

These two conditions give zero-inversion lasing of a three-level laser [5], first mentioned in [6].

Then, the arguments can be based on the equations for the number $n_0(t)$ of photons in the laser source and the number $n_1(t)$ of photons in the frequency converter [5]:

$$\dot{n}_1 = -\kappa_1 n_1 + \kappa n_0 \frac{\beta_1 n_1 \delta}{1 + \beta_1 n_1 \delta}, \tag{3.3}$$

$$\dot{n}_0 = -(\kappa_0(t) + \kappa) n_0 + \frac{\zeta r_0 \eta_2 n_0}{1 + \eta n_0}. \tag{3.4}$$

Here, the last equation is different from Eq. (2.1) in that besides the initial losses of the laser field, which are determined by the coefficient κ_0 and which also occurred in the absence of the converter, now the losses to excitation of the three-level medium with rate κ are also taken into account:

$$\kappa = \frac{\gamma_3 \beta_0 N_1}{1 + \beta_1 n_1}, \tag{3.5}$$

N_1 is the total number of three-level atoms in the resonator. The quantity

$$\beta_1 = 2g_1^2/\gamma_3^2$$

determines the saturation properties of the three-level medium on the transition (3–2), and g_1 is the dipole interaction constant between the pump wave and the transition (2–3).

In addition, we take into account the possibility that the lasing power of the laser source can change if the

average excitation rate r_0 of the medium is replaced by ζr_0.

Now, once again, assuming the modulation of the Q and of the lasing power of the initial laser source to be weak, we require the same thing for lasing at the new frequency:

$$n_1(t) = n_1 + \delta n_1(t), \quad n_1 \gg \delta n_1(t), \tag{3.6}$$

where n_1 is the stationary solution of Eq. (3.3):

$$-\kappa_1 n_1 + \kappa n_0 \frac{i_1\delta}{1+i_1\delta} = 0, \quad i_1 = \beta_1 n_1. \tag{3.7}$$

At saturation, when $i_1\delta \gg 1$,

$$n_0/n_1 = \kappa_1/\kappa. \tag{3.8}$$

Linearizing Eq. (3.3) with respect to the weak modulation of the input and output radiation, we obtain

$$\begin{aligned} \dot{\delta n_1} &= -\Gamma_1 \delta n_1 + \kappa \delta n_0, \\ \Gamma_1 &= \kappa_1\left(\frac{i_1}{1+i_1} + \frac{i_1\delta}{1+i_1\delta}\right) \longrightarrow 2\kappa_1. \end{aligned} \tag{3.9}$$

Taking into account the adiabatic approximation and the saturation factor, we obtain on the basis of Eq. (3.8)

$$\delta n_1 = \frac{\kappa}{2\kappa_1}\delta n_0. \tag{3.10}$$

Now we must estimate the relation between the lasing power of the laser source in the absence m_0 and in the presence n_0 of the frequency converter. Having in mind Eq. (3.4), it is easy to obtain

$$n_0 + \delta n_0(t) = \zeta \frac{\kappa_0}{\kappa+\kappa_0}\left(m_0 + \frac{\kappa_0}{\kappa+\kappa_0}\delta m_0\right). \tag{3.11}$$

Thus we see that the presence of a converter can, in principle, result in weaker modulation of the initial radiation and in a weaker signal as a whole. However, additional pumping can compensate the latter.

Now we can relate the characteristics of the secondary and primary emissions:

$$\frac{\delta n_1(t)}{n_1} = \frac{\kappa_0}{\kappa_0+\kappa}\frac{\delta m_0(t)}{2m_0}. \tag{3.12}$$

Here, it is important to underscore that under the chosen conditions the initial temporal variation was transferred from the initial frequency to the converted frequency without any deformations. We shall see below that even though the degree of modulation decreases as result of losses of the initial field on excitation of the three-level medium, it is nonetheless entirely realistic to obtain conditions under which the modulation of the light at the converted frequency will not be any more difficult to observe than the modulation at the initial frequency.

B. Signal/Noise Ratio at the Output of the Frequency Converter

The photocurrent spectrum at the output of the frequency converter can be calculated in the standard manner just as in [5] in the absence of modulation:

$$\begin{aligned} i_\omega^{(2)} = i_{\text{shot}}^{(2)}\Bigg[1 - 2\kappa_1^2 \\ \times \frac{(\kappa_0+\kappa)(\kappa_0-\kappa\xi_0)+\omega^2}{[\kappa_1(2\kappa_0+\kappa)-\omega^2]^2 + \omega^2(\kappa_0+\kappa+2\kappa_1)^2} \\ + \frac{\kappa_1}{n_1}\frac{|\delta n_1(\omega)|^2}{T}\Bigg]. \end{aligned} \tag{3.13}$$

Compared with the work cited, an additional information term of the same type as in Eq. (2.8) appears here. Once again, we shall determine the signal/noise ratio as the ratio of the third term to the sum of the first two terms. It should be remembered once again that because the informative modulation is introduced into the system adiabatically the spectral width of the detected signal (third term) is much smaller than the spectral width of the noise. Thus, to record the signal/noise ratio we can take the first two terms (the numerator of the ratio) at zero frequency.

Here we call attention to an important fact. According to Eq. (3.3), the number of photons $n_1(t)$ in the converter is proportional to the number of photons in the laser source n_0. This means, in turn, that the stationary number n_1 of photons and the modulation $\delta n_1(t)$ are also individually proportional to n_0. But then the informative term in Eq. (3.13) will also be proportional to n_0. Since the noise terms do not depend on n_0, we must conclude that the signal/noise ratio in the present case depends on the radiation power of the laser source, the dependence being all the stronger, the higher the power.

We shall now consider two limiting cases $\kappa_0 \gg \kappa$ and $\kappa_0 \ll \kappa$, which were called in [5] the cases of weak and strong coupling between lasers. In the first case the presence of the converter does not affect in any way the operation of the source. In the second case, however, absorption in the three-level medium is the main source of losses for the intraresonator laser field of the source.

For weak coupling the signal/noise ratio is independent of the statistics of the initial light:

$$SNR_\omega^{\text{out}} = \frac{2\kappa_1|\delta n_1(\omega)|^2}{n_1 T}, \tag{3.14}$$

which is entirely obvious, since the three-level system perceives any radiation as Poissonian radiation [5]. However, if the coupling is strong, then

$$SNR_\omega^{\text{out}} = \frac{\kappa_1|\delta n_1(\omega)|^2}{n_1 T(1+2\xi_0)}. \tag{3.15}$$

As we can see, the cases of Poissonian and sub-Poissonian laser sources are fundamentally different here: modulation of sub-Poissonian light can be observed for arbitrarily small degree of modulation, which cannot be said of modulation of Poissonian light.

To assess how the measurement possibilities have changed compared with the initial situation, we shall write the ratio of the signal/noise ratios at the output and input of the frequency converter, denoting this ratio by the letter F:

$$F = SNR_{\omega}^{\text{out}}/SNR_{\omega}^{\text{in}}. \tag{3.16}$$

It is easy to obtain that for weak coupling

$$F = \xi\frac{\kappa}{2\kappa_0}(1 + 2\xi_0) \tag{3.17}$$

and $F = 1$ for Poissonian light, if

$$\zeta = \frac{2\kappa_0}{\kappa} \gg 1. \tag{3.18}$$

Thus we can obtain the same measurement conditions at the converter output as at the converter input. For a sub-Poissonian light the signal/noise ratio at the converter output is always smaller.

For strong coupling

$$F = \zeta\frac{\kappa_0^2}{2\kappa^2} \tag{3.19}$$

and $F = 1$ obtains for

$$\zeta = \frac{2\kappa^2}{\kappa_0^2} \gg 1. \tag{3.20}$$

Therefore it can be concluded that, on the one hand, a three-level medium is entirely convenient and can serve for converting the frequency of the carrier wave and, the other hand, the detection conditions at the new frequency will be no worse than at the initial frequency.

REFERENCES

1. H. Ritsch, M. A. Marte, and P. Zoller, Europhys. Lett. **19**, 7 (1992).
2. M. O. Scully and W. E. Lamb, Phys. Rev. A **159**, 208 (1967).
3. Yu. M. Golubev and I. V. Sokolov, Zh. Éksp. Teor. Fiz. **87**, 408 (1984) [Sov. Phys. JETP **60**, 234 (1984)].
4. Y. Yamamoto, S. Mashida, and O. Nilson, Phys. Rev. A **34**, 4025 (1986); W. H. Richardson, Y. Yamamoto, and S. Mashida, Phys. Rev. Lett. **66**, 2867 (1991).
5. Yu. M. Golubev and G. R. Ershov, Zh. Éksp. Teor. Fiz. **114**, 1971 (1998) [JETP **87**, 1068 (1998)].
6. O. A. Kocharovskaya and Ya. I. Khanin, Pis'ma Zh. Éksp. Teor. Fiz. **48**, 581 (1988) [JETP Lett. **48**, 630 (1988)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 957–974.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1105–1124.

ATOMS, SPECTRA, RADIATION

Light Scattering by Extraordinarily Polarized Polaritons

T. V. Laptinskaya and A. N. Penin
Moscow State University, Moscow, 119899 Russia
e-mail: postmast@qopt.phys.msu.su
Received August 27, 1999

Abstract—The objective of this work is to investigate how the anisotropy of the interaction between dipole-active vibrations of a crystal lattice and infrared electromagnetic waves is manifested in the spontaneous parametric light scattering spectra of polaritons (Raman scattering by small angles). The case where scattering occurs by extraordinarily polarized polaritons—quasiparticles formed as a result of the coupling of the wave polarized in the symmetry plane of a biaxial crystal simultaneously with two phonons possessing orthogonal dipolar moments—is studied. A series of spectra of equilibrium fluctuations of the electromagnetic (infrared) field, each of which represents an intensity distribution in frequency–wave number coordinates for a fixed direction of the wave vector, are constructed on the basis of a scattering model that takes account of the tensor character of the permittivity and the quadratic and cubic susceptibilities of the crystal. Analysis of the computed spectra identified the basic laws and dependences which are determined by the anisotropy of the electromagnetic susceptibilities of various orders and made it possible to explain previous experimental results which cannot be interpreted on the basis of the generally accepted model of transversely polarized polaritons. A method is proposed for determining the contributions of the dipole-active vibrations of the crystal lattice to the permittivity and the quadratic and cubic susceptibilities, as well as the absorption of the material from the spectra of the extraordinarily polarized polaritons. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The main objective of this work is to examine the special features, which have not yet been investigated, of light scattering by polaritons (SP) under the conditions of strong anisotropy of the deformation potential and the dipole moment of optical phonons, to show that SP can be used to measure the dynamical characteristics of these phonons, and to revive interest in this phenomenon of nonlinear optics.

Spontaneous parametric (Raman) scattering by polaritons in noncentrosymmetric crystals was actively investigated at the end of the 1960s and in the 1970s. At that time the basic physical features and characteristics of the process were mainly studied. In addition, quite effective methods for determining the dynamical parameters (frequencies, damping constants, contributions to the linear and nonlinear susceptibilities) of optical phonons, as well as methods for measuring the optical characteristics of crystals in the frequency range from tens to several thousands of reciprocal centimeters were developed on its basis.

Detailed reviews of the theoretical and experimental works devoted to scattering by polaritons are presented in [1–4]. Scattering by polaritons was successfully used to investigate ferroelectric phase transitions and isotopic substitution processes. The influence of trace impurities as well as regular and irregular optical nonuniformities on scattering was studied in [5–7]. The method was used especially successively to study the anharmonicity of vibrations and the Fermi resonance and Fano antiresonance phenomena.

However, the effect of the anisotropy of the interaction of phonons with the IR field on the SP spectra has still not been investigated. It is obviously manifested whenever the polaritons participating in scattering are extraordinarily polarized. In such situations, ordinarily, only a complex dependence of the general form of the frequency-angular distribution of the scattering intensity on the orientation of the interacting waves is noted, and no attempt is made to explain it. This situation narrows the range of applicability of SP spectroscopy. After all, for many crystals the selection rules with respect to symmetry make it possible to observe scattering only by extraordinary polaritons.

Polaritons are the eigenstates of the electromagnetic field in a medium which are formed as a result of the coupling of the electromagnetic field (macrofield) with dipole-active phonons. In the process of a laser wave (pump waves with frequency ν_l and wave vector $\mathbf{k}_l$) scattering by polari tons, signal waves, whose frequencies ν_s lie in the visible range, are created. The form of the scattering spectrum depends on the dispersion of the permittivity and the quadratic and cubic susceptibilities at the frequencies of all three waves, but in the visible range the dispersion of the optical properties of a crystal is much weaker than in the infrared range and it has a smooth character, so that the polariton spectrum has a determining effect. From the two-dimensional frequency–scattering angle distribution $I(\nu_s, \theta)$ observed in the visible region, the spectrum $P(\nu, k)$ of the equilib-

rium fluctuations of the electromagnetic field at polariton frequencies $\nu = \nu_l - \nu_s$ can be easily reconstructed by using the relations

$$\mathbf{k} = \mathbf{k}_l - \mathbf{k}_s, \quad (1)$$

where $\mathbf{k}_s$ is the wave vector of the signal wave and $\mathbf{k}$ is the coordinate of the fluctuation spectrum. The dispersion law $I(\nu_s, \theta)$ for polaritons can be determined as the line joining the maxima of the frequency (for constant k) contours of the spectrum $P(\nu, k)$. If the frequency ν is close to the phonon frequency ν_j (in the resonance region), the dispersion of the polaritons is different from that of free electromagnetic waves in the medium [8].

Polaritons correspond to small angles θ; for large angles the SP lines pass into the Raman scattering (RS) lines of optical phonons (in what follows we shall assume that the wave number of the polaritons lies in the range

$$0 < |k| < 10\pi\nu\sqrt{\varepsilon_s},$$

where ε_s is the permittivity in the visible range; particles with larger values of the wave number will be called phonons).

The two-dimensional spectrum of equilibrium fluctuations of the electromagnetic field in the resonance region depends primarily on the ratio of the dynamical parameters of the phonon (contributions to the permittivity $\Delta\varepsilon$, the quadratic and cubic susceptibilities $\Delta\chi$ and $\Delta\Theta$, and the damping constant Γ) as well as the background values of the susceptibilities determined by the contributions of neighboring phonons and electronic states. Various methods for determining such parameters from the SP spectra have now been developed. They are used in special cases [4, 9, 10], but the most general approach is one based on the numerical simulation of the spectra $I(\nu_s, \theta)$ [or $P(\nu, k)$], since it is applicable irrespective of the ratio of the dynamical parameters of the vibrations [11]. Modern computational technology makes it possible to fit these parameters. The accuracy of such a procedure can be different depending on the form of the spectrum (e.g., how many phonons must be drawn into the analysis), but it is quite high. Thus, vibrations with oscillator strengths up to 10^{-7} can be detected and small contributions (10^{-14}–10^{-17} cm^3 erg^{-1}) to the cubic susceptibility can be detected. Such small vibrations are not observed in the RS spectra, and their dipole activity is too weak to study them by the infrared method [12, 13].

As a rule, when modeling spectra it is assumed that the polarization of the polaritons is strictly perpendicular to the direction of their wave vector, irrespective of the direction of propagation. However, in a biaxial crystal electromagnetic waves are strictly transverse only if they propagate in the symmetry plane of the ellipsoid of wave normals and are polarized perpendicular to this plane, while in a uniaxial crystal (irrespective of the direction of propagation) only one of the two possible waves (the ordinary wave) is transverse. Consequently, the cases where scattering by "purely ordinary" polaritons in an anisotropic crystal is observed are exceptional. Under real experimental conditions, for each SP spectrum the direction of propagation $\mathbf{k}_l$ and the polarization $\mathbf{e}_l$ of the pump wave are fixed and the wave vector $\mathbf{k}_s$ and polarization $\mathbf{e}_s$ of the signal wave can vary in the range 10°–12°. The orientation of the polariton wave vector $\mathbf{k}_p$ can vary by 90° and more because of the strong dispersion of the permittivity ε in the resonance regions. In an anisotropic crystal, the principal values of ε can differ by several orders of magnitude, and they can also be negative; correspondingly, the shape of the ellipsoid of wave normals can be subject to strong dispersion, so that the longitudinal component of the electric field can be much greater than the transverse component. Consequently, a model of scattering by ordinary polaritons that does not take account of the longitudinal field is unsuitable for interpreting the SP spectra of extraordinary (anisotropic) polaritons.

Extraordinary polaritons are formed when the electric field of the light wave possesses a nonzero projection on two (or three) axes of an anisotropic crystal and vibrations of various symmetry types are linearly coupled by this field. The dispersion of such polaritons, i.e., the dependence of the frequency on the magnitude of the wave vector $k_p = 2\pi\nu\sqrt{\varepsilon_p}$ for various fixed directions of the vector, was first described theoretically by Poulet in 1955 and then by Merten and Loudon (citations to these works are given in [14, 15, p. 180] and [16, p. 401]), the corresponding quasiparticles were called mixed-polarization polaritons. The dispersion $\nu(k_p)$ of electromagnetic waves near frequencies of nondegenerate vibrations in biaxial crystals were also studied in [17, 18]. It was assumed in the calculations that the damping constants of dipole-active vibrations Γ_j are negligibly small. The term "oblique polaritons" is used in [19–22], and "anisotropic polaritons" is used in [23].

The number of experimental works on observations of extraordinarily polarized (extraordinary or anisotropic) phonons and polaritons is small. As a rule, the angular dependences of the phonon frequencies and the intensity of Raman scattering by the phonons (for $k^2 \gg \nu_p^2$) or the dependence of the form of the curve $\nu(k_p)$ in the polariton region on the orientation of the sample are investigated. The works [20–22, 28] are devoted to experiments on scattering in lithium niobate, [19, 27] in lithium iodate, and [29] in potassium niobate; the contributions of the deformation potential and the macroscopic field are identified in [30], which is devoted to scattering in a beryllium sulfide crystal, on the basis of the measured angular dispersion of the RS intensity.

The theoretical description of the intensity of scattering by extraordinary polaritons is given in a general

form in, for example, [23–26]. However, the special features of the spectra for polariton values of the wave numbers near the resonance frequencies of the phonons have still not been analyzed; they have been calculated only in the limit $k^2 \gg \nu_p^2$, i.e., for phonons. This is because it is virtually impossible to obtain transparent analytical expressions describing scattering for arbitrary orientations of the wave vectors of the interacting waves relative to the symmetry elements of the crystal.

To determine the effect of the anisotropy of the deformation potential and dipole moment of the vibrations of the crystal lattice on the SP spectrum, in the present work the method of computer simulation using oscillator functions for giving the first-, second-, and third-order susceptibilities, was chosen. The computational algorithms were obtained on the basis of the expressions presented in [23].

In order for the anisotropy of the interaction of two phonons with macroscopic electromagnetic field to be manifested, it is sufficient that these vibrations be nondegenerate and possess dipole moments oriented in different directions and that the electric field vector of the macroscopic electromagnetic wave in the crystal possess a nonzero projection on these directions. Consequently, we do not limit the generality of the problem, but we decrease substantially the number of parameters by studying scattering in a uniaxial crystal or in the symmetry plane of a biaxial crystal. In general, a pair of polaritons can then participate in the process; one of the polaritons (the ordinary one) is polarized strictly perpendicular to the wave vector and the symmetry plane, and the vector of the other (extraordinary) polariton lies in this plane. These polaritons correspond to different phase velocities and therefore different scattering angles. Consequently, two branches separated by an angle can be distinguished in the SP spectrum, and scattering by mutually orthogonal polarized polaritons can be studied independently [31].

In the next section the computational formulas and the algorithm for calculating the spectrum $P(\nu, k)$ of equilibrium fluctuations of the IR field, whose unit polarization vector lies in the symmetry plane of a biaxial crystal, are presented.

Next, a series of model spectra $P(\nu, k)$ near the frequencies of two nondegenerate vibrations with orthogonally oriented dipole moments, constructed for different fixed directions of the wave vector of the polaritons using specific values of the dynamical parameters, are analyzed. In Section 3.1 the case where the anisotropy of the dipole moment is greater than the anisotropy of the deformation potential is studied. The spectra presented in Section 3.2 correspond to polaritons formed by a pair of vibrations with the deformation potential anisotropy predominating. The dynamical parameters were chosen so that the analysis would be most easily understood: it is assumed that the phonon damping constants Γ_j are less than the *LO–TO* splitting.

The spectrum of scattering by polaritons can be calculated using the formulas presented in Section 2, if, taking account of the real experimental conditions and the optical properties of a crystal, for each point (ν, k) the orientation $\mathbf{k}$ from the triangle (1) is found first. In Section 4, as an example, the computed spectrum is compared with the experimentally obtained spectrum for scattering in an iodic acid crystal (for the frequency range 800–900 cm^{-1}, which contains the stretching vibrations of the *IO* group). Scattering by polaritons in this crystal was first observed in 1972 [43–47], but up to now the numerous "anomalies" in the spectra have not been interpreted. In our view these features can be fully explained on the basis of the model of extraordinary polaritons.

The method used to determine the dynamical parameters of the vibrations of a crystal lattice on the basis of the spectra of scattering by anisotropic polaritons is discussed in Section 5.

2. SPECTRA OF THE EQUILIBRIUM FIELD FLUCTUATIONS OF POLARITONS POLARIZED IN THE SYMMETRY PLANE OF A CRYSTAL

Let us consider scattering by extraordinary polaritons in an orthorhombic crystal with symmetry axes x_1, x_2, and x_3. Let the polariton wave be produced as a result of the coupling of the macroscopic electromagnetic wave only with two phonons, whose dipole moments are mutually orthogonal and oriented along the x_1- and x_2-axes. The wave vector $\mathbf{k}_p$ and the unit polarization vector $\mathbf{e}_p$ of such a wave must lie in the x_1x_2 plane (Fig. 1). For definiteness, we assume that the pump and signal wave vectors are also parallel to the x_1x_2 plane, and the vector $\mathbf{k}_l$ makes with the x_1-axis an angle ϕ, while the vector $\mathbf{k}$ makes the angle ρ. Let the polarization unit vector of the pump $\mathbf{e}_l$ lie in the x_1x_2 plane, and let the unit vector $\mathbf{e}_s$ of the signal wave lie along the x_3-axis (Fig. 1). Then only two components χ_σ need be taken into account:

$$\chi_1 = \chi_{321}e_s^3e_l^2 + \chi_{311}e_s^3e_l^1 = \chi_{321}\cos\phi + \chi_{311}\sin\phi, \quad (2)$$

$$\chi_2 = \chi_{312}e_s^3e_l^1 + \chi_{322}e_s^3e_l^2 = \chi_{312}\sin\phi + \chi_{322}\cos\phi. \quad (3)$$

These components must be complex functions, just as the components of the permittivity

$$\chi_\sigma = \chi'_\sigma + i\chi''_\sigma,$$
$$\varepsilon_\sigma = \varepsilon'_\sigma + \varepsilon''_\sigma, \quad \sigma = 1, 2.$$

The intensity of scattering by polaritons in an anisotropic crystal is proportional to the imaginary part of the tensor Green's function for Maxwell's equations [8, 23]. Using the general expression for the scattering

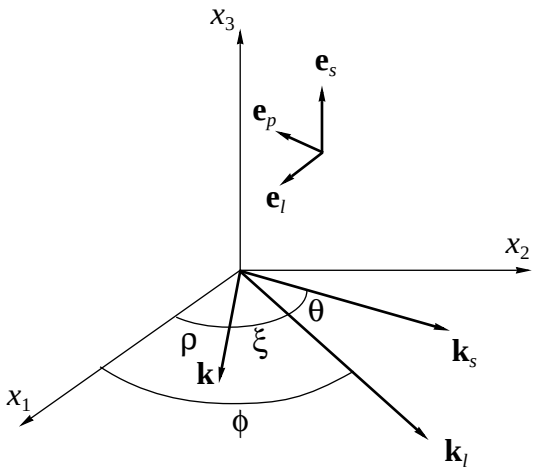

Fig. 1. Orientation of the polarization unit vectors of the pump wave $\mathbf{e}_l$, signal wave $\mathbf{e}_s$, polariton $\mathbf{e}_p$, and the wave vectors $\mathbf{k}_l$, $\mathbf{k}_s$, and $\mathbf{k}$ relative to the symmetry axes of a biaxial crystal.

form factor g, proposed in [23], we obtain the following system of formulas for the conditions chosen:

$$P(\nu, k) = Ag, \quad g = \frac{4\pi}{R^2 + I^2}(g' + g'') + \Theta_1 + \Theta_2. \quad (4)$$

Here A is a normalization constant, which is constant for the entire spectrum,

$$g' = (\chi_1')^2 Q_1 + (\chi_2')^2 Q_2 - 2I\mu^2 \chi_1' \chi_2' \cos\rho \sin\rho \quad (5)$$

contains only the real part of the quadratic susceptibility; the imaginary part of the quadratic susceptibility appears in the terms denoted by g'':

$$\begin{aligned} g'' = {} & 2\chi_1'\chi_1'' P_1 + 2\chi_2'\chi_2'' P_2 + (\chi_1'')^2 Q_1 + (\chi_2'')^2 Q_2 \\ & - 2\mu^2 \cos\rho \sin\rho [R(\chi_1''\chi_2' + \chi_1'\chi_2'') + I\chi_1''\chi_2'']. \end{aligned} \quad (6)$$

Here the variable $\mu = |k|/\nu_p$—the frequency normalized coordinate of the spectrum $|\mathbf{k}| = |\mathbf{k}_l - \mathbf{k}_s|$—is used. The contribution of the imaginary part of the cubic susceptibility is represented by two terms:

$$\Theta_1 = \gamma_{3322} e_s^3 e_s^3 e_l^2 e_l^2 + \gamma_{3321} e_s^3 e_s^3 e_l^2 e_l^1, \quad (7)$$

$$\Theta_2 = \gamma_{3311} e_s^3 e_s^3 e_l^1 e_l^1 + \gamma_{3321} e_s^3 e_s^3 e_l^2 e_l^1. \quad (8)$$

The functions R, I, P_σ, and Q_σ are given by the formulas

$$R = -\mu^2(\varepsilon_1' \cos^2\rho + \varepsilon_2' \sin^2\rho) + \varepsilon_1'\varepsilon_2' - \varepsilon_1''\varepsilon_2'', \quad (9)$$

$$I = -\mu^2(\varepsilon_1'' \cos^2\rho + \varepsilon_2'' \sin^2\rho) + \varepsilon_1''\varepsilon_2' + \varepsilon_1'\varepsilon_2'', \quad (10)$$

$$P_1 = R(\mu^2 \cos^2\rho - \varepsilon_2') - I\varepsilon_2'',$$

$$Q_1 = -R\varepsilon_2'' - I(\mu^2 \cos^2\rho - \varepsilon_2'),$$

$$P_2 = R(\mu^2 \sin^2\rho - \varepsilon_1') - I\varepsilon_1'',$$

$$Q_2 = -R\varepsilon_1'' - I(\mu^2 \cos^2\rho - \varepsilon_1').$$

The system of formulas described above makes it possible to calculate the dependence of the scattering intensity on the frequency and wave number irrespective of the value of the wave number in the "phonon" and in the "polariton" regions.

The system simplifies substantially for regions of the spectrum far from the characteristic phonon frequencies. Here $\chi' \gg \chi''$, so that $g'' \ll g'$, and the effect of the cubic susceptibility can also be neglected. In this case it follows from Eqs. (4) and (5) that the scattering intensity along the line

$$\mu^2 = \varepsilon_p'(\nu) = \frac{\varepsilon_1'\varepsilon_2'}{\varepsilon_1' \cos^2\rho + \varepsilon_2' \sin^2\rho}, \quad (11)$$

connecting the maxima of the scattering intensity k-contours in the ν–k plane is proportional to the factor

$$G' = \frac{g'}{R^2 + I^2} = \frac{(\chi_1'\varepsilon_1' \sin\rho)^2 + (\chi_2'\varepsilon_1' \cos\rho)^2 + 2\chi_1'\chi_2'\varepsilon_1'\varepsilon_2' \cos\rho \sin\rho}{\varepsilon_2''(\varepsilon_1')^2 \cos^2\rho + \varepsilon_1''(\varepsilon_2')^2 \sin^2\rho}. \quad (12)$$

As is well known, in a transparent anisotropic crystal the cosine of the angle β between the electric field vector and the x_1-axis is

$$\cos\beta = \frac{\varepsilon_2' \sin\rho}{\sqrt{(\varepsilon_1')^2 \cos^2\rho + (\varepsilon_2')^2 \sin^2\rho}}. \quad (13)$$

If it is assumed that the absorption in the nonresonance region is isotropic, i.e.,

$$\varepsilon_1'' = \varepsilon_2'' = \varepsilon'',$$

this quantity can be removed from the parentheses in Eq. (12) where the squared contraction of the quadratic susceptibility tensor with the field of the polariton wave, taking account of the transverse and longitudinal components the field, will remain.

We note a very important feature of the SP spectra, which is associated with the anisotropy. The last term in the numerator in Eq. (12) can be positive or negative. Let us assume that the components of the quadratic susceptibility tensor χ_1' and χ_2' have the same sign. Then its term is positive if the vectors $\mathbf{k}_p$ and $\mathbf{k}_l$ lie in the same or opposite quadrants of the x_1x_2 plane and negative if $\mathbf{k}_p$ and $\mathbf{k}_l$ lie in adjacent quadrants (see Fig. 1). Thus, for a definite orientation of the crystal the intensity can vanish in some frequency range. We recall that the spectra of ordinary polaritons also contain regions with zero intensity (so-called points of linearization of the crystal [1, 8]), but there they are due to the compensa-

tion (interference) of the electron–phonon contribution with the background value of the quadratic susceptibility. In this case the signal wave is extinguished (in some frequency range) for another reason. Each component χ_{ijk}, acting on the pump field and the polariton field, generates a field parallel to the x_3-axis. The sign and magnitude of this field (contraction) depend on the magnitude of the components and on the difference of the phases between the initial pump and polariton fields, as determined by the relative orientation of their wave vectors and the symmetry elements of the crystal. The intensity at the maximum of the *k*-contour of the spectrum of extraordinary polaritons is determined by the interference of the contributions of the components, and the regions of zero intensity can be termed points of phase compensation of the scattering intensity.

When describing scattering near the resonance frequencies of phonons the contribution of the imaginary parts of the quadratic and cubic susceptibilities cannot be neglected; the terms (6)–(8) make a large contribution to the SP intensity. In this case, qualitative conclusions about any particular feature of the spectra, associated with the anisotropy of the crystal lattice, can be drawn by constructing the series of spectra using specific dynamical parameters of the vibrations and analyzing the series.

Let the functions describing the dispersion of the principal values of the permittivity and the components of the quadratic and cubic susceptibilities be oscillatory. We shall confine our attention to a small range near the resonance frequencies ν_1 and ν_2 and we shall take into account only one oscillator for each polarization, making the assumption that the contributions of all vibrations in this section can be given by constant background values $\varepsilon_{0\sigma}$:

$$\varepsilon_\sigma(\nu) = \varepsilon'_{0\sigma} + \Delta\varepsilon_\sigma(\nu) f_\sigma + i\varepsilon''_{0\sigma}, \tag{14}$$

where

$$f_\sigma = \nu_\sigma^2(\nu_\sigma^2 - \nu^2 - i\nu\Gamma_\sigma)^{-1}. \tag{15}$$

The contractions χ_σ (5) can also be given in an oscillator form, since the angle ϕ for each spectrum is fixed:

$$\chi_\sigma = \chi_{0\sigma} + \Delta\chi_\sigma f_\sigma(\nu). \tag{16}$$

On the basis of the same considerations, the cubic susceptibility can be written as

$$\Theta_\sigma = \Delta\Theta_\sigma f_\sigma(\nu). \tag{17}$$

3. EFFECT OF THE ANISOTROPY OF THE PARAMETERS OF DIPOLE-ACTIVE PHONONS ON THE FLUCTUATION SPECTRA OF THE POLARITON FIELD

It is well known that the dispersion of extraordinary polaritons $\nu(k_p)$ depends on which parameter of the vibrational spectrum changes more strongly when the wave vector of the phonons rotates in a fixed plane—

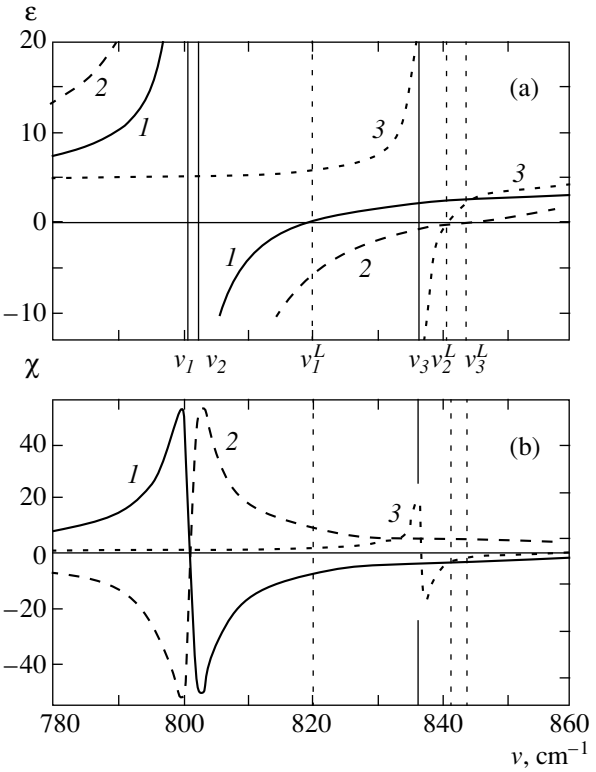

Fig. 2. Dispersion curves of (a) the real part of the principal values of the permittivity for $\Gamma_i = 0$ and (b) the real part of the components of the quadratic susceptibility in relative units for $\Gamma_i \neq 0$, $\nu_1 = 801$, $\nu_2 = 802$, and $\nu_3 = 837$ cm^{-1} are resonance frequencies of the vibration, $\nu_1^L = 820$, $\nu_2^L = 844$, and $\nu_3^L = 841$ cm^{-1} are longitudinal frequencies. These curves are used for constructing the model spectra.

the resonance frequency or the longitudinal–transverse splitting (dipole moment). It is said that in the first case the anisotropy of the deformation potential predominates while in the second case the dipole-moment anisotropy predominates [14–16]. The dynamical parameters of three vibrations, from which pairs corresponding to two types of anisotropy can be constructed, will be used below to model the SP spectra (see table). The resonance frequencies of the dipole-active phonons ν_1 and ν_2 in columns 1 and 2 of the table are almost identical; the contributions $\Delta\varepsilon_1$ and $\Delta\varepsilon_2$ are chosen in a manner so that in the x_1x_2 plane the dipole-moment anisotropy predominates:

$$\nu_1^L - \nu_1, \nu_2^L - \nu_2 > \nu_2 - \nu_1$$

(curves *1* and *2* in Fig. 2a). In order for the deformation potential anisotropy to predominate the difference between the frequencies ν_1 and ν_2 must be greater than the *LO–TO* splitting $\nu_1^L - \nu_1$. This type of interaction

Dynamical parameters of vibrations, used to calculate the SP spectra

Dynamical parameters of the vibrations	1	2	3
	vibration 1	vibration 2, dipole-moment anisotropy predominates	vibration 2, deformation potential anisotropy predominates
ν_σ, cm^{-1}	801	802	837
ν_σ^L, cm^{-1}	820	844	841
$\Delta\varepsilon_{0\sigma}$	0.184	0.478	0.044
$\varepsilon'_{0\sigma}$	3.84	4.45	4.58
$\Delta\chi_\sigma$	0.4	0.4	0.05
Γ_σ, cm^{-1}	3	3	1

can be studied using as parameters with index 2 the parameters from column 3 of the table. Then

$$\nu_2 > \nu_1^L > \nu_1$$

(curves *1* and *3* in Fig. 2a). It is assumed that the phonon values for the imaginary part of the permittivity are much smaller than for the real part:

$$\varepsilon''_{0\sigma} = 0.002.$$

The phonon damping constants Γ_σ were chosen to be small, not greater than the *LO–TO* splitting.

The quantities $\Delta\chi_\sigma$ are presented in relative units; we shall assume that $\chi'_{0\sigma} = 1$, and the background values of the imaginary part of the components of the quadratic susceptibility are negligibly small.

After $\Delta\varepsilon_\sigma$ and $\Delta\chi_\sigma$ are given, the contributions to the cubic susceptibilities cannot be chosen arbitrarily [8, 32]. We obtain for the components of the tensors of orthorhombic crystals from [33, 34]

$$4\pi(\Delta\chi_1)^2 = \Delta\varepsilon_1\Delta\Theta_1, \quad 4\pi(\Delta\chi_2)^2 = \Delta\varepsilon_2\Delta\Theta_2, \quad (18)$$

where

$$\begin{aligned}\Delta\chi_1 &= \Delta\chi_{321}e_s^3e_l^2 + \Delta\chi_{311}e_s^3e_l^1,\\ \Delta\chi_2 &= \Delta\chi_{312}e_s^3e_l^1 + \Delta\chi_{322}e_s^3e_l^2,\end{aligned} \quad (19)$$

but

$$\Delta\chi_{311} = \Delta\chi_{322} = 0,$$

and the relation between the cubic susceptibility and the RS tensor α_{ij} [15, 16] is determined by the relations

$$\Delta\Theta_1 = (\alpha_{ij}e_s^3e_l^2)^2, \quad \Delta\Theta_2 = (\alpha_{ij}e_s^3e_l^1)^2. \quad (20)$$

The background part of the cubic susceptibility can be neglected, since it is much smaller than the background part of the second-order susceptibility.

3.1. Predominance of the Dipole-Moment Anisotropy

We shall consider first the situation where the dipole-moment anisotropy predominates over the deformation potential anisotropy for the x_1x_2 plane: let the resonance frequencies of the vibrations forming the extraordinary polaritons be almost identical (ν_1 = 801 cm^{-1}, ν_2 = 802 cm^{-1}) and the dipole splittings different (ν_1^L = 820 cm^{-1}, ν_2^L = 844 cm^{-1}). The corresponding computed spectra of the fluctuations of the polariton field in the frequency–wave number coordinates $P(\nu, k)$ are presented in Figs. 3 and 4. For each spectrum, the orientation of the wave vector of the polaritons is fixed. It was assumed for the spectrum in Fig. 4a that the pump wave vector makes the angle ϕ = 90° with the x_1-axis; for Fig. 4b ϕ = 0°. For all other spectra ϕ = 45°, i.e., the growth components of the quadratic susceptibility contribute to the SP intensity [see Eq. (5)]. Each spectrum is normalized to the maximum intensity P_0 in the distinguished field ν, k, according to the formula

$$P = \log(999G/P_0 + 1);$$

the values of P, from 0 to 0.6 are marked white, the values from 0.6 to 3 are divided into 24 levels, a darker color corresponding to a higher level.

We shall now investigate how the permittivity anisotropy influences the spectrum $P(\nu, k)$. For this, it must be assumed that the quadratic susceptibility does not depend on the polarization of the polaritons in the range considered, i.e., $\chi_1 = \chi_2$. Let χ_1 and χ_2 also be independent of frequency in the oscillator model; this is possible if the phonon contributions to the quadratic susceptibility $\Delta\chi_\sigma = 0$ [see Eq. (16)]. It follows from the relations between the phonon contributions to the susceptibility of various orders (19) that in this case $\Delta\Theta_\sigma = 0$ and, corresponding to Eq. (20), the components of the tensor α_{ij} must also be zero. Such vibrations are active only in the dipole approximation, and they appear in the spectra of transversely polarized polaritons in the form of discontinuities of the dispersion branches.

Even though in this case the imaginary part of the quadratic susceptibility is zero [this follows from Eq. (16)], we cannot use the quite simple Eq. (12) to analyze the intensity distribution in frequency–wave number coordinates, since Eq. (11), describing the angular dependence of the effective value of the real part of the permittivity, is physically meaningless in the region of strong absorption.

The spectra calculated using Eqs. (4)–(10) for various fixed directions of the wave vector are presented in Fig. 3.

If the wave vector of the polaritons is directed along the x_2-axis (Fig. 3a), then the computed spectrum has two branches corresponding to the condition

$$\mu^2 = \varepsilon_1(\nu) \quad (\nu < 801 \text{ cm}^{-1} \text{ and } \nu > 820 \text{ cm}^{-1}),$$

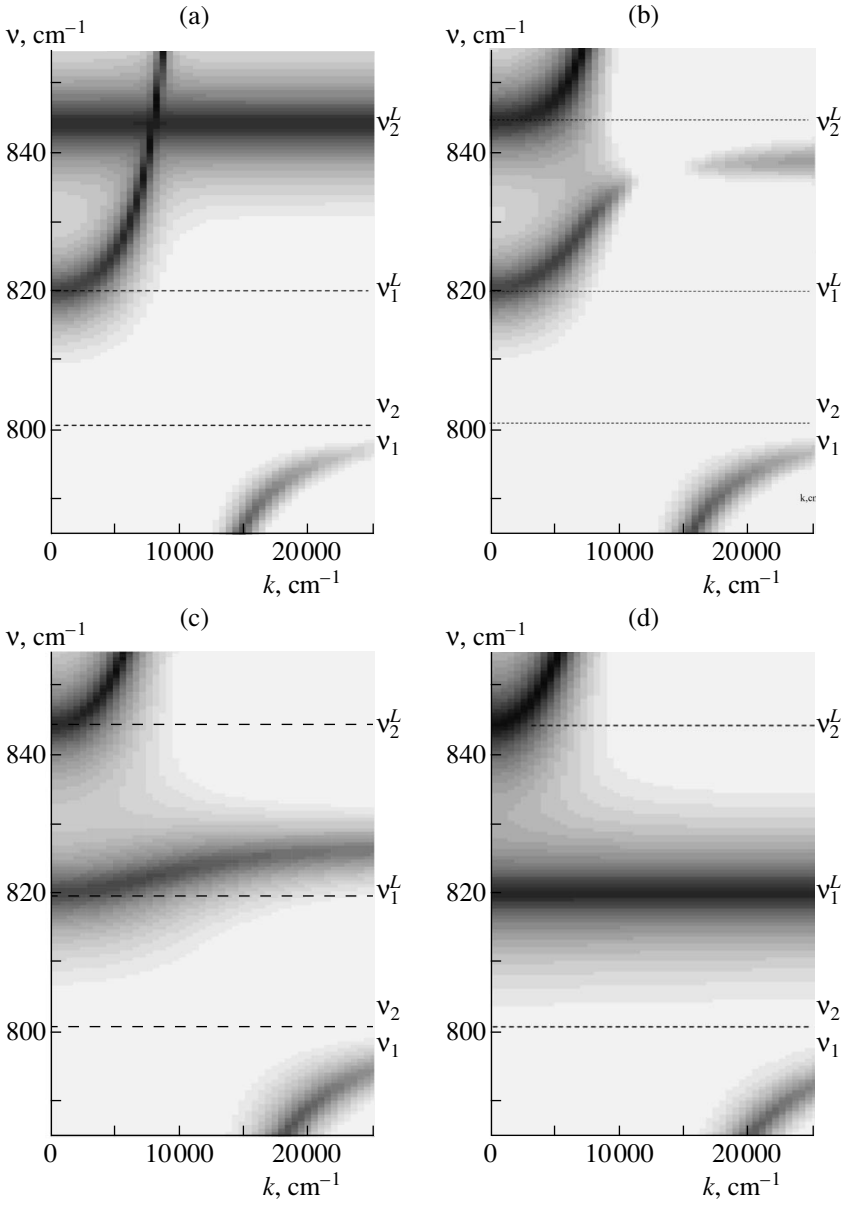

Fig. 3. Computed spectra of the equilibrium fluctuations of the polariton field for various fixed values of the angle ρ between the wave vector of the polaritons and the axis x_1. Dipole moment anisotropy predominates. For all spectra $\chi_1 = \chi_2 = 1$; ρ = (a) 90°, (b) 60°, (c) 30°, (d) 0°.

Note: The spectrum is constructed using a matrix of 7000 points. The fine intensity nonuniformities on the upper-frequency scattering branch are explained by the features of the graphical editor; they disappear when this fragment is constructed using a denser grid.

and a line at the frequency $\nu_2^L = 844$ cm^{-1}, parallel to the axis of the wave vectors. Indeed, ρ = 90°, we have from Eqs. (4)–(10)

$$G = 4\pi(G_1 + G_2), \tag{21}$$

where

$$G_1 = \frac{\varepsilon_1''[(\chi_1')^2 - (\chi_1'')^2] - 2(\varepsilon_1' - \mu^2)\chi_1'\chi_1''}{(\varepsilon_1' - \mu^2)^2 + (\varepsilon_1'')^2}, \tag{22}$$

$$G_2 = \frac{\varepsilon_2''[(\chi_2')^2 - (\chi_2'')^2] + 2(\varepsilon_2')\chi_2'\chi_2''}{(\varepsilon_2')^2 + (\varepsilon_2'')^2}. \tag{23}$$

The term G_1 does not contain χ_2. Hence, it corresponds to the unit polarization vector of the polaritons that is directed strictly along the x_1-axis (i.e., strictly transversely polarized polaritons), and it describes the branch the maxima of whose k-contours lie on the line $\mu^2 = \varepsilon_1'$. The term G_2 does not contain μ, and hence it does not depend on the matching conditions. The max-

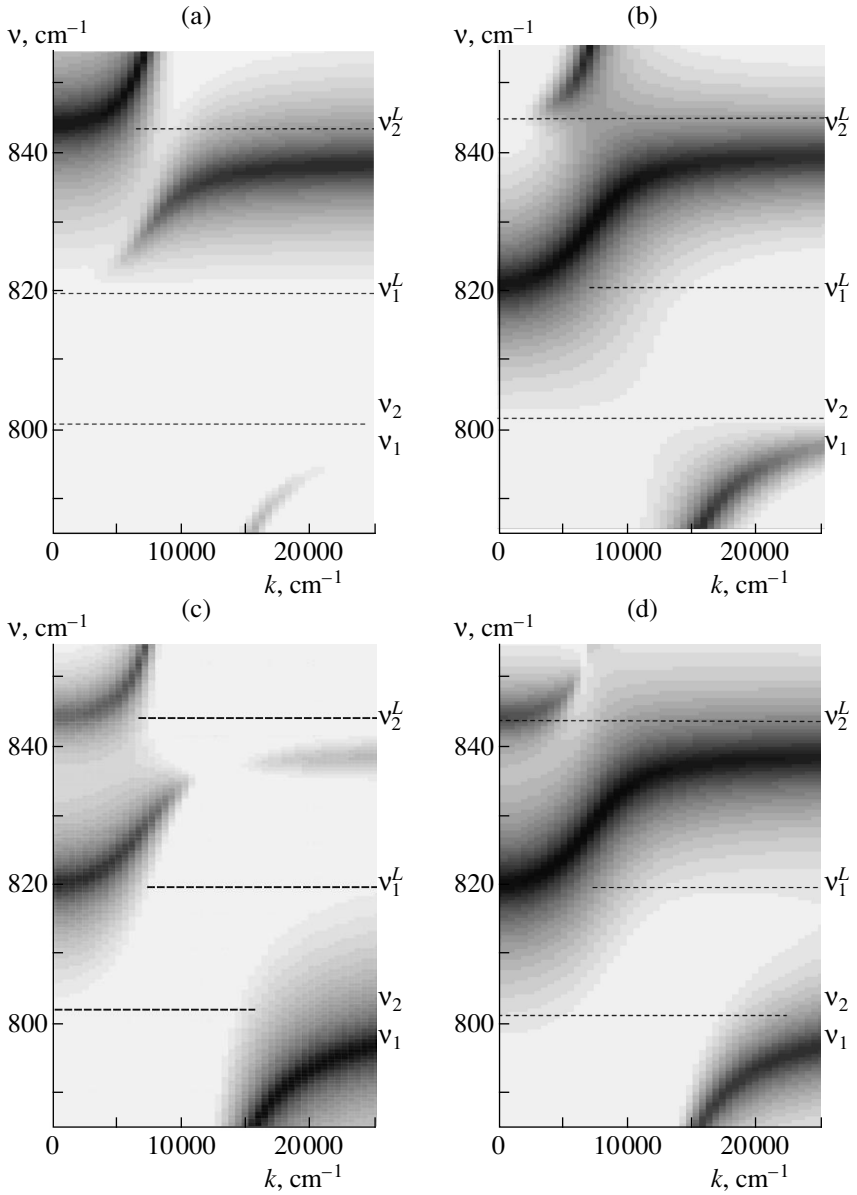

Fig. 4. Computed spectra for various values of the components of the quadratic susceptibility. Dipole moment anisotropy predominates. For all spectra the angle ρ between the wave vector of the polaritons and the x_1-axis is 60°. (a) $\chi_1 = 0$, $\chi_2 = 1$; (b) $\chi_1 = 1$, $\chi_2 = 0$; (c) dispersion of χ_1 and χ_2 is given by an oscillator function with parameters taken from the table (curve *1* in Fig. 2b); (d) the same function χ_2 is used, but for the function χ_1 the opposite sign of the contribution of the vibration $\Delta\chi_1$ was chosen (see curves *1* and *2* in Fig. 2b).

ima of the frequency contours of this branch lie on the straight line $\nu_2^L = 844$ cm^{-1}, where $\varepsilon_2'(\nu) = 0$. The term G_2 does not contain χ_1 and therefore describes the contribution of only polaritons polarized along the x_2-axis to scattering (in terms of [35] this is a fictitious longitudinal wave). The branches G_1 and G_2 cross: the longitudinally and transversely polarized polaritons do not interact. Far from the point of intersection with large values of the wave number ($\mu^2 \gg \varepsilon_1'$) the contribution of transversely polarized polaritons to scattering tends to zero. Thus the intensity at the maximum of scattering by a longitudinal phonon is

$$G_2^L = \frac{(\chi_2')^2 - (\chi_2'')^2}{\varepsilon_2''}. \qquad (24)$$

In this case the cubic susceptibility makes no contribution. Hence the interaction between a longitudinal phonon (or a fictitious longitudinal wave) and a laser wave is due only to the quadratic susceptibility. In the Raman scattering spectrum the line corresponding to the resonance frequency of the phonon will not appear

(Fig. 3a, $\nu = 801$ cm^{-1}, $|k| > 2300$ cm^{-1}), while the line at the longitudinal frequency will be intense.

If $\rho = 0°$ (Fig. 3d), then the scattering form factor is described by Eq. (21) where the two terms have the form (22) and (23) with indices 1 and 2 interchanged. Now G_1 gives the wave number independent line with a maxima of the frequency contours at $\nu = \nu_1^L = 820$ cm^{-1}—it corresponds to a longitudinal oscillation polarized along x_1; the intensity of scattering by it for large values of the wave number is determined completely by the component χ_1 and by absorption for a wave with corresponding polarization. The term G_2 describes the branch the line of whose maxima corresponds to the condition $\mu^2 = \varepsilon_2'$, i.e., dispersion of transversely polarized polaritons.

If the wave vector of the electromagnetic wave is not directed along one of the crystallographic axes, then the projections of the electric field on both axes are nonzero and two oscillations with orthogonal dipole moments contribute to the polariton wave. In this case the computed spectrum contains three dispersion branches, which cannot be separated according to polarization (Figs. 3b, 3c); in what follows, we shall call them the upper ($\nu > \nu_2^L$), middle ($\nu_1^L < \nu < \nu_2^L$), and lower ($\nu < \nu_2$) branches with respective frequency. In the region of existence of the middle branch the principal values of the permittivity have different signs. The end of this branch corresponding to large values of the wave number (the phonon section of the polariton branch) shifts to the frequency $\nu = 820$ cm^{-1} as the angle ρ decreases.

The curves $\varepsilon_p(\nu)$ connecting the maxima of the frequency contours of the scattering lines in the $\nu - k$ plane can be obtained from Eq. (11), if it is assumed that $\Gamma_{1,2} = 0$. They cannot lie in the interval 802–820 cm^{-1} for any values of ρ. The "phonon ends" of the polariton branches [the poles of the functions $\varepsilon_p(\nu)$] lie at the frequencies for which

$$\varepsilon_1 = -\varepsilon_2 \tan^2 \rho. \qquad (25)$$

Thus, the dispersion of the extraordinarily polarized polaritons $\varepsilon_p(\nu)$, obtained on the basis of our model, depends on the direction of the wave vector just as described in [14–16].

It was assumed above that both components of the quadratic susceptibility at any frequency equal 1, and the pump and polariton wave vectors lie in the same quadrant of the coordinate plane. Nonetheless, a dip in the scattering intensity is observed at the middle frequencies of the dispersion branches (Figs. 3b, 3c). Since the imaginary parts of the components of the permittivity and the function $\varepsilon_p(\nu)$ do not have singularities here, there is only one explanation for the dips: the contributions due to χ_1 and χ_2 to the signal field subtract out because ε_1' and ε_2' have different signs.

The dispersion of the scattering intensity in the spectra presented in Fig. 3 is determined to a greater extent by the change in the direction of the polarization unit vector of the polaritons than by the dispersion of the absorption. This direction cannot be determined from Eq. (13), since in the resonance region the imaginary part of the permittivity is two to three orders of magnitude greater than the real part. However, numerical modeling makes it possible to identify the singularities of the dispersion of the angle between $\mathbf{e}_p$ and $\mathbf{k}_p$.

Let us assume that only one of the components of quadratic susceptibility operates, that is χ_2, but $\chi_1 = 0$ (such conditions can be created in crystals belonging to certain symmetry classes: for example, let $\chi_{322} = \chi_{311} = 0$, and the pump wave vector is directed along the x_2-axis). The spectrum in Fig. 4a is constructed for $\rho = 60°$. Let us compare it with the spectrum in Fig. 3b: the dip in the intensity at the central frequency of the branch near $|k| = 1300$ cm^{-1} is absent, but the left end of this branch (where $k \longrightarrow 0$) is not seen in the scattering. For a different orientation of the wave vector of the polaritons, the general form of the spectrum, of course, will change, but it follows from model calculations that the intensity vanishes near the frequency $\nu_1^L = 820$ cm^{-1} irrespective of the angle ρ. This can be explained only by the fact that the field of the polariton waves does not have at this frequency a component along the x_2 axis (and participation of the other component is ruled out by the choice $\chi_1 = 0$). The spectrum calculated with $\rho = 90°$ will contain only a line due to a longitudinal vibration at the frequency 844 cm^{-1}. If, however, $\rho = 0°$, only the branch due to transversely polarized polaritons, which corresponds to the condition $\mu^2 = \varepsilon_2'(\nu)$, will be present.

One of the spectra, calculated assuming that $\chi_1 \neq 0$ while $\chi_2 = 0$, is presented in Fig. 4b ($\rho = 60°$). The dip in the intensity shifted from the middle to the upper branch, to the frequency 844 cm^{-1}. It exists here in the spectra calculated for any ρ, if $\chi_2 = 0$. For $\rho = 90°$ only the branch due to transversely (along the x_1-axis) polarized polaritons remains, and for $\rho = 0°$ only the line due to longitudinal vibrations at frequency 820 cm^{-1} remains.

Thus, irrespective of the orientation of the wave vector the polarization unit vector of extraordinary polaritons with $\nu = 844$ cm^{-1} is directed strictly along the x_1-axis, and for $\nu = 820$ cm^{-1} it is directed strictly along the x_2-axis.

Let us now assume that the vibrations are also active in Raman scattering, i.e., the resonance contributions to the components of the cubic and quadratic susceptibilities are different from zero (see table). In this case their dispersion is described by Eqs. (15)–(17). Let us com-

pare the spectrum in Fig. 4c with the spectrum in Fig. 3b. They were constructed for the same orientation of the wave vectors of the pump and the polaritons ($\rho = 60°$, $\phi = 45°$), so that the position of the dispersion branches in the $\nu - k$ plane is the same. The intensity distribution over the entire spectrum changed: in the lower-frequency branch in Fig. 4c it is now much larger than on the upper branch. This is due not only to the contribution of the cubic susceptibility but also the resonant growth of the real and imaginary parts of the components of the quadratic susceptibilities as $\nu \longrightarrow \nu_{1,2}$. We chose $\Delta\chi_{1,2}$ equal in magnitude and the functions χ_1 and χ_2 are also almost identical, so that once again we see the dip in the intensity in Fig. 4c on the middle branch, and it remains essentially unshifted.

The intensity distribution changes substantially, if the sign of $\Delta\chi_1$ is changed without changing the other parameters (Fig. 4d). In the first place, the point of phase compensation of the contributions of the components now lies on the upper branch and is absent on the middle branch. This is because in the frequency interval under study the components χ_1' and χ_2' have different signs (curves *1* and *2* in Fig. 2b). As a result, their contributions to scattering on the upper branch near $\nu = 860$ cm^{-1} are compensated, while on the middle branch they add, since here the main values of the permittivity also have different signs ($\varepsilon_2' < 0$). The total intensity of the lower-frequency branch is less than in Fig. 4c. This is because the real and imaginary parts of the components of the quadratic susceptibility have opposite signs here. The intensity increases, as a result of the contribution of both components of the cubic susceptibility, only for large values of the wave number, when $\nu \longrightarrow \nu_{1,2}$ (we note that in the spectrum in Fig. 3b the intensity decreases as frequency increases from 785 to 801 cm^{-1}).

3.2. Predominance of the Deformation Potential Anisotropy

Let the dynamical properties of the second vibration be given by the values in the third column of the table. Then the anisotropy of the deformation potential will predominate over the dipole-moment anisotropy: $\nu_2 > \nu_1^L > \nu_1$, i.e., the *LO–TO* gap for the second phonon, whose dipole moment is oriented along the x_2 axis (837–841 cm^{-1}), will lie above the longitudinal frequency of the first phonon (curves *1* and *3* in Fig. 2a). The computed spectra of the fluctuations of the polariton field for fixed orientations of the wave vector are presented in Figs. 5 and 6. The pump wave vector makes an angle of 90° with the x_1-axis for the spectrum in Fig. 6a and 0° for the spectrum in Fig. 6b; for all other spectra it is equal to 45°. The normalization and levels of intensity were chosen to be the same as in the preceding section (for the spectra in Figs. 3 and 4).

Just as in the preceding section, we shall investigate the role of the permittivity anisotropy, making the assumption that the components of the quadratic susceptibility are the same and the vibrations are inactive in Raman scattering.

For $\rho = 90°$ the scattering can be described by Eqs. (22) and (23). Figure 5a contains a branch, the position of the maxima of whose k-contours is determined by the condition $\mu^2 = \varepsilon_1'$ (the branch of polaritons polarized strictly along the x_1-axis). It intersects the line with the maxima of the frequency contours of the intensity at $\nu_2^L = 841$ cm^{-1} (the line of polaritons polarized longitudinally along the x_2-axis, transforming for large values of the wave number into the line of longitudinal phonons).

When the wave vector of the polaritons rotates by a small angle a local minimum forms at the intersection of the branches (at $k = 7500$ cm^{-1}, Fig. 5b) and the upper- and middle-frequency branches stand out. As the angle ρ decreases further (Figs. 5b, 5c) the "phonon section" of the middle branch shifts downwards to the frequency $\nu_2 = 837$ cm^{-1}, and at the same time its curvature changes slightly above the point $\nu_1^L = 820$ cm^{-1}. The "phonon section" of the lower branch moves upwards to ν_1^L; its curvature changes and approaches the middle branch. Both branches join at the point $\nu = 820$ cm^{-1}, $k = 11300$ cm^{-1}, when $\rho = 0°$ (Fig. 5d). In Fig. 5d we see a branch the position of whose maxima in the k-contours corresponds to the condition $\mu^2 = \varepsilon_2'(\nu)$, and it intersects a line with a maximum at the frequency 820 cm^{-1}. Just as in the case where the dipole moment predominates (see preceding section), for large values of the wave number the intensity here is determined by the ratio of the difference of the squared real and imaginary parts of the component χ_1 to the imaginary part of the permittivity ε_1'' (23); this is the same line of longitudinal polaritons as in the spectrum shown in Fig. 3d.

The "forbidden band" does not exist for the given form of the anisotropy of the phonon parameters for polaritons. If the line connecting the maxima of the frequency contours of the intensity is drawn, then the dependence of their form on the orientation of the wave vector is similar to that described in [14–16]. For phonons there is a "forbidden band": the "phonon sections" of the bottom and middle branches corresponding to large values of the wave number cannot lie in the interval $820 < \nu < 837$ cm^{-1}. The position of "phonon sections" is determined from the condition (25).

In contrast to the situation where the dipole-moment anisotropy predominates over the deformation-potential anisotropy, in the present case the main values of the permittivity have different signs not along the entire middle branch but only in the range $837 < \nu < 841$ cm^{-1}

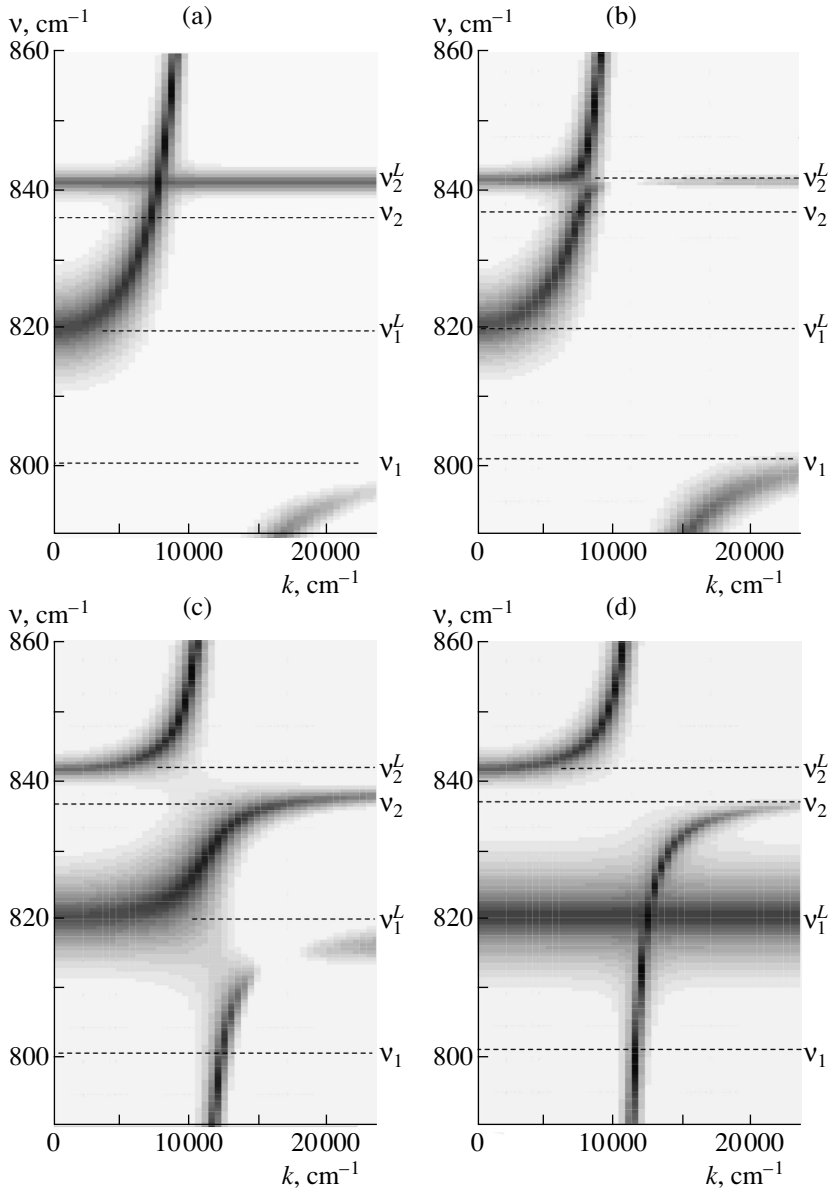

Fig. 5. Computed spectra for various fixed values of the angle ρ between the wave vector of the polaritons and the x_1-axis. Deformation potential anisotropy predominates. For all spectra $\chi_1 = \chi_2 = 1$; ρ = (a) 90°, (b) 70°, (c) 20°, (d) 0°. See also the remark for Fig. 3.

($\varepsilon_1 > 0$, $\varepsilon_2 < 0$) as well as on the lower branch at $801 < \nu < 820$ cm^{-1}, ($\varepsilon_1 < 0$, $\varepsilon_2 > 0$). Consequently, the sections with zero intensity are now present on the middle branch (Fig. 5b, $|k| = 10700$ cm^{-1}) and the bottom scattering branch (Fig. 5c, $|k| = 17500$ cm^{-1}): the contributions of the components of the quadratic susceptibility with constant and identical values of χ_1 and χ_2 can be compensated on both branches.

We shall now investigate the dispersion of the angle between the polarization unit vector of the polaritons and the wave vector according to the model SP spectrum just as was done in Section 3.1.

If $\chi_1 = 0$ and $\chi_2 = 1$, then the start of the middle-frequency branch (the region $k \longrightarrow 0$ for $\nu \longrightarrow \nu_1^L = 820$ cm^{-1}) is not seen in the scattering for any angles ρ (an example of such a spectrum for ρ = 30° is given in Fig. 6a). However, if $\chi_1 = 1$ and $\chi_2 = 0$, then for any ρ the left end of the upper branch is not seen (the region $k \longrightarrow 0$ for $\nu \longrightarrow \nu_2^L = 841$ cm^{-1}, an example is given in Fig. 6b). Hence, at the point $\nu = \nu_1^L$ the electric-field vector of the polariton wave is directed along the x_1-axis, and at the point $\nu = \nu_2^L$ it is directed along the x_2-axis, irrespective of the orientation of the wave vector.

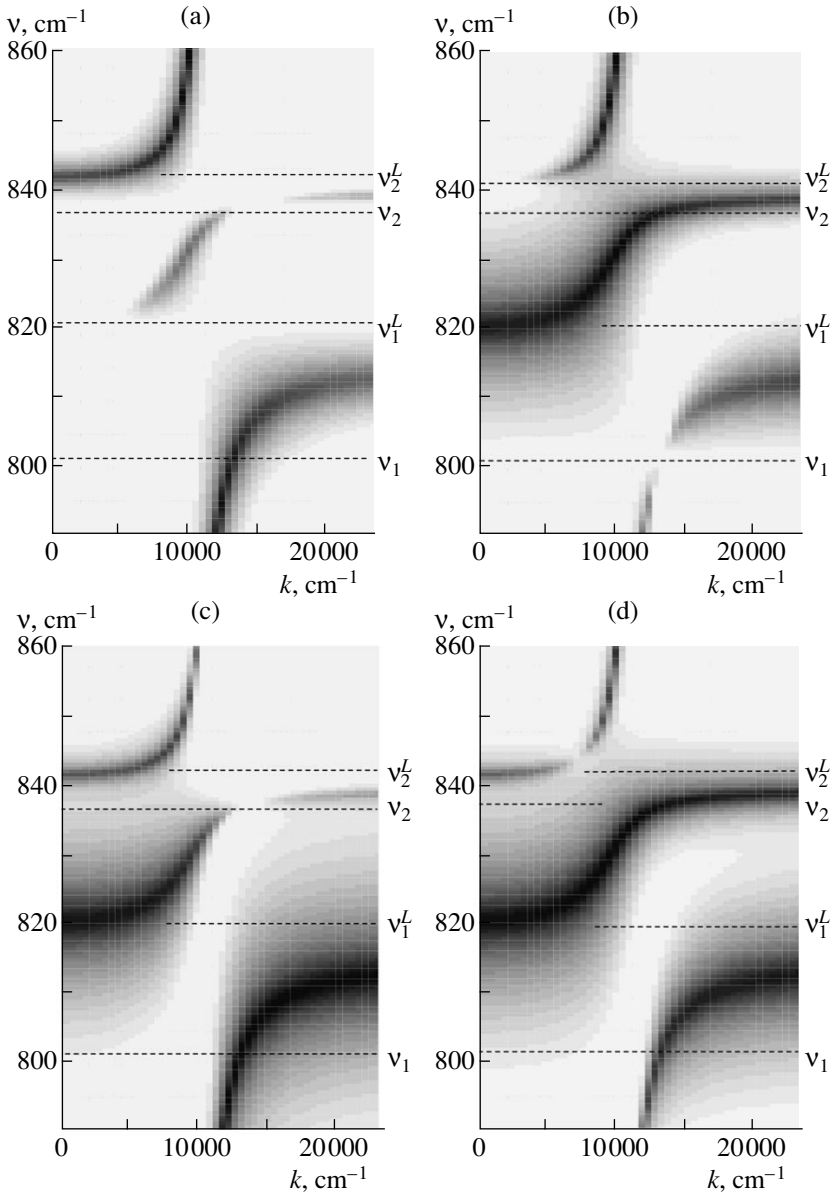

Fig. 6. Computed spectra for various values of the components of the quadratic susceptibility. Deformation potential anisotropy predominates. For all spectra the angle between the wave vector of the polaritons and the x_1-axis is 30°. (a) $\chi_1 = 0$, $\chi_2 = 1$; (b) $\chi_1 = 1$, $\chi_2 = 0$; (c) dispersion of χ_1 and χ_2 is given by an oscillator function with parameters taken from the table (curves *1* and *3* in Fig. 2b); (d) the same function χ_2 is used, but for the function χ_1 the opposite sign of the contribution of the vibration $\Delta\chi_1$ was chosen (see curves *2* and *3* in Fig. 2b). The presence of the fine nonuniformities is explained in the remark in Fig. 3.

The bottom and middle branches also contain two special points: for ε_2 at the frequency $\nu_1 = 801$ cm^{-1}, and for ε_1 at the frequency $\nu_2 = 837$ cm^{-1}. These points are also distinguished in the spectra of fluctuations of the polariton field. If the component of the field along x_1 (the spectrum in Fig. 6a) does not participate in the scattering, then a local intensity minimum is observed at the frequency 837 cm^{-1}. It exists in all spectra calculated for the angles $15° < \rho < 75°$. If the component along x_2 does not participate (Fig. 6b), then a similar minimum occurs on the lower branch at $\nu = 801$ cm^{-1}. Comparing the spectra in Figs. 6a and 6b, it can be inferred that the component of the field along the x_1-axis predominates at the frequency ν_2 and the component along x_2 predominates at the frequency ν_1.

Let us now consider the situation where the dipole-active vibrations also contribute to the cubic and quadratic susceptibilities: $\Delta\chi_1 = 0.4$ and $\Delta\chi_2 = 0.05$. In contrast to the case where the dipole-moment anisotropy predominates over the deformation-potential anisotropy, the frequencies of the maxima and minima of the nonlinear susceptibilities at $\rho \neq 0, 90°$ do not coincide

with the poles of the functions $\varepsilon_p(\nu)$ (see curves *1* and *3* in Fig. 2). Comparing the spectra calculated for different angles between the wave vectors of the polaritons and the x_1-axis, shows a quite complicated dependence of the general form of the intensity distribution on ρ. But, we present only one spectrum for the given parameters in Fig. 6c ($\rho = 30°$). We note that the intensity of the bottom branch is now greater than that of the middle and top branches (compare with Figs. 5b, 5c). There are two reasons for this: near the resonance frequency ($\nu_1 = 801$ cm^{-1}) the contributions of χ_1 and Θ_1 are large, and on the "phonon (high-frequency) section" of this branch the components of the quadratic susceptibility are also large, and the principal values of the permittivity have different signs (curves *1* and *3* in Fig. 2). The point of phase compensation of the contributions of the susceptibilities is located on the middle branch, since in the interval $820 < \nu < 837$ cm^{-1} χ_1 and χ_2 have different signs, while ε_1 and ε_2 have the same signs, and in the interval $837 < \nu < 841$ cm^{-1} both components of the quadratic susceptibility are negative, while the principal values of the permittivity have different signs.

The spectrum shown in Fig. 6d was constructed using the same value of $\Delta\chi_2$ but the sign of $\Delta\chi_1$ was changed (the dispersion of χ_1 is described now by curve *2*, and the dispersion of χ_2 is described by curve *3* in Fig. 2b). Here the middle branch is the most intense branch, since for $820 < \nu < 837$ cm^{-1} both components of the quadratic susceptibility are positive for positive ε_1 and ε_2 (curves *1* and *3* in Fig. 2a), while for $\nu > 837$ cm^{-1} ε_2 and χ_2 become negative simultaneously. The top branch possesses an intensity minimum at $k = 750$ cm^{-1}, which can be explained by the compensation of the contributions of the components because χ_1 is positive while χ_2 is negative.

4. EXPERIMENT

The experimentally observed spectra of scattering by polaritons are a two-dimensional dependence of the intensity on the frequency and scattering angle θ. To each value of θ there corresponds a wave number

$$|\mathbf{k}| = \sqrt{k_l^2 + k_s^2 - 2k_l k_s \cos\theta} \qquad (26)$$

(the range of variation of the magnitude of this spectra is limited below by the value $k_{\min} = ||k_l| - |k_s||$—the so-called point of collinear matching). The angle ξ between $\mathbf{k}$ and $\mathbf{k}_l$ (Fig. 1) is also determined from the triangle (1). Thus, the orientation of the wave coordinate in the observed spectrum cannot be constant and must be calculated for each point using the formula

$$\rho = \phi - \arcsin(|k_s|\sin\theta/|k|). \qquad (27)$$

In a small frequency range near the resonance frequency of the phonon the range of variation of ρ can reach 180°. Consequently, the form of the spectra of scattering by extraordinary polaritons depends strongly on the orientation of the sample. It is obvious that the observed frequency–angle distribution of the intensity can be compared with the computed distribution in some region in the ν–k plane only when the conditions (26) and (27) are taken into account for each point in the region.

As an example, we shall examine below a fragment of one of the SP spectra of the iodic acid crystal α-HIO_3. This crystal is widely used in nonlinear optics, its linear and nonlinear optical properties have been investigated in detail [39–42]. The Raman scattering and infrared absorption spectra, as well as their dependence on the orientation of the phonon wave vector, are described in [36]. At the beginning of the 1970s the first spectra of scattering by polaritons were obtained [43–45], and together with these spectra new information on higher order vibrations, bands of multiparticle states, and Fermi and Fano resonances, was also obtained [46–50]. The temperature dependence of the RS and SP spectra were investigated [51]. Experiments with this crystal demonstrated the great possibilities of SP spectroscopy in the investigation of lattice dynamics. However, they also revealed a large number of "anomalies": some sections of the spectrum, especially their sharp dependence on the orientation of the sample, could not be explained on the basis of the scattering model adopted.

The crystal belongs to symmetry class 222, and scattering by ordinarily polarized polaritons is impossible in it. At the same time a sample can be cut out and oriented so that the intensity is determined only by one or two components of the quadratic and cubic susceptibilities. This gives the minimum set of parameters for investigating the effect of crystal-lattice anisotropy on the SP spectrum [31].

To obtain the SP spectra in the coordinates frequency ν—scattering angle θ, we employed the standard experimental setup [1, 6] which made it possible to implement parallel information extraction using photographing system. This method makes it possible to obtain a much more data in a definite period of time compared with successive recording using photoelectronic devices.

A series of SP spectra of iodic acid crystals were obtained for various orientations of the samples. It was found that the sections of the spectra which we previously called "anomalous" could be explained by taking account of the tensor character of the linear and nonlinear susceptibilities and that modeling of the observed spectra was possible.

Figure 7a shows a fragment of one spectrum. The scattering plane is also the *xy* plane of the crystal (the designation of the axes corresponds to the ratio of the refractive indices for the visible region $n_x < n_y < n_z$), and the angle ϕ between the pump wave vector and the *y*-axis is 23°. The laser wave is polarized in the *xy* plane,

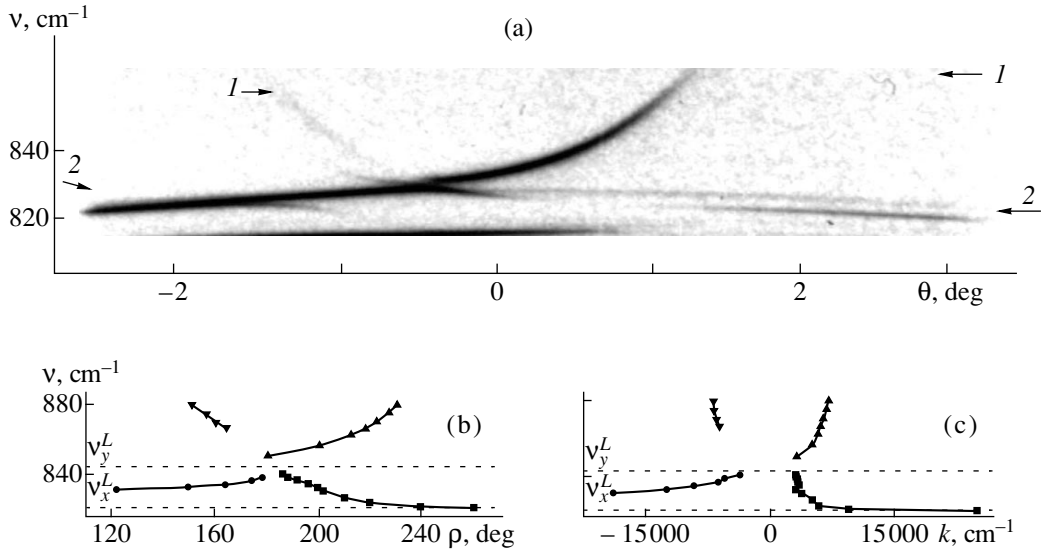

Fig. 7. (a) Fragment of an SP spectrum of an iodic acid crystal. The arrows mark the (*1*) upper-frequency and (*2*) middle-frequency scattering branches; (b) dispersion of the angle ρ between the *y*-axis and the wave vector of the polaritons, as obtained from this spectrum; (c) dispersion of the effective value of the wave number.

and the signal wave is polarized along the *c* axis. Scattering in this case is determined by two components of the quadratic susceptibility: χ_{zxy} and χ_{zyx}, so that only polaritons polarized in the *xy* plane can participate in the process. In the frequency range displayed in Fig. 7a, the polaritons are formed as a result of the coupling of the macroscopic electromagnetic wave with the stretching vibrations of the *IO* bond. The resonance frequencies of the B_x and B_y type phonons are almost identical, but the dipole moments are different: $\nu_x = 801$ cm^{-1}, $\nu_y = 802$ cm^{-1}, $\nu_x^L = 820$ cm^{-1}, and $\nu_y^L = 844$ cm^{-1}, i.e., dipole-moment anisotropy predominates in the *xy* plane [36].

The arrows in Fig. 7a mark the upper-frequency ($\nu >$ 844 cm^{-1}) and middle-frequency ($820 < \nu <$ 844 cm^{-1}) branches of the polariton dispersion. The spectrum is sharply asymmetric: the top branch is intense for positive scattering angles $\theta > -0.3°$ (right side of the spectrum) and is almost not seen for scattering in the other direction from the pump direction ($\theta < -0.3°$, left side of the spectrum). For the middle-frequency branch, conversely, the backward scattering is much more intense. The signs of the components of the quadratic susceptibility in the interval 800–880 cm^{-1} are constant, and the components vary monotonically (this follows from measurements according to a series of SP spectra for the given sample, which were obtained with different orientations of the sample). What is the nonmonotonic nature and asymmetry of the scattering intensity due to? It can be explained by the phase interference of the contributions of the components χ_{zxy} and χ_{zyx}: the products of these components make a contribution to the scattering intensity, being multiplied by the direction cosines of the wave vectors and the components of the real part of the permittivity.

Figure 7b shows the dispersion of the angle ρ between the wave vector of the polaritons and the *y*-axis for points connecting the maxima of the intensity of the frequency contours of the experimental spectra. The range of variation of the angle is almost 140°. The top and middle branches approach one another at $\rho = 180°$, and $|k| = 6900$ cm^{-1} (this is the minimum value of the wave vector for this spectrum—Fig. 7c). The wave vectors of the pump and the polaritons to the left of this point lie in adjoining quadrants of the coordinate plane; at the right of the point they lie in opposite quadrants.

In the region of existence of the top branch the principal values of the permittivity are positive. Consequently, if the pump and polariton wave vectors lie in opposite quadrants, the signal fields arising with the participation of the components χ_{zxy} and χ_{zyx} add (right side of the spectrum), and if they are located in adjoining quadrants, then they subtract (left side of the spectrum). In the region of the middle branch the signs of $\varepsilon_{x,y}$ are different. Consequently, this branch, conversely, is intense to the left of the point where $\rho = 180°$ and to the right of the point the intensity decreases rapidly with increasing angle, passing through zero at $\rho =$ 210° (this is the compensation point for the contributions of the components).

On the middle branch the magnitude of the wave vector of the polaritons undergoes the opposite dispersion: here the decrease in ε_p as a result of a rotation of **k** predominates over the normal dispersion growth of the components (Fig. 7c). We note that the anomalous

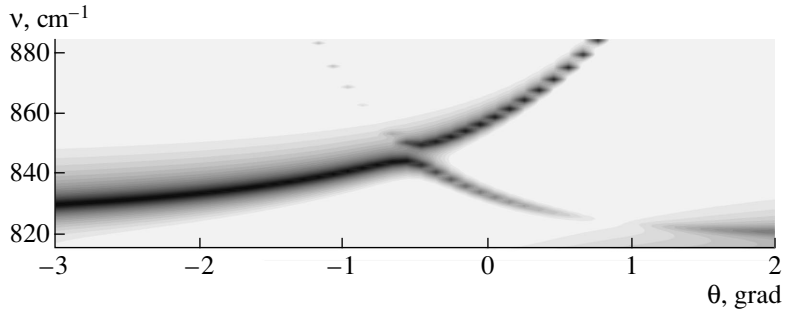

Fig. 8. Fragment of the spectrum of an iodic acid crystal, calculated taking account of the dispersion of the angle ρ. The known optical parameters were used. The orientation of the crystal was taken to be the same as for the spectrum in Fig. 7a.

dispersion in the spectrum of extraordinary polaritons can also exist for the same signs of the principal values of the permittivity. In the spectrum of ordinary polaritons, anomalous dispersion of $\varepsilon_p(\nu)$ is impossible.

The intensity distribution in the coordinates frequency–scattering angle, obtained for the given experimental conditions by means of calculations performed using Eqs. (4)–(10), (26), and (27), is presented in Fig. 8. The functions describing the dispersion of the quadratic susceptibility and the permittivity were given on the basis of measurements performed on a series of SP spectra observed at various angles ϕ between the pump wave vector and the *y*-axis. The principal values of the permittivity are given by six oscillators (see [37]); a rougher, single-pole approximation was used for the components of the quadratic susceptibility. Consequently, we attempted only to obtain qualitative agreement between the computed and observed spectra.

5. DISCUSSION

The numerical simulation of the spectra of scattering by polaritons with extraordinary polarization (anisotropic polaritons) was performed assuming that the susceptibilities are given by single-oscillator functions with small phonon damping constants. As a result, a number of features due to the anisotropy of particular dynamical parameters of the crystal lattice were identified (resonance frequencies, damping constants, phonon contributions to the first, second, and third order optical susceptibilities).

The effect of the anisotropy of the linear optical properties was shown for the example of the spectra of Raman-inactive vibrations assuming that the quadratic susceptibility does not depend on direction or frequency. This effect consists not only in the fact that the variation of the dispersion $\nu(k_p)$ changes when the direction of the wave vector changes but also in the presence of a longitudinal component of the polariton polarization, which can be much greater than the transverse component.

The manifestation of anisotropy of nonlinear optical properties depends on the symmetry type of the crystal, which corresponds to a definite form of the quadratic susceptibility tensor. The magnitude of its components at the frequencies of the dipole-active phonons depends in a definite manner on the components of the Raman scattering tensor and the phonon damping constants. Estimating the contributions of the individual components of the nonlinear susceptibilities to the intensity of scattering, it is necessary to take account of not only the polarization of the interacting waves but also the relative orientation of their wave vectors and the symmetry elements of the crystal and the signs of the principal values of the permittivity.

The main differences between the scattering spectra of extraordinary polaritons in the symmetry plane of noncentrosymmetric crystals (not only orthorhombic, but also uniaxial and monoclinic) from the spectra of ordinary polaritons are as follows.

1. In the observed spectra the frequency dependence of the scattering intensity is determined not only by the dispersion of the components of the quadratic susceptibility tensor but also by the frequency and scattering angle dependence of the direction of the wave vector of the polaritons (the angle $\rho(\nu_p, \theta)$ at each point of the spectrum for a specific experiment is determined by the matching conditions (1)). This dependence is also manifested when a single component operates. The relative orientation of the pump and polariton vectors and the symmetry elements of the crystal are also important.

2. If extraordinary polaritons are formed as a result of the coupling of an electromagnetic wave with two phonons with orthogonal dipole moments, then the spectra possessed three dispersion branches. When the dipole-moment anisotropy predominates over the deformation potential anisotropy, then along the entire middle-frequency branch one of the principal values of the permittivity ε_j is negative, and a similar section is also present on the lower-frequency branch. When the deformation potential anisotropy predominates, one value of ε_j changes sign on the middle-frequency branch and the other changes sign on the lower-frequency branch. The intensity of scattering by polaritons can vanish on sections where the contraction of the components of the quadratic susceptibility with polar-

ization unit vectors of the pump and the signal possess the same sign but the signs of the principal values of the permittivity are different.

3. Sections with anomalous dispersion which are not associated with absorption could exist in the scattering spectra for extraordinary polaritons. They are explained by the fact that the decrease of the effective value of the permittivity as a result of anisotropy as function of angle $\rho(\nu, \theta)$ predominates over the normal dispersion growth.

We also note one important feature of the phonon spectra (the region $k \gg \nu$). The intensity of scattering by a longitudinal phonon (for large values of the wave number) at the frequency ν_j^L is determined primarily by the ratio of the difference of the squared real and imaginary parts of the quadratic susceptibility $\chi_j = \chi_{ikj} e_s^i e_l^k e_p^j$ to the imaginary part ε_j. The contribution to the cubic susceptibility to scattering at the frequency ν_j^L is all the smaller, the farther away it is from the resonance frequency. A line due to scattering by a longitudinal phonon can also be observed when the line at the resonance frequency is absent because the vibration is Raman-inactive.

It is much more difficult to analyze the scattering by extraordinary polaritons than it is to analyze the spectra of transversely polarized polaritons. Nonetheless, the dynamical parameters of the vibrations can also be determined according to the spectra.

The measurements can be performed as follows. First, an orientation such that the intensity depends on the minimum number of components of the optical susceptibilities, for example, so that the scattering occurs in the symmetry plane of the crystal, must be chosen for observation. The effective values of the permittivity $\varepsilon_p(\nu)$ are determined from the maxima of the angular contours for several nonresonant sections of the spectra in a wide range of frequencies, covering the frequencies of the oscillations under study. Knowing these values for various orientations of the wave vector, the principal values of $\varepsilon_j(\nu)$ can be determined. Next, taking into account the maximum possible number of oscillators, using the same Raman scattering data, the contributions of phonons to the components of the permittivity $\Delta\varepsilon_j$ can be determined (this calculation for iodic acid is done in [37]).

To determine the phonon contributions to the quadratic susceptibility it is first necessary to find the background (slowly varying with frequency) value of the components. The effective value of the background quadratic susceptibility can be calculated after the intensity is measured at the maximum of the angular scattering contours on nonresonant sections. If the intensity is determined by a pair of components χ_{ikj}, then it is sufficient to perform measurements for a pair of differently oriented polaritons at the same frequency. The measurement error in this case is larger than for transversely polarized polaritons. The accuracy can be increased by observing the frequency shift accompanying rotation of the sample, when the intensity vanishes because of the compensation of the contributions of the components—this makes it possible to calculate the ratio of their background values. Next, it is necessary to find the sections where the signs of the components χ_{ikj} are different, and the points where each component vanishes because of compensation of the contribution of the vibrations and the background value. Sometimes the sample must be oriented so that only one component operates. As a rule, this happens when the wave vector of the pump or the polaritons is directed along the symmetry axis of the crystal. The reference points for the measurements can also be the longitudinal vibrational frequencies, where the direction of the polarization unit vector of the polaritons is constant and does not depend on the direction of the wave vector. Then the contributions of vibrations to the components of the quadratic susceptibility can be estimated. They will be determined more accurately, if several oscillators can be taken into account in the modeling of the function χ_{ikj}. In the cases where the parameters cannot be chosen uniquely, the intensity of the Raman scattering spectra for different positions of the sample can be measured on the same setup. It is also convenient to make a preliminary estimate of the damping constants Γ_j according to the Raman scattering spectra.

Given a set of preliminary approximate values, the spectra (taking account of the matching condition) can be calculated for scattering in different directions from the pump direction and the dynamical parameters of the oscillation under study can be refined, achieving agreement between the observed and computed intensity distributions.

The accuracy of such measurements and calculations must be estimated separately for each specific case, since it depends on many factors, for example, the number of characteristic vibrations in the spectrum of the crystal, anharmonicity, absorption and, first and foremost, the correct measurement of the spectral brightness of the scattering. However, irrespective of whether or not we observe scattering by ordinary or by anisotropic polaritons, this method is more accurate than other optical methods for measuring the dynamical parameters of vibrations.

6. CONCLUSION

The method described in this paper for measuring the components of the permittivity and the nonlinear susceptibility of crystals in the phonon frequency range and determining the dynamical parameters of these phonons is especially important for investigating the effect of external actions on the crystal lattice. For example, it is known that different components vary differently when the temperature is varied, electric or magnetic fields are applied, under pressure, and so on.

It is especially important to take account of the anisotropy of the interaction of lattice vibrations with an electromagnetic field when studying the kinetics of the structural transformations by the method of SP spectroscopy, since in the course of these processes not only the magnitudes of the dynamical parameters of the phonons but also the orientation of the symmetry elements of the crystal, the domain structure, and so on change.

Specific parameters of vibrations were used to construct a series of polariton spectra, but it should be noted that the choice of the ratios between them is not fundamental. For example, the spectra near phonons with large damping also can be calculated using the algorithm described above, and definite regularities in their variation as a function of the orientation of the wave vector can be found. Thus, the dynamical parameters were determined for the stretching vibration of the OH bond of the iodic acid crystal and quite good agreement was obtained between the observed and modeled spectra [38].

The central problem now is to simulate the spectra of scattering by extraordinary polaritons in the region of the bands of multiparticle states, taking account of the anisotropy of the density function of these states.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project nos. 98-02-16877 and 99-02-16418).

REFERENCES

1. Yu. N. Polivanov, Usp. Fiz. Nauk **126**, 185 (1978) [Sov. Phys. Usp. **21**, 805 (1978)].
2. V. S. Gorelik, Tr. Fiz. Inst. Akad. Nauk SSSR **132**, 15 (1982).
3. A. N. Penin and Yu. N. Polivanov, Tr. Inst. Obshch. Fiz. Akad. Nauk SSSR **2**, 3 (1986).
4. Yu. N. Polivanov, Tr. Inst. Obshch. Fiz. Akad. Nauk **43**, 3 (1993).
5. G. Kh. Kitaeva, S. P. Kulik, and A. N. Penin, Fiz. Tverd. Tela (St. Petersburg) **34**, 3440 (1992) [Sov. Phys. Solid State **34**, 1841 (1992)].
6. G. Kh. Kitaeva, A. A. Mikhaĭlovskiĭ, and A. N. Penin, Zh. Éksp. Teor. Fiz. **112**, 2001 (1997) [JETP **85**, 1094 (1997)].
7. G. H. Kitaeva, I. I. Naumova, A. A. Mikhailovsky, *et al.*, Appl. Phys. B **B66**, 201 (1998).
8. D. N. Klyshko, *Photons and Nonlinear Optics* (Nauka, Moscow, 1980).
9. T. A. Leskova, B. N. Mavrin, and Kh. E. Sterin, Fiz. Tverd. Tela (Leningrad) **18**, 3653 (1976) [Sov. Phys. Solid State **18**, 2127 (1976)].
10. M. V. Chekhova and A. N. Penin, J. Raman Spectrosc. **24**, 521 (1993).
11. O. A. Aktsipetrov, V. M. Ivanov, and A. N. Penin, Zh. Éksp. Teor. Fiz. **78**, 2309 (1980) [Sov. Phys. JETP **51**, 1158 (1980)].
12. V. L. Strizhevskiĭ and Yu. N. Yashkir, Opt. Spektrosk. **44**, 601 (1978) [Opt. Spectrosc. **44**, 349 (1978)].
13. V. M. Ivanov, T. V. Laptinskaya, and A. N. Penin, Dokl. Akad. Nauk SSSR **260**, 321 (1981) [Sov. Phys. Dokl. **26**, 859 (1981)].
14. R. Loudon, Adv. Phys. **13**, 423 (1964).
15. M. M. Sushchinsky, *Raman Spectra of Molecules and Crystals* (Nauka, Moscow, 1969; Israel Program for Scientific Translations, Jerusalem, 1973).
16. H. Poulet and J. P. Mathieu, *Vibrational Spectra and Symmetry of Crystals* (Gordon and Breach, Paris, 1970; Mir, Moscow, 1973).
17. L. Merten, Phys. Status Solidi **30**, 449 (1968).
18. S. K. Asava, Phys. Rev. B **2**, 2068 (1970).
19. W. Otaguro, E. Wiener-Avner, C. A. Arguello, and S. P. S. Porto, Phys. Rev. **4**, 4542 (1971).
20. G. Borstel and L. Merten, in *Proceedings of the International Conference on Light Scattering in Solids, Paris, 1971*, Ed. by M. Balkanski (Flammarion Science, Paris, 1971), p. 247.
21. R. Claus and H. W. Schrotter, in *Proceedings of the International Conference on Light Scattering in Solids, Paris, 1971*, Ed. by M. Balkanski (Flammarion Science, Paris, 1971), p. 255.
22. R. Claus, G. Borstel, E. Wiesendanger, and L. Steffan, Z. Naturforsch. A **27**, 1187 (1972).
23. V. L. Strigevskii and Ju. N. Jashkir, Phys. Status Solidi B **61**, 353 (1974).
24. A. S. Barker and R. Loudon, Rev. Mod. Phys. **44**, 18 (1972).
25. B. Unger and K. G. Shaack, Phys. Status Solidi B **48**, 285 (1971).
26. B. Bendow, Phys. Rev. B **2**, 552 (1971).
27. K. Kneipp, W. Wernske, H. E. Ponath, *et al.*, Phys. Status Solidi B **64**, 589 (1974).
28. V. A. Klimenko, I. I. Kondilenko, P. A. Korotkov, and G. S. Felinskiĭ, Ukr. Fiz. Zh. **26**, 1557 (1981).
29. F. X. Winter, E. Wiesendanger, and R. Claus, Phys. Status Solidi B **72**, 189 (1975).
30. B. Unger, Phys. Status Solidi B **49**, 107 (1972).
31. T. V. Laptinskaya and A. N. Penin, Izv. Akad. Nauk, Ser. Fiz. **63** (6), 1069 (1999).
32. P. N. Butcher, R. Loudon, and T. P. McLean, Proc. Phys. Soc. London **85**, 565 (1965).
33. D. N. Klyshko, Kvantovaya Élektron. (Moscow) **2**, 265 (1975).
34. V. L. Strizhevskiĭ, Zh. Éksp. Teor. Fiz. **62**, 1446 (1972) [Sov. Phys. JETP **35**, 760 (1972)].
35. V. M. Agranovich and V. L. Ginzburg, *Crystal Optics with Spatial Dispersion, and Excitons* (Nauka, Moscow, 1979; Springer-Verlag, New York, 1984).
36. M. Krauzman, M. Postollec, and J. P. Mathiev, Phys. Status Solidi B **60**, 761 (1973).
37. T. V. Laptinskaya, A. N. Penin, and M. V. Chekhova, Vestn. Mosk. Univ., Ser. 3: Fiz., Astron. **28** (3), 58 (1987).

38. T. V. Laptinskaya and A. N. Penin, in *Proceedings of XVI International Conference ICORS'98, Cape Town, South Africa, 1998*, p. 572.
39. J. E. Bjorkholm, IEEE J. Quantum Electron. **4** (11), 970 (1968).
40. A. N. Izrailenko, A. I. Kovrigin, and P. V. Nikles, Pis'ma Zh. Éksp. Teor. Fiz. **12** (10), 475 (1970) [JETP Lett. **12**, 331 (1970)].
41. G. F. Dobrzhanskiĭ, L. A. Kulevskiĭ, Yu. N. Polivanov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **12** (10), 505 (1970) [JETP Lett. **12**, 353 (1970)].
42. A. Naito and H. Inaba, Opto-Electron. **4**, 335 (1972).
43. D. N. Klyshko, V. F. Kutsov, A. N. Penin, *et al.*, Zh. Éksp. Teor. Fiz. **62** (5), 1947 (1972) [Sov. Phys. JETP **35**, 1014 (1972)].
44. V. F. Kitaeva, L. A. Kulevskiĭ, Yu. N. Polivanov, *et al.*, Dokl. Akad. Nauk SSSR **207** (6), 1322 (1972) [Sov. Phys. Dokl. **17**, 1189 (1972)].
45. V. A. Kiselev, V. F. Kitaeva, L. A. Kulevskiĭ, *et al.*, Zh. Éksp. Teor. Fiz. **62** (4), 1291 (1972) [Sov. Phys. JETP **35**, 682 (1972)].
46. V. F. Kitaeva, L. A. Kulevskiĭ, Yu. N. Polivanov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **16** (1), 23 (1972) [JETP Lett. **16**, 15 (1972)].
47. V. F. Kitaeva, L. A. Kulevskiĭ, Yu. N. Polivanov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **16** (10), 541 (1972) [JETP Lett. **16**, 383 (1972)].
48. G. M. Georgiev, A. G. Mikhaĭlovskiĭ, A. N. Penin, *et al.*, Fiz. Tverd. Tela (Leningrad) **16** (10), 2907 (1974) [Sov. Phys. Solid State **16**, 1882 (1974)].
49. Yu. N. Polivanov, Fiz. Tverd. Tela (Leningrad) **21** (6), 1884 (1979) [Sov. Phys. Solid State **21**, 1083 (1979)].
50. Yu. N. Polivanov, Pis'ma Zh. Éksp. Teor. Fiz. **30** (7), 415 (1979) [JETP Lett. **30**, 388 (1979)].
51. Yu. N. Polivanov and A. V. Shiryaeva, Kratk. Soobshch. Fiz., No. 11, 37 (1982).
52. T. V. Laptinskaya, A. G. Mikhaĭlovskiĭ, and A. N. Penin, Vestn. Mosk. Univ., Ser. 3: Fiz., Astron. **26** (4), 62 (1985).
53. M. V. Chekhova, T. V. Laptinskaya, A. N. Penin, *et al.*, Ferroelectr. Lett. Sect. **9** (3), 131, (1988).
54. M. V. Chekhova, T. V. Laptinskaya, and A. N. Penin, Opt. Commun. **73** (5), 361 (1989).
55. V. M. Ivanov, T. V. Laptinskaya, A. N. Penin, *et al.*, Fiz. Tverd. Tela (Leningrad) **31** (3), 68 (1989) [Sov. Phys. Solid State **31**, 388 (1989)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 975–982.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1125–1133.

PLASMA, GASES

Coherent Interaction of Ion and Electron Beams in Systems with Electron Cooling

V. V. Parkhomchuk and V. B. Reva*

Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia
**e-mail: reva@inp.nsk.su*

Received May 19, 2000

Abstract—A hydrodynamic approximation is used to study the behavior of dipole modes of the transverse oscillations of an ion beam in a storage ring with an electron cooling section. It is shown that in addition to the finite interaction time of the beams, instability may be caused by a specific interaction effect between the ion and electron beams in the magnetic field which leads to redistribution of energy between the various modes of the ion beam oscillations. In this case, the condition that the determinant of the transfer matrix for the cooling section does not exceed unity no longer guarantees the stability of the transverse coherent oscillations of the ion beam and all the eigenvalues of the complete matrix of the ion motion including the storage ring must be analyzed. Calculations of the stability of ion beam dipole oscillations are presented for the parameters of CELSIUS. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

By introducing damping in the oscillations of particles about an equilibrium orbit, electron cooling can easily control the beam parameters. By combining an ion beam and a high-intensity electron beam having a small momentum spread in a rectilinear storage section, it is possible to achieve efficient energy exchange between them. As it passes through the cooling section, the electron beam is lost to the collector, taking some of the ion beam thermal energy with it, which leads to an effective reduction in the transverse dimension and the momentum spread in the initial ion beam. At low ion beam intensities cooling can be successfully used in many devices [1]. However, on transition to higher beam intensities effects are observed which destroy the cooling. Dag Reistad observed this effect on CELSIUS and called it electron heating [2].

One of the mechanisms responsible for this effect may be coherent interaction of ion and electron beams in the cooling section [3]. We note that for the ions this system is open. Electrons can remove energy from the system. The work done during the reduction and thinning of the beams makes its own contribution to the energy balance.

In the present study we shall analyze the behavior of the transverse coherent oscillation modes of an ion beam in a storage ring allowing for a cooling section. Our study differs from previous studies on this topic first, by allowing for the finite interaction tine between the beams and second, by analyzing the influence of a finite magnetic field in the cooling section on the dynamics of the ion–electron interaction. Studies devoted to this topic [4, 5] generally use the approximation of an infinite magnetic field $B \longrightarrow \infty$. This has the result that the electrostatic interaction of the transverse oscillations of the electron and ion beams is not analyzed. In this case, the fast oscillations of the magnetized electron column have the frequency ω, of the order of the electron cyclotron frequency ω_{ce}, which is much higher than the characteristic frequencies of the ion motion, and the slow oscillations have the frequency ω_{pe} which is much lower. For real parameters of the problem this last condition is not always satisfied.

In order to analyze the situation, we shall use the hydrodynamic approximation, assuming that the ion and electron temperatures are zero and the beams undergo simultaneous coherent transverse motion. We shall also assume that the particle density in the beams is radially uniform and the beam radii are the same. The conducting wall is removed to infinity and image charges have no influence on the beam dynamics. The beams are matched for a time shorter than all the other characteristic times of the problem. Outside the cooling section the beam propagates in the storage ring with azimuthal focusing symmetry.

We shall first use a simple model to analyze the influence of the finite time of joint beam motion on the interaction of the electron and ion beams. We shall then study the behavior of the dipole oscillation mode of an ion beam in a storage ring with a cooling section and we shall calculate the instability growth rates for a specific device.

2. MODEL OF ION AND ELECTRON BEAM INTERACTION IN THE ABSENCE OF A MAGNETIC FIELD

By way of a simple example to illustrate the effect, we shall consider a model problem in which the ions and electrons move jointly in drift space in the absence

of a magnetic field. The equations of motion in a frame of reference moving with the beam may be written as

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} - eZ_i\mathbf{E} = 0, \quad (1)$$

$$m_e \frac{d^2\mathbf{r}_e}{dt^2} - e\mathbf{E} = 0, \quad (2)$$

where $\mathbf{r}_{i,e}$ are the position vectors of the ions and electrons and $\mathbf{E} = \mathbf{E}_e + \mathbf{E}_i$ is a superposition of the space charge fields of the electron and ion beams.

In the limit, where the transverse dimension of the beams is small compared with the perturbation wavelength, the electric fields will be described by the expressions

$$\mathbf{E}_e = -\frac{m_e}{e}\omega_{pe}^2(\mathbf{r} - \mathbf{R}_e),$$

$$\mathbf{E}_i = \frac{m_p A_i}{eZ_i}\omega_{pi}^2(\mathbf{r} - \mathbf{R}_i),$$

where

$$\omega_{pi}^2 = \frac{2\pi n_i Z_i^2 e^2}{\gamma m_p A_i}, \quad \omega_{pe}^2 = \frac{2\pi n_e e^2}{\gamma m_e}$$

are the plasma frequencies of the ions and electrons in a bounded plasma, $\mathbf{R}_{i,e} = (X_{i,e}, Y_{i,e})$ are the positions of the beam centers, $N_{i,e}$ are the ion and electron densities, $\mathbf{r}$ is the position vector, Z_i is the ion charge, A_i is the ion mass number, and γ is the relativistic factor. Expanding

$$\mathbf{r}_{i,e} = \mathbf{R}_{i,e} + \mathbf{r}'_{i,e}$$

in terms of the position of the beam center $\mathbf{R}_{i,e}$ and the position of a particle in it $\mathbf{r}'_{i,e}$, after integrating the equations of motion of the particles over the beam cross section S we obtain a system describing the dynamics of the centers:

$$\frac{d^2\mathbf{R}_i}{dt^2} + \omega_{ie}^2\mathbf{R}_i = \omega_{ie}^2\mathbf{R}_e, \quad (3)$$

$$\frac{d^2\mathbf{R}_e}{dt^2} + \omega_{ei}^2\mathbf{R}_e = \omega_{ei}^2\mathbf{R}_i, \quad (4)$$

where

$$\omega_{ie}^2 = \frac{2\pi e^2 Z_i n_e}{\gamma A_i m_p}, \quad \omega_{ei}^2 = \frac{2\pi e^2 Z_i n_i}{\gamma m_e}$$

are the frequencies of the beam oscillations in a space charge field of opposite sign. It can be seen from Eqs. (3) and (4) that in this case, the vertical and radial motion of the particles is independent and the four-dimensional problem is reduced to a two-dimensional one.

A consistent solution of the linear equations (3) and (4) under the initial conditions for the electrons

$$\mathbf{R}_e = 0, \quad \frac{d\mathbf{R}_e}{dt} = 0$$

for each component of the vector $\mathbf{R}_i$ may be written in the form

$$\begin{pmatrix} X_i \\ \frac{dX_i}{dt} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} X_i^0 \\ \frac{dX_i^0}{dt} \end{pmatrix} = A_{\text{cool}} \begin{pmatrix} X_i^0 \\ \frac{dX_i^0}{dt} \end{pmatrix}, \quad (5)$$

which links the position and velocity of the ion beam center "after" the cooling section with their values "before" this section using a linear transformation [6]. The elements of the matrix A_{cool} are determined by integrating the complete system of equations of motion (3) and (4) with unit initial conditions for the ions and zero initial conditions for the electrons. For the Y component of the beam motion all the equations are written similarly.

The solutions of the system (3) and (4) correspond to stable beam motion if all the moduli of the eigenvalues $|\lambda_k|$ of the matrix of motion (5) are less than or equal to unity. Otherwise, two physically different situations are possible. In the first we find $|\lambda_k| > 1$ but the determinant of the matrix

$$\det A_{\text{cool}} = \prod_k |\lambda_k|$$

remains equal to (or less than) unity. This implies that the ion energy is conserved (or decreases) but is redistributed between different modes of the ion beam oscillations. The possibility of energy being redistributed between the modes by means of elements positioned outside the cooling section and of beam stability being achieved depends on the specific physical conditions and requires a separate analysis. In the second situation when the determinant of the matrix is also greater than unity, it is impossible to achieve beam stability without introducing additional dissipative forces.

The results of an investigation of the matrix (5) may be summarized as follows. First, the moduli of the eigenvalues of the matrix for the cooling section are the same, i.e., $|\lambda_1| = |\lambda_2|$ and consequently

$$|\lambda_{1,2}|^2 = \det A_{\text{cool}}.$$

Second, for an arbitrary beam interaction time $\tau = L_{\text{cool}}/\gamma V_0$ we have

$$\det A_{\text{cool}} = \frac{\omega_{ie}^4 + \omega_{ei}^4 + 2\omega_{ie}^2\omega_{ei}^2\cos(\omega_0\tau)}{(\omega_{ie}^2 + \omega_{ei}^2)^2} + \frac{\omega_{ie}^2\omega_{ei}^2}{(\omega_{ie}^2 + \omega_{ei}^2)^2}\sin(\omega_0\tau)\omega_0\tau, \quad (6)$$

where $\omega_0^2 = \omega_{ie}^2 + \omega_{ei}^2$. A graph of this function is plotted in Fig. 1. The first term in this expression may be interpreted as the transfer of energy from the ions to the electrons in their interaction process and its value is strictly less than unity. The second term allows for the binding energy which occurs when the beams are matched. For a short interaction time ($\omega_0\tau \ll 1$) when the relative displacement of the beams during their joint motion is small, the difference between the energy of the electrostatic interaction between the beams at the beginning and end of the cooling section can be neglected. The ions transfer some of their energy of transverse motion to the electrons and the amplitude of the ion beam oscillations decreases. However, for interaction times $\tau \sim \omega_{ie}^{-1}$ a situation arises where the discontinuity of the ion coupling with the electrons leads to an increase in the total energy of the system some of which is transferred to the transverse motion of the ion beam, and the amplitude of the oscillations increases.

We write the energy of the complete electron + ion system:

$$W = n_i W_i + n_e W_e + W_{ie} = \frac{n_i m_i \dot{\mathbf{R}}_i^2}{2} + \frac{n_e m_e \dot{\mathbf{R}}_e^2}{2} + 2\pi n_i n_e e^2 (\mathbf{R}_i - \mathbf{R}_e)^2.$$

The first term in this expression is the energy of the ion motion, the second is that of the electrons, and the third allows for the work required to create the ion + electron system. Since the electron velocity at the beginning of the cooling section is zero, an increase in the electron energy will subsequently lead to a reduction in the coherent fluctuations of the ions, i.e., to "cooling." The existence of the third term changes the situation and growth of the perturbations becomes possible.

3. DIPOLE OSCILLATIONS OF AN ION BEAM IN A STORAGE RING WITH AN ELECTRON COOLING SECTION

3.1. Cooling Section

We shall now consider a situation more consistent with real conditions. The electron and ion beams propagate in the direction of the external magnetic field. The characteristic relationship between the parameters of the problem is:

$$\omega_{ce} \gg \omega_{pe}, \omega_{ci}, \omega_{pi},$$

where

$$\omega_{ce} = \frac{eB}{m_e c}, \quad \omega_{ci} = \frac{eZ_i B}{A_i m_p c}$$

are the cyclotron frequencies of the electrons and ions, respectively. The equations for the transverse motion of the ion beam center in a reference frame moving with the beam are written in the following form:

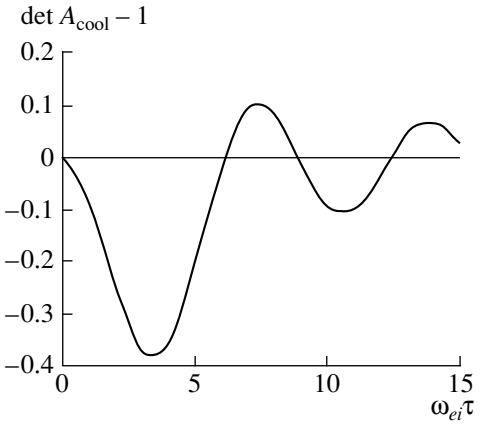

Fig. 1. Determinant of the cooling section matrix for joint motion of the beams without a magnetic field ($\omega_{ie}\tau = 1$).

$$\frac{d^2 X_i}{dt^2} - \omega_{ci}\frac{dY_i}{dt} + \omega_{ie}^2 X_i = \omega_{ie}^2 X_e, \tag{7}$$

$$\frac{d^2 Y_i}{dt^2} + \omega_{ci}\frac{dX_i}{dt} + \omega_{ie}^2 Y_i = \omega_{ie}^2 Y_e. \tag{8}$$

In this case the vertical and radial beam motion ceases to be independent. However, introducing the new variables

$$z_i = X_i + iY_i, \quad z_e = X_e + iY_e,$$

we can transform the system describing the combined behavior of the electron and ion beams to the simpler form:

$$\frac{d^2 z_i}{dt^2} + i\omega_{ci}\frac{dz_i}{dt} + \omega_{ie}^2 z_i = \omega_{ie}^2 z_e, \tag{9}$$

$$\frac{d^2 z_e}{dt^2} - i\omega_{ce}\frac{dz_e}{dt} + \omega_{ei}^2 z_e = \omega_{ei}^2 z_i. \tag{10}$$

In the calculations of A_{cool} we can use the fact that the determinants of the matrices

$$J = A + iB, \quad R = ((A, B), (-B, A))$$

(where A and B are arbitrary matrices with real coefficients) are related by $\det R = |\det J|^2$.

Since the electron motion is strongly magnetized, Eq. (10) describing the motion of the electron beam center may be reformulated using the drift approximation

$$\mathbf{v}_e = \frac{c\mathbf{E} \times \mathbf{B}}{B^2}$$

as follows:

$$\frac{dz_e}{dt} + i\Lambda z_e = i\Lambda z_i, \tag{11}$$

where $\Lambda = \omega_{ei}^2/\omega_{ce}$ is the drift rotation frequency of the electrons in the ion space charge field.

Solving the system (9) and (11) in the limit of short interaction times

$$\tau \ll \omega_{ci}^{-1}, \omega_{ie}^{-1}, \Lambda^{-1},$$

we obtain the following expressions for the eigenvalues:

$$|\lambda_1|^2 = 1 + \frac{\omega_{ie}^2\Lambda}{L}\tau^2, \quad |\lambda_2|^2 = 1 - \frac{\omega_{ie}^2\Lambda}{L}\tau^2. \tag{12}$$

It can be seen that for this case the eigenvalues of the various modes differ. In addition to a damped mode, there is a mode which grows for an arbitrarily short interaction time. The electron beam acts as an "intermediary" transferring energy from one mode to another. An increase in the space-charge electric field of each component ($E_{i,e} \propto N_{i,e} \propto \omega_{ei,ie}$) leads to an increase in the instability growth rate. An increase in the magnetic field reduces this value.

The determinant of the matrix in this model is given by

$$\det A_{\text{cool}} = 1 - \frac{1}{12}\omega_{ie}^2\Lambda(2\Lambda - \omega_{ci})\tau^4. \tag{13}$$

An interesting characteristic of this expression is the lack of dependence on the electron beam density. The electron beam density changes the magnitude of the difference $\det(A_{\text{cool}})$ from unity but not the sign.

This expression supplements the results obtained in [7] assuming that the magnetic field has no influence on the ion dynamics in the cooling section (i.e., $\omega_{ci} \longrightarrow 0$). Allowance for this influence narrows the region in which the determinant of the matrix of the cooling section does not exceed unity, by imposing the constraint:

$$4\pi A_i m_p c^2 n_i > B^2$$

(assuming that the beam interaction time is short). In addition, as will be shown in Section 3.2, the condition

$$\det A_{\text{cool}} = 1$$

is necessary but not sufficient for stable ion motion.

The redistribution of energy between the ion modes can also be demonstrated by reducing the system (9) and (10) to the form of coupled oscillations. We make the change of variables

$$a_i = \frac{dz_i}{dt} - i\Omega_{1i}z_i, \quad b_i = \frac{dz_i}{dt} - i\Omega_{2i}z_i, \tag{14}$$

$$a_e = \frac{dz_e}{dt} - i\Omega_{1e}z_e, \quad b_e = \frac{dz_e}{dt} - i\Omega_{2e}z_e, \tag{15}$$

where

$$\Omega_{1i,2i} = \frac{\omega_{ci} \pm \sqrt{\omega_{ci}^2 + 4\omega_{ie}^2}}{2},$$

$$\Omega_{1e,2e} = \frac{-\omega_{ce} \pm \sqrt{\omega_{ce}^2 + 4\omega_{ei}^2}}{2}$$

are the partial oscillation frequencies of the ion and electron beams, respectively, a_i and b_i are the oscillation amplitudes of the ion modes, and a_e and b_e are the oscillation amplitudes of the electron modes. Without limiting the generality we shall take the partial frequencies such that

$$\Omega_{2i} - \Omega_{1i} > 0, \quad \Omega_{2e} - \Omega_{1e} < 0.$$

The system (9) and (10) can then be transformed as follows:

$$\frac{da_i}{dt} - i\Omega_{2i}a_i = \frac{i\omega_{ie}^2 b_e}{\Omega_{2e} - \Omega_{1e}}, \tag{16}$$

$$\frac{db_i}{dt} - i\Omega_{1i}b_i = \frac{i\omega_{ie}^2 b_e}{\Omega_{2e} - \Omega_{1e}}, \tag{17}$$

$$\frac{db_e}{dt} - i\Omega_{1e}b_e = \frac{i\omega_{ei}^2(b_i - a_i)}{\Omega_{2i} - \Omega_{1i}}. \tag{18}$$

The second electron mode corresponds to fast electron motion at a frequency of the order of ω_{ce} so that within the limits of our analysis its influence on the dynamics of the slow ion modes can be neglected.

We shall consider the case when each ion mode a_i and b_i interacts independently with the electron mode b_e:

$$\frac{da_i}{dt} - i\Omega_{2i}a_i = \frac{i\omega_{ie}^2 b_e}{\Omega_{2e}}, \tag{19}$$

$$\frac{db_e}{dt} - i\Omega_{1e}b_e = -\frac{i\omega_{ei}^2 a_i}{\Omega_{2i} - \Omega_{1i}}. \tag{20}$$

The equations for interaction of the modes b_i and b_e are written similarly. The solution of this system of equations with the initial conditions

$$a_i = a_i^0, \quad b_e = 0$$

gives the following result:

$$|a_i|^2 = |a_i^0|^2 \times \left[1 + \frac{\Delta^2}{(\Omega_{2i} - \Omega_{1e})^2 - \Delta^2}\sin^2\left(\frac{(\omega_+ - \omega_-)t}{2}\right)\right], \tag{21}$$

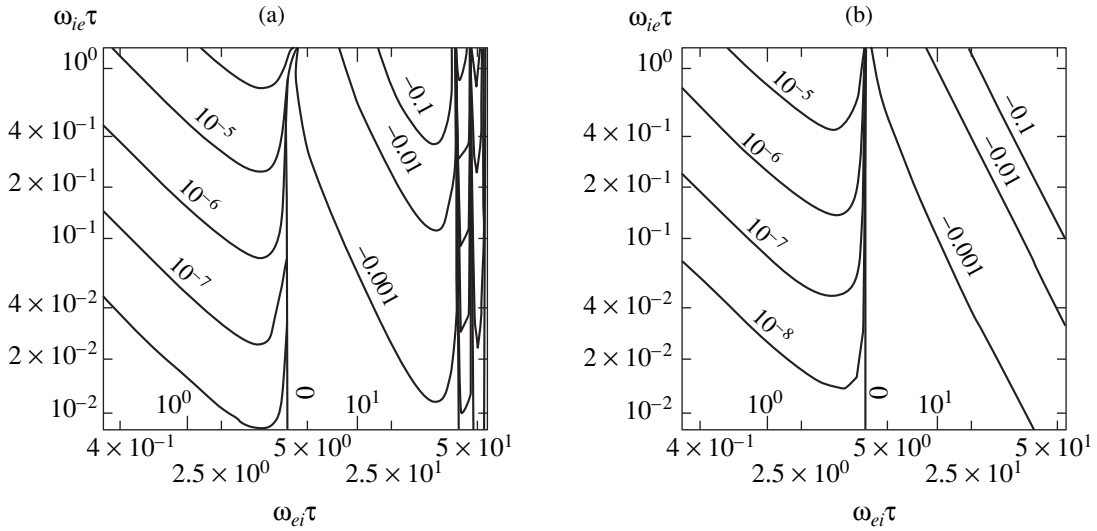

Fig. 2. Determinant of the cooling section matrix. Diagram (a) gives the value of det $A_{\text{cool}} - 1$ calculated using the complete systems of Eqs. (9) and (10) and diagram (b) gives that calculated using the formulas for a short ion–electron beam interaction time (13).

$$|b_i|^2 = |b_i^0|^2 \times \left[1 - \frac{\Delta^2}{(\Omega_{2i} - \Omega_{1e})^2 - \Delta^2} \sin^2\left(\frac{(\omega_+ - \omega_-)t}{2}\right)\right], \quad (22)$$

where

$$\Delta^2 = \frac{4\omega_{ie}^2\omega_{ei}^2}{|\Omega_{2e}|(\Omega_{2i} - \Omega_{1i})},$$

and

$$(\omega_+ - \omega_-)^2 = (\Omega_{2i} - \Omega_{1e})^2 - \Delta^2$$

for the interaction of modes a_i and b_e

$$(\omega_+ - \omega_-)^2 = (\Omega_{1i} - \Omega_{1e})^2 - \Delta^2$$

for the interaction of modes b_i and b_e.

It can be seen that the ion beam oscillation mode b_i loses energy and $|b_i/b_i^0| \le 1$ for any ion beam interaction time. When the ion mode a_i interacts with the electron beam, its energy increases with time and only at times when

$$\sin((\omega_+ - \omega_-)t/2) = 0,$$

does the energy of the mode a_i reach its initial value which it had before the onset of interaction.

In order to determine to what extent these effects may be significant in the physics of real systems, we shall present calculations made using the complete system of Eqs. (9) and (10) for the parameters corresponding to those of CELSIUS: length of cooling section $L_{\text{cool}} = 250$ cm, ion velocity $V_0 = 9 \times 10^9$ cm/s, and magnetic field $B = 500$ G. Figures 2 and 3 show contours of the matrix determinant $\det A_{\text{cool}} - 1$ and $|\lambda_{\max}| - 1$ in the plane of the parameters $\omega_{ie}\tau$ and $\omega_{ei}\tau$ for the matrix of the cooling section corresponding to the drift of ions in the cooling section. It can be seen that for the determinant the value $\omega_{ei}\tau \approx 4$ divides the plane of the parameters into two regions. For $\omega_{ei}\tau < 4$ the matrix determinant is greater than unity, which corresponds to instability whereas for $\omega_{ei}\tau > 4$ the determinant is much smaller than unity. In the region $\omega_{ei}\tau > 40$ we observe a rapid exchange of zones of stability and instability. The maximum eigenvalue of the matrix of the cooling section behaves completely differently for the same parameters. Its value is always much greater than unity and increases monotonically with increasing parameter $\omega_{ie}\omega_{ei}\tau^2 \propto (n_i n_e)^{1/2}$. For comparison we also plot contours obtained using Eq. (12) for short interaction times. It can be seen that for low values of ω_{ie} and ω_{ei} these fairly accurately describe the qualitative pattern of instability behavior as a function of the parameters of the cooling section. We also note that the difference from unity for the eigenvalues of the matrix is substantially greater than the difference from unity for the determinant.

3.2. Storage Ring with Cooling Section

In order to determine the complete dynamics of the beam in the storage ring, we need to supplement the matrix of the coolant section with the matrix describing the ion beam motion in the storage ring. For dipole oscillations this will be the well-known Twiss matrix:

$$\begin{pmatrix} \cos(2\pi\nu) & \beta\sin(2\pi\nu)/V_0 \\ -V_0\sin(2\pi\nu)/\beta & \cos(2\pi\nu) \end{pmatrix}, \quad (23)$$

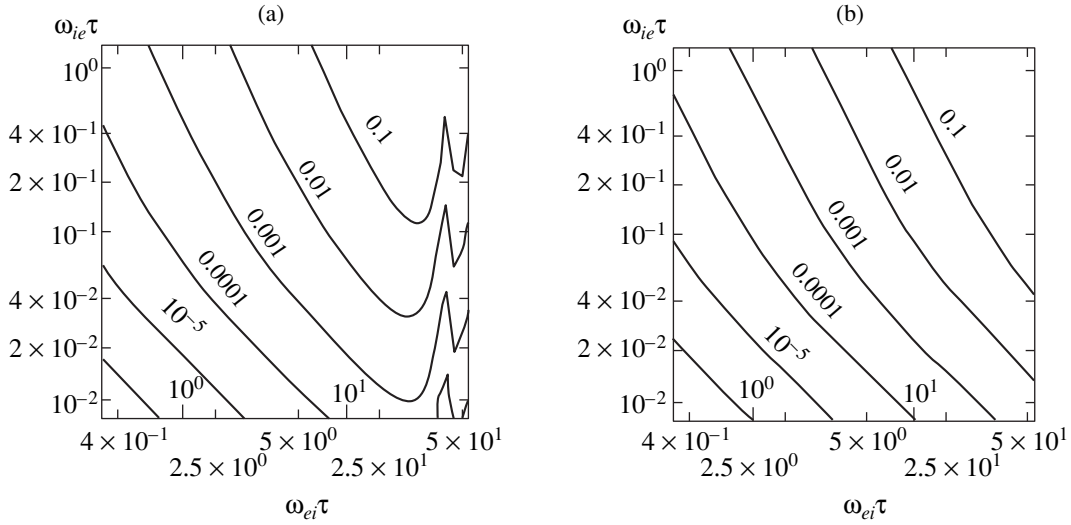

Fig. 3. Modulus of the maximum eigenvalue of the cooling section matrix $|\lambda_{max}| - 1$. Diagram (a) was calculated using the complete system of Eqs. (9) and (10) and diagram (b) was calculated using the formulas for a short ion–electron beam interaction time (13).

where β is the betatron function of the storage ring, ν is the betatron tune.

The following sequence of operations was performed to correctly match the cooling section with the storage ring. Before beginning to propagate in the cooling section, the ion beam was displaced longitudinally over the length $-L_{cool}/2$ by using a matrix inverse to the drift section matrix. After passing through a cooling section of length L_{cool}, this operation was repeated before using the Twiss matrix. The matrix of the complete storage ring was thereby converted into a matrix corresponding to the storage ring without a cooling section.

For the dipole mode and a short beam interaction time, the eigenvalues of the resultant matrix can be calculated analytically in the simple form:

$$|\lambda_1|^2 = 1 + \frac{1}{2}\frac{\beta}{\gamma V_0}\omega_{ie}^2\Lambda\tau^2,$$
$$|\lambda_2|^2 = 1 - \frac{1}{2}\frac{\beta}{\gamma V_0}\omega_{ie}^2\Lambda\tau^2, \quad (24)$$

where allowance is made for an inverse transition to the laboratory frame using the transformation

$$\begin{pmatrix} 1 & 0 \\ 0 & 1/\gamma \end{pmatrix} A_{cool} \begin{pmatrix} 1 & 0 \\ 0 & \gamma \end{pmatrix}. \quad (25)$$

Expressions (24) can also be rewritten in the following form, having collated the definitions for Λ, ω_{ie}, and τ:

$$\lambda_{1/2}^2 = 1 \pm 2\pi^2\frac{n_e n_i}{\gamma^5}(\beta r_e r_i L_{cool}^3)\frac{c^4}{V_0^4}\left(\frac{V_0}{\omega_{ce}L_{cool}}\right), \quad (26)$$

where

$$r_e = \frac{e^2}{m_e c^2}, \quad r_i = \frac{(z_i e)^2}{A_i m_p c^2}$$

are the classical radii of the electron and ion, respectively.

It can be seen from (24) that allowance for the storage ring section slightly changes the relationship between the maximum and minimum eigenvalues of the cooling section matrix but as before, keeps one of these greater than unity, which is slightly unusual for accelerator physics. Generally, if the parameters of the elements forming the storage ring channel do not go beyond certain limits, it is possible to obtain stable motion of the various particles characterized by the condition $|\lambda_k| \equiv 1$ for all k. For the dynamics of the collective dipole oscillation mode this is not the case and although the amplification of the focusing (reduction in the β function) and an increase in the energy of the cooled ions leads to a reduction in the growth rate of this instability, it does not suppress it completely.

In order to explain this effect we propose the following approximate physical model. Although these beam oscillations are electrostatic and curl $\mathbf{E} = 0$, when the vector fields are averaged over fast oscillations ($\omega \sim \omega_{ce}$) a nonpotential vortex component may appear [8]. This can occur if the field $\mathbf{E}$ is noncollinear to the ion displacement induced by it, which is readily achieved in the presence of a longitudinal magnetic field. In the presence of nonzero curl $\mathbf{E}$, the ion motion in the cooling section is equivalent to the motion in a structure described by the following equation of motion:

$$\frac{d^2 z_i}{dt^2} + i\omega_{ci}\frac{dz_i}{dt} + (\omega^2 + i f_c^2)z_i = 0. \quad (27)$$

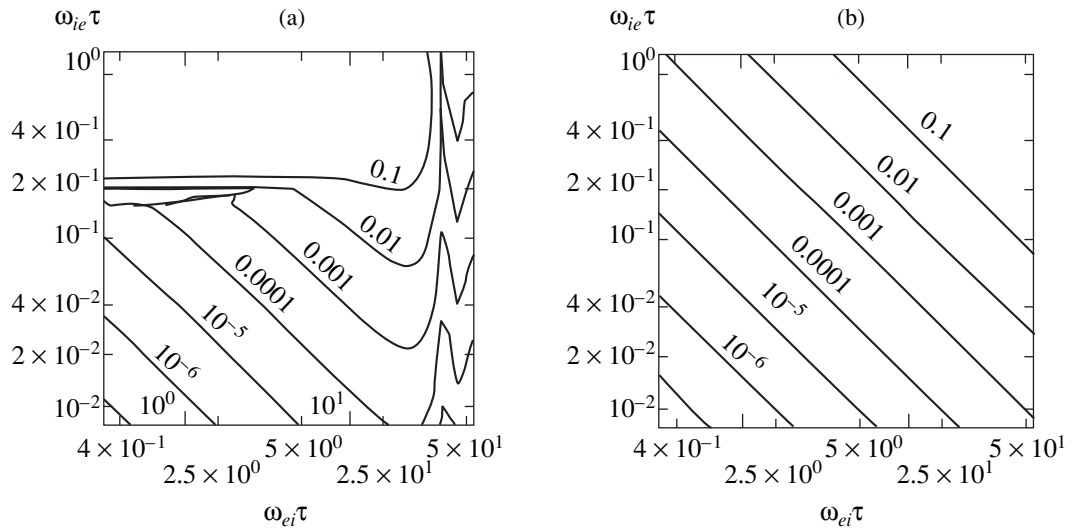

Fig. 4. Maximum eigenvalue $|\lambda_{\max}| - 1$ of the complete matrix of the storage ring ($\nu = 1.45$, $\beta_x = \beta_y = 900$ cm, $V_0 = 9 \times 10^9$ cm/s). The ion motion in the cooling section was calculated (a) using the complete system of Eqs. (9) and (10) and (b) using the formulas for a short ion–electron beam interaction time (24).

The term $i f_c^2 z_i$ is associated with the additional transverse nonpotential field

$$\mathbf{F} = m_p f_c^2(-y, x).$$

It can be shown that the influence of this section on the ion motion in the storage ring leads to the appearance of eigenvalues having a modulus which differs from unity and no choice of accelerator parameters ν and β can make this equal to unity. The value of the matrix determinant is then identically equal to unity. In the limit of a short drift time in this section, the eigenvalues of the complete matrix of this structure are

$$|\lambda_1|^2 = 1 + \frac{1}{2}\frac{\beta}{\gamma V_0} f_c^2 \tau, \quad |\lambda_2|^2 = 1 - \frac{1}{2}\frac{\beta}{\gamma V_0} f_c^2 \tau. \quad (28)$$

However, unlike the situation obtained when the ions drift in the cooling section, in this model the effect is first order with respect to the time of interaction with the structure.

In order to study the beam dynamics in a storage ring with a moderately long interaction time we shall again use the CELSIUS parameters. We first use (24) to estimate the possible magnitude of the effect. For the working parameters, electron beam current $I_e \approx 0.5$ A, proton beam current $I_i \approx 10$ mA, beam radii $D_i = D_e \approx 2$ cm, we have the following values:

$$\omega_{ie}\tau \approx 0.3, \quad w_{ei}\tau \approx 1.6.$$

The corresponding moduli of the eigenvalues are $\lambda_{1/2} \approx 1 \pm 7 \times 10^{-4}$ which is fairly close to the Landau damping decrement. For this device its value is of the order of 10^{-3} which implies mixing of the fluctuations in phase space caused by a spread of the betatron oscillation frequencies over 10^3 ion revolutions in the storage ring. The similarity between the instability growth rate and the Landau decrement implies that as the charge density in the ion or electron beams increases further, the ensuing fluctuations of the ion beam dipole oscillations will be amplified as it passes through the cooling section.

The results of numerical calculations for arbitrary values of $\omega_{ie}\tau$ and $\omega_{ei}\tau$ using the system of Eqs. (9) and (10) are plotted in Fig. 4. As for the case of a short interaction time, allowance for the storage ring does not significantly alter the maximum eigenvalue of the matrix describing the ion dynamics. We note that at low ion densities additional constraints are imposed on the electron current density (the threshold value $\omega_{ie}\tau \approx 2 \times 10^{-1}$) because the frequency shift of the betatron oscillations in the space charge field is too large.

It can be seen from these estimates and numerical calculations for an arbitrary interaction time that the effect associated with energy redistribution between modes may be much larger than the effect caused by the determinant of the cooling section matrix differing from unity. Thus, we propose the following mechanism for the buildup of instability. If the instability growth rate associated with one of the eigenvalues exceeding unity is smaller than the decrement associated with the frequency spread of the betatron oscillations (Landau damping), the particle energy will be redistributed between various collective modes within a single revolution. On average over many revolutions the change in the ion energy will be determined by the relationship between all the eigenvalues of the cooling section matrix, i.e., its determinant. If the instability growth rate begins to exceed the Landau decrement, over a single revolution a fluctuation which has not had time to mix in phase space, acquires an additional energy

increment as it passes through the cooling section, and within a number of revolutions

$$N \sim (|\lambda_k| - 1)^{-1}$$

the beam is lost.

4. CONCLUSIONS

The results of the analysis presented above show that interaction of an electron beam with ions over a finite time may lead to instability of the coherent ion oscillations in a storage ring. In the hydrodynamic approximation two physically different effects are possible.

The first effect is associated with the determinant of the cooling section matrix differing from unity, i.e., with a change in the total energy of the ion beam. We note that the finite time of interaction between the ions and electrons leads to instability. The entry of ions into the electron interaction section followed by the disruption of this interaction are sources of energy transfer to the ion transverse motion.

The second effect is associated with the redistribution of energy between the various ion beam oscillation modes in the cooling section. In this case the magnetic field has an important influence. Even a relatively low value $B \approx 100$ G leads to the buildup of instability which is characterized by an increase in the amplitude of one oscillation mode as a result of a decrease in the other. In the limit of a short ion beam drift time in the cooling section, allowance for the storage ring does not alter the situation, keeping one oscillation mode growing and the other decaying. Thus, the condition that the determinant of the transfer matrix does not exceed unity no longer guarantees the stability of the transverse coherent ion oscillations in systems with electron cooling and it is necessary to analyze all the eigenvalues of the complete matrix of the ion motion including the storage ring.

ACKNOWLEDGMENTS

The authors are grateful to K.V. Lotov, D.V. Pestrikov, and A.N. Skrinskiĭ for useful discussions of the problem.

REFERENCES

1. V. V. Parkhomchuk and A. N. Skrinsky, Rep. Prog. Phys. **54**, 921 (1991).
2. D. Reistad *et al.*, in *Workshop on Beam Cooling and Related Topics, Switzerland, Montreux, 1994*, CERN 94-03, p. 183.
3. V. V. Parkhomchuk, Nucl. Instrum. Methods Phys. Res. A **441**, 9 (2000).
4. N. S. Dikanskiĭ and D. V. Pestrikov, *Physics of Intense Beams in Accumulators* (Nauka, Moscow, 1989), p. 269.
5. A. V. Burov, Part. Accel. **57**, 131 (1997).
6. V. V. Parkhomchuk, in *Proceedings of Conference on Space Charge Effect in Formation of Intense Low Energy Beams, Joint Institute for Nuclear Research, Dubna, 1999*, p. 153.
7. P. R. Zenkevich and A. E. Bolshakov, Nucl. Instrum. Methods Phys. Res. A **441**, 36 (2000).
8. G. M. Zaslavskiĭ and R. Z. Sagdeev, *Introduction to the Nonlinear Physics* (Nauka, Moscow, 1988), p. 53.

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 983–990.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1134–1142.

FLUIDS

Bipolarons in a KCl Melt

G. N. Chuev

Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino, Moscow oblast, 142290 Russia
e-mail: gena@impb.psn.ru

Received April 24, 2000

Abstract—A theory of two excess electrons in alkali halide melts is developed using variational estimates of path integrals. As a result of the strong screening, the average field generated by the ions has little influence on the electrons and the problem reduces to a study of a bipolaron type of free energy functional. The behavior of this functional is determined as a function of the thermodynamic and structural characteristics of the melt. Variational bipolaron calculations are made using the approximation of uncorrelated electrons and using Kohn–Sham theory to allow for electron–electron correlations. The results of the calculations using Kohn–Sham theory agree with the data obtained by quantum molecular dynamics and show that a correct choice of trial wave function which allows explicitly for the correlation of two electrons is required to obtain a correct estimate of bipolaron stability. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

An investigation of the bipolaron, two excess electrons forming a bound state or quasi-molecule, is one of the most interesting and intriguing problems in the physics of condensed media. The formation of an F' center (two electrons bound at an anionic vacancy) in alkali halide crystals has been fairly well studied experimentally [1]. Various experimental data are available on the possible formation of bipolarons in polar liquids [2]. It has been shown that bipolarons may play an important role in the mechanism for superconductivity [3]. Various experimental observations have been made to support the existence of bipolarons in molecular organic polymers [4].

Recently, the bipolaron problem has started to be extensively studied by numerical calculations based on quantum molecular dynamics methods [5–10]. Calculations of bipolaron states using the Car–Parinello method stimulated numerous studies on quantum molecular dynamics. According to these studies, bipolaron states are stable in water [8] and in metal–ammonia solutions [9, 10] at metal concentrations of approximately 1% and, as the metal concentration increases further, the bipolarons form clusters. However, the complexity and cumbersome nature of the calculations has prevented sufficiently detailed full-scale investigations of bipolarons from being made.

From the theoretical point of view, the behavior of bipolarons in a liquid has been less well studied than the behavior of a solvated electron or polaron. Various studies [11–15] have been made using the continuum model and these have generally all been devoted to investigating the criterion for the existence of bipolarons. Studies have also been reported where bipolaron states were investigated using a semicontinuous approximation (see [16] and references therein) for various liquids (water, ammonia) and polar matrices. According to these studies, in all these media bipolaron states are stable and energetically more favorable than two single-electron states. At present, no experimental data are available to confirm these calculations.

In the present study we apply a method [17, 18] developed recently to investigate the behavior of self-trapped electron states in alkali halide melts to the bipolaron problem. On the basis of this method we use variational estimates of the partition function to obtain the free energy functional of a bipolaron and then determine the bipolaron behavior as a function of the structural and thermodynamic parameters of the melt. This functional is investigated by variational methods using the simplest multiplicative approximation for a bipolaron wave function in the uncorrelated electron approximation. Allowance for electron–electron correlations will be made using local density functional theory (LDA) [19, 20].

As a clear example we describe calculations of bipolaron states in a KCl melt. This system was chosen because the relationships for the free-energy functional are simplest for alkali halide melts as a result of the strong screening. In addition numerical data obtained using the Car–Parinello method are available for a bipolaron in KCl [5–7].

2. FORMULATION OF THE PROBLEM

We shall consider two excess electrons situated in a melt of an alkali halide salt such as KCl. This system is a liquid electrolyte. In the statistical approach the problem reduces to calculations of a grand partition function which contains the configuration integral over the

ion coordinates $\mathbf{R}^{\{N\}} = \mathbf{R}_{i-}^{\{N/2\}}\,\mathbf{R}_{i+}^{\{N/2\}}$ and path integrals over the coordinates of both electrons $r_1(\tau)$ and $r_2(\tau)$:

$$\Xi = \sum_{N\geq 0}^{\infty} \frac{[(2\pi M)^{-3/2}\beta^{3/2}V\exp(\beta\mu)]^N}{N!} \times \int D[\mathbf{r}_1(\tau)]\int D[\mathbf{r}_2(\tau)]\int d\mathbf{R}^{\{N\}} \times \exp[-\beta W(\mathbf{r}_1, \mathbf{r}_2, \mathbf{R}^{\{N\}})]. \quad (1)$$

Here we used a system of units in which $\hbar = 1$, $m = 1$, $e = 1$ and also the following notation: V is the volume of the system, $\beta = (k_B T)^{-1}$ is the reciprocal temperature, μ and M are the chemical potential and mass of the ions, respectively. In order to simplify the calculations we shall assume that the chemical potentials (μ_+, μ_-) and the ion masses (M_+, M_-) are equal, i.e., $\mu_+ = \mu_- = \mu$ and $M_+ = M_- = M$, and the ion charge is ± 1, i.e., the melt is considered to be a symmetric 1–1 electrolyte.

In Eq. (1), the interaction energy $W(\mathbf{r}_1, \mathbf{r}_2, \mathbf{R}^{\{N\}})$ contains the ion interaction potential $U_{ii}(\mathbf{R}^{\{N\}})$ and the action for the bipolaron S_2:

$$W(\mathbf{r}_1, \mathbf{r}_2, \mathbf{R}^{\{N\}}) = U_{ii}(\mathbf{R}^{\{N\}}) + \beta^{-1}S_2. \quad (2)$$

In turn the action for the bipolaron S_2 contains the kinetic terms ($\propto \dot{r}^2$), the Coulomb repulsion of two electrons ($\propto |r_1 - r_2|^{-1}$), and also the interaction with the liquid particles:

$$S_2 = \int_0^{\beta} d\tau \left[\frac{\dot{r}_1^2(\tau) + \dot{r}_2^2(\tau)}{2} + \frac{1}{|\mathbf{r}_1(\tau) - \mathbf{r}_2(\tau)|} + \sum_i^N \{ u_{\pm}(\mathbf{r}_1(\tau) - \mathbf{R}_{i\pm}) + u_{\pm}(\mathbf{r}_2(\tau) - \mathbf{R}_{i\pm}) \} \right], \quad (3)$$

where $u_{\pm}(\mathbf{r} - \mathbf{R}_{i\pm})$ are the corresponding electron–anion and electron–cation interaction potentials, and $\mathbf{R}_{i+}$ (or $\mathbf{R}_{i-}$) is the coordinate of the ith anion or cation. The choice of signs in Eqs. (1)–(3) depends on the polarity of the ion charge.

We can postulate that the interaction potential of the liquid particles U_{ii} includes the Coulomb interaction $u_q \propto R^{-1}$ and the short-range component $U_s(R)$:

$$U_{ii}(\mathbf{R}^{\{N\}}) = U_s(\mathbf{R}^{\{N\}}) + \frac{1}{2}\sum_{i\neq j} \pm u_q(\mathbf{R}_{i\pm} - \mathbf{R}_{j\pm}). \quad (4)$$

The short-range component $U_s(R)$ can then be approximated by the hard-sphere potential:

$$U_s(R \geq \sigma) = 0, \quad U_s(R \leq \sigma) = 0, \quad (5)$$

where σ is the hard-sphere diameter for the ions. To simplify the calculations, we shall assume that the diameters of the hard spheres are the same for all the ions. The electron–ion interaction is more complex and is not generally local at short distances. However, to simplify the estimates we shall use the pseudopotential approximation which may include the two types of interaction noted above. We shall separate the potentials for the electron–liquid particle interaction as in (4). The potential of the electron–negative ion interaction $u_{e-}(r)$ includes the Coulomb repulsion $u_q(r)$ and the hard-sphere potential $u_{es}(r < d_-)$ which allows for the effect of excluded volume

$$u_{e-}(r) = u_q(r) + u_{es}(r), \quad (6)$$

whereas the electron–cation potential $u_{e+}(r)$ is a purely Coulomb potential at large distances and is constant at distances shorter than some characteristic value d_+:

$$u_{e+}(r \geq d_+) = -u_q(r), \quad u_{e+}(r < d_+) = u_{e+}(d_+). \quad (7)$$

This last effect simulates the influence of the polarization of the cation nucleus and is frequently used in numerical calculations [5–7]. In general, the parameter d_+ can also be determined from quantum-chemical calculations, for example, using the pseudopotential method [21].

In order to find the grand partition function Ξ, we need to calculate the complex multidimensional integrals in (1) and then sum the series over N. Various statistical methods can be used for this purpose. Note that the ions create a complex potential field for the excess electrons and in principle, the KCl melt can be considered to be an ensemble of classical charged particles of a particular size in some external field. Two excess electrons form the source of this external field. The long- and short-range components of this external field are determined by the following functionals:

$$U_{e2}(\mathbf{R}) = \frac{1}{2\beta}\int_0^{\beta} d\tau [u_q(\mathbf{R} - \mathbf{r}_1(\tau)) - u_{e+}(\mathbf{R} - \mathbf{r}_1(\tau)) + u_q(\mathbf{R} - \mathbf{r}_2(\tau)) - u_{e+}(\mathbf{R} - \mathbf{r}_2(\tau))]. \quad (8)$$

$$U_{s2}(\mathbf{R}) = \frac{1}{\beta} \times \int_0^{\beta} [u_{es}(\mathbf{R} - \mathbf{r}_1(\tau)) + u_{es}(\mathbf{R} - \mathbf{r}_2(\tau))] d\tau. \quad (9)$$

Thus, the problem can be reduced to estimating the grand partition function for a liquid electrolyte in a certain potential field which includes long- and short-range components. Having obtained this estimate, we can make self-consistent calculations of the electron density distribution for the excess electrons which induce this field.

3. EFFECTIVE FUNCTIONAL FOR A BIPOLARON

The grand partition function for classical ions in an external field can be estimated by transforming this sum into a continuous field integral. The method of transforming the partition function into a field integral is used in plasma theory [22] and to investigate electrolytes [23, 24]. As a result of this transformation (see Appendix 1) we obtain a relationship for the grand partition function as a continuous field integral Ψ:

$$\Xi = \Xi_0 \int D[\Psi] \int D[\mathbf{r}_1(\tau)] \int D[\mathbf{r}_2(\tau)] \exp[-\beta\Omega]. \quad (10)$$

Here $\Omega(\{\Psi, \mathbf{r}_1(\tau), \mathbf{r}_2(\tau)\})$ is the thermodynamic potential for a bipolaron:

$$\Omega(\{\Psi, \mathbf{r}_1(\tau), \mathbf{r}_2(\tau)\}) = T_1 + T_2 + \frac{1}{|\mathbf{r}_1(\tau) - \mathbf{r}_2(\tau)|} + (\Psi - U_{e2}) * \frac{u_q^{-1}}{2} * (\Psi - U_{e2}) - \beta^{-1} A(\{\Psi, \mathbf{r}_1(\tau), \mathbf{r}_2(\tau)\}). \quad (11)$$

The symbol * denotes convolution integration:

$$y * x \equiv \int x(\mathbf{R}) y(\mathbf{R} - \mathbf{r}) d\mathbf{r}.$$

The last term in expression (11) reflects the changes in the ion distribution caused by the excess electrons:

$$A(\Psi, \mathbf{r}_1(\tau), \mathbf{r}_2(\tau)) = f_2 * \rho + \frac{1}{2!} f_2 * \rho^2 h_s * f_2, \quad (12)$$

where $f_2(\mathbf{r}_1, \mathbf{r}_2)$ is the generalized Mayer function for a bipolaron:

$$f_2 = \frac{1}{2}(\exp[\beta\Psi] + \exp[-\beta\Psi - \beta U_{s2}] - 2). \quad (13)$$

Here ρ is the average density of the melt, and $h_s(r)$ is the complete correlation function for hard spheres. In general, Eq. (12) also contains third- and higher-order irreducible correlation functions of hard spheres (see Appendix 1) although these correlation functions can be neglected if this system is not close to the phase transition point.

An advantage of Eq. (11) is that terms associated with short- and long-range interactions are explicitly isolated in it. We shall show below that in some cases, this separation can be used to obtain analytic estimates for the bipolaron free energy.

The field integral can be estimated using the saddle-point method which determines the average field $\tilde{\Psi}(R)$:

$$\partial\Omega(\Psi = \tilde{\Psi})/\partial\Psi = 0.$$

In general, the average field $\tilde{\Psi}$ can be related to the binary bipolaron–cation and bipolaron–anion correlation functions $g_{b\pm}(\mathbf{r})$ using the Poisson–Boltzmann equation. The derivation of these relationships is given in Appendix 1. Note that in our system there is a small parameter $(r_e\kappa)^{-1}$ where r_e is the average radius of the electron density distribution and $\kappa = (4\pi\rho\beta)^{1/2}$ is the reciprocal Debye length. For a bipolaron $r_e \simeq 4$–5 Å and under normal conditions in a KCl melt when $T \simeq 1000$ K and $\rho \simeq 3 \times 10^{-2}$ Å^{-3} we obtain the estimate $(\kappa r_e)^{-1} \simeq 0.06$. Consequently, in the zeroth approximation we can expand the thermodynamic functional (11) in terms of the small parameter $\Omega(\{\tilde{\Psi}, \mathbf{r}_1(\tau), \mathbf{r}_2(\tau)\}) = \Omega_0 + \Omega_1 + \ldots$ and confine ourselves to the zeroth term of this series $\Omega_0(\{\mathbf{r}_1, \mathbf{r}_2\})$ which depends on the electron coordinates:

$$\Omega_0 = T_1 + T_2 + \frac{1}{|\mathbf{r}_1(\tau) - \mathbf{r}_2(\tau)|} + \Omega_s - \frac{1}{8\pi} \int [\nabla U_{e2}(R, \mathbf{r}_1, \mathbf{r}_2)]^2 d\mathbf{R}, \quad (14)$$

where Ω_s is the component associated with the short-range repulsion:

$$\Omega_s = -\rho\beta^{-1}(f_{s2} + \frac{\rho}{2} f_{s2} * h_s * f_{s2}), \quad (15)$$

$$f_{s2}(R) = \exp[-\beta U_{s2}(R)] - 1. \quad (16)$$

The derivation of Eqs. (14) and (15) and the estimate $\Omega_1(\{\mathbf{r}_1(\tau), \mathbf{r}_2(\tau)\})$ are given in Appendix 2.

In order to determine the bipolaron free energy, we need to calculate the continuous integral of the functional $\Omega_0(\{\mathbf{r}_1, \mathbf{r}_2\})$ which only depends on the electron coordinates. Estimates of this functional may be obtained in terms of the two-particle Green's function which is related to the bipolaron wave functions:

$$G(\mathbf{r}_1, \mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_2) = \sum \phi_i(\mathbf{r}_1, \mathbf{r}_2)\phi_i(\mathbf{r}_1, \mathbf{r}_2) \exp[-\beta E_i], \quad (17)$$

where E_i and ϕ_i are the total energies and wave functions for the ith bipolaron state. In general, these functions also depend on the spin coordinate. However, we shall confine ourselves to the case when the electron spins are opposed and we shall subsequently only take into account the dependence of these wave functions on the spatial coordinates. Optimizing the effective functional yields the nonlinear Schrödinger equation for $\phi_i(\mathbf{r}_1, \mathbf{r}_2)$. If the bipolaron ground state is not degenerate, i.e., $\beta|E_0 - E_i| \gg 1$, we can only consider the contribution associated with this ground state. The estimate of the continuous integral is then reduced to averaging over the electron density distribution of the ground state $\phi_0^2(\mathbf{r}_1, \mathbf{r}_2)$. As a result of this procedure, the potentials which depend on the electron path are replaced in the continuous integral by the average potentials, i.e.,

$U_{e2} \longrightarrow \langle U_{e2}\rangle$, $U_{s2} \longrightarrow \langle U_{s2}\rangle$. Finally, we obtain the effective thermodynamic potential for the bipolaron:

$$\Omega_{\text{eff}} = T_1 + T_2 + \iint \frac{n_2(\mathbf{r}_1, \mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2}{|r_1 - r_2|} + \langle U_{e2}\rangle * \frac{u_q^{-1}}{2} * \langle U_{e2}\rangle + \Omega_s(\langle U_{s2}\rangle), \tag{18}$$

where we used the following notation:

$$T_i = -\frac{1}{2}\int d\mathbf{r}_1 d\mathbf{r}_2 \phi_0(\mathbf{r}_1, \mathbf{r}_2)\nabla_i^2 \phi_0(\mathbf{r}_1, \mathbf{r}_2), \tag{19}$$

$$\langle U_{e2}(\mathbf{R})\rangle = \frac{1}{2} \times \sum_i \int [u_q(\mathbf{R} - \mathbf{r}_i) - u_{e+}(\mathbf{R} - \mathbf{r}_i)]n(\mathbf{r}_i)d\mathbf{r}_i, \tag{20}$$

$$\langle U_{s2}(\mathbf{R})\rangle = \sum_i \int u_{es}(\mathbf{R} - \mathbf{r}_i)n(\mathbf{r}_i)d\mathbf{r}_i. \tag{21}$$

Here $n(\mathbf{r})$ and $n_2(\mathbf{r}_1, \mathbf{r}_2)$ are the single- and two-particle electron density distribution functions and these are related to the bipolaron wave function by:

$$n(\mathbf{r}_i) = \int d\mathbf{r}\phi_0^2(\mathbf{r}, \mathbf{r}_i), \quad n_2(\mathbf{r}_1, \mathbf{r}_2) = \phi_0^2(\mathbf{r}_1, \mathbf{r}_2). \tag{22}$$

The contribution Ω_s associated with short-range repulsion leads to the formation of a region of reduced anion density similar to that accompanying the formation of an F center in alkali halide crystals. The simplest approximation of this contribution may be written as [18]

$$\Omega_s \approx \frac{4\pi\rho}{3\beta} r_e^3(n). \tag{23}$$

Assuming that the average scale of variation of the short-range component of the electron–cation potential is much smaller than the characteristic dimension of the electron density distribution, i.e., $d \ll r_e$, we can simplify expression (18) and, converting from the thermodynamic potential to the free energy, we obtain the final expression for the bipolaron free energy:

$$F_{\text{eff}}(n_1, n_2) \approx T_1 + T_2 + \iint \frac{n_2(\mathbf{r}_1, \mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2}{|\mathbf{r}_1 - \mathbf{r}_2|} - \frac{1}{2}\sum_{ij}\iint \frac{n(\mathbf{r}_i)n(\mathbf{r}_j)d\mathbf{r}_j d\mathbf{r}_i}{|\mathbf{r}_i - \mathbf{r}_j|} + a_+^2 \int n^2(\mathbf{r})d\mathbf{r} + \frac{4\pi\rho}{3\beta} r_e^3, \tag{24}$$

where $a_+^2 = \int [u_q(r) + u_{e+}(r)]d\mathbf{r}/2$ is the square of the characteristic dimension of the cation nucleus. Thus, the effective functional of the total bipolaron energy depends on the single- and two-particle electron density distributions and only on two parameters associated with the structural characteristic of the cation (a_+^2) and the thermodynamic state of the medium ($\rho\beta^{-1}$).

The expression (24) obtained by us for the total bipolaron energy in a KCl melt is similar to the known expressions. For instance, the first three terms on the right-hand side of (24) correspond to the functional for a bipolaron obtained by Pekar for the static permittivity $\epsilon_0 = \infty$ (strong screening limit $\kappa^{-1} \longrightarrow 0$) and the high-frequency permittivity $\epsilon_\infty = 1$ (classical medium). The fourth term on the right-hand side of (24) is associated with the short-range repulsion and was analyzed using continuous bipolaron models in [13, 14]. Contributions similar to the last term in (24) are associated with the formation of a cavity and were obtained in semicontinuous theories of bipolarons in a polar liquid [16].

4. VARIATIONAL ESTIMATES OF THE BIPOLARON FUNCTIONAL

Further study of the bipolaron involves minimizing the effective functional (24) or solving the two-particle Schrödinger equation. The prospects for proceeding successfully along this path are determined by the choice of approximation for the two-particle wave function $\phi_0(\mathbf{r}_1, \mathbf{r}_2)$.

We shall begin our study with the simplest approximation, based on assuming that the two electrons are uncorrelated when the bipolaron wave function is approximated as a product of the single-particle wave functions:

$$\phi_0(\mathbf{r}_1, \mathbf{r}_2) = \phi_0(\alpha\mathbf{r}_1)\phi_0(\alpha\mathbf{r}_2), \tag{25}$$

where $\alpha \propto r_e^{-1}$ is the variational parameter. The trial wave functions can be taken to be Gaussian or Coulomb wave functions:

$$\phi_G(r) \propto \exp[-\alpha^2 r^2],$$

$$\phi_C(r) \propto (1 + \alpha r)\exp[-\alpha r]$$

and the parameter α can then be sought by varying the free-energy functional. As a result of minimizing $F_{\text{eff}}(\alpha)$ we obtain the nonlinear algebraic equation for α:

$$\alpha = C_0 - C_2 a_+^2 \alpha^2 + C_s\rho/\beta\alpha^4, \tag{26}$$

where C_0, C_2, and C_s are constants determined by the choice of trial wave function. In this particular approximation the average electron density is defined as $n(\mathbf{r}) = 2\phi^2(\alpha\mathbf{r})$. Ultimately we obtain the relationship for the total bipolaron energy:

$$F[n(\mathbf{r})] \approx T + \langle U_{\text{eff}}\rangle + \iint \frac{\phi_0^2(\mathbf{r})\phi_0^2(\mathbf{r}')d\mathbf{r}d\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|}, \tag{27}$$

where we used the notation $\langle U_{\text{eff}}\rangle$ for the average potential energy of the electron–ion interaction:

$$\langle U_{\text{eff}}\rangle = -\frac{1}{2}\int\int\frac{n(\mathbf{r})n(\mathbf{r}')(d\mathbf{r})d\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|} - a_+^2\int n^2(\mathbf{r})d\mathbf{r} + \frac{4\pi\rho}{3\beta}r_e^3(n), \tag{28}$$

whereas

$$T = -\int d\mathbf{r}\phi(\mathbf{r})\nabla^2\phi(\mathbf{r})$$

is the bipolaron kinetic energy.

Another method of allowing for many-electron effects is to use the local density functional theory [19, 20] for two spin-paired electrons. In this approach the multiparticle electron problem is reduced to calculation of the single-particle wave functions of the electrons and the electron–electron interaction is approximated by a functional which depends on the average electron density $n(\mathbf{r}) = 2\phi^2(\mathbf{r})$. Following this theory we write the functional of the total system energy in the form [20]

$$W(n(\mathbf{r}), \mathbf{R}^{\{N\}}) = U_{ii}(\mathbf{R}^{\{N\}}) + \sum_i^N\int u_{e\pm}(\mathbf{r}-\mathbf{R}_{i\pm})n(\mathbf{r})d\mathbf{r} + W_e[n(\mathbf{r})],$$
$$W_e[n(\mathbf{r})] = T_1 + T_2 + \frac{1}{2}\int\int d\mathbf{r}d\mathbf{r}'\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{xc}[n(\mathbf{r})], \tag{29}$$

where

$$T_i = -\frac{1}{2}\int d\mathbf{r}\phi_i(\mathbf{r})\nabla^2\phi_i(\mathbf{r})$$

is the kinetic energy of the ith electron. The last term in (29) is the exchange correlation energy E_{xc} which includes the exchange (ϵ_x) and correlation (ϵ_c) components per particle in a homogeneous electron gas:

$$E_{xc}[n(\mathbf{r})] = \int d\mathbf{r}n(\mathbf{r})[\epsilon_x(\mathbf{r}) + \epsilon_c(\mathbf{r})]. \tag{30}$$

Following [25], we shall approximate these by

$$\epsilon_x(\mathbf{r}) = -\frac{C_p}{r_s(\mathbf{r})},$$
$$\epsilon_c(\mathbf{r}) = \frac{\gamma_p}{1+\beta_1 r_s(\mathbf{r})^{1/2} + \beta_2 r_s(\mathbf{r})}, \tag{31}$$

where $r_s(\mathbf{r})$ is the average radius of the electron density distribution: $r_s^{-3}(\mathbf{r}) = 4\pi n(r)/3$.

Performing transformations for the partition function and taking into account all the reasoning put forward above, we finally obtain an expression for the functional of the total bipolaron energy in terms of the average electron density:

$$F_{LDA}(n) = T + \langle U_{\text{eff}}\rangle + \frac{1}{2}\int\int\frac{n(\mathbf{r})n(\mathbf{r}')d\mathbf{r}d\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|} + E_{xc}[n(\mathbf{r})], \tag{32}$$

where $\langle U_{\text{eff}}(n(r))\rangle$ is the average potential energy of the electron–ion interaction defined by Eq. (28). The variation of this functional in terms of $\phi_0(r)$ gives the equations for the single-particle wave functions. We can then proceed as above, specifically taking $\phi_0(\alpha r)$ in a certain form and seeking the parameter α by varying (32).

5. RESULTS

A qualitative analysis of (24) shows that the bipolaron functional includes various factors. Coulomb interaction associated with the polaron effect and the Hartree potential mainly determines the bipolaron energy. The contribution proportional to $4\pi\rho r_e^3(n)/\beta$ is associated with the formation of a cavity; it leads to a reduction in the average bipolaron radius and helps to increase the average bipolaron binding energy $\Delta E = E - 2E_1$ (where E_1 is the total energy of a single electron solvated in KCl) as a result of entropy effects. The contribution proportional to $a_+^2\int n^2(\mathbf{r})d\mathbf{r}$ leads to opposite effects. Competition between these two contributions determines the stability conditions for the bipolaron. Allowance for electron–electron correlations using the local density functional method also leads to an increase in the entropy factor as a result of effects associated with the formation of an exchange correlation hole [20] and ultimately helps to increase ΔE.

We performed variational calculations for these types of trial functions and determined the bipolaron energy characteristics and its average radius in a KCl melt. For the calculations we used the values $T = 1000$ K, $\rho = 2.4\times10^{-2}$ Å^{-3} and $a_+ = 3.5$ au. Two types of calculations were made: using the uncorrelated electron approximation when the bipolaron wave function was approximated as a product of single-particle wave functions (25) and using the local density functional approximation (29). The results of these calculations are given in Table 1. For comparison this table also gives similar variational estimates for a single electron.

It can be seen that the variational calculations show good agreement with the numerical estimates and broadly give a correct estimate of the bipolaron energy and structural characteristics. The characteristic size of a bipolaron in a KCl melt is around 3.5 Å and is slightly greater than the characteristic size of a single electron solvated in KCl. The formation of a bipolaron state from spin-paired electrons is similar to the formation of an F center. The total bipolaron energy is approximately –(2.2–2.7) eV. However, choosing a multiplicative trial wave function (25) has the result that in the

Energy and structural characteristics of a bipolaron and polaron in a KCl melt

	Polaron		Bipolaron		
			uncorrelated electron approximation		local density approximation
Trial function	$\phi_G(r)$	$\phi_C(r)$	$\phi_G(r)$	$\phi_C(r)$	$\phi_G(r)$
r_e, Å	3.03	3.02	3.61	3.62	3.55
T, eV	0.093	1.035	1.308	1.443	1.358
$-\langle U_{\text{eff}}\rangle$, eV	1.961	1.979	7.003	7.198	7.126
E_{ee}, eV			3.886	4.098	3.958
E_x, eV					7.918
$-E_{\text{xc}}$, eV					4.585
$-F$, eV	1.032	0.094	1.810	1.657	2.434

uncorrelated electron approximation the bipolaron energy is lower than the energy of two unbound polarons, i.e., the bipolaron state is unstable and decays into two isolated polarons. This sharply contradicts the data obtained by quantum molecular dynamics [6] according to which the singlet bipolaron state is stable. However, variational calculations using the local density method based on Kohn–Sham theory give a result similar to the variational calculations [6]. Compared with the uncorrelated electron approximation the characteristic radius and the total bipolaron energy are reduced. The bipolaron binding energy ΔE is around 0.4 eV which also agrees with the data obtained by the quantum molecular dynamics method.

We can therefore conclude that in order to obtain a correct estimate of the bipolaron stability and its range of existence, we need to make a correct choice of trial wave function which explicitly allows for the correlation of the two electrons. These numerical calculations indicate the characteristic feature already noted on several occasions [16, 9] that bipolaron calculations require accurate allowance for electron–electron interactions and that the stability of a bipolaron is extremely sensitive to these interactions.

ACKNOWLEDGMENTS

To conclude, the author would like to thank N.L. Leonova for refinements and checking the numerical calculations.

APPENDIX 1

Method of Collective Variables

The electric charge density $\rho_q(\mathbf{R})$ at an arbitrary point $\mathbf{R}$ may be expressed as a sum of the densities of the positive and negative charges $\rho_\pm(\mathbf{R})$:

$$\rho_\pm(\mathbf{R}) = \sum_i^{N/2} \pm\delta(\mathbf{R} - R_{i\pm}), \quad \rho_q(\mathbf{R}) = \rho_+(\mathbf{R}) + \rho_-(\mathbf{R}).$$

Essentially $\rho_q(\mathbf{R})$ is a collective variable. Its Fourier transform has the meaning of the mode of a fluctuation wave having the wave vector $\mathbf{k}$ for the charge. The long-range interaction between the particles can be expressed in terms of these modes. Finally the partition function may be expressed in the functional form

$$\Xi \propto \int D[\mathbf{r}_1(\tau)] \int D[\mathbf{r}_2(\tau)] \int D[\rho_q] J(R\rho_q) \times \exp[-W(r, \rho_q)],$$

$$W = \beta[T_1 + T_2 + U_{e2} * \rho_q + \rho_q * \frac{u_q}{2} * \rho_q + U_s(\rho_q) + U_{s2} * \rho_-],$$

where $J(R\rho_q)$ is the Jacobian of the transition from $\mathbf{R}^{\{N\}}$ variables to ρ_q collective variables. Note that in this last relationship the long-range contribution to W which is proportional to $U_{e2} * \rho_q + \rho_q * u_q * \rho_q/2$ depends quadratically on ρ_q. This can be expressed in terms of Gaussian functional integrals if an inverse operator exists for the potential $u_q(R)$. For Coulomb interaction such an inverse operator exists $u_q^{-1}(R) = 1/4\Delta(\mathbf{r})$ so that the exponential function can be Fourier transformed from the quadratic form [26]:

$$\exp\left[\frac{1}{2}\rho_q * u_q * \rho_q\right] = \left\{\int D[\Psi] \exp\left[-\frac{1}{2}\Psi * u_q^{-1} * \Psi\right]\right\}^{-1} \times \int D[\Psi] \exp\left[-\frac{1}{2}\Psi * u_q^{-1} * \Psi + \rho_q * \Psi\right].$$

Finally we transform the grand partition function Ξ into the continuous field integral Ψ:

$$\Xi = \Xi_0 \int D[\Psi] \exp[-\beta\Omega],$$

$$\Omega = T_1 + T_2 + (\Psi - U_{er}) * \frac{u_q^{-1}}{2} * (\Psi - U_{er}) I(\Psi, U_{er}, U_s),$$

$$I(\Psi, U_{es}, U_s) = \sum_{N \geq 0}^{\infty} d\mathbf{R}^{\{N\}} \frac{z^N}{N!} \times \prod_i^N \exp[\beta(\pm\Psi(R_{i\pm}) - U_s(\mathbf{R}^{\{N\}}) - U_{es}(R_{i-})].$$

We introduce the n-particle correlation functions $\rho_s^{(n)}(\mathbf{r}_1, \ldots, \mathbf{r}_n)$ for hard spheres:

$$\rho_s^{(n)}(\mathbf{r}_1, \ldots, \mathbf{r}_2) = \Xi^{-1} \sum_N^{\infty} \frac{z^N}{(N-n)!} \times \int \exp[-\beta U_S] d\mathbf{R}^{\{N-n\}}.$$

Using the expression for the Mayer function in this last relationship, we perform a Mayer transformation for the configurational component $I(\Psi, U_{es}, U_s)$:

$$I(\Psi, U_{es}, U_s) = 1 + f_2 * \rho_s^{(1)} + \frac{1}{2!} f_2 * \rho_s^{(2)} * f_2 \ldots + \frac{1}{n!} f_2 * \rho_s^{(n)} * \ldots f_2.$$

We shall assume that $\rho_s^{(1)} \equiv \rho$ is the average density of the melt, and $\rho_s^{(2)} = \rho^2 + \rho^2 h_s(r)$ is the second-order correlation function for hard spheres which is related to $h_s(r)$, the total correlation function for hard spheres which is determined by a standard method [27]. Then, neglecting all the irreducible correlation functions of the third order and above, we obtain

$$I(\Psi, U_{es}, U_s) = 1 + \sum_{k=1}^{\infty} \frac{(f_2 * \rho)^k}{k!} + \sum_{k=2}^{\infty} \frac{1}{k!} \left(\frac{1}{2!} f_2 * \rho^2 h_s * f_2 \right)^k.$$

Transforming this relation into exponential form, we finally obtain Eq. (11).

In general, the potentials $U_{ii}(r)$ and $u(r)$ are separated into short- and long-range components by different methods. For a representation in terms of collective variables it is sufficient for the potential $u_q(r)$) at short distances to be a fairly smooth function (belonging to the L_2 class of functions).

APPENDIX 2

Screened Potentials

In order to relate the field $\tilde{\Psi}$ and the electron-induced external field, we shall analyze the bipolaron–anion and bipolaron–cation correlation functions $g_{b\pm}(r)$ which describe the probability of finding the corresponding ion at the distance $\mathbf{r}$ from the localization center:

$$g_{b+}(r) = \frac{\delta\Omega_{\text{eff}}}{\delta\langle u_{e+}\rangle} = (1 + \rho h_s * f_2)\exp[\beta\tilde{\Psi}],$$

$$g_{b-}(r) = \frac{\delta\Omega_{\text{eff}}}{\delta\langle u_{e-}\rangle} = (1 + \rho h_s * f_2)\exp[-\beta\tilde{\Psi} - \beta\langle U_{s2}\rangle].$$

Using this function, we obtain the Poisson–Boltzmann equation for the average field $\tilde{\Psi}$:

$$u_q^{-1} * (\tilde{\Psi} - U_{e2}) = \frac{\rho}{2}(g_{b+} - g_{b-}).$$

This equation may be written in the integral form:

$$\tilde{\Psi} = U_{e2} - \rho\frac{u_q}{2}(g_{b+} - g_{b-}).$$

The field $\tilde{\Psi}$ is sometimes called the screened potential [27] since it determines the electrostatic field in the system and is associated with charge screening of the external field. If we confine ourselves to quadratic terms with respect to the field Ψ in the thermodynamic potential (11) (random phase approximation), the continuous field integral will be Gaussian:

$$\Omega = \Omega_0 - A * \Psi + \frac{1}{2}\Psi * B * \Psi,$$

where Ω_0 is determined by formula (14) and the operators A and B are defined as follows:

$$A = u_q^{-1} * U_{e2} + \frac{\rho}{2}(1 + ph_s * f_{s2}) f_{s2},$$

$$B = u_q^{-1} + \frac{\rho\beta}{2}(1 + ph_s * f_{s2})[1 + \exp(-\beta U_{s2})].$$

The continuous field integral has a Gaussian form and by calculating it for the Fourier transform of the field $\tilde{\Psi}(k)$, we obtain $\tilde{\Psi}(k) = A(k)/B(k)$. Finally we obtain a definitive expression for the thermodynamic potential:

$$\Omega = \Omega_0 - \frac{1}{2}A * \tilde{\Psi} = \Omega_0 + \Omega_1.$$

It can be shown that $\Omega_1 \propto \alpha^2\kappa^{-2}$. In general, we can also obtain corrections which include the terms $\propto\Psi^3$, and so on.

REFERENCES

1. A. M. Stoneham, *Theory of Defects in Solids: the Electronic Structure of Defects in Insulators and Semiconductors* (Clarendon Press, Oxford, 1975; Mir, Moscow, 1979).

2. N. R. Kestner, in *Electron–Solvent and Anion-Solvent Interactions*, Ed. by L. Kevan and B. C. Webster (Elsevier, Amsterdam, 1976).
3. A. S. Aleksandrov and N. F. Mott, Rep. Prog. Phys. **57**, 1289 (1994).
4. J. C. Scott, P. Pfluger, M. T. Krounby, and G. B. Street, Phys. Rev. B **28**, 2140 (1983).
5. A. Selloni, M. Parrinello, R. Car, and P. Carevali, J. Phys. Chem. **91**, 4947 (1987).
6. E. S. Fois, A. Selloni, M. Parrinello, and R. Car, J. Phys. Chem. **92**, 3268 (1988).
7. E. S. Fois, A. Selloni, and M. Parrinello, Phys. Rev. B **39**, 4812 (1989).
8. H. P. Kaukonen, R. N. Barnett, and U. Landman, J. Chem. Phys. **97**, 1365 (1992).
9. G. Martyna, Z. Deng, and M. L. Klein, J. Chem. Phys. **98**, 555 (1993).
10. Z. Deng, G. Martyna, and M. L. Klein, Phys. Rev. Lett. **71**, 267 (1993).
11. S. I. Pekar, *Investigation on Electronic Theory of Crystals* (Gostekhizdat, Moscow, 1951).
12. V. L. Vinetskiĭ and M. Sh. Giterman, Zh. Ésp. Teor. Fiz. **33**, 730 (1957) [Sov. Phys. JETP **6**, 560 (1957)].
13. H. Hiramoto and Y. Toyozawa, J. Phys. Soc. Jpn. **54**, 245 (1985).
14. D. Emin and M. S. Hilery, Phys. Rev. B **39**, 6575 (1989).
15. V. L. Vinetskiĭ, O. Meredov, and V. A. Yanchuk, Teor. Éksp. Khim. **25**, 631 (1989).
16. D. F. Feng, K. Feuki, and L. Kevan, J. Chem. Phys. **58**, 3281 (1973).
17. G. N. Chuev, Zh. Éksp. Teor. Fiz. **115**, 1463 (1999) [JETP **88**, 807 (1999)].
18. G. N. Chuev and V. V. Sychyov, J. Chem. Phys. **112**, 4707 (2000).
19. S. L. Lundquist and N. H. March, *Theory of the Inhomogeneous Electron Gas* (Plenum, New York, 1983; Mir, Moscow, 1987).
20. R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989).
21. M. Boulahbak *et al.*, J. Chem. Phys. **108**, 2111 (1998).
22. Stu Samuel, Phys. Rev. D **18**, 1916 (1978).
23. A. L. Kholodenko and A. L. Beyerlein, Phys. Rev. A **34**, 3309 (1986).
24. A. L. Kholodenko and C. Qian, Phys. Rev. B **40**, 2477 (1989).
25. I. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
26. M. V. Fedoryuk, *Saddle Point Approximation* (Nauka, Moscow, 1977).
27. I. R. Yukhnovskiĭ and M. F. Golovko, *Statistical Theory of Classical Equilibrium Systems* (Naukova Dumka, Kiev, 1980).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 5, 2000, pp. 991–999.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 5, 2000, pp. 1143–1152.

Structure and Dynamics of the $He_2^*(a^3\Sigma_u^+)$ Molecular Complex in Condensed Phases of Helium

S. G. Kafanov, A. Ya. Parshin[a], and I. A. Todoshchenko

Kapitza Institute for Physical Problems, Russian Academy of Sciences, Moscow, 117973 Russia
[a] e-mail: parshin@kapitza.ras.ru

Received July 4, 2000

Abstract—An investigation is made of the absorption spectra of triplet metastable helium molecules in the $a^3\Sigma_u^+$ state in liquid ^{4}He and ^{3}He at various pressures and in dense ^{3}He gas. An analysis of the spectrum corresponding to the $a^3\Sigma_u^+ \longrightarrow c^3\Sigma_g^+$ transition confirms the conclusion that there is a microscopic bubble surrounding the molecule in liquid helium. A simple approximation is proposed for the wave function of the valence electron of the molecule and the parameters of the bubble are determined for various experimental conditions. The coefficient of molecular recombination in liquid ^{3}He and ^{4}He was determined experimentally at various pressures and in dense cold ^{3}He gas. The results show good agreement with the theory of mutual recombination limited by molecular diffusion under conditions of strong van der Waals interaction. It is shown that in the condensed phases of helium the polarization of the molecules under the action of the magnetic field does not lead to suppression of their mutual recombination, and this is confirmed experimentally. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Numerous experimental and theoretical studies have been devoted to the neutral triplet excitations of helium. The lowest triplet atomic (2^3S) and molecular ($a^3\Sigma_u^+$) states are metastable having intrinsic lifetimes of approximately 8000 s [1, 2] and 15 s [3, 4] and energies of 19.82 and 17.86 eV, respectively. When helium is excited by fast particles, an appreciable fraction of the energy is dissipated in the formation of triplet atoms and molecules. As the helium density increases, the ratio of the steady-state molecular concentration to the concentration of excited atoms increases [5] which can be attributed to an increase in the probability of three-body collisions when a triplet atom may capture an unexcited atom and form a dimer [6]. In dense helium gas ($n \geq 3 \times 10^{20}$ cm^{-3}) [5] and in liquid helium [7], triplet molecules in the $a^3\Sigma_u^+$ state are the predominant type of neutral excitations.

The main mechanism for the loss of triplet molecules in condensed helium [5, 7, 8] and triplet atoms in the low-density gas [9] is their mutual recombination which takes place via a Penning ionization scheme

$$\begin{aligned} He_2^*(a^3\Sigma_u^+) + He_2^*(a^3\Sigma_u^+) &\longrightarrow 2He(1^1S) + He_2^+ + \bar{e}, \\ He^*(2^3S) + He^*(2^3S) &\longrightarrow He(1^1S) + He^+ + \bar{e}. \end{aligned} \quad (1)$$

The characteristic lifetime of the molecules decreases as their concentration increases $\tau = 1/(\alpha n)$ (α is the mutual recombination coefficient) and is a few milliseconds at concentrations of approximately 10^{13} cm^{-3}.

In condensed helium the excimer decay process is limited by diffusion:

$$\alpha = 4\pi D R_I, \quad (2)$$

where R_I is the characteristic distance between the molecules for which the ionization takes place with a probability of the order of unity. This distance is determined from the condition that the characteristic diffusion time R_I^2/D and the characteristic time of reaction (1) are equal.

Detailed calculations of the interaction between a triplet helium atom and a surrounding liquid were made in [10]. The calculated shift of the absorption line for the $2^3S \longrightarrow 2^3P$ transition relative to the position at low pressure showed good agreement with the experiment [7]. Similar calculations have not yet been made for triplet molecules.

In the present study we give the molecular absorption spectra corresponding to the $a^3\Sigma_u^+ \longrightarrow c^3\Sigma_g^+$ and $a^3\Sigma_u^+ \longrightarrow b^3\Pi_g$ transitions measured in liquid ^{3}He and ^{4}He at various pressures and in dense cold ^{3}He gas. The integrated intensities of the spectra were used to determine the ratio of the oscillator strengths of the $a^3\Sigma_u^+ \longrightarrow c^3\Sigma_g^+$ and $a^3\Sigma_u^+ \longrightarrow b^3\Pi_g$ transitions whose value

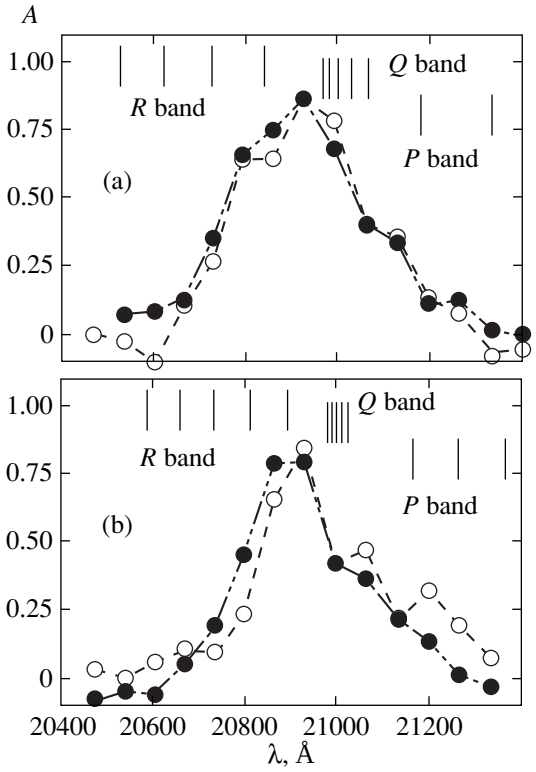

Fig. 1. Absorption spectra of molecules corresponding to the $a^3\Sigma_u^+ \longrightarrow b^3\Pi_g$ transition: (a) in liquid ^{4}He (○—1.0 atm, 2.1 K; ●—23.9 atm, 1.9 K); (b) in liquid ^{3}He (○—1.0 atm, 1.8 K; ●—14.4 atm, 1.8 K). The spectra are normalized to the absorption at the maximum.

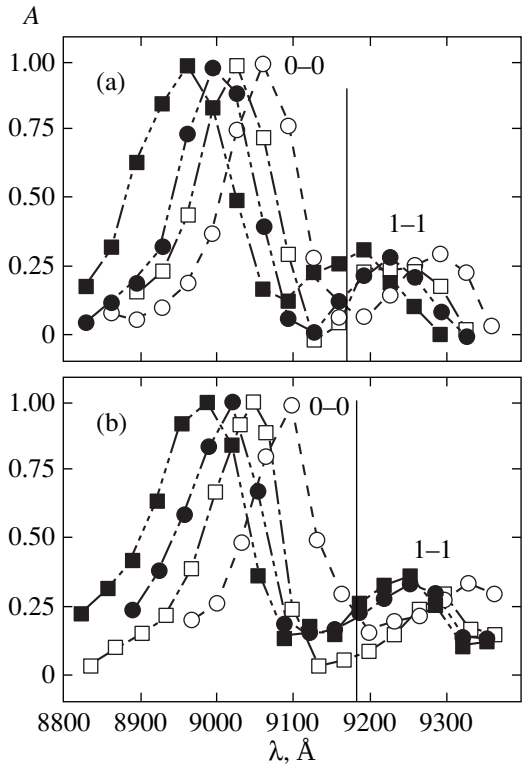

Fig. 2. Absorption spectra of molecules corresponding to the $a^3\Sigma_u^+ \longrightarrow c^3\Sigma_g^+$ transition: (a) in liquid ^{4}He (○—0.05 atm, 2.1 K; □—6.3 atm, 2.1 K; ●—10.1 atm, 2.0 K; ■—23.9 atm, 1.9 K; (b) in liquid ^{3}He (○—4.8 atm, 1.8 K; □—14.4 atm, 1.8 K; ●—23.9 atm, 1.7 K; ■—31.5 atm, 1.7 K). The vertical line gives the wavelength corresponding to the (0–0) $a^3\Sigma_u^+ \longrightarrow c^3\Sigma_g^+$ transition in vacuum. The spectra are normalized to the absorption at the maximum.

shows good agreement with the calculated value. Quantitative data were obtained on the van der Waals coefficients of the following pair interactions:

$$\mathrm{He}_2^*(a^3\Sigma_u^+) - \mathrm{He}_2^*(a^3\Sigma_u^+),$$

$$\mathrm{He}_2^*(a^3\Sigma_u^+) - \mathrm{He}(1^1S),$$

$$\mathrm{He}_2^*(c^3\Sigma_g^+) - \mathrm{He}(1^1S).$$

We calculated the interaction between a molecule and surrounding helium which leads to a shift and broadening of the absorption lines, using a model of the "bubble" formed by the molecule in liquid helium and we determined the bubble radius under various conditions. The molecular spectra in the gas were described using the standard theory of line broadening in the binary approximation [10, 11].

The theory of diffusion-limited mutual recombination was extended to the case of strong van der Waals interaction. Good agreement was observed between the calculated data and the experimental data obtained in the present study and in [7]. The coefficient of recombination calculated using this model does not depend on the magnetic field and this is confirmed experimentally.

2. ABSORPTION SPECTRA AND THEIR INTERPRETATION

The method of generating molecules and measuring the absorption is similar to that described in [12]. The molecules are formed as a result of the recombination of positive ions and electrons injected into the helium from tungsten tips. Light from a halogen lamp passing through a mechanical chopper and a monochromator is fed along a quartz light guide into an experimental cell from which it is extracted to a photodetector using another light guide. The photodetector signal is amplified and demodulated using a lock-in amplifier. The excimer concentration was modulated at low frequency by periodically varying the current through the cell and the signal from the lock-in amplifier was demodulated using a computer.

Typical absorption spectra observed in liquid ^{3}He and ^{4}He are shown in Figs. 1 and 2. To within experi-

mental accuracy, the $a^3\Sigma_u^+ \longrightarrow b^3\Pi_g$ absorption line does not shift with varying pressure whereas the $a^3\Sigma_u^+ \longrightarrow c^3\Sigma_g^+$ line undergoes an appreciable displacement in the short-wavelength direction and becomes broader as the pressure and particle density increase. Line broadening of the $a^3\Sigma_u^+ \longrightarrow c^3\Sigma_g^+$ transition was observed as a function of temperature (Fig. 3). It is important to note that in the given temperature range at constant pressure, the variation of the helium density is within 1.5% so that we can reliably talk of temperature-induced broadening of the line. Unlike the absorption line width, its shift relative to the vacuum position does not depend on temperature, which suggests that the position of the $a^3\Sigma_u^+ \longrightarrow c^3\Sigma_g^+$ line may be used as an indicator of the static interaction between the molecule and the environment, neglecting the temperature fluctuations which lead to additional broadening.

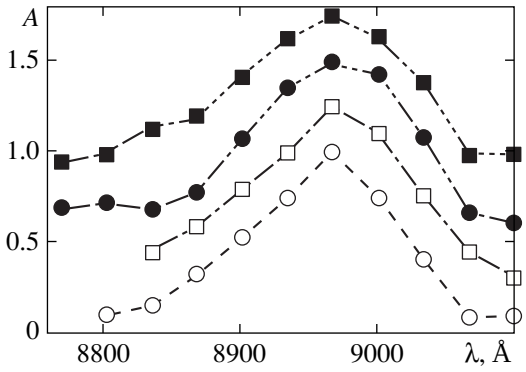

Fig. 3. Line broadening of the $a^3\Sigma_u^+ \longrightarrow c^3\Sigma_g^+$ transition in liquid ^{4}He as a function of temperature at 23.9 atm (○—1.55 K; □—1.75 K; ●—1.95 K; ■—2.1 K; T_λ = 1.88 K). The spectra are normalized to the maximum and are shifted along the ordinate by 0.25 relative to each other.

In order to describe the pressure dependence of the (0–0) $a^3\Sigma_u^+ \longrightarrow c^3\Sigma_g^+$ line shifts, we used a model which assumes that a microscopic bubble surrounds the molecule in liquid helium. This approach was used in [10] which was devoted to the $He_2^*(2^3S)$ metastable triplet helium atoms. It is assumed that the bubble is formed as a result of the repulsion of an excited electron from the surrounding helium atoms. This mechanism may lead to the formation of a bubble around the metastable molecule since the size of the outer electron orbit is comparable with the interatomic distance in liquid helium [13].

The equilibrium bubble radius R_0 is determined by minimizing the total energy of the complex $E(R)$ which consists of the total energy of the interaction between the molecule and the surrounding helium atoms E_{ma}, the potential energy of the cavity in the liquid pV, the potential energy at the bubble interface E_{sur}, and the kinetic energy of the molecule E_m which is associated with the oscillations of the molecule in the bubble.

We shall make an assumption which will be justified by the following calculations, that the size of the bubble is considerably greater than the internuclear distance in the molecule (around 1 Å [14]). To a first approximation the interaction of the molecule with the surroundings reduces to the repulsion of the outer electron from the helium atoms at short distances and van der Waals interaction of the molecule with the atoms. In the widely used optical model the energy of the interaction between an electron and a helium atom is written in the form

$$\varepsilon_{ea} = \frac{2\pi\hbar^2 a_0}{m_e}|\psi(\mathbf{R})|^2,$$

where a_0 is the scattering length of the electron on the helium atom in the pseudopotential approximation (a_0 = 0.62 Å) and $\psi(\mathbf{R})$ is the electron wave function [15]. Calculation of the electron wave function of a helium molecule is a complex theoretical problem. However, we are merely interested in its behavior at comparatively large distances from the nuclei. Calculations made for the $A^1(\Sigma_u^+)$ and $C^1(\Sigma_g^+)$ singlet molecular states by Guberman and Goddard [16] show that at distances greater than 5 Bohr radii the wave functions of the outer electron are accurately approximated by hydrogen-like functions of the $2S$ and $2P_0$ type, respectively with the effective charges of the molecular core $Z(A^1)$ = 1.08 and $Z(C^1)$ = 0.69. We shall assume that the wave functions of an excited electron in the $a^3\Sigma_u^+$ and $c^3\Sigma_g^+$ states are also essentially $2S$ and $2P_0$ hydrogen-like functions and the effective core charges will be fitting parameters.

We do not know of any experimental or theoretical data on the coefficient of the van der Waals interaction between a molecule and a helium atom. In order to estimate this we can use the following simple reasoning: the dipole moment of a molecule is mainly determined by the outer electron whose characteristic frequency of motion is low compared with the frequencies of electrons in the 1^1S ground state of the helium atom. Assuming that the motion of the outer electron relative to the molecular core is classical, at each instant the energy of the interaction between a molecule and a helium atom is $-\alpha\mathbf{E}^2(\mathbf{R})/2$ where $\mathbf{E}(\mathbf{R})$ is the electric field induced by the molecule and α is the polarizability of the helium atom in the ground state (α = 1.383 au [17]).

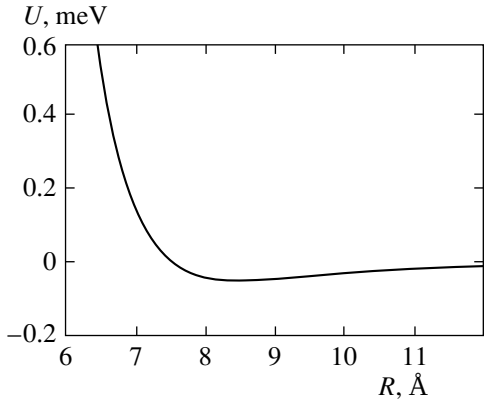

Fig. 4. Energy of interaction between an $He_2^*(a^3\Sigma_u^+)$ molecule and a 1^1S helium atom at large distances.

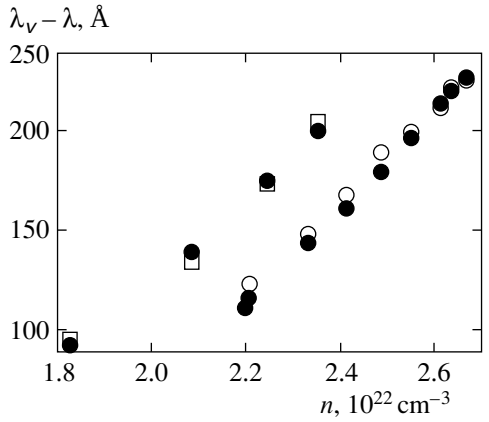

Fig. 5. Shift of the maximum of the (0–0) $a^3\Sigma_u^+ \longrightarrow c^3\Sigma_g^+$ absorption line relative to the vacuum position in liquid helium [□—^{3}He, ○—^{4}He, ●—fitting using bubble model (see text)].

Averaging over the state of the outer electron in the molecule, we can easily obtain

$$\varepsilon_{\mathrm{vdW}}(2S) = -\frac{\alpha\bar{e}^2\langle r^2\rangle_{2S}}{R^6},$$

$$\varepsilon_{\mathrm{vdW}}(2P_0) = -\frac{\alpha\bar{e}^2\langle r^2\rangle_{2P_0}}{R^6}\left(\frac{4}{5}+\frac{3}{5}\cos^2\Theta\right),$$

where the angular brackets denote averaging over the state. We note that the expression for $\varepsilon_{\mathrm{vdW}}(2P_0)$ is the same as the result of the approximate quantum-mechanical calculations [18]. Figure 4 gives the interaction potential between a molecule in the $a^3\Sigma_u^+$ state and a helium atom in the ground state. It can be seen that the characteristic scale of the region of major variation in the potential is smaller than the interatomic distance in the liquid. Thus, the helium density was subsequently assumed to be equal to its value at infinity everywhere outside the bubble and

$$E_{ma} = n\int\limits_{R>R_0}\{\varepsilon_{ea}(\mathbf{R})+\varepsilon_{\mathrm{vdW}}(\mathbf{R})\}d^3R.$$

We write the energy associated with the presence of a liquid–vacuum interface at the bubble surface in the form $E_{\mathrm{sur}} = 4\pi R_0^2\gamma$. The value of γ was assumed to be equal to $\sigma_0(n/n_0)$, where σ_0 and n_0 are the coefficients of surface tension at the liquid–saturated vapor interface and the liquid density at the saturated vapor pressure, respectively, when $T \longrightarrow 0$. When calculating the kinetic energy of the molecule we assumed that the configuration of the surrounding liquid remains the same under the molecular oscillations since the attached mass of the bubble is considerably greater than the molecular mass and the characteristic frequencies of the bubble oscillations are relatively low. Under this assumption the change in the potential energy of the interaction of the molecule with the surroundings when the molecule is displaced by $\delta\mathbf{r} = \{\delta x, \delta y, \delta z\}$ relative to the center of the bubble is

$$\delta E(\delta\mathbf{r}) = n\int\limits_{R>R_0}\{\varepsilon_{ea}(\mathbf{R}-\delta\mathbf{r})+\varepsilon_{\mathrm{vdW}}(\mathbf{R}-\delta\mathbf{r})\}d^3R. \quad (3)$$

Expanding (3) as a series in powers of δx, δy, and δz as far as quadratic terms and integrating, we find the frequencies of the molecular oscillations in the corresponding directions. For a bubble radius of 6–7 Å these frequencies correspond to temperatures of 10–15 K and consequently the kinetic energy of the molecule is simply the energy of its zero-point oscillations.

In the adiabatic approximation the frequency shift of the transition is $\Delta\omega = (E_c(R_0) - E_a(R_0))/\hbar$, where $E_a(R_0)(E_c(R_0))$ is the total energy of the "bubble + molecule in state $a^3\Sigma_u^+(c^3\Sigma_g^+)$" complex, and R_0 is the equilibrium radius of the bubble formed by the molecule in the initial state $a^3\Sigma_u^+$. In our model, the shift of the absorption line only depends on the unknown effective charges of the molecular core Z_a and Z_c in the initial and final states which were determined by fitting the experimental values. It can be seen from Fig. 5 that the proposed model accurately describes the interaction between the molecule and the surrounding liquid.

The values of $Z_a = 1.04 \pm 0.05$ and $Z_c = 0.78 \pm 0.04$ thus determined are close to the effective charges Z_A and Z_C determined for, respectively, the $A^1(\Sigma_u^+)$ and $C^1(\Sigma_g^+)$ singlet states from Guberman and Goddard's calculation[16]. The bubble radius varies between 7 Å at low pressure and 6.4 Å at pressures close to solidification.

The absorption spectra corresponding to the $a^3\Sigma_u^+ \longrightarrow c^3\Sigma_g^+$ transition were also measured in cold ^{3}He gas at

densities of 1.3×10^{21}–1.1×10^{22} cm^{-3} (Fig. 6). In order to describe the observed line shifts relative to the vacuum position we used the standard theory of line broadening in the binary limit (see, e.g., [11]) assuming that the energy of the interaction between the molecule and the surroundings can be reduced to the sum of the energies of two-particle interactions between the molecule and isolated atoms. Then, in the adiabatic approximation the frequency dependence of the absorption intensity is given by

$$I(\omega) = \int_{-\infty}^{\infty} e^{i\omega\tau}\varphi(\tau)d\tau,$$

where

$$\varphi(\tau) = \exp\left[-\int\left\{1 - \exp\left[-\frac{i\tau(U_c(\mathbf{R}) - U_a(\mathbf{R}))}{\hbar}\right]\right\}n(\mathbf{R})d^3\mathbf{R}\right]. \quad (4)$$

Here, ω is the frequency shift, $U_a(\mathbf{R})$ and $U_c(\mathbf{R})$ are the energies of the interaction of the molecule with an isolated atom in the initial and final states, $n(\mathbf{R})$ is the coordinate distribution function of the helium atoms. Taking into account the slope of the interaction potential $U_a(R)$ and the smallness of the van der Waals minimum compared with temperature (see Fig. 4), we approximated the real potential by an infinite wall located at a distance $R_{\min}$ from the molecule. The value of $R_{\min}$ is determined from the condition $U_a(R_{\min}) = T$ (classical turning point) and depends weakly on temperature. If three-body "molecule + atom + atom" collisions are neglected, we can easily calculate the coordinate distribution function of the atoms:

$$n(R < R_{\min}) = 0,$$

$$n(R > R_{\min}) = n_\infty\left(1 - \exp\left\{-\frac{2mT(R - R_{\min})^2}{\hbar^2}\right\}\right).$$

Figure 7 shows measured shifts of the maximum of the $a^3\Sigma_u^+ \longrightarrow c^3\Sigma_g^+$ absorption line from the vacuum position in ^{3}He at various densities and results of calculations using the bubble model and in the binary approximation. On comparing the experimental data with the calculations, we can conclude that at densities $\gtrsim 1.5 \times 10^{22}$ cm^{-3} the molecule is localized in a bubble (for comparison, the critical ^{3}He density is 8.3×10^{21} cm^{-3} [19]).

3. KINETICS OF MOLECULAR DECAY

Triplet $He_2^*(a^3\Sigma_u^+)$ molecules are the longest-lived neutral excitations in condensed helium and thus the question of the processes leading to their decay is of considerable interest. The dominant mechanism of excimer loss in liquid and dense gaseous helium is the binary Penning ionization reaction (1) [5, 7, 8]. In cases where the total electron spin of the interacting triplets does not exceed unity, reaction (1) takes place to the extent of exchange interaction between the outer electrons of the reacting particles and its rate at maximum convergence (2.5–3 Å [20]) is $W_{ex} \approx 10^{14}$ s^{-1} [20, 21]. If the electron spin of the reacting triplets is two, ionization as a result of total spin-conserving exchange interaction is forbidden since the spin of the reaction products does not exceed unity. In this case, a nonzero Penning ionization probability only occurs when weak spin dipole interaction is taken into account. The rate W_{ex-d} of the reaction taking place by this exchange dipole mechanism is seven orders of magnitude lower than W_{ex} [4]. Konovalov and Shlyapnikov [4] predict that recombi-

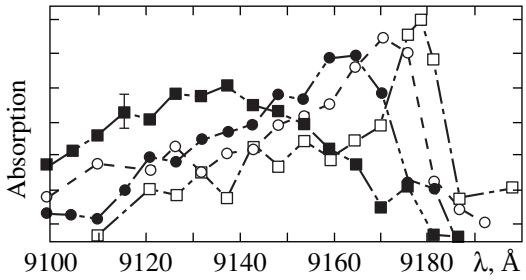

Fig. 6. Absorption spectra of molecules corresponding to the $a^3\Sigma_u^+ \longrightarrow c^3\Sigma_g^+$ transition in cold ^{3}He gas at various densities (□—0.7 atm, 2.9 K, 0.0064 g/cm^3; ○—1.8 atm, 4.2 K, 0.017 g/cm^3; ●—2.2 atm, 4.2 K, 0.033 g/cm^3; ■—2.1 atm, 3.4 K, 0.057 g/cm^3).

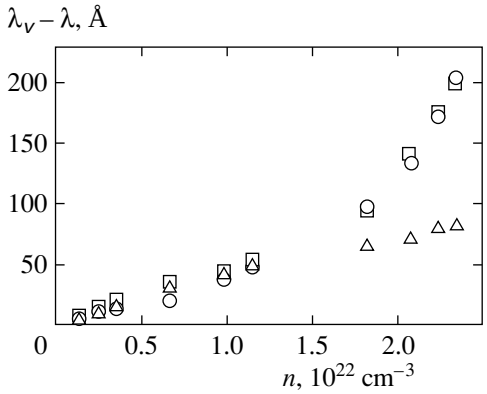

Fig. 7. Shift of the $a^3\Sigma_u^+ \longrightarrow c^3\Sigma_g^+$ absorption line relative to the vacuum position in liquid and gaseous ^{3}He at various densities [○—experiment, □—calculations using bubble model, △—calculations using binary approximation (see text)].

nation will be appreciably suppressed when the triplets are strongly polarized.

The dynamics of the loss of molecules as a result of mutual recombination may be described by the following relationships:

$$\begin{aligned} \dot{n}_\uparrow &= -\alpha_{\uparrow\uparrow} n_\uparrow^2 - \alpha_{\uparrow 0} n_\uparrow n_0 - \alpha_{\uparrow\downarrow} n_\uparrow n_\downarrow, \\ \dot{n}_0 &= -\alpha_{0\uparrow} n_0 n_\uparrow - \alpha_{00} n_0^2 - \alpha_{0\downarrow} n_0 n_\downarrow, \\ \dot{n}_\downarrow &= -\alpha_{\downarrow\uparrow} n_\downarrow n_\uparrow - \alpha_{\downarrow 0} n_\downarrow n_0 - a_{\downarrow\downarrow} n_\downarrow^2, \end{aligned} \tag{5}$$

where $n_\uparrow$, n_0, and $n_\downarrow$ are the concentrations of molecules having different spin projections and α_{ij} are the recombination coefficients. If, following Konovalov and Shlyapnikov [4], we assume that free molecular diffusion takes place, the rate of recombination is determined by

$$\alpha = 4\pi D R_I, \tag{6}$$

where the ionization radius is determined by the condition $R_I^2/D \approx 1/W(R_I)$ ($W(R)$ is the probability of recombination event (1) per unit time). The molecular diffusion coefficient can be estimated in the τ approximation using the calculated bubble radius (see previous section). Typical values of the diffusion coefficient in a normal liquid are around 10^{-5} cm^2/s, increasing in a superfluid liquid as the density of the normal component decreases. As a result of an exponential decrease in the recombination probability W with distance, the characteristic ionization radius varies weakly as the diffusion coefficient varies. Under our experimental conditions the ionization radius of a spin-allowed reaction was 7–10 Å. The characteristic time of an ionization reaction by the exchange–dipole mechanism is several orders of magnitude greater than the diffusion time at all distances and in the approximation under study we find $\alpha_{\uparrow\uparrow} \ll \alpha_{\uparrow\downarrow}, \alpha_{\uparrow 0}$.

However, the assumption of free diffusion is not consistent with the real situation because of the presence of strong van der Waals interaction between the molecules. Allowance for this interaction yields the conclusion that molecules having converged to a distance at which the van der Waals energy is comparable with the temperature, do not diverge but form a bound state having a short intermolecular distance (≈3 Å) and react within times much shorter than the diffusion convergence time even when the total molecular spin is two. The van der Waals capture radius is [22]

$$\begin{aligned} R_{\mathrm{vdW}} &= \left[\int_{R_I}^{\infty} \exp\left(-\frac{C_6}{r^6 T}\right)\frac{dr}{r^2}\right]^{-1} \\ &= \left[\frac{1}{6}\left(\frac{T}{C_6}\right)^{1/6} \int_0^{C_6/(TR_I^6)} e^{-t} t^{-5/6} dt\right]^{-1}. \end{aligned} \tag{7}$$

We use the following formula to estimate the van der Waals coefficient (see, e.g. [23])

$$C_6 = \frac{3}{2}\frac{\hbar e^4}{m^2}\sum_{k,k'} \frac{f_{ak} f_{ak'}}{(\omega_{ak} + \omega_{ak'})\omega_{ak}\omega_{ak'}}, \tag{8}$$

where f_{ak} and ω_{ak} are the oscillator strength and frequency of the $a^3\Sigma_u^+ \longrightarrow k$ transition, respectively, and summation is performed over all possible transitions. The oscillator strength of the $a^3\Sigma_u^+ \longrightarrow b^3\Pi_g$ transition is known, $f_{ab} = 0.205$ [5]. The oscillator strength of the $a^3\Sigma_u^+ \longrightarrow c^3\Sigma_g^+$ transition can easily be calculated using the matrix elements of the dipole moment operator calculated by Yarkony [24] and the calculations give $f_{ac} = 0.307$.

Using the well-known relationship (see, e.g., [25])

$$f_{kk'} = \frac{c^2 m}{\pi \bar{e}^2}\int \sigma(\nu) d\nu,$$

where $f_{kk'}$ is the oscillator strength of the $k \longrightarrow k'$ transition, $\sigma(\nu)$ is the cross section for absorption of light at frequency ν, and the integral is taken along the entire absorption line corresponding to this transition, we obtain from the integrated intensities of our measured spectra: $f_{ac}/f_{ab} = 1.5 \pm 0.2$ which agrees with the calculated data.

Having retained only the principal terms with k, $k' = b, c$ in the sum (8), we obtain the lower constraint on the van der Waals coefficient $C_6^{\min} = 6020$ au. For the upper constraint on C_6 we assume $f_{ae} = 1 - f_{ab} - f_{ac}$ (the $a \longrightarrow d$ transition is parity-forbidden), $f_{ak} = 0$ for $k \neq b, c, e$, which gives $C_6^{\max} = 7940$ au. Thus, we have $C_6 = 7000 \pm 1000$ au.

Bearing in mind that $R_I \lesssim 10$ Å we have $C_6/(TR_I^6) \gg 1$ and integration in (7) can be extended to infinity, which gives

$$R_{\mathrm{vdW}} = \frac{6}{\Gamma(1/6)}\left(\frac{C_6}{T}\right)^{1/6} = 15\text{–}18 \text{ Å} \tag{9}$$

at temperatures of 1.5–4.2 K. Thus, at low temperatures we find $R_{\mathrm{vdW}} > R_I$ so that the van der Waals capture radius should be taken as the characteristic ionization radius:

$$\alpha = 4\pi D R_{\mathrm{vdW}}. \tag{10}$$

Thus, in this model the polarization of the molecules has no influence on their decay dynamics.

Formula (10) obtained in the diffusion approximation holds for short mean free paths $l \ll R_{\mathrm{vdW}}$. If the molecule is situated in a liquid, it is localized in a bub-

ble and direct estimates in the τ-approximation give $l \approx R_0/6$ (R_0 = 6.4–7.0 Å is the bubble radius); for a superfluid liquid we have $l \approx (R_0/6)(n/n_{\text{norm}})$ where n_{norm} is the density of the normal component. If the molecule is situated in a low-density gas, we find $l \approx 2/\sigma n$ where the cross section for scattering of a helium atom at a molecule is $\sigma = 1\text{–}2 \times 10^{-14}$ cm^2 at 2–4 K (the factor 2 appears as a result of a difference between the atomic and molecular masses which has the result that for a molecule a single collision with an atom is insufficient to reverse its momentum). Thus, we find that the result (10) is valid in normal liquid helium, in superfluid ^{4}He at temperatures above 1.7 K, and also in gases at densities $n \gg 10^{21}$ cm^{-3}, and under all the conditions listed above polarization does not influence the recombination of excimers.

We shall now consider the opposite limit of long mean free paths $l \gg R_{\text{vdW}}$. The coefficient of molecular recombination in this case is determined by

$$\alpha = \sigma v_T, \tag{11}$$

where σ is the cross section of reaction (1) and v_T is the thermal velocity. The limit of long mean free paths obtains in superfluid helium at ≲1.3 K and in gas at densities $\ll 10^{21}$ cm^{-3}. In superfluid helium as the molecules converge to distances $R \lesssim 2R_0$ (R_0 is the bubble radius), a bound state of two molecules localized in a single bubble forms so that the cross section is $\simeq \pi(2R_0)^2$ and the recombination coefficient also does not depend on the polarization of the molecules.

We note that an electron–ion pair formed as a result of the excimer ionization reaction (1) may recombine to form a "secondary" molecule. If the probability of this process γ is not low, the recombination coefficient (10), (11) should be multiplied by $(1 - \gamma)$. The experimental setup in the present study can be used to estimate γ. The lifetime of the molecules in liquid helium under the conditions used to observe the absorption spectra was several milliseconds which is much shorter than the characteristic vibrational and rotational relaxation times (~300 ms and 15 ms, respectively [8]). Consequently, the molecular distribution over excited vibrational and rotational states corresponds to the probability of the formation of a molecule in a particular excited state. However, the characteristic relaxation times of the He_2^+ and He_3^+ molecular ions are relatively short as a result of the absence of an excited electron which suppresses the interaction of the ion core with the surroundings in the case of a molecule. Hence, molecules generated in highly excited rotational states are formed as a result of recombination of an electron–ion pair which occurred when the helium atom was ionized, and attachment of one or two atoms to an atomic ion. Such a pair is comparatively short-lived (the characteristic distance of maximum separation during expansion is $R \lesssim 10^{-5}$ cm, the time of convergence and pair recombination is $\tau_{ei} \sim R^3/(\mu\bar{e}) \lesssim 10^{-7}$ s) and the ion core does not have time to relax.

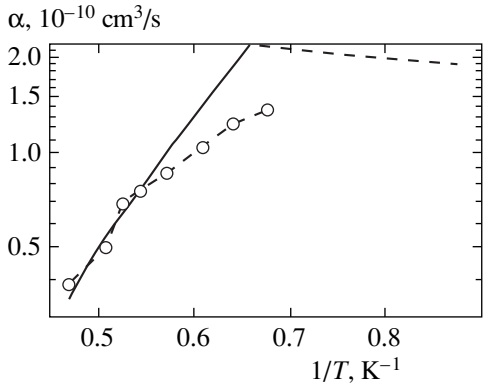

Fig. 8. Temperature dependence of the coefficient of recombination of molecules in superfluid ^{4}He: ○—Fitzsimmons experimental data [7], corrected allowing for the calculated [5] oscillator strength f_{ab}; the solid curve gives the calculations using the diffusion approximation, the dashed curve gives the calculations using the long mean free path limit (see text).

An analysis of the spectra shows that the fraction of molecules formed in this process is (20 ± 10)%. Thus, using the measured spectra we can estimate the probability γ of the formation of a single molecule in a mutual recombination event involving two molecules: let n_1 and n_2 be the concentrations of unexcited (primary) and excited (secondary) molecules and then the rate of generation of secondary molecules is $\gamma\alpha(n_1 + n_2)^2$ and their rate of recombination is $\alpha n_2(n_1 + n_2)$. Consequently in the steady-state regime we have $\gamma = n_2/(n_1 + n_2) \approx 0.2$ and allowance for the factor $(1 - \gamma)$ in estimates of the recombination coefficient is needlessly accurate.

The calculated values (10) and (11) and the Fitzsimmons experimental data [7] measured in superfluid ^{4}He at temperatures of 1.4–2.1 K are compared in Fig. 8.

Thus, the suppression of recombination by a strong magnetic field can only be observed in a gas in the long mean free path regime. The condition $l \gg R_{\text{vdW}}$ is equivalent to the absence of three-body "molecule + molecule + atom" collisions. Then, as they converge, these

Table 1

Phase	T, K	P, atm	$\mu_B H/(kT)$
Superfluid ^{4}He	2.12	0.05	3.5
Superfluid ^{4}He	1.76	28.1	4.2
Superfluid ^{4}He	1.43	0.004	5.2
20% ^{3}He–^{4}He solution	1.8	14.4	4.1
^{3}He, liquid	1.8	14.4	4.1
^{3}He, liquid	1.8	31.5	4.1
^{3}He, gas	2.9	0.7	1.4

Table 2

Phase	*T*, K	*P*, atm	α, cm^3/s (±15%) experiment	α, cm^3/s calculated	l/R_{vdW}
Liquid ^{4}He	2.10	1.0	3.8×10^{-11}	3.6×10^{-11}	0.086
Liquid ^{4}He	1.98	14.4	3.2×10^{-11}	2.9×10^{-11}	0.072
Liquid ^{4}He	1.76	28.1	2.8×10^{-11}	3.2×10^{-11}	0.080
Liquid ^{4}He	1.76	1.0	1.1×10^{-10}	0.9×10^{-10}	0.23
Liquid ^{3}He	1.64	14.4	3.3×10^{-11}	2.9×10^{-11}	0.086
Gas ^{3}He	3.0	0.7	4.3×10^{-10}	8.5×10^{-10}	0.60

molecules do not form bound states and react with a certain probability which depends on their polarization. The corresponding reaction cross sections for triplet atoms were calculated in [20]: $\sigma(\uparrow\downarrow) \approx \sigma(\uparrow 0) \approx 3.2 \times 10^{-14}$ cm^2, $\sigma(\uparrow\uparrow) \approx 2.9 \times 10^{-15}$ cm^2 at low temperatures.

Unfortunately our method of generating molecules cannot operate at fairly low helium densities since, when $n \lesssim 10^{21}$ cm^{-3}, the electron mobility increases rapidly [15] and breakdown occurs in the cell. Table 1 gives all the experimental conditions used to study the influence of molecular polarization on the recombination coefficient. The molecular lifetimes were 200 ms.

Simple estimates using standard spin–lattice relaxation theory (see, for example [26]) for molecules in helium of appropriate density give longitudinal relaxation times on microsecond scales as a result of "spin–axis" interaction, which suggests that the polarization of the molecules in our experiments is close to equilibrium. Under our conditions it was impossible to observe molecular polarization using the Zeeman effect because in the fairly strong magnetic fields for which multiplet splitting could become appreciable in the absorption spectra, optical transitions accompanied by a change in the spin projection M_S are forbidden as a result of the Paschen–Back effect.

Under all the experimental conditions listed above no influence of the magnetic field on excimer decay was observed, which is in complete agreement with the theory.

In order to determine the numerical values of the recombination coefficient under various experimental conditions, the experimental time dependences of the absorption signal were fitted using the binary reaction equation

$$A(t) = \left(\frac{1}{A(t_0)} + \frac{\alpha(t-t_0)}{\sigma_0 V/S}\right)^{-1},$$

where $A(t) = n(t)\sigma_0 V/S$, σ_0 is the cross section for absorption of light determined from the oscillator strength of the transition and the integrated intensity of the spectrum, V is the volume in which absorption takes place, S is the area of the light beam, and the recombination coefficient α was the fitting parameter. The unknown effective volume V was obtained by comparing our data with the results [7] under similar conditions. The values of the recombination coefficient thus determined and those calculated using formula (10) are given in Table 2.

All the experimental data agree with the calculations within measurement error, except for the coefficient of recombination measured in a gas, where the criterion for the validity of the diffusion approximation ceases to be satisfied. Assuming that the cross section of the ionization reaction is approximately equal to the cross section for the reaction of two triplet atoms at low temperatures (see above), in the long mean free path limit we obtain $\alpha \approx 4.8 \times 10^{-10}$ cm^3/s at 3.0 K which is in good agreement with the experimental value.

4. CONCLUSIONS

The position and shape of the absorption line corresponding to the $a^3\Sigma_u^+ \longrightarrow c^3\Sigma_g^+$ molecular transition exhibits a strong dependence on the helium density which means that the interaction of the $He_2^*(a^3\Sigma_u^+)$ molecule with the surroundings can be studied using optical measurements. By analyzing the spectra obtained under various experimental conditions, we established that at above-critical densities the molecules are localized in microscopic bubbles similar to the localization of excess electrons. The size of this complex, unlike a bubble, formed by an electron varies weakly with pressure.

We obtained estimates of the coefficients of the van der Waals interaction between a molecule and a ground-state helium atom:

$$C_6(He_2^*(a^3\Sigma_u^+) - He(1^1S)) \approx 54 \text{ au},$$

$$C_6(He_2^*(c^3\Sigma_g^+) - He(1^1S)) \approx 68(4/5 + 3/5\cos^2\Theta) \text{ au}.$$

We observed appreciable broadening of the absorption line in ^{4}He as a function of temperature. The natural oscillation frequencies of the bubble which are easily estimated correspond to temperatures around 3 K and we ascribe the observed broadening to the excitation of vibrational degrees of freedom of the bubble.

We obtained an estimate of the coefficient of van der Waals interaction of the molecules $C_6 = 7000 \pm 1000$ au. Allowance for the strong attraction of molecules at large distances yields the conclusion that the diffusion-limited rate of excimer recombination does not depend on the molecular polarization. Calculations using the proposed model show good agreement with all the available experimental values measured under conditions when the diffusion approximation is valid (normal ^{3}He and superfluid ^{4}He at temperatures above 1.7 K). The recombination coefficient measured in cold ^{3}He gas agrees with the data [5] obtained in ^{4}He at similar densities and shows good agreement with the results of the theoretical calculations [20] for an extremely low-density gas.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 97-02-16360) and by the INTAS (grant no. 96-0610).

REFERENCES

1. G. W. E. Drake, Phys. Rev. A **3**, 908 (1971).
2. H. W. Moos and J. R. Woodworth, Phys. Rev. Lett. **30**, 775 (1973).
3. D. B. Kopeliovich, A. Ya. Parshin, and S. V. Pereverzev, Zh. Éksp. Teor. Fiz. **96**, 1122 (1989) [Sov. Phys. JETP **69**, 638 (1989)].
4. A. V. Konovalov and G. V. Shlyapnikov, Zh. Éksp. Teor. Fiz. **100**, 521 (1991) [Sov. Phys. JETP **73**, 286 (1991)].
5. D. W. Tokaryk, R. L. Brooks, and J. L. Hunt, Phys. Rev. A **48**, 364 (1993).
6. B. Brutschy and H. Haberland, Phys. Rev. A **19**, 2232 (1979).
7. J. W. Keto, F. L. Soley, M. Stockton, and W. A. Fitzsimmons, Phys. Rev. A **10**, 872 (1974).
8. V. B. Eltsov, S. N. Dzhosyuk, A. Ya. Parshin, and I. A. Todoshchenko, J. Low Temp. Phys. **110**, 219 (1998).
9. J. C. Hill, L. L. Hatfield, N. D. Stockwell, and G. K. Walters, Phys. Rev. A **5**, 189 (1972).
10. A. P. Hickman, W. Steets, and Neal F. Lane, Phys. Rev. B **12**, 3705 (1975).
11. P. W. Anderson, Phys. Rev. **86**, 809 (1952).
12. V. B. El'tsov, A. Ya. Parshin, and I. A. Todoshchenko, Zh. Éksp. Teor. Fiz. **108**, 1657 (1995) [JETP **81**, 909 (1995)].
13. W. Lichten, M. V. McCusker, and T. L. Vierima, J. Chem. Phys. **61**, 2200 (1974).
14. K.-P. Huber and G. Herzberg, *Molecular Spectra and Molecular Structure* (Van Nostrand, New York, 1979; Mir, Moscow, 1984), Part 1.
15. V. B. Shikin, Usp. Fiz. Nauk **121**, 457 (1977) [Sov. Phys. Usp. **20**, 226 (1977)].
16. S. L. Guberman and W. A. Goddard, III, Phys. Rev. A **12**, 1203 (1975).
17. A. A. Radtsig and B. M. Smirnov, *Reference Data on Atoms, Molecules, and Ions* (Atomizdat, Moscow, 1980; Springer-Verlag, Berlin, 1985).
18. J. Callaway and E. Bauer, Phys. Rev. A **140**, 1072 (1965).
19. B. A. Wallace and H. Meyer, Phys. Rev. A **5**, 953 (1972).
20. M. W. Müller, A. Mertz, M.-W. Ruf, *et al.*, Z. Phys. D **21**, 89 (1991).
21. B. C. Garrison and W. H. Miller, J. Chem. Phys. **59**, 3193 (1973).
22. R. M. Noyes, in *Progress in Reaction Kinetics*, Ed. by G. Porter (Pergamon, New York, 1961), Vol. 1.
23. Yu. S. Barash, *Van der Waals Forces* (Nauka, Moscow, 1988), p. 31.
24. D. R. Yarkony, J. Chem. Phys. **90**, 7164 (1989).
25. M. A. El'yashevich, *Atomic and Molecular Spectroscopy* (Fizmatlit, Moscow, 1962), p. 193.
26. I. V. Aleksandrov, *Theory of Magnetic Relaxation. Relaxation in Liquids and in Solid Non-metallic Paramagnets* (Nauka, Moscow, 1975), p. 270.

Translation was provided by AIP