Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 1010–1021.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1161–1174.

SOLIDS
Electronic Properties

Magnetic-Field Oscillations of the Critical Current and the Pinning Force in PbTe/PbS Semiconductor Superlattices and $YBa_2Cu_3O_{7-\delta}$ Films

A. I. Fedorenko*, V. V. Zorchenko, A. Yu. Sipatov, O. A. Mironov, and S. V. Chistyakov

Kharkov State Polytechnical University, Kharkov, 310002 Ukraine
**e-mail: fedor@kpi.kharkov.ua*

Received November 12, 1999

Abstract—The magnetic-field dependences of the critical current I_c and the pinning force F_p in single-crystal semiconductor PbTe/PbS superlattices on KCl and $YBa_2Cu_3O_{7-\delta}$ films in magnetic fields oriented perpendicular to the plane of the samples or parallel to the current were investigated. Oscillations of I_c and F_p were observed for superlattices with a parallel orientation of the magnetic field and with two directions of the field for the $YBa_2Cu_3O_{7-\delta}$ film. A model was proposed for the vortex structures which correspond to extrema of the pinning force in superlattices. It was shown that the single points (extrema and points of inflection) of the field dependences of F_p for superlattices and a $YBa_2Cu_3O_{7-\delta}$ film appear for critical values of the magnetic field equal to $H_{cr} = (p/q)H_0$, where p and q are integers, $H_0 = 838.37$ Oe, and $H_0 = ch/2\pi eR_0^2$ is determined by the length $R_0 = 88.607$ nm, somewhat less than the reciprocal of the Rydberg constant $R_\infty = me^4/4\pi\hbar^3c$ ($R_\infty^{-1} = 91.127$ nm). It was inferred on the basis of published data that the temperature and magnetic-field dependences of the properties of superlattices and HTSC materials follow general laws. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It has been shown in the last few years that high-temperature superconductors (HTSCs) belonging to the bismuth and thallium families [1] and also, apparently, $YBa_2Cu_3O_{7-\delta}$ [2] can be regarded as Josephson superlattices, where stacks of two (or three) CuO_2 planes in a unit cell, which are separated by an interlayer from the metal atoms, play the role of the superconducting layers. The Josephson coupling between the stacks of CuO_2 layers in neighboring unit cells is manifested in specific behavior of the current–voltage characteristics (IVCs) of samples when a current flows perpendicular to the CuO_2 planes (hysteresis, the existence of a set of branches of the IVC, and so on) as well as in oscillations of the critical current when a magnetic field is applied parallel to the layers [1, 2]. It was inferred that the Josephson coupling in $YBa_2Cu_3O_{7-\delta}$ can exist even between the nearest-neighbor CuO_2 planes separated by an yttrium layer [2]. In this connection it is of interest to investigate the properties of artificial superlattices consisting of alternating layers of materials with different superconducting and electrophysical properties.

Only a relatively few of the many works on superlattices concern critical currents. The critical currents and pinning force have been determined for superlattices of the type *S*/*S*' (Nb/Ta [3], Nb/NbTi [4], and Nb/NbZr [5]), *S*/*N* (Nb/Cu [6], Nb/CuX (X = Ge, Mn) [7], NbTi/Ti and Nb/Ti [4], Pb–Bi/Cr [8]), and *S*/*I* (Nb/Si [9], Nb/NbO_x [10], Nb/Al_2O_3 [4, 11], $Nb/Al–AlO_x/Nb$ [12], NbN/AlN [13], Pb/Ge [14]) as well as for superlattices consisting of the alloy Pb–Bi with sinusoidal modulation of the Bi concentration [15] and superlattices based on HTSC materials (YBCO/PrBCO [16], EuBCO/PrBCO [17] and $Nd_{1.83}Ce_{0.17}CuO_x$/YBCO) [18]. The main result of these works is the conclusion that the pinning force can be substantially increased in superconductors by using interlayers with weak superconducting properties. Josephson effects have also been observed in superlattices of the type *S*/*N* [6] and *S*/*I* [9, 12]).

In the present paper the behavior of the magnetic-field dependences of the critical currents I_c and volume pinning force F_p in semiconductor superlattices PbTe/PbS [19] with regular networks of misfit dislocations is examined. Such superlattices are physical and, to a large extent, structural analogs of HTSCs [19– 23]. Oscillations of $I_c(H)$ and $F_p(H)$ for a magnetic field oriented parallel to the current and monotonic decay of $I_c(H)$ in a magnetic field perpendicular to the layers of the superlattices were observed. It is shown that the field behavior of I_c and F_p in these superlattices and in epitaxial $YBa_2Cu_3O_{7-\delta}$ films are similar, and certain general characteristics of the oscillations of the pinning force, which make it possible (on the basis of the data from previous investigations) to talk about a similarity of properties and mechanisms of superconductivity in these superlattices and HTSCs, are found.

1063-7761/00/9006-1010$20.00

2. SAMPLE PREPARATION AND MEASUREMENT PROCEDURE

The samples were prepared in a vacuum chamber with an oil-free pumping system ($P \sim 10^{-4}$–10^{-5} Pa) by thermal evaporation of lead chalcogenides from tungsten boats, followed by condensation on a (001) KCl at 520–570 K. The layer thicknesses and condensation rates were monitored with a quartz resonator. The samples consisted of 10 PbTe–PbS bilayers (PbS is the layer closest to the substrate). The investigations showed [19] that the lead chalcogenide films grow on each other in layer-by-layer mode according to the Frank–van der Merwe mechanism, as a result of which single crystal layers with a low density of growth dislocations (10^7–10^8 cm^{-2}) and mosaic blocks, repeating the blocks of the KCl substrate (~100 μm) are formed. Regular square networks of misfit dislocations, oriented along the ⟨110⟩ directions with Burgers vector $(a/2)\langle 110\rangle$ and period 5.2 nm [19] form on the (001) PbTe–PbS interfaces.

Analysis of X-ray diffraction curves on the basis of the standard $\vartheta - 2\vartheta$ scanning scheme showed that the superlattices are periodic and possess sharp interfaces, which is indicated by well-defined satellite reflections around the Bragg reflections and alongside the primary peak. The period of the superlattice and the thickness of the layers were determined (to within an accuracy of 0.1 nm) from the distance between the satellites [24].

Previous investigations have shown that the superconducting properties of superlattices are determined not only by the thicknesses of the layers but also substantially by their stoichiometry [21]. Only superlattices obtained from the standard, nonoptimized, PbS charge, giving for 500-nm thick single-layer PbS films a Hall charge-carrier density $n_H \sim 10^{19}$ cm^{-3} and mobility $\mu_H \sim 10^2$ cm^2/(V s) (at 78 K), were used in the present experiments. The composition of the PbTe charge was optimized by repeated sublimation, which gave the best stoichiometry for a given growth temperature [21].

Samples in the form of a double Hall cross, which were obtained by condensing layers of lead chalcogenides through a mask with an appropriate geometry, were used to measure the IVCs and the critical currents by the four-contact method. The samples were 10 mm long and 1 mm wide, and the distance between the potential contacts was 3 mm. The measurements were performed in a helium cryostat ($T \geq 1.5$ K) using dc current and in a constant magnetic field H oriented perpendicular or parallel to the layers in the direction of the current (the deviation from the orientation parallel to the layers did not exceed 0.1°). The magnetic field was produced by superconducting solenoids and determined to within ~0.1%. The working currents fell within the ohmic sections of the IVCs of the samples. The temperature of the samples was maintained constant to within 0.005 K. The critical current I_c was determined from the IVC as the current at which the voltage on the potential contacts reached 10 μV. The character of the dependence $I_c(H)$ remained the same with an order of magnitude lower voltage level 1 μV as the criterion.

The geometric and superconducting parameters of the samples are indicated in table. It is evident from the table that one of the samples (no. 4) had several PbTe and PbS layers with thicknesses substantially different from those of other similar layers.

3. EXPERIMENTAL RESULTS AND DISCUSSION

Figure 1 shows the field dependences of the critical currents for two samples in a magnetic field perpendicular ($H^\perp$) and parallel ($H^\parallel$) to the layers. Such dependences are typical for the superlattices investigated at temperatures quite far from the superconducting transition temperature T_c (in our case $T < T_c/2$, with the exception of sample no. 4). They have two characteristic features. First, a sharp drop of $I_c(H)$ is seen for all experimental superlattices even in fields of the order of several or tens of oersteds (the discrepancy in I_c ($H = 0$) in the same figure is due to the degradation of the samples heated up to room temperature in the interval between measurements of $I_c(H^\parallel)$ and $I_c(H^\perp)$ and with a substantial time interval between these measurements). Such a sharp drop in $I_c(H)$ can be understood if it is assumed that the lower critical fields $H_{c1\perp}$ and $H_{c1\parallel}$ are weak for superlattices and weak Josephson-type couplings exist between mosaic blocks of superlattices, which form as a result of the interaction of the lead chalcogenide films with the atmosphere and the high chemical activity of the block boundaries. This conjecture was advanced previously in [25], where the absorption of microwave radiation in PbTe/PbS superlattices was investigated in dc and ac magnetic fields, and it also follows from the behavior of the amplitude of the diamagnetic response of a stack of superlattices as a function of the perpendicular magnetic field [26]. In the latter case the setup at the Physicotechnical Institute, the National Academy of Sciences of Ukraine (Donetsk) was used to determine the magnetic susceptibility of bulk superconductors according to the shift of the frequency of a resonance loop when the experimental sample. In this case, a stack of eight superlattices on KCl substrates with equal PbTe and PbS layer thicknesses ($d_{PbTe} = d_{PbS} = 30$ nm) oriented so that the superlattice layers were perpendicular to the coil axis was inserted into the measuring coil. The measurements were performed at a frequency of 67 Hz in a modulation field $h = 0.2$–2.2 Oe. It was found that the diamagnetic response amplitude L (the inductance of the measuring loop) has the form $L \propto (h - h_{c2})^t$ as $L \longrightarrow 0$, where the value of t is close to 1.7—the critical exponent characteristic for power-law behavior of a number of parameters of Josephson superconductors in the percolation state, specifically, granular and block HTSC films [27]. The quantity h_{c2} can be identified as the second critical field of films as Josephson media. It is at least three orders of magnitude less than the value of

Basic Parameters of the Samples

Basic parameters	Sample no.			
	1	2	3	4
$d_{PbS} + d_{PbTe}$, nm	16 + 14.5	16 + 14.5	18 + 17	18 + 17
T_c, K	3.7	4.2	5.2	3
Measurement temperature, K	1.61	1.71	1.65	1.7
$J_c(H = 0)$, A/cm^2	1774	3117	923	816
$H_\perp^{max}$ (exp.), Oe	210	420	1395	167
$H_\perp^{max}$ (theor.), Oe	209.59	419.19	1397.28	167.67
$H_\perp^{max}/H_0$	1/4	1/2	5/3	1/5
$H_{c2\perp}^{max}$, Oe	620	1700	4100	480
$F_p^\perp$, dynes/cm^3	2323	9640	8239	612
H_1^{max} (exp.), Oe	90	840	720	120
H_1^{max} (theor.), Oe	93.15	838.37	718.6	119.77
H_1^{max}/H_0	1/9	1	6/7	1/7
$F_p^\parallel \times 10^{-3}$, dynes/cm^3	18.2	56.9	13.27	6.2
$L_\perp$, nm	337.2	238.4	130.9	378.5
$L_\parallel$, nm	753.3	80.7	81.6	468.9

Note: The layer thicknesses in sample no. 4 (in order of deposition) were 18, 17, 18, 20, 17, 33.5, 17, 17 nm + 6 periods (18 nm PbS + 17 nm PbTe). The thickness of the first PbS layer in sample no. 3 was 20 nm. The highest values of $J_c(H = 0)$, taking into account the aging of the samples, are presented; $H_\perp^{max}$ the field corresponding to the maximum of the pinning force $F_p^\perp$; H_1^{max} the field of the first maximum of $F_p^\parallel$; $F_p^{\perp,\parallel}$ the pinning force at the first maximum; $H_{c2\perp}^{max}$ the critical field in which $J_c \longrightarrow 0$; $L_\perp$ the distance between vortices in a triangular lattice with $H^\perp = H_\perp^{max}$; $L_\parallel$ in the plane of a superlattice with $H^\parallel = H_1^{max}$ for a structure of type *a*, shown below in Fig. 6.

$H_{c2\perp}$ for superlattices. Using the data obtained, the upper and lower limits of the penetration depth $\lambda_\parallel$ for a magnetic field parallel to the layers of the superlattices can be roughly estimated. In Josephson-coupled media h_{c2} is always less than H_{c1} (the Josephson couplings are destroyed when the field penetrates into the granules). Consequently, $H_{c1\perp} = h_{c2}$ can be used to estimate the upper limit of $\lambda_\parallel$ and $H_{c1\perp} = 40h_{c2}$ can be used to estimate the lower limit (as in HTSCs). Using the well-known relation

$$\frac{H_{c2\perp}}{H_{c1\perp}} = \frac{2(\lambda_\parallel/\xi_\parallel)^2}{\ln(\lambda_\parallel/\xi_\parallel)},$$

where $\xi_\parallel$ is the coherence length parallel to the superlattice layers, for temperature 2.4 K ($H_{c2\perp} = 4$ kOe, $\xi_\parallel = 6$ nm, $h_{c2} = 2.82$ Oe) we obtain $H_{c1\perp} \sim 3$–100 Oe and $\lambda_\parallel \sim 30$–320 nm, which agrees with the characteristic scale of the decay region of $I_c(H)$.

In addition, the critical current density J_c for PbTe/PbS superlattices [20, 21] is small even at relatively low temperatures: several orders of magnitude lower than for the typical Nb/Ta-type superlattices [3] and HTSC epitaxial films and of the same order of magnitude as for polycrystalline HTSC samples. On the other hand, estimates of J_c from point-contact measurements [21] give $J_c \approx (1–7) \times 10^7$ A/cm^2, corresponding to the best film samples of HTSCs. This discrepancy serves as an additional confirmation of the presence of weak links between blocks in PbTe/PbS superlattices.

The second feature is the oscillatory behavior of the derivatives dI_c/dH with a magnetic field applied parallel to the layers. In addition (just as for sample no. 1 in Fig. 1a), at low temperatures, where superconductivity is localized in the PbTe layers [20, 21], clear oscillations of $I_c(H^\parallel)$ can sometimes be seen (for sample no. 4 with layer thickness errors, two maxima were observed).

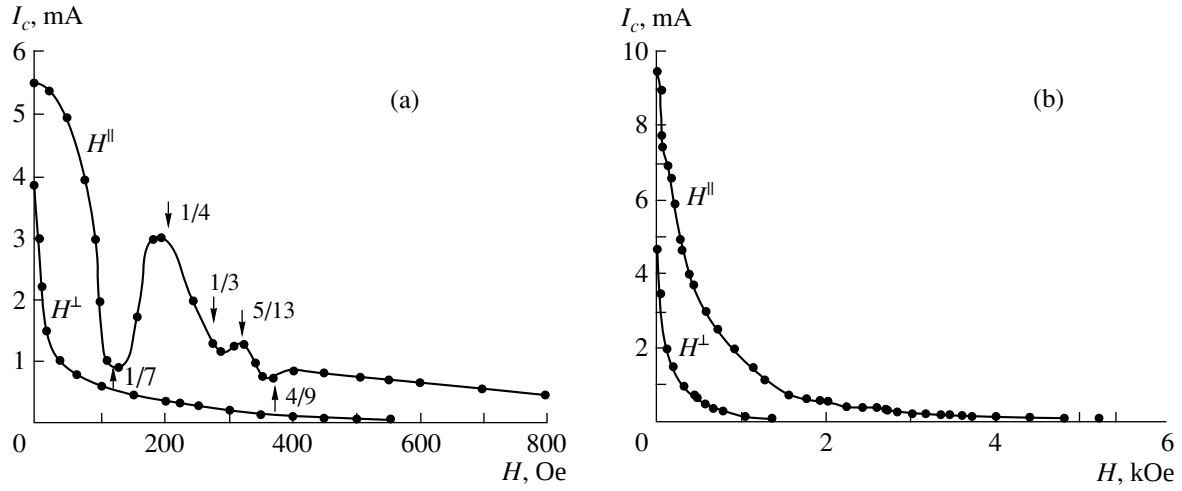


Fig. 1. Field dependences of the critical currents in longitudinal ($H^{\parallel}$) and transverse ($H^{\perp}$) magnetic fields for samples: (a) no. 1, T = 1.61 K; (b) no. 2, T = 1.71 K.

It should be kept in mind that the dependences $I_c(H^{\parallel})$ were determined in a magnetic field parallel to the direction of the measuring current (the so-called force-free configuration). Formally, the current does not exert a Lorentz force, proportional to $\mathbf{J} \times \mathbf{B}$, on the vortices. This force is a result of the curvature of the flux lines, and the theory for this case is still far from completion [28]. In addition, in layered superconductors with a small deviation of the magnetic field from the plane of the layers it is possible to have a situation where the flux lines are rectilinear segments, lying in neighboring layers and connected by step-like sections of the vortices (kinks) [29]; this gives rise to the appearance of a Lorentz force. It has not been ruled out that such a possibility occurs in our measurements.

In what follows, we shall talk about the volume pinning force $F_p = (1/c)|\mathbf{J}_c \times \mathbf{B}|$, in calculating which the magnetic field H will be used instead of the magnetic induction B. This substitution is justified in fields much greater than H_{c1}, since it has virtually no effect on the result, and it is often used. For a magnetic field parallel to the current, we shall use the quantity $F_p = (1/c)(J_c^{\parallel} H^{\parallel})$ for the pinning force.

Let us consider first the behavior of the pinning force in a field $H^{\perp}$. It is well known that for type-II superconductors the density of the pinning force F_p far from H_{c1} can often be described well by a scaling law [28]

$$f = F_p / F_p^{\max} = Ah^k(1-h)^n, \qquad (1)$$

where $F_p^{\max}$ is the maximum pinning force, $h = H/H_{c2}^{\max}$, and $H_{c2}^{\max}$ is the upper critical field corresponding to $J_c \longrightarrow 0$; k and n are integers (or half-integers for k), which depend on the pinning mechanism; and, A is a normalization factor. The curve (*1*) has the typical bell shape, and the position of its maximum on the h-axis is determined by the values of k and n, i.e., by the pinning mechanism. Figure 2 shows the data for four samples in the reduced coordinates f–h. It is evident from this figure that for the present samples the dependences $f(h)$ are close to functions of the form (1). Curves with $k = 1/2$, $n = 1$ and $k = 1$, $n = 2$ are represented by the solid lines *1* and *2* in Fig. 2.

For three of our samples (nos. 1, 3, and 4) the maximum of $F_p^{\perp}$ occurs almost at the value $h = 1/3$, while for sample no. 2 the maximum occurs at $h = 0.25$. The

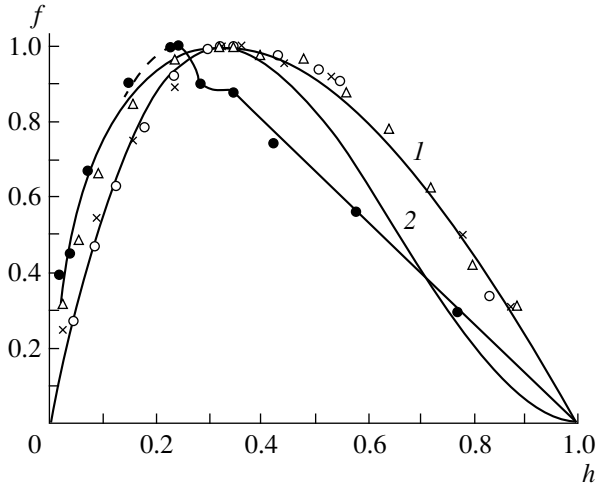


Fig. 2. The dependence of the reduced pinning force f on the relative magnetic field h for samples no. 1 (△), no. 2 (●), no. 3 (×), and no. 4 (○). The temperatures at which the measurements were performed are indicated in table. Here $f = F_p^{\perp}/F_{p\perp}^{\max}$, $h = H_{\perp}/H_{c2\perp}^{\max}$. The solid lines *1* and *2* are, respectively, the dependences $h^{1/2}(1-h)$ and $h(1-h)^2$. The dashed line passing into a solid line shows the section of the curve $h(1-h)^3$ before the maximum in $f(h)$.

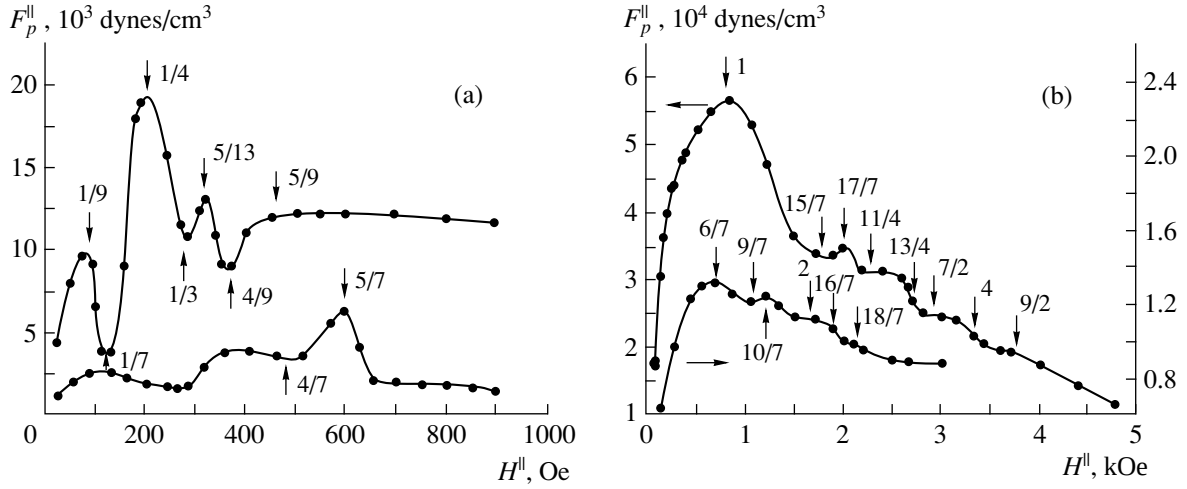


Fig. 3. The pinning force $F_p^{\parallel}$ in the longitudinal magnetic field $H^{\parallel}$ for samples: (a) no. 1 (top curve), no. 4 (bottom curve) and (b) no. 2 (top curve), no. 3 (bottom curve).

form of the dependence $f(h)$ beyond the maximum is the same for samples nos. 1, 3, and 4 ($k = 1/2$, $n = 1$); as far as we know, such a dependence has not been observed for other types of superlattices. Before the maximum the behavior of $F_p^{\perp}$ is described best by the relation (1) with $k = 1$, $n = 2$ for samples nos. 3 and 4 and $k = 1/2$, $n = 1$ for samples nos. 1 and 2; but, for sample no. 2 the curve $f(h)$ moves sharply upwards at $h \approx 0.15$, approaching the dependence (1) with $k = 1$, $n = 3$. After the maximum the sample no. 2 showed a virtually linear dependence for $F_p^{\perp}(h)$.

These facts show that the pinning mechanism probably changes after the maximum. It is known from [28] that pinning on the grain boundaries corresponds to a normalized curve $F_p(h)$ with a maximum at $h = 1/3$ ($k = 1$, $n = 2$; $k = 1/2$, $n = 1$), and a maximum at $h = 0.2$–0.85 corresponds to pinning on dislocations. Consequently, the most likely mechanism in fields $H^{\perp} > H_{\perp}^{\max}$ for samples with a maximum at $h = 1/3$ is pinning on the boundaries of ~200–300 nm regions of coherent scattering (CS) of X-rays. Estimates of the distances between vortices in fields corresponding to the maximum of $F_p^{\perp}(h)$ attest to this mechanism (see table). These distances were found to be close to the sizes of the blocks of the CS regions and much shorter than the average distance between the growth dislocations inside the blocks.

For sample no. 2, the maximum of the pinning force at $h = 0.25$ could be due to the fact that in this sample the process of enlargement of the blocks of CS regions, accompanied by a decomposition of dislocation clusters at the boundaries of CS regions and emergence of dislocations into the volume of the blocks, does not go to completion at the growth stage of the film. Then one can talk about pinning on individual dislocations, resulting in a maximum of $F_p^{\perp}$ at $h = 0.25$. We also note that for our samples the maximum of $F_p^{\perp}$ is reached even in fields of the order of several hundreds of oersteds, and the field $H_{c2}^{\max}$, corresponding to $J_c \longrightarrow 0$, is appreciably less than H_{c2}. The small magnitude of the pinning force is not surprising, considering the high degree of perfection of the crystal structure of the superlattices. The behavior of the curves $F_p^{\parallel}(H)$ for a collinear orientation of the external magnetic field $H^{\parallel}$ and the current was more unexpected. The corresponding data are shown in Fig. 3. The dependences presented are oscillatory even for samples in which there are no oscillations of $J_c(H)$. This is easy to understand, since the condition $J_c + (dJ_c/dH)H = 0$ is necessary in order for extrema to appear in $F_p(H)$; this condition can be satisfied for several values of H with oscillatory behavior of dJ_c/dH, as happens for our samples.

It should be noted that the appearance of maxima in the field dependence $J_c(H)$ has been noted for different types of superlattice (here we do not have in mind oscillations of $J_c(H)$ in Josephson superlattices [6, 9, 12], which are of the same nature as for single Josephson contacts in a longitudinal field). For a perpendicular orientation of the magnetic field for superlattices Nb/Ta [3] the peak in $J_c^{\perp}(H)$ for a sample with period 1.95 nm, which appeared at $T/T_c \leq 0.84$ near $h \approx 0.7$, was explained by a two-dimensional collective pinning of vortices, while for Pb/Ge superlattices [14] with weak pinning of vortices the appearance of a maximum in $J_c^{\perp}(H)$ can be equally explained either on the basis of

the model of "decoupling" of the superconducting layers or the model of melting of a vortex lattice. A single peak in the dependence $J_c(H)$ was observed for Nb/Ta [3], Nb/Zr [5], and Nb/NbO_x [10] superlattices with the field orientation along the layers. This peak seems to be due to the commensurability of the vortex lattice and the period of the superlattice. An explicit manifestation of commensurability was found for superlattices consisting of the alloy Pb–Bi with sinusoidal variation of the Bi concentration [15]: three maxima were observed in $J_c^{\parallel}(H)$. Besides these structures, commensurability effects in the form of two sharp kinks in the temperature dependences $H_{c2}^{\parallel}(T)$ have been observed in V/Ag superlattices [30].

Analysis of the appearance of extrema in $F_p^{\parallel}(H)$ for our superlattices showed that if the field $H_1^{\max}$ of the first maximum is used as the unit of measurement for the magnetic field, then the singular points (extrema and inflection points) in the curves $F_p^{\parallel}(H)$ appear at $H/H_1^{\max}$ equal to either integers or rational numbers. The dependence $F_p^{\parallel}(H)$ for sample no. 1 is shown schematically in such coordinates at the top of Fig. 4. This suggests that the singular points of the curves $F_p^{\parallel}(H)$ should correspond to definite configurations of vortex lattices, each configuration being obtained by restructuring of the preceding lattice formed in weaker fields. Possible configurations of vortices for sample no. 1 are shown in Fig. 4. Here a section of a film perpendicular to its surface is shown schematically (the horizontal lines represent the top and bottom boundaries of the sample), and repeated fragments-cells of vortex structures (rectangular "boxes") are singled out. The filled dots represent the centers of the vortices (the axes of the vortices are perpendicular to the plane of the figure). The numbers on the left side of the rectangles-cells indicate the magnetic field of the singularity of $H/H_1^{\max}$ in units of $F_p^{\parallel}(H)$; the numbers on the right indicate the number of vortices per cell. The letters *a*, *b*, *c*, and *d* represent four types of characteristic vortex structures which were encountered in all of the experimental samples. We note that the horizontal dimensions *b*1, *c*2, and *d*1 of the cells in Fig. 4 differ somewhat from the cell size (*a*) (approximately by factors of 0.972, 1.083, and 0.938, respectively).

In Fig. 5, the axes of the vortices are tied to the superlattice layers schematically (with a distortion of the real scales). It was assumed that the vortices will be located in PbS interlayers, since at sufficiently low temperatures, as in our case, the order parameter will be completely localized in the PbTe layers [20, 21] so that the arrangement of the vortices in the PbS layers is energetically favorable. Similar schemes of the configuration of vortices can also be presented for other samples. The distances L_n between the vortices in one of the PbS layers in structures of the type *a*–*d* with n = 1, 2, 3, and 5 rows of vortices are much larger than the characteristic thicknesses of the layers of the superlattices (see table). At singular points of $F_p^{\parallel}(H)$ the length $L_n = n\phi_0/tH$ (t is the thickness of the sample, $\phi_0 = ch/2e$) are either equal or differ from one another by less than a factor of 2 (for Fig. 4 the ratios $L_n/L_1 = nH_1^{\max}/H_n$ are 1, 14/9, 8/9, 1, 39/45, 5/4, and 1).

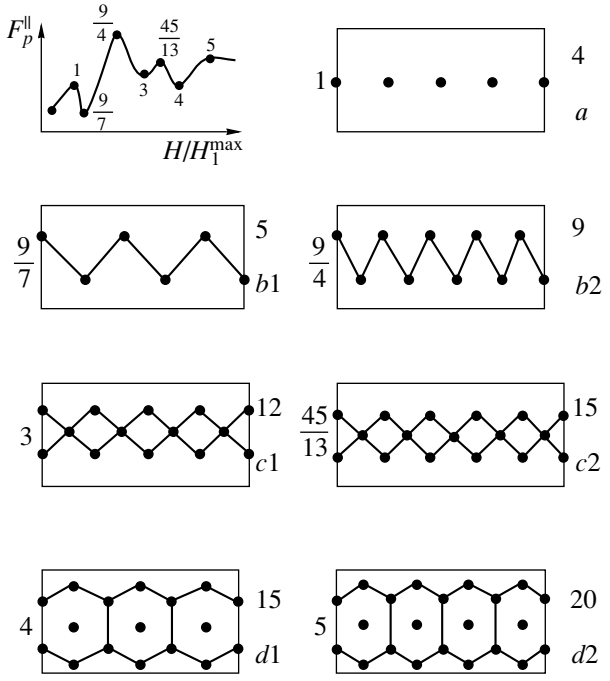


Fig. 4. Schematic illustration of the dependence of the pinning force $F_p^{\parallel}$ on the longitudinal magnetic field $H^{\parallel}$, taken in units of the field at the first maximum $H_1^{\max}$, and repeating fragments—cells of the vortex lattices (rectangles), corresponding to the characteristic features of the function $F_p^{\parallel}(H)$ for sample no. 1. The horizontal size of a cell for the first maximum of $F_p^{\parallel}$ is chosen arbitrarily, and the length of the remaining cells are matched with the sizes of the given cell.

We took account of the following circumstances in constructing the vortex lattices in Fig. 4. It has been shown theoretically [31–33] and experimentally [32, 33] that in a longitudinal field for film samples with thickness t less than the penetration depth $\lambda_{\parallel}$ vortices will interact strongly with the Meissner currents, and this interaction aligns vortices along the center of the sample until their mutual repulsion induces a lateral displacement of vortices and the instability of the one-dimensional configuration, leading with increasing field to the appearance of a two-dimensional collection

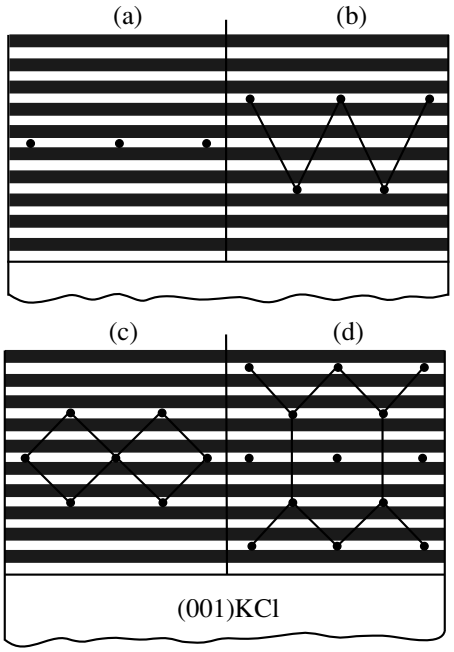


Fig. 5. Schematic illustration of the transverse section of samples by a plane perpendicular to the direction of the longitudinal magnetic field $H^{\|}$, and the arrangement of vortices in the layers of superlattices. The dots represent the centers of the vortices. The dark bands represent the PbTe layers.

of vortices in the form of two parallel rows of vortices (the structure (*b*) in Fig. 4). Such an instability (onset of a transition to a two-dimensional configuration) leads to the appearance of a maximum for the pinning force [31], expulsion of magnetic flux out of the sample [32], and a perpendicular component of the magnetic moment M_z of the sample with a small deviation of the magnetic field from the plane of the sample [33]. The last two effects have been confirmed experimentally on Nb/Cu superlattices [32, 33]. The calculation and experimental data [33] have shown that the distance between the rows of vortices and from each row to the nearest surface of the sample is $t/3$. As the field increases, the two-row configuration of vortices in turn becomes unstable and the restructuring of the vortex lattice continues, which gives a configuration of vortices with distances t/n for a $(n-1)$-row vortex structure [33] (see also [34], where numerical calculations of the deformation of such a vortex structure under the action of a current perpendicular to the magnetic field were performed). The maxima of the pinning force should correspond to the moments when the corresponding configurations become unstable. Since for our samples, apparently, $t \sim \lambda_{\|}$, the above remarks are also applicable to our case. By analogy to the vanishing of the peaks in $M_z(H)$ for Nb/Cu superlattices [33], the number of maxima and points of inflection in the curves $F_p^{\|}(H)$ should decrease as the temperature approaches T_c, but such investigations have not yet been performed.

We note that in the construction of the vortex structures shown in Fig. 4 we immediately switched from a three-row structure of the type *c* to a five-row structure of the type *d*, omitting the four-row stage, which should occur according to the theory of [33]. However, in this theory a superlattice is treated simply as a uniform anisotropic superconductor, neglecting the periodicity of the pinning potential in it. In our case the four-row configuration will not possess a plane of symmetry (parallel to the surface) relative to the center of the sample (as happens for structures *a*, *c*, and *d*, if the PbS layer closest to the KCl substrate is assumed to be a continuation of the substrate), and in order to obtain it the vortices in previously formed rows must be displaced through the superconducting PbTe layers, which is energetically unfavorable (the transitions *b*–*c* and *c*–*d* can be realized without such displacements by introducing additional rows of vortices in the corresponding PbS layers free of vortices; see Fig. 5). In addition, the distance between vortices in a PbS layer in a four-row structure will be 5/4 times shorter than in the corresponding five-row structure, and therefore the electromagnetic interaction energy of the vortices will be much higher in the four-row structure, since the interaction of the vortices lying in the neighboring PbS layers will be screened by the PbTe layers, where the order parameter is mainly concentrated. Consequently, it seems to us that the five-row configuration *d* is more likely.

For sample nos. 2 and 3 (Fig. 3b) the curves $F_p^{\|}(H)$ have several step-like sections, each of which seems to be strongly deformed and degenerated into a minimum–maximum pair. Then the second plateau for these samples evidently corresponds to a nine-row configuration of a vortex lattice, since the same considerations are applicable to the seven-row structure as to the four-row structure.

The variants presented in Fig. 4 of vortex structures are not the only possible ones. For certain layered structures at sufficiently low temperatures so that the order parameter is localized in layers of one type, motion of vortices perpendicular to the layers can be blocked [35]. In this case it can be expected that as the magnetic field increases, the vortices located in one plane will not be able to pass into neighboring planes, and consequently restructuring from the structure *a* into the structure *b* will be impossible. Then, as new vortices are introduced, the number of rows of vortices must increase by at least 2 (for example, the transition *a*–*c*), i.e., the maxima in Fig. 4 should correspond to structures with 1, 3, 5, and 9 rows of vortices, which form ordered symmetric configurations. In this variant the distance between the vortices in the PbS layers will increase substantially with the number of rows of vortices, and consequently when new rows appear, "rarefaction" of the vortices in previously formed rows should occur and vortices should leave the sample. It is difficult to indicate unequivocally which of the variants considered above is realized in our case, though the first variant seems to be more likely.

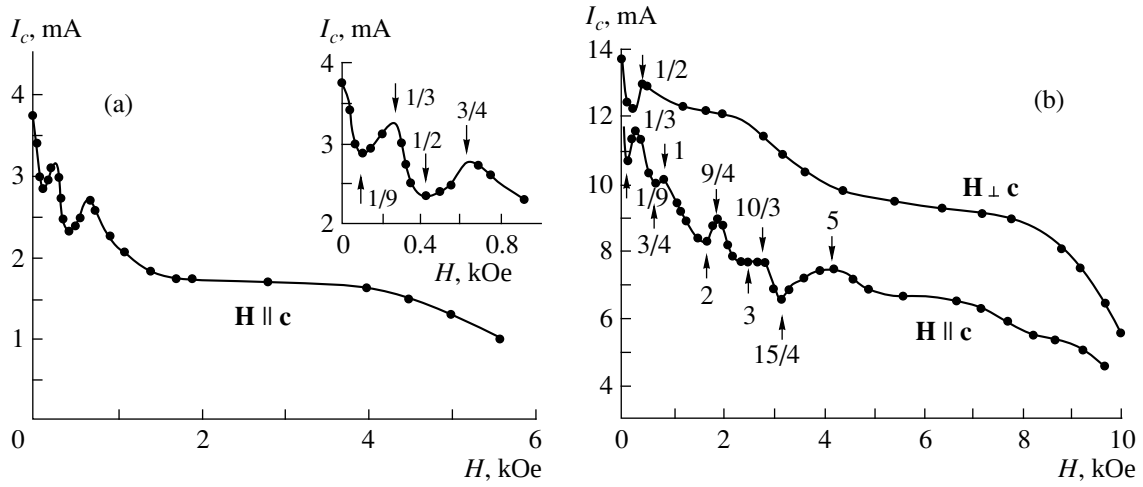


Fig. 6. Critical currents of a 1-μm thick and 0.1-mm wide $YBa_2Cu_3P_{7-\delta}$ epitaxial film (the *c*-axis is perpendicular to the surface) versus the magnetic field at (a) $T = 85$ K and (b) 78 K. Inset: Initial section of the curve on an enlarged scale.

Oscillations of $J_c(H)$ and $F_p(H)$ are observed not only in superlattices but also in HTSC materials, for which, besides the extrema which sometimes appear in relatively strong fields, the appearance of a maximum (or several maxima) in $J_c(H)$ in weak fields is characteristic [36–38]. Our investigations of a 1-μm thick $YBa_2Cu_3O_{7-\delta}$ epitaxial film ($T_c = 88.4$ K), obtained at the Institute of Applied Physics (Nizhniĭ Novgorod) by laser evaporation, the results of which are presented in Fig. 6, also attest to oscillatory behavior of $J_c(H)$ and $F_p(H)$. In contrast to PbTe/PbS superlattices, oscillations of $J_c(H)$ for $YBa_2Cu_3O_{7-\delta}$ films can be seen for orientations of the magnetic field parallel (**H** ⊥ **c**) and perpendicular (**H** ∥ **c**) to the substrate [37].

Analysis of the field dependences of the pinning force for PbTe/PbS superlattices and $YBa_2Cu_3O_{7-\delta}$ films showed that the singular points of $F_p(H)$ are attained in fields $H_{cr} = (p/q)H_0$, where p and q are integers and $H_0 = 838.37$ Oe (see below). In Figs. 1, 3, 4, and 6 the critical fields H_{cr} are marked by arrows with numbers indicating the magnitude of H_{cr} in units of H_0. We note that in the general case the extrema of the function $I_c(H)$ are slightly shifted from their positions corresponding to H_{cr}, since the extrema in $I_c(H)$ and $F_p(H)$ do not coincide with one another. The maxima of the pinning force $F_p^{\perp}(H)$ for superlattices also turned out to be virtually coincident with H_{cr} ($p/q = 1/4$, 1/2, 5/3, and 1/5; see table). The values $p/q = 1/9$, 1/3, 1, and 2 for positions of the extrema of $I_c(H)$ for a $YBa_2Cu_3O_{7-\delta}$ film (Fig. 6) are also encountered in Fig. 3 for PbTe/PbS superlattices. In Fig. 3 of [37] the pinning force for $YBa_2Cu_3O_{7-\delta}$ films in a field parallel to the *c* axis possesses extrema for $p/q = 2$, 7/3, 6, 23/3 (bottom curve) and 7/3, 11/3, 5, and 7 (top curve), while for a $Bi_2Sr_2CaCu_2O_8$ single crystal [38] the first three extrema of J_c with current flowing parallel to the *c* axis (**H** ⊥ **c**) appeared in fields close to $(3/7)H_0$, $(6/7)H_0$, and $(12/7)H_0$. We note that the values $p/q = 6/7$, 2, and 5 coincide with the values which we obtained for superlattices and the $YBa_2Cu_3O_{7-\delta}$ film.

In addition, the previously observed magnetic-field oscillations of the resistance $R(H)$ and the derivative dR/dH for samples of the single-crystal superlattice Bi/Sb on mica with triangular networks of misfit dislocations [39], whose appearance was attributed tentatively to the Aharonov–Bohm effect, also fit very well into the scheme of a discrete set of values of $H_{cr} = (p/q)H_0$. This follows from an analysis of hundreds of oscillations of $R(H)$ and dR/dH in fields up to 50 kOe for samples with different thicknesses of the Bi and Sb layers. Specifically, the first extrema of $R(H)$ in Fig. 1 of [39] appear in a perpendicular magnetic field with $p/q = 2/7$, 8/7, 2, and 15/7.

In analyzing the experimental data, the position of the singular points of $F_p(H)$ was determined more accurately using well-known numerical methods. The magnetic fields obtained in this manner for the singular points of $F_p(H)$ differed from the corresponding values of H_{cr} by an amount not exceeding ~0.2–0.4% of H_{cr}. This discrepancy seems to be due to the error in the measurement of the magnetic fields, the uncertainties introduced by the numerical methods, and the replacement of the magnetic induction by the magnetic field H in the pinning force F_p. The latter force was probably the reason for the appreciable (~3.4%) deviation of the field H_1^{max} for sample no. 1 (see table) from the value $(1/9)H_0$.

The appearance of a discrete ordered set of the magnetic fields in which extrema of the electrophysical characteristics of such diverse objects as a PbTe/PbS and Bi/Sb superlattices and HTSC films are attained shows that in these objects we are dealing with unusual states which, in a certain sense, are universal.

The value of a "quantum" of the magnetic field H_0 was calculated (and then used to analyze the experimental data) from the following considerations. Previously, it was established that the IVCs for island films and tunneling barriers in which metallic island interlayers are imprinted [40] possess singularities in the form of current (or voltage) jumps at voltages on the sample which are multiples of U_0 = 16.25 mV, and a set of inflection points was observed in the temperature dependences of the resistance of metallic island films at temperatures which are multiples of T_0 = 5.238 K ($kT_0 = (1/36)eU_0$), and at such singular temperature points the conductance of the samples and the slopes of the linear sections of the IVCs are quantized in units of e^2/h. The unusual behavior of the temperature dependences of the resistance with extrema and inflection points at $T = nT_0$ (n is an integer) was then also observed for layered compositions consisting of lead chalcogenides [22, 41]. For three-layer PbS–PbTe–PbS films on KCl [42], besides a set of inflection points, giant oscillations of the resistance with the distances between the extrema being multiples of $(1/7)T_0$ were also observed. The singular temperature points in the temperature dependences $R(T)$ appeared at temperatures $(p/q)T_0$ (p and q are integers), specifically, at q = 1–5, 7. The resistance of such 10×1 mm sandwiches at the singular temperature points was found to be equal to one of the quantized values $(m/n)(h/e^2)$. Analysis of our data as well as the published data showed that singular temperature points of the type $T_{cr} = (p/q)T_0$ are also characteristic for the temperature dependences of the superconducting gap [19, 21], the critical current, and the derivative $(1/H_{c2})dH_{c2}/dT$ of PbTe/PbS multilayers [20] and for different physical characteristics of HTSC materials [22].

We note that the critical temperature T_c for superlattices is the same as one of the singular temperature points T_{cr} [22, 42], if T_c is determined according to the inflection point on the curve of the resistive superconducting transition. The values presented for T_c in the table for superlattices nos. 1–4 (3.7, 4.2, 5.2, and 3 K) were determined according to the point at the middle of the resistive transitions. If the inflection point corresponding to the curves $R(T)$ is taken as the criterion T_c, then these values were essentially identical to the singular temperature points, equal to (in units of T_0) 5/7 (3.742 K), 4/5 (4.191 K), 1 (5.238 K), and 4/7 (2.993 K).

The published data on HTSC materials show that the known values of T_c (according to the criterion of zero resistance or center of a superconducting transition) for them also lie near temperatures which are multiples of T_0. For example, the temperatures 89, 110, and 125 K are often encountered. There are reports of even higher critical temperatures: 162 K (zero resistance) [43] and 220 K [44]. These temperatures are close to, respectively, nT_0 with n = 17 (89.05 K), 21 (110 K), 24 (125.72 K), 31 (162.38 K), and 42 (220 K). The deviation of T_c from nT_0 (just as for PbTe/PbS superlattices) seems to be due to the fact that the center of the resistive superconducting transition does not coincide with the inflection point of the corresponding curve $R(T)$ near the transition. Nine peaks in the temperature dependences of the real and imaginary parts of the magnetic susceptibility have been observed in the system TlBaCaCuO [45] at the temperatures of approximately 318, 304, 296, 284, 276, 260, 240, 200, and 130 K; these values are close to the values (in units of T_0) $60\frac{5}{7}$ (318.03 K), 58 (303.81 K), $56\frac{1}{2}$ (295.96 K), $54\frac{1}{5}$ (283.91 K), $52\frac{2}{3}$ (275.89 K), $49\frac{2}{3}$ (260.16 K), $45\frac{4}{5}$ (239.91 K), $38\frac{1}{5}$ (200.1 K), and $24\frac{4}{5}$ (129.91 K). Many other examples of the manifestation of a series of singular temperature points can be presented for various properties of HTSCs, including the critical currents (see, for example, the data in [46] for critical currents, parallel and perpendicular to the c axis, in $Bi_2Sr_2CaCuO_8$ single crystals).

The potential difference U_0 can be represented in the form [40] $U_0 = e/R_0$, where the length R_0 = 88.607 nm is somewhat less than (by 2.8%) the reciprocal of the Rydberg constant $R_\infty = (me^4/4\pi\hbar^3c)$ (R_∞^{-1} = 91.127 nm). The "quantum" of the magnetic field $H_0 = \phi_0/\pi R_0^2$ ($\phi_0 = ch/2e$ is the magnetic flux quantum), whose value was indicated above (838.37 Oe), can be calculated using the length R_0. It deemed it useful to present the computed value of H_0 with two significant digits (though this exceeds the accuracy with which H_0 can be determined, taking into account the measurement error in U_0 [40]), since someday the value of H_0 could be determined more accurately from measurements of the magnetic-field properties of PbTe/PbS superlattices and other objects (for example, Bi/Sb superlattice) in strong fields up to 50–100 kOe, for which the uncertainty in the decimal places for H_0 can lead to an appreciable discrepancy between the experimental and computed values of the positions of the singular points in the corresponding dependences.

We note that the singular points in the field dependence of the pinning force at $H = H_{cr} = (p/q)H_0$ corresponds to "quantization" of the area of the sample $(q/p)\pi R_0^2$ per vortex. Thus, it can be inferred that in PbTe/PbS superlattices and HTSC superconducting films the position of the singular points in the curves $F_p(H)$ is determined by the characteristic properties of the vortex lattice and not by the structural defects present in the sample, though the details of the transition between the singular points should depend on the defect structure of the samples. In polycrystalline HTSC films, the large-angle grain boundaries and other defects will probably play a decisive role in determin-

ing the pinning force, thereby masking the effects observed in the present work.

A different explanation is also possible, in principle, for the appearance of extrema in the magnetic-field dependences of the critical currents and pinning force in HTSCs and PbTe/PbS superlattices. According to this explanation, the maxima of I_c and F_p could be associated with a transition of the system of vortices from one structural and phase state into another, for example, with melting of the vortex lattice, transition of the system of vortices from a three-into a two-dimensional state, and so on (see the review in [47]). For a $YBa_2Cu_3O_{7-\delta}$ film and **H** ∥ **c** at temperatures near T_c, the relation (5.5) from the review [47] can be used to describe the melting curve $B_m(T)$:

$$B_m(T) \approx \beta_m \frac{c_L^4}{\mathrm{Gi}} H_{c2}(0)\left(1-\frac{T}{T_c}\right)^2, \qquad (2)$$

where $\beta_m \approx 5.6$, Gi is the Ginzburg number, $c_L \approx 0.1$–0.2 is the number appearing in the Lindeman criterion for melting, and $H_{c2}(0)$ is the upper critical field determined by linear extrapolation of the section of $H_{c2}(T)$ near T_c to zero temperature. For $YBa_2Cu_3O_{7-\delta}$, for purposes of estimation we can use $\mathrm{Gi} \sim 10^{-2}$, $c_L \approx 0.2$, and then $\beta_m c_L^4/\mathrm{Gi} \sim 1$ [47]. For the $YBa_2Cu_3O_{7-\delta}$ film prepared at the Institute of Applied Physics by the same method as our film, the slope of the upper critical field near T_c was $\left|dH_{c2}^{\perp}/dT\right| = 11.3$ kOe, if T_c is determined according to the center of the resistive superconducting transition [48]. Using this value for our film we obtain $H_{c2}(0) \approx 10^6$ Oe, which gives for $T_c = 88.4$ K the values $B_m(85\ \mathrm{K}) \approx 0.148$ T and $B_m(78\ \mathrm{K}) \approx 1.38$ T. These fields are higher than the fields for the last maxima in Fig. 6 by approximately factors of 2.35 and 3.3, respectively. We note, however, that the published values of $\left|dH_{c2}^{\perp}/dT\right|$ for $YBa_2Cu_3O_{7-\delta}$ single crystals are lower, the lowest, apparently, being 0.46 T/K [49]. Even for such a small slope only the last maximum in Fig. 6a penetrates somewhat into the region beyond the melting curve $B_m(T)$. Such a low value of $\left|dH_{c2}^{\perp}/dT\right|$ is uncharacteristic for epitaxial $YBa_2Cu_3O_{7-\delta}$ films. Consequently, melting of the vortex lattice is hardly possible in the range of fields where maxima of $I_c(H)$ appeared. According to the *B*–*T* phase diagram for layered structures (Fig. 39 in [47]), there are no singular lines below the melting curve near T_c. Consequently, it seems that the appearance of maxima of $I_c^{\perp}(H)$ in our case cannot be explained by any known phase transitions in a vortex structure.

In [50] an *H*–*T* diagram was constructed for a $YBa_2Cu_3O_{7-\delta}$ single crystal for the case **H** ∥ **c**. The curves $H_{c2}(T)$, $H_{c2}^{\max}(T)$, and $H_m(T)$ as well as the magnetic fields $H_{\perp}^{\max}$ in which a maximum of $F_p^{\perp}$ was attained were plotted in the diagram. It was found that the ratio $H_{c2}^{\max}/H_m$ is essentially equal to 2 in the entire temperature range of measurements (from T_c up to 70 K). The curves $f(h)$ in Fig. 4 in [50] for various temperatures have the same characteristic feature as the curve for sample no. 2 in Fig. 2, i.e., in fields above $H_{\perp}^{\max}$ a rapid drop of the pinning force followed by emergence on a plateau-like section, beyond which the rate of decay of $F_p^{\perp}$ increases sharply, was observed. The field H_m lies approximately at the center of the plateau. We note in this connection that the fields in which maxima of $I_c(H)$ were observed in Fig. 6 lie on the ascending section of the curves $F_p^{\perp}(H)$, and they are appreciably less than the field at the absolute maximum $H_{\perp}^{\max}$; this is an additional argument in favor of the absence of melting of a vortex lattice in the range of fields which are of interest to us. By analogy to the $YBa_2Cu_3O_{7-\delta}$ film it can be inferred that this behavior of $f(h)$ for PbTe/PbS superlattices (Fig. 2) is also due to the melting of a vortex lattice. The change in the form of the functions $f(h)$ after the maximum for samples nos. 2, 3, and 4 could also be due to melting processes.

If this hypothesis is adopted for PbTe/PbS superlattices, then the fields at which vortex lattices melt in a longitudinal magnetic field can be estimated using the scaling relation (5.17) from [47]:

$$B_m(T, \theta) = (1/\varepsilon_\theta) B_m^{\perp}(T),$$

where $\varepsilon_\theta^2 = \varepsilon^2 \cos^2\theta + \sin^2\theta$, θ is the angle of inclination of the magnetic field away from the plane of the layers, and $\varepsilon = \xi_{\perp}/\xi_{\parallel} = H_{c2}^{\perp}/H_{c2}^{\parallel}$ is the anisotropy parameter. For our superlattices at $T/T_c \approx 0.5$, we have $1/\varepsilon = 5$–7 [21]. Using the data from the table and taking $h_m \approx 0.3$ for superlattice no. 2 and $h_m \approx 0.5$ for all other superlattices, we obtain for the minimum value $1/\varepsilon = 5$ for fields $H_m^{\parallel}$ the estimates 1.55, 2.55, 10.25, and 1.2 kOe for superlattices nos. 1, 2, 3, and 4, respectively. These values are more than a factor of 2 greater than the fields of the last features in the curves $F_p^{\parallel}(H)$, marked by arrows in Fig. 3; sample 2 is an exception. Although it is not obvious that the scaling relation for $B_m(T, \theta)$ is applicable in our case of a film with thickness less than or of the order of the penetration depth $\lambda_{\parallel}$ of a magnetic field under conditions of strong localization of the order parameter in the PbTe layers ($T/T_c \sim 0.5$), it can probably be assumed that, just as in the case **H** ∥ **c** for the $YBa_2Cu_3O_{7-\delta}$ film, the features in the curves $F_p^{\parallel}(H)$ appear in fields below $H_m^{\parallel}$. Detailed investigations of the magnetic-field dependences of the IVCs at various

temperatures are required in order to clarify the mechanism leading to the formation of the extrema in $F_p^{\parallel}$ and $F_p^{\perp}$.

In connection with the quantization effects described above, we note that in a number of works singularities in the resistivity and anomalies in the dielectric response along the *c* axis at critical values $x = 4^{-n}$ and 2×4^{-n} ($n = 1$, 2, 3) have been observed for $La_{2-x}Sr_xCuO_4$ films (see [51] and the literature cited there). In the opinion of the authors of [51] this attests to Wigner crystallization of hole pairs in two-dimensional conducting layers with macroscopic quantum features similar to the quantum Hall effect. Similar ideas concerning pairing of free carriers and the relation between T_c and their density and the geometric characteristics of the crystal structure of HTSC materials and PbTe/PbS superlattices have been advanced in [21, 22].

4. CONCLUSIONS

The results presented above attest to the fact that certain universal laws governing the behavior of the field and temperature dependences of various properties are manifested in HTSC materials and PbTe/PbS superlattices. The appearance of quantization of the potential difference, resistance, temperature, and magnetic field in objects with different composition and structure, such as, metallic island films, Bi/Sb superlattices, and layered compositions consisting of lead chalcogenides and HTSCs, shows that phenomena of this kind can be observed in a wider class of objects and a wider temperature range than the quantum Hall effect and can hardly be explained on the basis of existing models of the quantum Hall effect and high-temperature superconductivity. The fact that the length R_0, which determines the values of U_0, T_0, and H_0, is close to R_∞^{-1} is probably not accidental. This circumstance could be the key to the explanation of the phenomena described. Careful analysis of the existing data on the behavior of the temperature and field dependences of various properties of HTSC materials from the standpoint of the manifestation of quantization effects in them is required in order to clarify the nature of these effects and the mechanism of high-temperature superconductivity.

ACKNOWLEDGMENTS

We thank P.N. Mikheenko for measuring and interpreting the diamagnetic response of superlattices.

REFERENCES

1. R. Kleiner, F. Steinmeyer, G. Kunkel, and P. Müller, Phys. Rev. Lett. **68**, 2394 (1992); R. Kleiner and P. Müller, Phys. Rev. B **49**, 1327 (1994); R. Kleiner, P. Müller, H. Kohlstedt, *et al.*, Phys. Rev. B **50**, 3942 (1994).
2. D. C. Ling, G. Yong, J. T. Chen, and L. E. Wenger, Phys. Rev. Lett. **75**, 2011 (1995); J. T. Chen, D. C. Ling, and L. E. Wenger, Czech. J. Phys. **46**, Suppl., Part S3, 1257 (1996).
3. P. R. Broussard and H. T. Geballe, Phys. Rev. B **37**, 68 (1988).
4. M. Ikebe, Y. Obi, H. Kujishiro, and H. Fujimori, Czech. J. Phys. **46**, Suppl., Part S2, 719 (1996).
5. Y. Kuwasawa, T. Yamaguchi, T. Tosaka, *et al.*, Physica C (Amsterdam) **169**, 39 (1990); T. Nojima, M. Kinoshita, and Y. Kuwasawa, Physica C (Amsterdam) **206**, 387 (1993); Y. Kuwasawa and T. Nojima, Czech. J. Phys. **46**, Suppl., Part S2, 745 (1996).
6. V. M. Krasnov, V. A. Oboznov, V. V. Ryazanov, and N. F. Pedersen, Phys. Rev. B **50**, 1106 (1994); V. M. Krasnov, N. F. Pedersen, V. A. Oboznov, and V. V. Ryazanov, Phys. Rev. B **49**, 12969 (1994).
7. C. N. Hoff and J. A. Cowen, Czech. J. Phys. **46**, Suppl., Part S2, 723 (1996).
8. W. E. Yetter, E. J. Kramer, and D. G. Ast, J. Low Temp. Phys. **49**, 227 (1982).
9. S. N. Song and J. B. Ketterson, Physica B (Amsterdam) **165/166**, 479 (1990); P. Lobotka, I. Vavra, S. Gazi, and J. Derer, Czech. J. Phys. **46**, Suppl., Part S2, 701 (1996).
10. V. I. Dedyu, A. H. Lykov, and S. L. Prishchepa, Zh. Éksp. Teor. Fiz. **97**, 872 (1990) [Sov. Phys. JETP **70**, 488 (1990)]; A. N. Lykov and V. I. Zdravkov, Pis'ma Zh. Tekh. Fiz. **17** (8), 73 (1991) [Sov. Tech. Phys. Lett. **17**, 306 (1991)].
11. Y. Obi, M. Ikebe, and H. Fujimori, Jpn. J. Appl. Phys. **31**, 1334 (1992); M. Ikebe, Y. Obi, H. Fujishiro, and H. Fujimori, Jpn. J. Appl. Phys. **32**, 55 (1993).
12. I. P. Nevirkovets, H. Kohlstedt, G. Hallmans, and C. Heiden, Supercond. Sci. Technol. **6**, 146 (1993); H. Kohlstedt, G. Hallmans, I. P. Nevirkovets, *et al.*, IEEE Trans. Appl. Supercond. **3**, 2197 (1993).
13. J. M. Murduck, D. W. Capone, II, I. K. Schuller, *et al.*, Appl. Phys. Lett. **52**, 504 (1988); K. E. Gray, R. T. Kampwirth, D. W. Capone, and J. M. Murduck, IEEE Trans. Magn. **25**, 2060 (1989).
14. D. Neerinck, K. Temst, M. Dhalle, *et al.*, Physica B (Amsterdam) **165/166**, 473 (1990); D. Neerinck, K. Temst, M. Baert, *et al.*, Phys. Rev. Lett. **67**, 2577 (1991).
15. H. Raffy, J. C. Renard, and E. Guyon, Solid State Commun. **11**, 1679 (1972); H. Raffy, E. Guyon, and J. C. Renard, Solid State Commun. **14**, 427 (1974).
16. S. J. Pennycook, M. F. Chisholm, D. E. Jesson, *et al.*, Phys. Rev. Lett. **67**, 765 (1991); H. Obara, A. Sawa, and S. Kosaka, Phys. Rev. B **49**, 1224 (1994); H. Obara, M. Anderson, L. Fabrega, *et al.*, Phys. Rev. Lett. **74**, 3041 (1995).
17. M. Vélez, E. M. González, J. I. Martín, and J. L. Vicent, Phys. Rev. B **54**, 101 (1996).
18. R. Gross, A. Gupta, E. Olsson, *et al.*, Appl. Phys. Lett. **57**, 203 (1990).
19. O. A. Mironov, B. A. Savitskiĭ, A. Yu. Sipatov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **48**, 100 (1988) [JETP Lett. **48**, 106 (1988)]; I. K. Yanson, N. L. Bobrov,

L. F. Rybal'chenko, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **49**, 293 (1989) [JETP Lett. **49**, 335 (1989)].

20. O. A. Mironov, S. V. Chistyakov, I. F. Skrylev, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **50**, 300 (1989) [JETP Lett. **50**, 334 (1989)].
21. N. L. Bobrov, L. F. Rybal'chenko, V. V. Fisun, *et al.*, Fiz. Nizk. Temp. **16**, 1531 (1990) [Sov. J. Low Temp. Phys. **16**, 862 (1990)].
22. O. A. Mironov, V. V. Zorchenko, A. Yu. Sipatov, *et al.*, Defect Diffus. Forum **103–105**, 473 (1993).
23. I. M. Dmitrenko, N. Ya. Fogel', V. G. Cherkasova, *et al.*, Fiz. Nizk. Temp. **19**, 747 (1993) [Low Temp. Phys. **19**, 533 (1993)]; N. Ya. Fogel', V. G. Cherkasova, A. Yu. Sipatov, *et al.*, Fiz. Nizk. Temp. **20**, 1142 (1994) [Low Temp. Phys. **20**, 897 (1994)].
24. S. S. Borisova, I. F. Mikhaĭlov, and L. P. Shpakovskaya, Kristallografiya **31**, 651 (1986) [Sov. Phys. Crystallogr. **31**, 384 (1986)].
25. I. M. Zaritskiĭ, A. A. Konchits, S. P. Kolesnik, *et al.*, Sverkhprovodimoct': Fiz., Khim., Tekh. **4**, 1400 (1991); O. A. Mironov, S. V. Chistyakov, L. M. Zaritskiĭ, *et al.*, Physica C (Amsterdam) **180**, 196 (1991).
26. P. N. Mikheenko, O. A. Mironov, S. V. Chistyakov, *et al.*, in *Proceedings of the XXVI All-USSR Conference on Low Temperature Physics, Donetsk, 1990*, Vol. 1, p. 289.
27. P. England, T. Venkatesan, X. D. Wu, and A. Inam, Phys. Rev. B **38**, 7125 (1988).
28. A. M. Campbell and J. E. Evetts, *Critical Currents in Superconductors* (Taylor and Francis, London, 1972; Mir, Moscow, 1975).
29. D. Feinberg and C. Villard, Phys. Rev. Lett. **65**, 919 (1990); V. I. Ivlev, Yu. N. Ovchinnikov, and V. L. Pokrovsky, Mod. Phys. Lett. B **5**, 73 (1991); L. N. Bulaevskii, M. Ledvij, and V. G. Kogan, Phys. Rev. B **46**, 366 (1992).
30. K. Kanoda, H. Mazaki, T. Yamada, *et al.*, Phys. Rev. B **33**, 2052 (1986); K. Kanoda, H. Mazaki, N. Hosoito, and T. Shinjo, Phys. Rev. B **35**, 6736 (1987).
31. S. Takacs, Czech. J. Phys., Sect. B **38**, 1050 (1988).
32. J. Guimpel, L. Civale, F. de la Cruz, *et al.*, Phys. Rev. B **38**, 2342 (1988).
33. S. H. Brongersma, E. Verweij, N. J. Koeman, *et al.*, Phys. Rev. Lett. **71**, 2319 (1993).
34. G. Carneiro, Phys. Rev. B **57**, 6077 (1998).
35. L. S. Levitov, Phys. Rev. Lett. **66**, 224 (1991); D. I. Karasyov and V. L. Pokrovsky, J. Phys.: Condens. Matter **4**, L225 (1992); M. Tachiki, S. Takahashi, and K. Sunaga, Phys. Rev. B **47**, 6095 (1993).
36. J. R. Thompson, J. Brynestad, D. M. Kroeger, *et al.*, Phys. Rev. B **39**, 6652 (1989).
37. S. Y. Dong and H. S. Kwok, Phys. Rev. B **48**, 6488 (1993).
38. K.-H. Yoo, D. H. Na, Y. K. Park, and J. C. Park, Phys. Rev. B **49**, 4399 (1994).
39. B. A. Aminov, S. Sh. Akhmedov, Do Thi Sham, *et al.*, Fiz. Nizk. Temp. **16**, 939 (1990) [Sov. J. Low Temp. Phys. **16**, 548 (1990)].
40. V. V. Zorchenko, V. P. Sapelkin, and A. A. Udovenko, Fiz. Tverd. Tela (Leningrad) **30**, 2349 (1988) [Sov. Phys. Solid State **30**, 1354 (1988)]; Fiz. Tverd. Tela (Leningrad) **32**, 905 (1990) [Sov. Phys. Solid State **32**, 534 (1990)].
41. C. H. Grigorov, V. V. Zorchenko, D. A. Litvinov, and V. P. Sapelkin, Ukr. Fiz. Zh. **35**, 708 (1990); Vopr. At. Nauki Tekh., Ser.: Yad.-Fiz. Issled. (Teor. Éksp.) **4** (12), 88 (1990).
42. A. I. Fedorenko, V. V. Zorchenko, A. Yu. Sipatov, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **41**, 1693 (1999) [Phys. Solid State **41**, 1551 (1999)].
43. R. S. Liu, P. T. Wu, J. M. Liang, and L. J. Chen, Phys. Rev. B **39**, 2792 (1989).
44. H. D. Yostarndt, M. Galffy, A. Freimuth, and D. Wohlleben, Solid State Commun. **69**, 911 (1989).
45. Min-Guang Zhao and Li-Ming Cheng, in *Proceedings of the International Magn. Conference, Dig. INTERMAG'89, Washington, DC, March 28–31, 1989*, p. BQ2.
46. J. H. Wang, D. N. Zheng, G. H. Chen, *et al.*, Mater. Sci. Eng., B **1**, 161 (1988); N. Mros, A. Yurgens, D. Winkler, *et al.*, Czech. J. Phys. **46**, Suppl., Part S3, 1273 (1996).
47. G. Blatter, M. V. Feigel'man, V. V. Geshkenbein, *et al.*, Rev. Mod. Phys. **66**, 1125 (1994).
48. S. V. Gaponov, G. G. Kaminskiĭ, E. B. Klyuenkov, *et al.*, Zh. Éksp. Teor. Fiz. **95**, 2191 (1989) [Sov. Phys. JETP **68**, 1266 (1989)].
49. T. K. Worthington, W. J. Gallagher, and T. R. Dinger, Phys. Rev. Lett. **59**, 1160 (1987).
50. Yuuji Horie, Shun-ichi Kuroumaru, Bai-ru Zhao, *et al.*, Physica C (Amsterdam) **170**, 513 (1990).
51. M. Sugahara, Jpn. J. Appl. Phys. **31**, L324 (1992); M. Sugahara, X.-Y. Han, H.-F. Lu, *et al.*, Czech. J. Phys. **46**, Suppl., Part S2, 1165 (1996); Jpn. J. Appl. Phys. **35**, 1221 (1996).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 1022–1026.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1175–1180.

SOLIDS
Electronic Properties

On the Effect of the Cubic Anharmonicity of the Interatomic Interaction on the Low-Temperature Phase of a Peierls System

A. L. Semenov
Ul'yanovsk State University, Ul'yanovsk, 432700 Russia
e-mail: semenov@quant.ulsu.ru
Received November 17, 1999

Abstract—The low-temperature phase of a Peierls system is studied theoretically taking into account the cubic anharmonicity of the interatomic interaction. It is shown that at a transition into the semiconductor phase a uniform deformation of the system occurs simultaneously with the atoms approaching one another in pairs. The cubic anharmonicity of the interatomic interaction (with a negative anharmonicity constant) produces a large increase in the band gap in the electronic spectrum and the order parameters—the reduced amplitude of the static phonon at the edge of the Brillouin zone and the relative uniform deformation of the atomic chain—of the metal–semiconductor phase transition. An interpretation of the experimental data on the metal–semiconductor phase transition in vanadium dioxide is given on the basis of the results obtained. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is well known that a chain of equidistant interacting atoms, each atom containing one outer electron, at temperature T below a critical value T_0 is unstable with respect to the atoms approaching one another in pairs and the formation of a band gap in the electronic spectrum at the Fermi level [1]. This instability, called in the literature the Peierls instability, has been observed experimentally in a variety of quasi-one-dimensional materials: complexes of platinum with mixed valence [1], TCNQ salts [1], TaS_3 [2, 3], blue bronze $K_{0.3}MoO_3$ [2, 3], $(NbSe_4)_{10/3}I$ [2, 3], $NbSe_3$ [3], VO_2 [4, 5], and others [1–4].

In the theoretical analysis of the Peierls model, the interatomic interaction is usually divided into two components. The first component, which is of a covalent nature, is due to the overlapping of the electronic wave functions of neighboring atoms and is calculated on the basis of a microscopic theory. The second component includes the remaining interatomic interaction and is described phenomenologically by a Taylor series expansion in power of the displacements of the atoms from their equidistant position, the thermodynamically equilibrium state in the high-temperature metallic phase. Ordinarily, quadratic terms are retained in the expansion (harmonic approximation) [1–12]. At the same time, if the displacements of the atoms at a Peierls-type structural phase transition are sufficiently large (for example, for vanadium dioxide the displacements of the vanadium atoms are about 10% of the interatomic distance [4]), the anharmonic terms of the interatomic interaction start to play a substantial role and they can increase the accuracy of the calculations substantially and, in a number of cases, lead to fundamentally new physical results.

In the present work the low-temperature semiconductor phase of a Peierls system is investigated taking into account the cubic anharmonicity of the interatomic interaction. It is shown that together with pairwise approach of the atoms toward one another at a metal–semiconductor phase transition, a uniform deformation of the atomic chain (stretching or compression) also occurs. In addition, on account of the cubic anharmonicity of the interatomic interaction, the displacements of the atoms and the band gap in the electronic spectrum can be much greater than (several-fold) the analogous quantities calculated on the basis of the harmonic approximation; this will give a more adequate description of the existing experimental results.

An interpretation of the experimental data on the metal–semiconductor phase transition in vanadium dioxide is given on the basis of the theory developed [4].

2. ELECTRONIC SPECTRUM OF THE SYSTEM

Let us consider a chain of atoms, each atom containing a single outer electron. We write the Hamiltonian of the electronic subsystem in the tight-binding approximation as [1]

$$H = \sum_n B_{n,n+1}(a_n^+ a_{n+1} + a_{n+1}^+ a_n), \qquad (2.1)$$

where n is the number of the atom in the chain, $B_{n,n+1}$ is the overlap integral of the wavefunctions of neighboring atoms with the numbers n and $n+1$, and a_n^+ and a_n are operators creating and annihilating an electron at the nth atom.

1063-7761/00/9006-1022$20.00

When the distance $r_{n,n+1}$ between neighboring atoms is several times greater than the effective radius R of the atomic wave function of an electron, the overlap integral $B_{n,n+1}$ can be determined approximately by the relation [13]

$$B_{n,n+1} \propto \exp(-r_{n,n+1}/R). \quad (2.2)$$

Having in mind a uniform deformation of the chain and the possibility of pairwise approach of neighboring atoms toward one another, we write the distance between neighboring atoms as

$$r_{n,n+1} = r_0 + R\zeta + (-1)^{n+1}R\xi, \quad (2.3)$$

where r_0 is the interatomic distance in the metallic phase, ζ is a parameter characterizing the change in the length of the chain, and ξ is a parameter characterizing the pairwise approach of the atoms in a chain toward one another (the reduced amplitude of a static phonon at the edge of the Brillouin zone). Substituting the relation (2.3), the overlap integral $B_{n,n+1}$ becomes

$$B_{n,n+1} = b\exp(-\zeta + (-1)^n\xi), \quad (2.4)$$

where b is the overlap integral of the $3d$ wavefunctions of an electron for nearest-neighbor atoms in the metallic phase with $\xi = \zeta = 0$. The phases of the atomic wavefunctions for the Hamiltonian (2.1) are chosen so that b in the expression (2.4) is real and positive.

We shall diagonalize the Hamiltonian (2.1) using Bogolyubov's canonical-transformations method [14]. We switch to collective second-quantization Fermi operators, c_k and c_k^+, using the formula

$$a_n = \frac{1}{\sqrt{N}}\sum_k c_k e^{ikn}, \quad (2.5)$$

where N is the number of atoms in the chain, $k = -\pi + 2\pi l/N$, $l = 1, \ldots, N$, and $c_{k+2\pi} = c_k$. In the new operator representation the Hamiltonian (2.1) becomes

$$H = \sum_k 2be^{-\zeta}(c_k^+ c_k \cosh\xi\cos k + ic_k^+ c_{k-\pi}\sinh\xi\sin k). \quad (2.6)$$

We now perform in the Hamiltonian (2.6) another canonical transformation to the Fermi operators α_k and α_k^+ in accordance with the formula

$$c_k = \frac{\alpha_k + i\varphi_k\alpha_{k-\pi}}{\sqrt{1+\varphi_k^2}}. \quad (2.7)$$

The function φ_k in equation (2.7) is chosen so that in the new variables α_k and α_k^+ the Hamiltonian is diagonal

$$H = \sum_k \varepsilon_k \alpha_k^+ \alpha_k. \quad (2.8)$$

Substituting the expression (2.7) into the Hamiltonian (2.6) and equating the off-diagonal elements to 0, we find φ_k and the dispersion law ε_k:

$$\varphi_k = \frac{\cosh\xi\cos k - \mathrm{sgn}(\cos k)\sqrt{\cos^2 k + \sinh^2\xi}}{\sinh\xi\sin k}, \quad (2.9)$$

$$\varepsilon_k = 2be^{-\zeta}\mathrm{sgn}(\cos k)\sqrt{\cos^2 k + \sinh^2\xi}. \quad (2.10)$$

It is evident from the relation (2.10) that the spectrum ε_k with $\varepsilon \neq 0$ possesses two bands, the lower band being completely filled and the upper band being empty (semiconductor phase) in the ground state. For $\xi = 0$ the spectrum (2.10) is a single half-filled band (metallic phase).

3. EQUILIBRIUM EQUATIONS

We now represent the potential energy U for the interatomic interaction of the chain of atoms as follows:

$$U = \sum_{n=1}^{N-1} W(r_{n,n+1}) + F_e, \quad (3.1)$$

where

$$F_e = \mu N - k_B T\sum_k \ln\left[1 + \exp\left(\frac{\mu - \varepsilon_k}{k_B T}\right)\right] \quad (3.2)$$

is the covalent component of the potential energy and is due to overlapping of the electronic wavefunctions of the neighboring atoms, μ is the chemical potential of the electrons, k_B is Boltzmann's constant, and T is the temperature. The summation over k in the expression (3.2) extends over all possible single-electron states taking spin degeneracy into account. The relation (3.2) is the free energy F_e of the electronic subsystem and depends on the relative arrangement of the atoms in the chain.

We now write the first term on the right-hand side of the expression (3.1), describing the total potential energy of the interatomic interaction of a chain of atoms minus the covalent component (3.2), phenomenologically in terms of a Taylor series expansion in the displacements of the atoms in the chain from their equidistant positions:

$$W(r) = \alpha_1 b\frac{r - r_0}{R} + \frac{\alpha_2 b}{2}\left(\frac{r - r_0}{R}\right)^2 + \frac{\alpha_3 b}{6}\left(\frac{r - r_0}{R}\right)^3, \quad (3.3)$$

where α_i is the dimensionless expansion coefficient of the ith term.

The thermodynamically equilibrium values of the order parameters of the metal–semiconductor phase transition—ξ and ζ—in the mean-field approximation can be determined from the conditions for a minimum of the potential energy U (3.1):

$$\frac{\partial U}{\partial \xi} = 0, \quad \frac{\partial U}{\partial \zeta} = 0. \quad (3.4)$$

Substituting the expressions (3.1) and (3.2) into the relation (3.4), we obtain the equations describing the thermodynamic equilibrium of a Peierls system:

$$b(\alpha_2\xi + \alpha_3\xi\zeta) + \frac{1}{N}\sum_k f_k\frac{\partial\varepsilon_k}{\partial\xi} = 0, \tag{3.5}$$

$$b\left[\alpha_2 + \alpha_2\zeta + \frac{\alpha_3}{2}(\xi^2+\zeta^2)\right] + \frac{1}{N}\sum_k f_k\frac{\partial\varepsilon_k}{\partial\zeta} = 0. \tag{3.6}$$

Here

$$f_k = \left[1 + \exp\left(\frac{\varepsilon_k - \mu}{k_BT}\right)\right]^{-1} \tag{3.7}$$

is the Fermi–Dirac distribution. Substituting the relations (3.7) and (2.10) into equations (3.5) and (3.6) and switching from summation over k to $\int\nu(E)dE$, where

$$\nu(E) = \frac{2N|E|}{\pi\sqrt{(4b^2e^{-2\zeta}\cosh^2\xi - E^2)(E^2 - 4b^2e^{-2\zeta}\sinh^2\xi)}} \tag{3.8}$$

is the electron density of states corresponding to the electronic spectrum (2.10), we find

$$\alpha_2\xi + \alpha_3\xi\zeta - \frac{2GB}{\pi b}\int_G^B \frac{\tanh(E/2k_BT)dE}{\sqrt{(B^2-E^2)(E^2-G^2)}} = 0, \tag{3.9}$$

$$\alpha_1 + \alpha_2\zeta + \frac{\alpha_3}{2}(\xi^2+\zeta^2) + \frac{2}{\pi b}\int_G^B \frac{E^2\tanh(E/2k_BT)dE}{\sqrt{(B^2-E^2)(E^2-G^2)}} = 0. \tag{3.10}$$

For convenience, the following notation has been introduced in equations (3.9) and (3.10):

$$G \equiv 2be^{-\zeta}\sinh\xi, \tag{3.11}$$

$$B \equiv 2be^{-\zeta}\cosh\xi. \tag{3.12}$$

We shall determine the constants α_1 and α_2 in the expansion (3.3), which appear in the equations (3.9) and (3.10) describing the thermodynamic equilibrium of a Peierls system, on the basis of the following conditions. From equation (3.10) and the condition

$$\xi(T_0) = \zeta(T_0) = 0$$

we have

$$\alpha_1 = -\frac{2}{\pi b}\int_0^{2b} \frac{E\tanh(E/2k_BT_0)}{\sqrt{4b^2-E^2}} = dE. \tag{3.13}$$

Calculating the integral in equation (3.13), we obtain approximately

$$\alpha_1 = -\frac{4}{\pi}\left[1 - \frac{1}{6}\left(\frac{k_BT_0}{b}\right)^2\right]. \tag{3.14}$$

The critical temperature T_0 of a Peierls-type structural phase transition in the mean-field approximation is found from the equation

$$\left.\frac{\partial^2U}{\partial\xi^2}\right|_{\xi=\zeta=0,\,T=T_0} = 0. \tag{3.15}$$

Hence we obtain, using the relations (3.1)–(3.3) (see also equation (3.9)), an expression for the constant α_2:

$$\alpha_2 = \frac{8b}{\pi}\int_0^{2b} \frac{\tanh(E/2k_BT_0)}{E\sqrt{4(b^2-E^2)}}dE. \tag{3.16}$$

Calculating the integral in equation (3.16), we obtain approximately

$$\alpha_2 = \frac{4}{\pi}\ln\frac{2be}{k_BT_0}. \tag{3.17}$$

Thus, we have obtained the equations (3.9) and (3.10) describing the thermodynamic equilibrium which form a closed system for determining the temperature dependences of the order parameters $\xi(T)$ and $\zeta(T)$ of a Peierls system, taking into account the cubic anharmonicity of the interatomic interaction. The constants α_1 and α_2 in the expansion (3.3), appearing in the equations of thermodynamic equilibrium (3.9) and (3.10) of a Peierls system, are determined using equations (3.14) and (3.17).

4. THE ORDER PARAMETERS ξ AND ζ OF A METAL-SEMICONDUCTOR PHASE TRANSITION IN A PEIERLS SYSTEM AT LOW TEMPERATURES

We shall analyze the equations of thermodynamic equilibrium (3.9) and (3.10) for the order parameters ξ and ζ of a metal–semiconductor phase transition for low temperatures, where

$$k_BT \ll be^{-\zeta}\sinh\xi. \tag{4.1}$$

Then, we can set in equations (3.9) and (3.10)

$$\tanh\frac{E}{2k_BT} \simeq 1. \tag{4.2}$$

Calculating the integrals in equations (3.9) and (3.10), we obtain approximately

$$\alpha_2\xi + \alpha_3\xi\zeta - \frac{4e^{-\zeta}\sinh\xi}{\pi}K\left(\frac{1}{\cosh\xi}\right) = 0, \tag{4.3}$$

$$\alpha_1 + \alpha_2\zeta + \frac{\alpha_3}{2}(\xi^2 + \zeta^2) + \frac{4e^{-\zeta}\cosh\xi}{\pi}E\left(\frac{1}{\cosh\xi}\right) = 0, \quad (4.4)$$

where $K(x)$ and $E(x)$ are, respectively, the complete normal Legendre elliptic integrals of the first and second kinds. Expanding the elliptic integrals $K(x)$ and $E(x)$ in series near the point $x = 1$ [15], and $\sinh\xi$ and $\cosh\xi$ in a Taylor series at the point $\xi = 0$, we find from equations (4.2) and (4.3) in the approximation $\xi < 1$

$$\alpha_2 + \alpha_3\zeta - \frac{4e^{-\zeta}}{\pi}\left[\ln\left(\frac{4}{\xi}\right)\left(1 + \frac{2}{3}\xi^2\right) - \frac{\xi^2}{6}\right] = 0, \quad (4.5)$$

$$\alpha_1 + \alpha_2\zeta + \frac{\alpha_3}{2}(\xi^2 + \zeta^2) + \frac{4e^{-\zeta}}{\pi} \times \left\{1 + \frac{\xi^2}{2}\left[\ln\left(\frac{4}{\xi}\right) + \frac{1}{2}\right]\right\} = 0. \quad (4.6)$$

Eliminating from equations (4.5) and (4.6) the unknown coefficient α_3, we obtain, specifically,

$$\alpha_1 + \alpha_2\zeta + \frac{4e^{-\zeta}}{\pi}\left\{1 + \frac{\xi^2}{2}\left[\ln\left(\frac{4}{\xi}\right) + \frac{1}{2}\right]\right\} + \frac{\xi^2 + \zeta^2}{2\zeta} \times \left\{\frac{4e^{-\zeta}}{\pi}\left[\ln\left(\frac{4}{\xi}\right)\left(1 + \frac{2}{3}\xi^2\right) - \frac{\xi^2}{6}\right] - \alpha^2\right\} = 0. \quad (4.7)$$

The formula (4.7) is a relation between the order parameters ζ and ξ of a metal–semiconductor phase transition at low temperatures T, satisfying the condition (4.1).

5. NUMERICAL ESTIMATES, COMPARISON WITH EXPERIMENT, AND DISCUSSION OF THE RESULTS

We shall compare with experiment for vanadium dioxide, which possesses a one-dimensional electronic conduction band. This band is formed as a result of the overlapping of the 3*d* electronic wave functions of the vanadium atoms arranged in chains parallel to the *c* crystal axis [4]. In the high-temperature metallic phase the width of the one-dimensional electronic conduction band is $E_0 = 4b \simeq 1.1$ eV, and the distance between the nearest-neighbor vanadium atoms along the chain is $r_0 = 0.285$ nm. A metal– semiconductor phase transition is observed in VO_2 at the critical temperature $T_0 = 340$ K. The band gap in the electronic spectrum in the low-temperature phase is $E_g \simeq 0.6$eV and the distances between the nearest-neighbor vanadium atoms along the chain alternate: $r_1 = 0.265$ nm and $r_2 = 0.312$ nm. Since $r_1 + r_2 > 2r_0$, the vanadium atoms are displaced relative to the axis of the chain, and the chain itself has a zig-zag shape.

We shall calculate the effective radius R of the atomic 3*d* electronic wavefunction using the experimental data. We obtain from the relations (2.3) and (2.10)

$$\frac{E_g}{E_0} = \exp\left(-\frac{r_2 - r_1}{2R}\right)\sinh\left(\frac{r_2 - r_1 - 2r_0}{2R}\right). \quad (5.1)$$

Hence we find, using the experimental data for vanadium dioxide ($T_0 = 340$ K, $E_g = 0.6$ eV, $E_0 = 4b = 1.1$ eV, $r_0 = 2.85$ Å, $r_1 = 2.65$ Å, $r_2 = 3.12$ Å), the approximate value $R \simeq 0.41$ Å. This result agrees with the general notion that the localization radius of the *d* and *f* wavefunctions of transition-metal compounds is small [4].

Using the relation (2.3) we calculate the experimental values of the order parameters ξ and ζ of the system to be

$$\xi = \frac{r_2 - r_1}{2R} \simeq 0.57, \quad (5.2)$$

$$\zeta = \frac{r_2 + r_1 - 2r_0}{2R} \simeq 0.09. \quad (5.3)$$

Substituting in equations (3.14) and (3.17) the characteristic numerical values of the parameters for VO_2, we find the dimensionless coefficients α_1 and α_2 in the expansion (3.3) of the repulsive part of the potential energy of the interatomic interaction in the Taylor series in the displacements of the atoms from their equidistant position in the chain:

$$\alpha_1 \simeq -1.27, \quad \alpha_2 \simeq 5.02. \quad (5.4)$$

Using the experimental value of ξ (5.2) and the values of the parameters α_1 and α_2 (5.4) of the theory and using equations (4.5)–(4.7) we obtain the theoretical values of the order parameter ζ of the metal–semiconductor phase transition and the cubic anharmonicity coefficient α_3:

$$\zeta \simeq 0.38, \quad \alpha_3 \simeq -7.90. \quad (5.5)$$

Thus, the theoretical value of the parameter ζ (5.5) characterizing the elongation of a chain of atoms is the same as the experimental value (5.3) in order of magnitude. The difference seems to be due to the small number of terms in the expansion (3.3) of the repulsive part of the interatomic interaction energy and the influence of vanadium atoms from neighboring chains and the nearby oxygen atoms on the chain of the vanadium atoms in VO_2.

Substituting in equation (4.1) the numerical values characteristic for vanadium dioxide, we find the condition of applicability of the approximation (4.2) used in the derivation of equations (4.3)–(4.7):

$$T \ll 1000 \text{ K}.$$

We note that the coefficients α_2 and α_3 in the expansion (3.3) can be estimated independently using the formula

$$\alpha_2 \simeq \frac{ER^2 r_0}{b}, \tag{5.6}$$

where E is Young's modulus of the crystal along the c crystallographic axis, and the relation [16]

$$\alpha_3 \approx -\frac{2\alpha_T r_0 \alpha_2^2 b}{k_B R}, \tag{5.7}$$

where α_T is the linear thermal expansion coefficient of the material along the c crystallographic axis.

Substituting in the expressions (5.6) and (5.7) the characteristic numerical values $E = 3 \times 10^{12}$ din/cm^2 and $\alpha_T \approx 10^{-5}$ we obtain the numerical values of the coefficients: $\alpha_2 \simeq 5$ and $\alpha_3 \simeq -10$, which are close to the corresponding values (5.4) and (5.5) calculated on the basis of the theory examined in this paper.

CONCLUSIONS

In summary, in the present paper a one-dimensional model, taking into account the cubic anharmonicity of the interatomic interaction, was proposed for a Peierls-type metal–semiconductor phase transition. It was shown that at a phase transition into the semiconductor state one-dimensional deformation (elongation or compression) of the chain of atoms is also observed together with the pairwise approach of the atoms in a chain toward one another (dimerization).

An interpretation based on the results obtained was given for an experiment on the metalsemiconductor phase transition in vanadium dioxide [4]. The numerical values of the effective radius R of the atomic $3d$ electronic wavefunction, the reduced parameter ζ characterizing the elongation of the atomic chain, the dimensionless coefficient of elasticity α_2, and the cubic anharmonicity constant α_3, all calculated on the basis of the proposed theory, agree with the experimental data and with independent estimates.

We note that for most known quasi-one-dimensional materials (TCNQ salts, TaS_3, $TaSe_3$, NbS_3, $NbSe_3$, $K_{0.3}MoO_3$, $(NbSe_4)_{10/3}I$, $(NbSe_4)_2I$, and others [3]) the experimentally measured value of the band gap $E_g(T \ll T_0)$ in the electronic spectrum is several-fold greater than the value computed on the basis of the harmonic and mean-field approximations [3]:

$$E_g(T \ll T_0) \simeq 3.5 k_B T_0. \tag{6.1}$$

Taking into account the effect of fluctuations (which goes beyond the scope of the mean-field approximation) improves the agreement between theory and experiment [17]. But it seems that this does not completely solve the problem, since even a weak nonuniformity of the system leads to a strong suppression of the fluctuations [1]. In the present paper it was shown that another important factor, which can increase E_g several-fold compared with the value (6.1), is the cubic anharmonicity of the interatomic interaction (going beyond the scope of the harmonic approximation).

Cubic anharmonicity of the interatomic interaction also engenders a uniform elongation of a Peierls system at a transition into the low-temperature semiconductor phase. As a result, the unit cell of a three-dimensional crystal, consisting of a collection of mutually parallel Peierls atomic chains in a matrix consisting of atoms of a different kind, undergoes substantial deformation, specifically, rhombohedral [18] or monoclinic [4] distortion. In this case Peierls chains themselves can assume a zig-zag shape [4].

REFERENCES

1. L. N. Bulaevskiĭ, Usp. Fiz. Nauk **115**, 263 (1975).
2. S. N. Artemenko, A. F. Volkov, and S. V. Zaĭtsev-Zotov, Usp. Fiz. Nauk **166**, 434 (1996) [Phys. Usp. **39**, 403 (1996)].
3. G. Grüner, Rev. Mod. Phys. **60**, 1129 (1988).
4. A. A. Bugaev, B. P. Zakharchenya, and F. A. Chudnovskiĭ, in *Phase Transition Matal–Semiconductor and Its Application* (Nauka, Leningrad, 1979), p. 28.
5. A. L. Semenov, Fiz. Tverd. Tela (St. Petersburg) **36**, 1974 (1994) [Phys. Solid State **36**, 1079 (1994)].
6. K. F. Berggren and B. A. Huberman, Phys. Rev. B **18**, 3369 (1978).
7. I. M. Suslov, Pis'ma Zh. Éksp. Teor. Fiz. **39**, 547 (1984) [JETP Lett. **39**, 670 (1984)].
8. V. B. Stopachinskiĭ and I. M. Suslov, Zh. Éksp. Teor. Fiz. **91**, 314 (1986) [Sov. Phys. JETP **64**, 183 (1986)].
9. A. L. Semenov, Fiz. Tverd. Tela (St. Petersburg) **40**, 2113 (1998) [Phys. Solid State **40**, 1915 (1998)].
10. A. L. Semenov, Zh. Éksp. Teor. Fiz. **111**, 1398 (1997) [JETP **84**, 774 (1997)].
11. A. L. Semenov, Zh. Éksp. Teor. Fiz. **114**, 1407 (1998) [JETP **87**, 764 (1998)].
12. A. L. Semenov, Zh. Éksp. Teor. Fiz. **115**, 1297 (1999) [JETP **88**, 716 (1999)].
13. O. Madelung, *Introduction to Solid State Theory* (Springer-Verlag, Berlin, 1978; Nauka, Moscow, 1985).
14. N. N. Bogolyubov and N. N. Bogolyubov, Jr., *Introduction to Quantum Statistical Mechanics* (Nauka, Moscow, 1984; World Scientific, Singapore, 1982).
15. H. B. Dwight, *Tables of Integrals and Other Mathematical Data* (Macmillan, London, 1961; Nauka, Moscow, 1983).
16. A. I. Ansel'm, in *Introduction to the Theory of Semiconductors* (Nauka, Moscow, 1978), p. 191.
17. M. V. Sadovskiĭ, Fiz. Tverd. Tela (Leningrad) **16**, 2504 (1974) [Sov. Phys. Solid State **16**, 1632 (1974)].
18. B. A. Volkov and L. A. Fal'kovskiĭ, Zh. Éksp. Teor. Fiz. **85**, 2135 (1983) [Sov. Phys. JETP **58**, 1239 (1983)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 1027–1034.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1181–1189.

SOLIDS
Electronic Properties

Kinetic Effects in an $La_{0.8}Ba_{0.2}MnO_3$ Single Crystal

N. G. Bebenin*, **, R. I. Zaĭnullina*, V. V. Mashkautsan*, V. S. Gaviko*, V. V. Ustinov*, Ya. M. Mukovskiĭ*, and D. A. Shulyatev*****

**Institute of Metal Physics, Ural Division, Russian Academy of Sciences, Yekaterinburg, 620219 Russia*
***e-mail: bebenin@imp.uran.ru*
****Moscow Institute of Steel and Alloys, Moscow, 117936 Russia*
Received December 2, 1999

Abstract—Results are presented of a complex study of the magnetic and resistive properties, the Hall effect, the thermal emf, and the longitudinal Nernst–Ettingshausen effect of an $La_{0.8}Ba_{0.2}MnO_3$ single crystal at temperatures between 77 and 400 K. A maximum was observed near the Curie temperature T_c on the temperature dependences of the resistivity, the thermal emf, and the normal Hall coefficient. It was established that the Hall mobility remains constant near T_c. It is shown that these anomalies in the kinetic properties are attributable to a change in the position of the mobility edge relative to the Fermi level. A semiphenomenological theory is put forward to quantitatively describe the temperature and magnetic-field dependences of the resistivity and thermal emf of lanthanum manganites near the phase-transition temperature. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Lanthanum manganites $La_{1-x}D_xMnO_3$, where D = Ca, Sr, Ba, are attracting interest because of the colossal magnetoresistance (CMR) effect, see the reviews [1–4]. The initial oxide $LaMnO_3$ is an antiferromagnetic insulator. Doping with divalent metals changes the antiferromagnetic ordering to ferromagnetic and lowers the resistivity. At a certain concentration x_c, which depends on the type of divalent ion and the heat treatment, a metal–insulator concentration transition takes place with the result that at $T < T_c$ (T_c is the Curie temperature) the conductivity becomes metallic. In the paramagnetic phase however, the temperature dependence of the resistivity ρ usually remains semiconducting (with $d\rho/dT < 0$) even for $x > x_c$; in other words, a metal–insulator temperature transition takes place near T_c. Metallic conductivity at all temperatures has only been observed in $La_{1-x}Sr_xMnO_3$ single crystals for $0.3 \le x \le 0.4$ [5].

In manganites having a divalent ion concentration close to x_c (both with $x < x_c$ and $x \ge x_c$) a sharp resistivity peak is observed near T_c and which is reduced and shifted toward higher temperatures on application of a magnetic field. The temperature T_R at which the resistivity peak is observed is frequently considered to be the metal–insulator transition temperature because at $T = T_R$ "metallic" behavior (with $d\rho/dT > 0$) is replaced by semiconducting behavior. The magnetoresistance $\Delta\rho/\rho$ also has a sharp peak near T_c and may attain extremely high ("colossal" values) if the magnetic field is fairly strong. There are reports of similar singularities on the temperature dependences of the thermal emf, see for example [6, 7]. The nature of these anomalies has yet to be clarified despite numerous studies of CMR materials. Following the publication of [8] the CMR effect is frequently attributed to the existence of strong electron–phonon interaction leading to the formation of small-radius polarons. Many authors stress the inherent tendency of manganites to form inhomogeneous states and to undergo phase separation [1, 9–13]. Some studies (for example, [14, 15]) draw attention to the existence of strong disorder in manganites, both nonmagnetic and magnetic, the latter increasing abruptly near the Curie point. Existing data on the properties of CMR manganites cannot be used to distinguish between the different scenarios since only results of measurements of the resistivity and magnetoresistance are given in most cases and in only a few cases is the thermal emf given. Data on the Hall effect [16–30] mainly refer to thin-film samples [16–25] and in many cases, the method of determining the normal Hall coefficient from results of measurements of the Hall resistivity ρ_H is of dubious validity In [26] the Hall effect was studied in $La_{1-x}Sr_xMnO_3$ single crystals with $x > x_c = 0.17$ but outside the range where CMR is observed. Studies of the magnetic properties are confined to determining the Curie temperature and in many cases the question of the interrelationship between the magnetic and kinetic characteristics is not even addressed.

In the present paper we report results of a complex study of the magnetic, electrical, galvanomagnetic, and thermomagnetic properties of $La_{0.8}Ba_{0.2}MnO_3$ single crystals. This crystal was selected for the study because of the lack of data on the properties of single crystals of this type of manganite. The results suggest various conclusions which in our view are important for under-

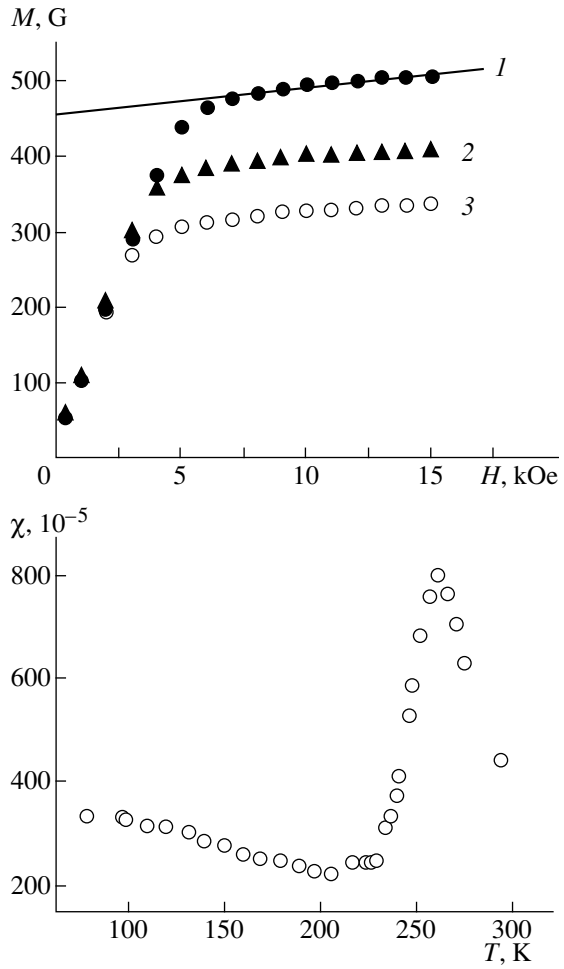


Fig. 1. (a) Magnetization isotherms of an $La_{0.8}Ba_{0.2}MnO_3$ single crystal at T = (*1*) 78 K, (*2*) 206 K, and (*3*) 231 K. (b) Temperature dependence of the susceptibility of the paraprocess.

standing the transport processes in all manganites with colossal magnetoresistance.

2. SAMPLES AND METHOD

Single crystals of $La_{0.8}Ba_{0.2}MnO_3$ were grown by the floating zone method described in [31]. Samples for X-ray analyses and measurements of the magnetic and kinetic properties were cut from the ingot.

The crystal structure and composition were investigated using a DRON-3 diffractometer using Cr K_α radiation. The temperature dependence of the X-ray spectra was studied in the range 80–300 K under heating. The spectra obtained at room temperature were processed using the FullProf code [32].

The resistivity ρ, magnetoresistance $\Delta\rho/\rho$, thermal emf coefficient S, magnetothermal emf (longitudinal Nernst–Ettingshausen effect), and the Hall resistivity ρ_H were measured using the same sample in the form of an $8 \times 3 \times 0.6$ mm wafer whose long side was cut parallel to the rhombohedral c-axis. The magnetization curves were recorded using a vibrating magnetometer in a magnetic field having the same direction as in the measurements of the Hall effect. The resistivity of the sample was measured by a standard four-contact method. The thermal emf was measured at a temperature difference of around 3 K produced by a heater positioned near one edge of the sample. The temperatures of the sample ends were monitored using two copper–constantan thermocouples. The thermal emf and the Hall emf were measured by a potentiometric method. In order to eliminate side effects the Hall emf was measured for two directions of the magnetic field and the current in the sample. In all cases, a magnetic field of up to 15 kOe was directed perpendicular to the plane of the wafer. Indium contacts were deposited on the samples using an ultrasonic soldering iron.

The sample was cooled to the required temperature without applying a magnetic field.

3. RESULTS OF MEASUREMENTS

At temperatures $T < 185$ K the diffraction patterns of $La_{0.8}Ba_{0.2}MnO_3$ only exhibited lines of the orthorhombic phase *Pbnm*, at $T > 196$ K only lines of the rhombohedral phase $R\bar{3}c$ were present, and at 185 K < T < 196 K lines of both phases coexisted. Thus, in this particular manganite a first-order structural phase transition takes place near T = 190 K. The magnetic and kinetic properties exhibited hysteresis characteristics in this range and these are analyzed in detail in [33]. However, the change in the crystal structure has no significant influence on the general profile of the temperature dependences of the magnetization and kinetic coefficients and consequently any changes associated with the structural transition will subsequently be neglected.

The general form of the magnetization curves is typical of ferromagnets, see Fig. 1. The Curie temperature T_c determined by the method of thermodynamic coefficients is 251 K. The paramagnetic temperature determined from the curve giving the temperature dependence of the reciprocal susceptibility in the paramagnetic range was considerably higher: Θ = 260 K, which indicates that this single crystal possesses magnetic inhomogeneity.

The spontaneous magnetization $M_s(T)$ and the susceptibility of the paraprocess $\chi(T)$ were determined by linearly extrapolating the magnetization curve in the range of fields between 7 and 15 kOe to $H = 0$. Treating the temperature dependence $M_s(T)$ as following the law $M_s(T) = M_s(0) - \text{const}\,T^{3/2}$ yielded the estimate $M_s(0) \approx$ 490 G which corresponds to the magnetic moment of the manganese ion $3.3\mu_B$ and is considerably lower than the value of $3.8\mu_B$ obtained from the doping level. The temperature dependence $\chi(T)$ has a well-defined peak at $T = \Theta$. However, instead of a drop in the susceptibility with decreasing temperature typical of ferromag-

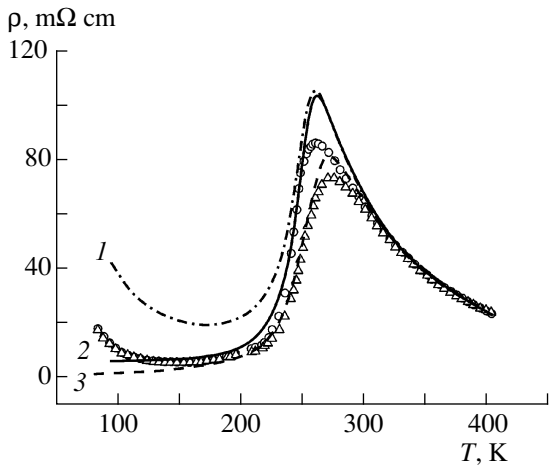


Fig. 2. Temperature dependences of the resistivity ρ of an $La_{0.8}Ba_{0.2}MnO_3$ single crystal at $H = 0$ (○) and $H =$ 10 kOe (△). Curves *1–3* give the calculated values: (*1*) ρ(*T*) for $H = 0$ neglecting the inhomogeneous distribution of the magnetization; (*2*) and (*3*) ρ(*T*) for $H = 0$ and $H = 10$ kOe, respectively allowing for the inhomogeneous distribution of the magnetization.

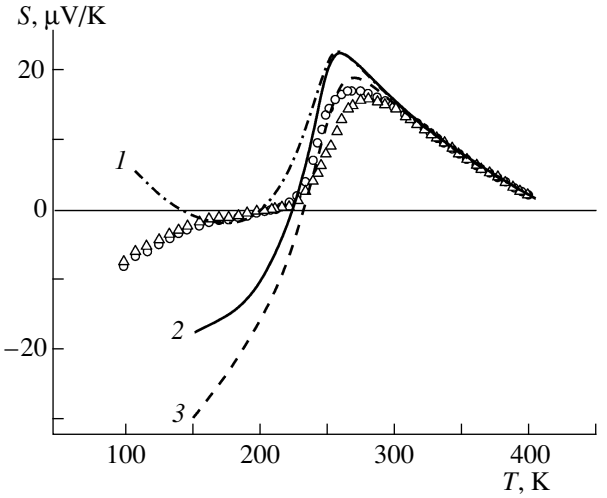


Fig. 3. Temperature dependences of the thermal emf coefficient *S* of an $La_{0.8}Ba_{0.2}MnO_3$ single crystal at $H = 0$ (○) and $H = 10$ kOe (△). Curves *1–3* give the calculated values: (*1*) *S*(*T*) for $H = 0$ neglecting the inhomogeneous distribution of the magnetization; (*2*) and (*3*) *S*(*T*) for $H = 0$ and $H = 10$ kOe, respectively allowing for the inhomogeneous distribution of the magnetization.

nets, the curve $\chi(T)$ exhibits a continuous increase in susceptibility in the range $T < 200$ K. Consequently, in addition to ferromagnetic regions, the sample also contains nonferromagnetic regions which determine the temperature dependence of χ at $T < 200$ K. In order to estimate the volume of these regions we plotted the curves $M(H^{-1})$ and found that the saturation magnetization M_{sat} is of the order of 560 G, i.e., $3.8\mu_B$ per Mn ion. Thus, the volume of the nonferromagnetic regions is around 10% of the total sample volume. A similar estimate was obtained previously (by a different method) for an $La_{0.9}Sr_{0.1}MnO_3$ single crystal [34].

Figure 2 gives temperature dependences of the resistivity ρ(*T*) at $H = 0$ and $H = 10$ kOe. As the temperature increases from 80 to 150 K, the resistivity decreases and then acquires a metallic character ($d\rho/dT > 0$). From 210 K the resistivity increases steeply and at $T_R = 260$ K reaches a maximum after which it becomes "semiconducting" ($d\rho/dT < 0$). The magnetic field lowers the resistivity and shifts its peak toward higher temperatures.

Figure 3 shows temperature dependences of the thermal emf in zero field *S*(0) and in a field $H = 10$ kOe, *S*(*H*). At low temperatures both coefficients are negative and their modulus decreases with increasing temperature. At $T \approx 210$ K the coefficients *S*(0) and *S*(*H*) simultaneously change sign and begin to increase rapidly with increasing temperature, where $S(0) > S(H)$. The coefficient *S*(0) reaches a maximum at $T \approx 270$ K and then begins to decrease smoothly.

Figure 4 gives field dependences of the Hall resistivity $\rho_H = R_0B + R_sM$ for various temperatures, where R_0 and R_s are the normal and anomalous (spontaneous) Hall coefficients, *B* is the induction in the sample, and *M* is the magnetization [35]. For this particular sample geometry the demagnetization factor is close to one so that we can assume $B = H$. In the ferromagnetic region in weak fields we observe an abrupt dependence of ρ_H on *H* whereas for $H > 6$ kOe the dependence $\rho_H(H)$ is linear. Figure 5 gives temperature dependences of the coefficients R_0 and R_s calculated using the curves $\rho_H(H)$ and *M*(*H*) as in our earlier studies [27–30]. At low temperatures the normal Hall coefficient is negative. At $T =$ 200 K the coefficient R_0 becomes positive, increases rapidly with temperature, and reaches a maximum at $T = 261$ K. The anomalous Hall coefficient is negative at all temperatures; no extremum was observed on its temperature dependence.

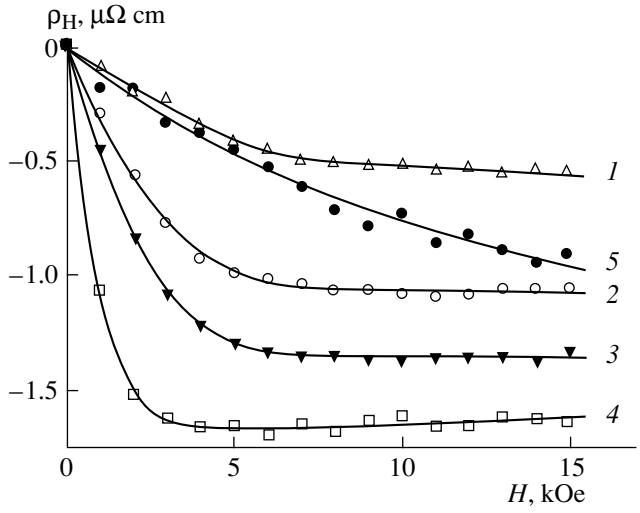


Fig. 4. Isotherms of the Hall resistivity ρ_H at $T =$ (*1*) 96 K, (*2*) 206 K, (*3*) 224 K, (*4*) 241 K, and (*5*) 279 K.

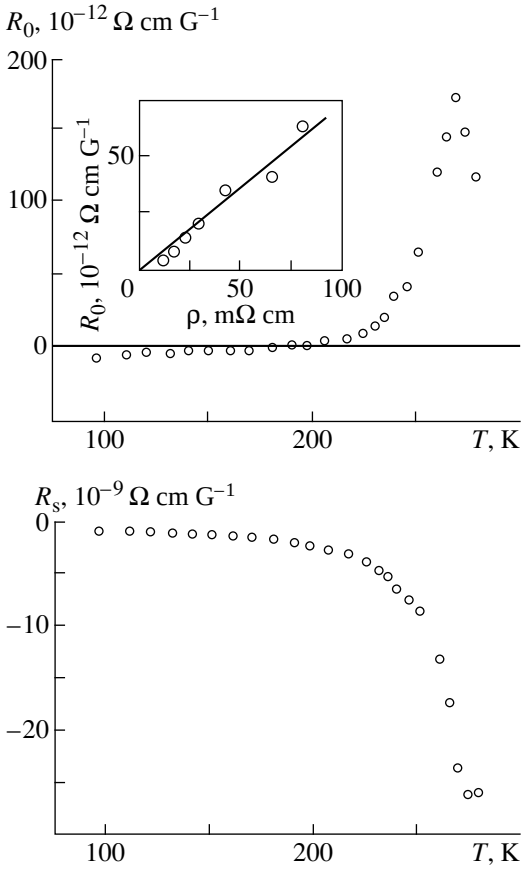


Fig. 5. Temperature dependences of the normal (R_0) and anomalous (R_s) Hall coefficients. The inset shows the dependence of the normal Hall coefficient on the resistivity ρ.

4. THEORY

We shall now propose a theoretical model. We shall start from the assumption that any change in the kinetic characteristics of manganites with $x \sim x_c$ may be described as the result of a change in the position of the hole mobility edge ε_c relative to the Fermi hole energy ε_F. The value of ε_c is determined by the disorder in the system which may be either magnetic, i.e., determined by spin fluctuations or nonmagnetic caused by the presence of impurity atoms, vacancies, and so on. It was shown in [14, 36] that in the narrow-gap *s–d* model, which is now usually called the double exchange model, the growth of fluctuations as the phase transition point is approached from the low-temperature region leads to an appreciable shift of ε_c toward higher hole energies. If the mobility edge intersects the Fermi level, a metal–insulator transition takes place.

A simple semiphenomenological theory was proposed in [15, 37] to describe the change in the position of the mobility edge near the Curie point. It is assumed that near T_c when the relative magnetization m is low, ε_c is a linear function of m^2: $\varepsilon_c = \Delta_0 - \Delta_1 m^2$, where Δ_0 and Δ_1 are constants to be determined experimentally. Following [38], we write the conductivity σ in the form

$$\sigma = \int_{\varepsilon_c}^{\infty} \left(-\frac{\partial f}{\partial \varepsilon}\right) \sigma(\varepsilon) d\varepsilon, \qquad (1)$$

where $f(\varepsilon)$ is the Fermi function. It was noted in [39] that if ε_c is close to ε_F, we can assume that $\sigma(\varepsilon) \approx \sigma(\varepsilon_c) = \sigma_m$, where σ_m is the minimum metallic conductivity so that $\sigma \approx \sigma_m f(\varepsilon_c)$. Consequently, any change in the conductivity is caused by changes in the carrier concentration in delocalized states. For the resistivity near T_c we obtain

$$\rho = \rho_0[\exp(E^\rho/T) + 1], \qquad (2)$$

where $\rho_0 = \sigma_m^{-1}$, $E^\rho = \varepsilon_c - \varepsilon_F = \Delta_0 - \varepsilon_F - \Delta_1 m^2$. If $\Delta_0 - \varepsilon_F > 0$ in the paramagnetic region for $H = 0$ the resistivity has an activational character with the activation energy $E_0^\rho = \Delta_0 - \varepsilon_F$. If the magnetic field is nonzero or the temperature drops below T_c, the activation energy decreases causing a drop in the resistivity. If $E_0^\rho < \Delta_1$, at a certain temperature $T_{MI} < T_c$ a metal–insulator transition takes place and as the temperature T drops further, the temperature dependence of the resistivity is determined by the change in the carrier mobility rather than their concentration. The boundary between these regimes can be taken to be the temperature at which $\rho = \rho_0$.

We shall now consider the thermal emf. The formula obtained by Mott [39] neglects the interaction of carriers with other elementary excitations. Thus, Fritzsche [40] proposed a more general expression for S which can be conveniently written in the form

$$S = \frac{1}{\sigma}\frac{k_B}{e}\int_{\varepsilon_c}^{\infty} \frac{\varepsilon - \varepsilon_F - W(\varepsilon, T)}{T}\left(-\frac{\partial f}{\partial \varepsilon}\right)\sigma(\varepsilon) d\varepsilon, \qquad (3)$$

where k_B is the Boltzmann constant and e is the modulus of the electron charge. In the absence of drag effects $W \geq 0$; for $W = 0$ which corresponds to the case of elastic scattering [41, 42], expression (3) yields the Mott formula. For our purposes the integral containing the function W can be estimated using the mean-value theorem which leads to the appearance of the term $(k_B/e)(\overline{W}/T)$ in the expression for S where $\overline{W} = W(\bar{\varepsilon}, T)$, $\bar{\varepsilon} > \varepsilon_c$; if $y = E^\rho/T \gg 1$, then $\bar{\varepsilon} = \varepsilon_c$. Assuming in the remaining integral $\sigma(\varepsilon) = \sigma(\varepsilon_c)$, for the metal–insulator transition region we obtain

$$S = \frac{k_B}{e}\left[\frac{E^S}{T} + (e^y + 1)\ln(1 + e^{-y})\right], \qquad (4)$$

where $E^S = E^\rho - \overline{W}$.

In manganites with colossal magnetoresistance, the value of $\overline{W}$ in the paramagnetic region is usually greater than zero and close to E_0^ρ which is usually interpreted as evidence of the existence of strong electron–phonon coupling leading to the formation of small-radius polarons [6, 43]. An analysis of our experimental results made in the following section shows that the magnetic rather than the phonon system plays the major role in the formation of $\overline{W}$. At first glance, this contradicts the idea that carriers activated at the mobility edge behave in many respects as free carriers. This is because unlike amorphous materials, in manganites the magnetic disorder determining the position of the mobility edge is dynamic rather than static. Consequently, the propagation of carriers having an energy of the order ε_c is accompanied by a cloud of magnetic excitations which results in a nonzero value of $\overline{W}$. As the magnetic disorder decreases, i.e., as m increases, the value of $\overline{W}$ should decrease. We shall assume that $\overline{W} = W_0 - W_1 m^2$, where W_0 and W_1 are positive constants. It will be shown below that this assumption agrees with the experiment.

The concept of a cloud of perturbations accompanying an electron is valid as long as the radius of the cloud is substantially shorter than the distance between the mobile carriers. On transition to the metallic region, the distance between neighboring holes in manganites undergoing a metal–insulator transition becomes of the order of magnitude of the distance between neighboring manganese ions, which should lead to destruction of the cloud and an appreciable drop in $\overline{W}$. Then, the usual relationship $S \propto T$ remains valid for the metallic region.

It will be shown that the inequality $y > 2$ is satisfied for the $La_{0.8}Ba_{0.2}MnO_3$ single crystal under study. In this case, we can assume that

$$\rho = \rho_0 \exp[(E_0^\rho - \Delta_1 m^2)/T], \tag{5}$$

and set $(e^y + 1)\ln(1 + e^{-y}) \approx 1$ in expression (4). Then, carriers whose role in the conductivity is negligible may make a significant contribution to the thermal emf. For simplicity we shall assume that this contribution does not depend on T in the temperature range of interest, which yields the following expression:

$$S = S_0 + \frac{k_B}{e}\frac{E_0^S - E_1^S m^2}{T}, \tag{6}$$

where $E_0^S = E_0^\rho - W_0$, $E_1^S = \Delta_1 - W_1$.

In the temperature range where activation predominates at the mobility edge the normal Hall coefficient should be positive since the Aharonov–Bohm loop includes four manganese ions and the Hall mobility μ_H should be of the order of 0.1 $cm^2/(V\,s)$ [38].

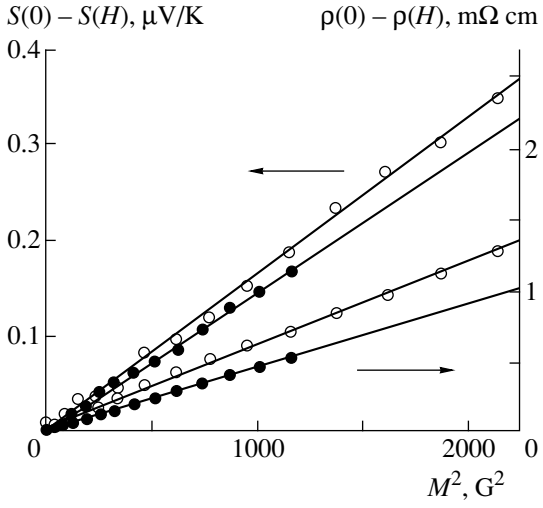


Fig. 6. Dependences of the differences between the resistivity $\rho(0) - \rho$ (H = 10 kOe) and the thermal emf $S(0) - S$ (H = 10 kOe) on the square of the magnetization for T = (○) 316 K and (●) 335 K.

5. DISCUSSION OF RESULTS

We shall first consider the range $T > 200$ K where peaks are observed for the resistivity, thermal emf, and coefficient R_0. The normal Hall coefficient and thermal emf are positive in this range. As the temperature increases from 210 to 250 K, the resistivity and R_0 increase fivefold, the coefficient R_0 being directly proportional to ρ, see inset to Fig. 5. Consequently, the Hall mobility remains the same at $\mu_H = 0.07$ $cm^2/(V\,s)$. We can conclude that at $T > 210$ K the kinetic effects are determined by holes activated at the mobility edge.

At $T > 250$ K the normal Hall coefficient ceases to be proportional to the resistivity. In our view this may be because R_0 and R_s are determined with an appreciable error in this range where ρ varies substantially on application of a magnetic field, in our case when $T > 250$ K. However, we note that the Hall mobility calculated for $T > 250$ K is also close to 0.1 $cm^2/(V\,s)$. Hence the activation at the mobility edge near T_c plays a leading role both below and above the phase transition point.

In order to determine the parameters appearing in formulas (5) and (6), we used data for $T > 300$ K. From the temperature dependences of the resistivity and the thermal emf at $H = 0$ we obtained: $\rho_0 = 1.5$ mΩ cm, $E_0^\rho = 1120$ K, $S_0 = -38$ μV/K, and $E_0^S = 190$ K. In order to find Δ_1 and E_1^S we constructed dependences of $\Delta\rho = \rho(0) - \rho(H)$ and $\Delta S = S(0) - S(H)$ on the square of the magnetization at fixed temperatures. We can see from Fig. 6 that Δρ and ΔS are proportional to M^2 which is consistent with the formulas in the preceding section. Using the saturation magnetization $M_{sat} = 560$ G we obtained $\Delta_1 = 1100$ K and $E_1^S = 180$ K.

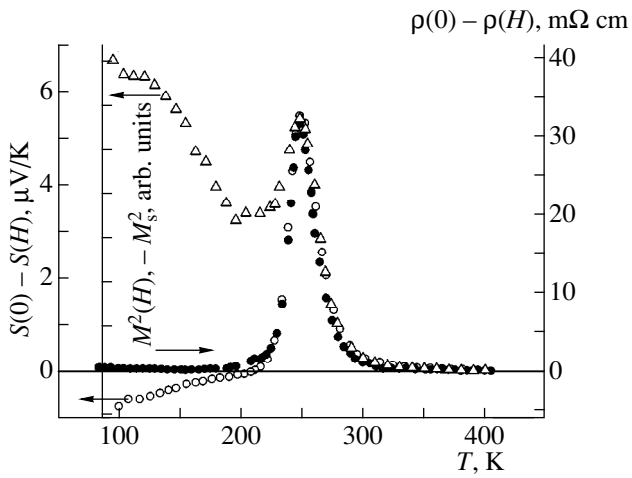


Fig. 7. Comparison between the temperature dependences of the resistivity $\rho(0) - \rho$ (H = 10 kOe), the thermal emf $S(0) - S$ (H = 10kOe), and the square of the magnetization M^2 (H = 10 kOe) – M_s^2.

Temperature dependences of $\Delta\rho$, ΔS, and $\Delta M^2 = M^2(H) - M_s^2$ are plotted in Fig. 7 for H = 10 kOe. We can see that near the resistivity peak the profiles of the curves $\Delta\rho(T)$ and $\Delta S(T)$ are almost indistinguishable and are similar to the curve of $\Delta M^2(T)$ for $T > 210$ K, i.e., where the ferromagnetic regions make the dominant contribution to the susceptibility. This confirms our assumption that the changes in the resistivity and the thermal emf are determined by the change in the square of the magnetization not only in the paramagnetic region but also near the Curie point.

Using the values obtained for ρ_0, E_0^ρ, Δ_1, E_0^S, and E_1^S and the experimental values of the spontaneous magnetization, we calculated the temperature dependence of the resistivity and the thermal emf for $H = 0$. The results are given by the dot-dash curves in Figs. 2 and 3. The calculated curves *1* accurately reproduce the profile of the peaks although they lie above the experimental points in the ferromagnetic region. In our view, this difference can be attributed to the magnetic and electrical inhomogeneities of the sample. It has been noted that approximately 10% of our sample volume is occupied by nonferromagnetic inclusions. It follows from the results of measurements of the resistivity of $La_{1-x}Sr_xMnO_3$ single crystals [5] that for samples with $x \sim x_c$ the resistivity depends weakly on x in the far paramagnetic region but depends very strongly on it when $T < T_c$. Bearing in mind that the difference in the susceptibility is reduced when $T > T_c$, we can assume that the sample is in comparatively homogeneous electric and magnetic states if the temperature is considerably higher than T_c. Thus, we determined all the parameters from the results of measurements in the paramagnetic region. On transition to the magnetically ordered phase, the material becomes magnetically and electrically inhomogeneous. Since the current flows predominantly in the ferromagnetic regions possessing low resistivity, in formulas (5) and (6) we need to substitute the magnetization of these ferromagnetic regions rather than the experimentally determined average over the sample. This circumstance can be taken into account most easily by substituting into (5) and (6) values of the magnetization 10% higher than the experimental values. In this case the role of inhomogeneity is exaggerated in the paramagnetic region but this is unimportant since the magnetization in this region is low at the magnetic fields used. Results of calculations of $\rho(T)$ and $S(T)$ allowing for this correction for $H = 0$ and $H = 10$ kOe are given by the solid (*2*) and dashed (*3*) curves in Figs. 2 and 3. For the resistivity the theoretical and experimental curves are very similar whereas for the thermal emf the agreement between theory and experiment is slightly inferior although in this case, the peak profile and its change on application of the magnetic field can be reproduced quite satisfactorily. We stress that all the parameters used in the calculations were obtained from independent experiments and no additional fitting was carried out.

It follows from this reasoning that the abrupt decrease in the resistivity on transition to the magnetically ordered state is caused by a reduction in the activation energy; in this case the material continues to be in the dielectric state. This is consistent with the fact that at all temperatures the resistivity of this single crystal is several times higher than ρ_0.

We can see that the inhomogeneity of the material must be taken into account for a quantitative comparison between theory and experiment. Quite clearly, allowing for this inhomogeneity by eliminating the nonferromagnetic regions from the analysis while implicitly assuming that the ferromagnetic regions are electrically homogeneous is a very rough approximation since in general the ferromagnetic regions are also inhomogeneous. It may well be that the difference (up to 10–20%) between the experimental and theoretical curves for 250 K < T < 300 K (see Figs. 2, 3) arises because the inhomogeneity must be taken into account more accurately in this region.

In the literature attention has repeatedly been drawn to the fact that in lanthanum manganites there is a large difference between the activation energies of the resistivity and the thermal emf in the paramagnetic region and this difference depends weakly on H. On this basis it is concluded that for $T > T_c$ interaction of carriers with lattice vibrations predominates and charge transport is achieved by hopping of small-radius polarons (single-site polarons). However, the internal parameter is not the magnetic field but the magnetization. In our case we have $\overline{W} = E^\rho - E^S = (930–920m^2)$ K. The strong dependence of $\overline{W}$ on m indicates that magnetic interactions play a dominant role while the weak

dependence of $\overline{W}$ on H is caused by the weak magnetic susceptibility in the paramagnetic range.

The strong dependence of $\overline{W}$ on the magnetization implies that an electron in any Mn site also senses the spin of neighboring sites. Consequently, in these materials an electron excited to the mobility edge moves surrounded by a "cloud" of magnetic excitations. The radius of this cloud (in other words the characteristic radius of the carrier wave function) is no less than the distance between the nearest manganese ions so that the cloud covers many lattice sites. This conclusion agrees with the results of recent theoretical studies [44, 45].

We shall now consider temperatures $T < 200$ K. In this range the normal Hall coefficient is negative which suggests that three-site processes predominate [38]. Bearing in mind the substantial difference between the calculated and experimental curves of $\rho(T)$ and $S(T)$, the change in the dependence of the thermal emf on the magnetic field (see Fig. 7), and the rapid increase in resistivity at $T < 120$ K, we can conclude that at low temperatures conduction is mainly achieved by hopping between localized states and not by activation at the mobility edge. This conclusion is completely consistent with general reasoning on the role of various conduction mechanisms in disordered materials [38] and with data from optical experiments [46, 47] which indicate that various types of localized states are present in materials with colossal magnetoresistance. A specific feature of manganites is that the parameters describing the hopping conduction at $T \ll T_c$ should depend on the magnetic state but differently from the mobility edge near the Curie temperature. As far as the authors are aware, no such calculations have yet been made.

6. CONCLUSIONS

Near the Curie temperature single crystals of lanthanum manganites close to the metal–insulator concentration transition exhibit the same singularities on curves giving the temperature and magnetic-field dependences of the resistivity. We can thus assume that the results obtained from our study of an $La_{0.8}Ba_{0.2}MnO_3$ single crystal are general and reflect the physics of transport processes in all lanthanum manganite single crystals with $x \sim x_c$ regardless of the type of doping. We can briefly formulate these as follows.

Near the Curie temperature the dominant conduction mechanism is activation at the mobility edge. Colossal magnetoresistance and other anomalies of the kinetic coefficients are caused by shift of the mobility edge on application of a magnetic field and/or a change in temperature which lead to changes in the carrier concentration in the delocalized states. The steep drop in the resistivity accompanying a transition to the magnetically ordered state does not imply a transition to metallic conductivity, as is usually assumed, but is caused by a drop in the activation energy. From this it follows that the temperature at which the resistivity is highest cannot generally be considered to be the metal–insulator transition temperature.

Far from T_c hopping between localized states can make a substantial contribution to the kinetic coefficients.

Carriers activated at the mobility edge move accompanied by a cloud of different excitations among which magnetic excitations are the most important.

The inhomogeneity of the materials plays an important role in forming the kinetic properties of these manganites, in particular determining the temperature dependence of the kinetic coefficients near the temperature of maximum resistivity.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 97-02-16008).

REFERENCES

1. É. L. Nagaev, Usp. Fiz. Nauk **166**, 833 (1996) [Phys. Usp. **39**, 781 (1996)].
2. A. P. Ramírez, J. Phys.: Condens. Matter **9**, 8171 (1997).
3. Y. Tokura and Y. Tomioka, J. Magn. Magn. Mater. **200**, 1 (1999).
4. J. M. D. Coey, M. Viret, and S. von Molnar, Adv. Phys. **48**, 167 (1999).
5. A. Urushibara, Y. Moritomo, T. Arima, *et al.*, Phys. Rev. B **51**, 14103 (1995).
6. M. Jaime, M. B. Salamon, K. Pettit, *et al.*, Appl. Phys. Lett. **68**, 1576 (1996).
7. V. Crespi, Li Lu, Y. X. Jia, *et al.*, Phys. Rev. B **53**, 14303 (1996).
8. A. J. Millis, P. W. Littlewood, and B. I. Shraiman, Phys. Rev. Lett. **74**, 5144 (1995).
9. A. Moreo, S. Yunoki, and E. Dagotto, Science **283**, 2034 (1999).
10. D. Khomskii, cond-mat/9909349; submitted to Physica B (2000).
11. M. Yu. Kagan, D. I. Khomskii, and M. V. Mostovoy, Eur. Phys. J. B **12**, 217 (1999).
12. L. P. Gor'kov, Usp. Fiz. Nauk **168**, 665 (1998) [Phys. Usp. **41**, 589 (1998)].
13. M. Uehara, S. Mori, C. H. Chen, and S.-W. Cheong, Nature **399**, 560 (1999).
14. R. Allub and B. Alascio, Solid State Commun. **99**, 613 (1996).
15. N. G. Bebenin and V. V. Ustinov, J. Phys.: Condens. Matter **10**, 6301 (1998).
16. J. E. Núñez-Reguero, D. Gupta, and A. M. Kadin, J. Appl. Phys. **79**, 5179 (1996).
17. P. Wagner, D. Mazilu, L. Trappeniers, *et al.*, Phys. Rev. B **55**, R14721 (1997).
18. G. Jakob, F. Martin, W. Westerburg, and H. Adrian, Phys. Rev. B **57**, 10252 (1998).
19. G. Jakob, W. Westerburg, F. Martin, and H. Adrian, Phys. Rev. B **58**, 14966 (1998).

20. G. Jeffrey Snyder, M. R. Beasley, T. H. Geballe, *et al.*, Appl. Phys. Lett. **69**, 4254 (1996).
21. M. Jaime, H. T. Hardner, M. B. Salamon, *et al.*, Phys. Rev. Lett. **78**, 951 (1997).
22. P. Mandal, A. Poddar, A. G. M. Jansen, *et al.*, Phys. Status Solidi **165**, 219 (1998).
23. M. Ziese and C. Srinitiwarawong, Europhys. Lett. **45**, 256 (1999).
24. P. Matl, N. P. Ong, Y. F. Yan, *et al.*, Phys. Rev. B **57**, 10248 (1998).
25. P. Mandal, K. Barner, L. Haupt, *et al.*, Phys. Rev. B **57**, 10256 (1998).
26. A. Asamitsu and Y. Tokura, Phys. Rev. B **58**, 47 (1998).
27. N. G. Bebenin, R. I. Zaĭnullina, V. V. Mashkautsan, *et al.*, Zh. Éksp. Teor. Fiz. **113**, 981 (1998) [JETP **86**, 534 (1998)].
28. R. I. Zaĭnullina, N. G. Bebenin, V. V. Mashkautsan, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **40**, 2085 (1998) [Phys. Solid State **40**, 1889 (1998)].
29. N. G. Bebenin, R. I. Zaĭnullina, V. V. Mashcautsan, *et al.*, Phys. Status Solidi A **175**, 665 (1999).
30. R. I. Zaĭnullina, N. G. Bebenin, V. V. Mashkautsan, *et al.*, submitted to Fiz. Tverd. Tela (St. Petersburg) (2000).
31. D. Shulyatev, S. Karabashev, A. Arsenov, and Ya. Mukovskii, J. Cryst. Growth **198/199**, 511 (1999).
32. J. Rodríguez-Carvajal, Physica B (Amsterdam) **192**, 55 (1993).
33. V. E. Arkhipov, N. G. Bebenin, V. P. Dyakina, *et al.*, submitted to Phys. Rev. B (2000).
34. V. E. Arkhipov, V. S. Gaviko, A. V. Korolyov, *et al.*, J. Magn. Magn. Mater. **196/197**, 539 (1999).
35. S. V. Vonsovskiĭ, *Magnetism* (Nauka, Moscow, 1971; Wiley, New York, 1974).
36. E. M. Kogan and M. I. Auslender, Phys. Status Solidi B **147**, 613 (1988).
37. N. G. Bebenin and V. V. Ustinov, J. Magn. Magn. Mater. **196/197**, 451 (1999).
38. N. F. Mott and E. A. Davis, *Electronic Processes in Non-Crystalline Materials* (Clarendon Press, Oxford, 1971; Mir, Moscow, 1974).
39. N. F. Mott, Philos. Mag. **31**, 217 (1975).
40. H. Fritzsche, Solid State Commun. **9**, 1813 (1971).
41. H. Böttger and V. V. Bryksin, Phys. Status Solidi B **78**, 9 (1976).
42. H. Böttger, Phys. Status Solidi B **78**, 415 (1976).
43. M. Jaime, H. T. Hardner, M. B. Salamon, *et al.*, Phys. Rev. Lett. **78**, 951 (1997).
44. T. Mizokawa, D. I. Khomskii, and G. A. Sawadtzky, cond-mat/9912021.
45. M. Korotin, T. Fujiwara, and V. Anisimov, cond-mat/9912456.
46. N. G. Bebenin, N. N. Loshkareva, Yu. P. Sukhorukov, *et al.*, Solid State Commun. **106**, 357 (1998).
47. N. N. Loshkareva, Yu. P. Sukhorukov, E. A. Neĭfel'd, *et al.*, Zh. Éksp. Teor. Fiz. **117**, 440 (2000) [JETP **90**, 389 (2000)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 1035–1041.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1190–1197.

SOLIDS
Electronic Properties

Anomalous Hall Effect in Magnetic Sandwiches with a Dielectric Spacer

A. V. Vedyaev, O. A. Kotel'nikova, N. G. Pugach*, and N. V. Ryzhanova

Moscow State University, Moscow, 119899 Russia
**e-mail: pugach@magn.phys.mus.su*

Received December 7, 1999

Abstract—Quantum-statistical calculations are presented for the anomalous Hall effect in a magnetic sandwich with a tunnel junction across a thin dielectric spacer. The tunneling current flows across the junction perpendicular to the plane of the layers while the Hall component of the current lies in this plane. The Kubo formalism and the Green's functions are used to calculate the contribution of skew scattering to the Hall conductivity. The classical size effect in the Hall conductivity of this structure is studied and two new effects are observed. One is associated with the dependence of the effective electric field in the magnet on the transparency of the dielectric potential barrier for electrons when the current flows perpendicular to the layers of the structure and may be called "geometric". The other occurs as a result of the influence of the strong electric field in the dielectric on the electron motion in the adjacent magnetic layers.

1. INTRODUCTION

Recently considerable attention has been paid to studies of multilayer magnetic structures: sandwiches and superlattices. It has been observed that these structures possess giant magnetoresistance [1–3] and this effect has been extensively studied theoretically [4–6]. Models have been developed to calculate the anomalous Hall effect in thin magnetic films [7, 8] and multilayer structures [9, 10]. It is generally assumed that the Hall emf in ferromagnets has two components: normal and anomalous. The constant of the anomalous Hall effect may differ in order of magnitude from the constant of the normal effect and may even have a different sign. In [9] the side jump mechanism was studied for the anomalous Hall effect and the authors concluded that the Hall conductivity (associated with the off-diagonal component of the electrical conductivity tensor) is proportional to the ordinary conductivity (diagonal component) with a coefficient determined by the average spin–orbit interaction. Another mechanism responsible for the anomalous Hall effect, skew scattering, involved allowing for the three-point character of the conductivity which resulted in the absence of a simple relationship between the Hall conductivity and the ordinary conductivity [10]. A study of this mechanism for the anomalous Hall effect showed that when the classical size effect is taken into account, the Hall constant depends nonlinearly on the thickness of the magnetic layers and even in the limit of large layer thickness differs when the primary current flows parallel and perpendicular to the plane of the layers. This difference depends on which type of electron, *s* or *d*, makes the main contribution to the anomalous Hall effect [10].

Interest is now being focused on studying multilayer magnetic structures with a tunnel junction across a thin dielectric layer in which the electric current is produced by electron tunneling through the potential barrier formed as a result of a difference between the band structures of the metal and the dielectric [11–16]. In these magnetic structures the potential profile for electrons belonging to subbands with different directions of spin varies as the relative orientation of the magnetizations of neighboring ferromagnetic layers varies, which naturally influences their electrical resistance. For this reason these structures were called "magnetic valve" structures and the effect similar to the giant magnetoresistance was called tunneling magnetoresistance. Experimental studies of these structures were started some time ago [17] but an appreciable tunneling magnetoresistance at room temperature, which makes these compounds potentially useful for applications in microelectronics, was only observed comparatively recently [11, 13, 14].

The Slonszewski quantum-mechanical model [18] neglecting electron scattering and a quantum-statistical model using the Kubo formalism [19, 20] have been proposed to describe this phenomenon. It has been shown that the magnetoresistance of tunnel-junction magnetic structures is mainly attributable to *d*-electrons, in contrast to structures with a paramagnetic metal spacer where the more mobile *s*-electrons play the most important role.

In the present paper we put forward calculations of the anomalous Hall effect produced as a result of skew scattering of spin-polarized *d*-electrons caused by spin–orbit interaction in a magnetic sandwich with a tunnel junction. As far as we are aware, these are the first calculations of this type. The calculations are based on the Kubo formalism using the method of Green's functions and a model developed in [10] to describe the anomalous

1063-7761/00/9006-1035$20.00

Hall effect in magnetic sandwiches with a nonmagnetic metal spacer.

2. MODEL

We shall consider a sandwich consisting of two ferromagnetic layers of thickness a_1 and a_3 separated by a thin dielectric spacer of thickness a_2. The spacer should be sufficiently thin (of nanometer order) so that the tunneling current, which decreases exponentially with increasing layer thickness, can be observed. The tunneling current flows perpendicular to the plane of the layers (z-axis) under the action of an external electric field E_z produced by a voltage applied to the two magnetic layers. The magnetization vectors lie in the plane of the layers (we assume that one of these is directed along the z-axis) and may change their relative orientation from parallel to antiparallel depending on the external magnetic field. The Hall conductivity ρ is then measured along the x-axis and the Hall field is denoted by E_x.

Electrons from the s- and d-bands of the metal layers contribute to the tunneling current. However, since skew scattering as a result of spin–orbit interaction can be considered to be negligible for s-electrons, the d-electrons make the dominant contribution to the formation of the anomalous component of the Hall field. In addition, the d-band has stronger spin splitting which results in a large conductivity difference between the different spin channels. This is taken into account by introducing the values of the Fermi quasimomentum $k_{F\uparrow}$ and $k_{F\downarrow}$ for the d-subband electrons with projections of the spin magnetic moment parallel and antiparallel to the magnetization, respectively. Thus, we shall subsequently consider a system of d-electrons using the model of an electron gas inserted in a structure with a spin-dependent potential profile in the z direction, allowing for spin-dependent skew scattering in the bulk of the ferromagnetic layers.

The ferromagnetic layers are considered to be sufficiently thick so that interference of electron waves reflected from their external boundaries can be neglected, i.e., the size effect associated with quantization of the z-projection of the electron quasimomentum in the two potential wells of the metal layers can be neglected. Allowance for this quantum size effect is outside the scope of our formulated problem.

The structure of the magnetic sandwich is spatially inhomogeneous in the z direction perpendicular to its plane but the configuration-averaged quantities are translationally invariant in the xy plane. Thus, we use a mixed coordinate-momentum $\mathbf{k}$–z representation where $\mathbf{k}$ is the projection of the quasimomentum in the xy plane and z is the coordinate along the z-axis. The electron wave function is a plane wave with the damping $\gamma_\alpha = \hbar^2 k_{F\alpha}/m l_\alpha$ as a result of scattering in the magnetic layers and decays exponentially inside the dielectric. Here m is the effective mass of the d-band electrons, l_α is the mean free path of an electron having the spin projection $\alpha = \uparrow$ or $\alpha = \downarrow$, and $\hbar$ is Planck's constant. The origin on the z-axis is located at the center of the dielectric layer a_2. Then taking into account the conditions for continuity of the Green's function and its derivative at the layer interface and also the conditions for suppression of spatial correlations at infinity, the Green's functions of this problem may be written in the form

$$G_{\mathbf{k}}(z,z') = \frac{ma_0}{i\hbar^2 p_1}\left\{\exp\left[ip_1\left(z+\frac{a_2}{2}\right)\right]-\frac{q+ip_1}{q-ip_1}\right.$$
$$\left.\times\exp\left[-ip_1\left(z+\frac{a_2}{2}\right)\right]\right\}\exp\left[-ip_1\left(z'+\frac{a_2}{2}\right)\right]$$
$$\text{for } -a_1-a_2/2\le z'<z\le -a_2/2,$$
$$G_{\mathbf{k}}(z,z') = \frac{2ma_0\exp[-ip_1(z+a_2/2)]}{\hbar^2(q-ip_1)(q-ip_3)\exp(qa_2)}$$
$$\times\left\{(q-ip_3)\exp\left[-q\left(z'-\frac{a_2}{2}\right)\right]\right. \qquad (1)$$
$$\left.+(q+ip_3)\exp\left[q\left(z'-\frac{a_2}{2}\right)\right]\right\}$$
$$\text{for } -a_1-a_2/2\le z\le -a_2/2\le z'\le a_2/2.$$
$$G_{\mathbf{k}}(z,z') = -\frac{4qma_0}{\hbar^2(q-ip_1)(1-ip_3)\exp(qa_2)}$$
$$\times\exp\left[-ip_1\left(z+\frac{a_2}{2}\right)\right]\exp\left[ip_3\left(z'-\frac{a_2}{2}\right)\right]$$

for

$$-a_1-a_2/2\le z\le -a_2/2,\quad a_2/2\le z'\le a_2/2+a_3.$$

In these expressions the z-projection of the electron quasimomentum in the ferromagnetic layers a_1 and a_3 is denoted as

$$p_{1,3} = c_{1,3}+id_{1,3} = \sqrt{k_{F1,3}^2-\mathbf{k}^2+2mi\gamma_{1,3}/\hbar^2},$$

and inside the dielectric layer this is an imaginary quantity:

$$iq = i\sqrt{Q^2+\mathbf{k}^2},$$

where $\hbar^2Q^2/2m$ is the height of the potential barrier above the Fermi level, the subscripts 1 and 3 indicate the $\uparrow$ or $\downarrow$ spin projection of electrons in the selected spin channel on the direction of magnetization in the particular ferromagnetic layer a_1 or a_3, respectively and a_0 is the crystal lattice constant. We neglected here small terms with respect to $\exp(-qa_2)$.

The diagonal components of the conductivity tensor $\sigma_{zz}(z, z')$ and $\sigma_{xx}(z, z')$ are calculated using the Kubo formula [19]. The average value $\sigma_{xx}(z, z')$ of the diagonal conductivity in the layer plane when the coordinates z and z' belong to the same ferromagnetic layer is the same of the conductivity of the bulk sample. If z and z' lie in different layers of the tunnel junction structure, the diagonal conductivity is exponentially small.

The expression for the off-diagonal (Hall) conductivity is written as [10]

$$\sigma_{xz}(z, z') = \frac{\hbar e^2}{\pi a_0^4}\sum_{\mathbf{k},\alpha} v_x \tilde{G}_{\mathbf{k}}^{\alpha-}(z, z') v_{z'} \tilde{G}_{\mathbf{k}}^{\alpha+}(z', z), \quad (2)$$

where $\mathbf{k}$ is the projection of the electron quasimomentum on the xy layer plane,

$$v_x = \frac{\hbar k_x}{m}, \quad v_{z'} = -\frac{i\hbar}{m}\left(\frac{\overleftarrow{\partial}}{\partial z'} - \frac{\overrightarrow{\partial}}{\partial z'}\right)$$

are the components of the velocity operator in the $\mathbf{k}$–z representation, the arrows indicate that the derivative with respect to the coordinate is taken of the function to the left or right of the velocity operator; $\tilde{G}_{\mathbf{k}}^{\alpha+}(z, z')$ and $\tilde{G}_{\mathbf{k}}^{\alpha-}(z, z')$ are the advanced and retarded Green's functions of the d-electrons in the presence of a random potential $V_{A(B)}$ responsible for scattering. These functions can be obtained from the Dyson equation in the first approximation with respect to spin–orbit interaction H_{SO}^{α} which is much smaller than the Fermi energy:

$$\tilde{G}_{\mathbf{k}}^{\alpha}(z, z') = G_{0\mathbf{k}}^{\alpha}(z, z') + \sum_{\tilde{\mathbf{k}}}\int G_{0\mathbf{k}}^{\alpha}(z, z'') \times [T^{\alpha}(z'') + H_{SO}^{\alpha}(z'')] G_{0\tilde{\mathbf{k}}}^{\alpha}(z'', z') dz'', \quad (3)$$

where $G_{0\mathbf{k}}^{\alpha}(z, z')$ are the Green's functions of the system neglecting scattering and $T^{\alpha}(z'')$ is the matrix of scattering at impurities (zeroth order with respect to H_{SO}^{α}) which is calculated in the coherent potential approximation for the random potential $V_{A(B)}$. We shall consider a binary ferromagnetic alloy A_xB_{1-x} so that the potential at each site has the value ε_A with the probability x and ε_B with the probability $1 - x$. Allowing for spin–orbit interaction with the constants λ_A and λ_B the scattering potential at each site may be written in the form

$$V_{A(B)} = \varepsilon_{A(B)} + H_{SO}^{A(B)} = \varepsilon_{A(B)} + i\lambda_{A(B)}(\mathbf{M}\cdot[\mathbf{v}\times\mathbf{v}'])a_0^2 m^2/\hbar^2,$$

where $\mathbf{M} = (0, M, 0)$ is the magnetization vector. Substituting the Green's functions $\tilde{G}_{\mathbf{k}}^{\alpha}(z, z')$ (3) into the expression for the conductivity (2) and averaging over the random impurity distribution, we obtain an expression for the three-point Hall conductivity allowing for skew scattering as a result of spin–orbit interaction:

$$\sigma_{xz}^{\alpha}(z, z'', z') = \frac{e^2 m^2 M}{\pi\hbar a_0^2}\sum_{\mathbf{k},\tilde{\mathbf{k}}} v_x \{ G_{\mathbf{k}}^{\alpha-}(z, z'')[\mathbf{v}\times\tilde{\mathbf{v}}]_y \times G_{\tilde{\mathbf{k}}}^{\alpha-}(z'', z') v_{z'} G_{\mathbf{k}}^{\alpha+}(z, z'') G_{\tilde{\mathbf{k}}}^{\alpha+}(z'', z') i\langle\lambda^{\alpha}(z'')T^{\alpha+}(z'')\rangle$$
$$- G_{\mathbf{k}}^{\alpha-}(z, z'') G_{\tilde{\mathbf{k}}}^{\alpha-}(z'', z') v_{z'} G_{\mathbf{k}}^{\alpha+}(z, z'')[\mathbf{v}\times\tilde{\mathbf{v}}]_y \times G_{\tilde{\mathbf{k}}}^{\alpha+}(z'', z') i\langle\lambda^{\alpha}(z'')T^{\alpha-}(z'')\rangle \}.$$

This contains the Green's functions with allowance for attenuation of the electron states as a result of scattering by impurities (1).

We introduce the notation $\delta = \varepsilon_A - \varepsilon_B$. For the calculations we need to assume that skew scattering as a result of spin–orbit interaction of electrons having different spin projections on the magnetization axis ($\uparrow$ and $\downarrow$) takes place in different directions: $\lambda^{\uparrow} = -\lambda^{\downarrow} = \lambda_A - \lambda_B$. For small δ and λ^{α} we can write in the second Born approximation

$$\mathrm{Im}\langle\lambda^{\alpha}(z'')T^{\alpha}(z'')\rangle = -x(1-x)(1-2x)\delta^2\lambda^{\alpha}\mathrm{Im}\sum_{\mathbf{k}} G_{\mathbf{k}}^{\alpha}(z'', z'').$$

We shall now analyze the motion of spin-polarized electrons across the tunnel junction in the direction of the z axis.

The expression for the primary current is written in the form

$$j_z^{\alpha}(z) = \int\sigma_{zz}^{\alpha}(z, z')E_z^{\alpha}(z')dz' + \iint\sigma_{zx}^{\alpha}(z, z'', z')E_x^{\alpha}(z')dz'dz''. \quad (4)$$

Here $E_z^{\alpha}(z')$ and $E_x^{\alpha}(z')$are the accelerating and Hall components of the effective electric field for electrons of spin α. In this definition the effective field corresponds to the electrochemical potential gradient.

The second term on the right-hand side of expression (4) may be neglected because of the smallness of the spin–orbit interaction responsible for the appearance of the Hall component, compared with the z-component of the current. Then, assuming that the main drop in the voltage applied to the sandwich takes place at the dielectric layer (since its resistance is much higher than the metal layers), from the condition of current continuity $\partial j_z^{\alpha}(z)/\partial z = 0$ we can find relationships between the electric fields in the different layers

neglecting interference effects, which do not depend on the coordinate z:

$$\frac{k_1^2 l_1}{3} E_{1z}^{\alpha} = \frac{k_3^2 l_3}{3} E_{3z}^{\alpha} = \frac{16 a_2}{(Q^2 + k_1^2)(Q^2 + k_3^2)} E_{2z} \times \int q^2 c_1 c_3 \exp(-2qa_2) k dk,$$

where E_{1z}^{α} and E_{3z}^{α} are the components of the electric field in the ferromagnetic layers, and E_{2z} is the component in the dielectric. The Hall component of the electric field is obtained from the condition that the current in the direction of the x axis is zero in each of the ferromagnetic layers:

$$\int\limits_{z \in a_1} j_x(z) dz = \int\limits_{z \in a_1} \sum_{\alpha} j_x^{\alpha}(z) dz = 0,$$
$$\int\limits_{z \in a_3} j_x(z) dz = \int\limits_{z \in a_3} \sum_{\alpha} j_x^{\alpha}(z) dz = 0. \quad (5)$$

For example, for the current j_x in the x direction in the ferromagnetic layer a_1 we have from equation (5)

$$\int\limits_{z \in a_1} dz \int \sum_{z'} \sum_{\alpha} \sigma_{xx}^{\alpha}(z, z') E_x(z') dz'$$
$$+ \int\limits_{z \in a_1} dz \int\int \sum_{z'z''} \sum_{\alpha} \sigma_{xz}^{\alpha}(z, z'', z') E_z^{\alpha}(z') dz'' dz' = 0.$$

The effective electric field E_x does not depend on the spin since in this structure there is no chemical potential gradient along the x axis. Note that in a nonmagnetic dielectric the Hall field E_{2x} is negligible because of the absence of free electrons. We shall also assume that both the diagonal $\sigma_{xx}^{\alpha}(z, z')$ and off-diagonal $\sigma_{xz}^{\alpha}(z, z'', z'$ components of the conductivity tensor are exponentially small if the coordinates z, z', or z'' lie in different ferromagnetic layers. The expression for the Hall field in the magnetic layer a_1 then has the form

$$E_{1x} = -\sum_{\alpha} \left\{ E_{1z}^{\alpha} \int\int\int\limits_{z, z', z'' \in a_1} \sigma_{xz}^{\alpha}(z, z'', z') dz'' dz' dz \right.$$
$$\left. + E_{2z}^{\alpha} \int\int\int\limits_{z, z'' \in a_1, z' \in a_2} \sigma_{xz}^{\alpha}(z, z'', z') dz'' dz' dz \right\} \quad (6)$$
$$\times \left[\sum_{\alpha} \int\int\int\limits_{z, z' \in \alpha_1} \sigma_{xx}^{\alpha}(z, z') dz' dz \right]^{-1}.$$

Here the second term in braces reflects the influence of the strong electric field E_{2z} in the dielectric spacer on the Hall component of the field in the neighboring metal layer. The Hall field in the other ferromagnetic layer E_{3x} is calculated by analogy with the field (6).

The value which can be measured experimentally is the Hall resistivity averaged over the thickness of the entire magnetic sandwich which is given in the form

$$\rho = \left| \frac{a_1 |E_{1x}| \pm a_3 |E_{3x}|}{(a_1 + a_2 + a_3) j_z} \right|,$$

where the "+" sign is taken for a parallel configuration of metal layer magnetizations and the "–" sign is taken for an antiparallel configuration because in this case the Hall field in the ferromagnetic layers is in opposite directions,

$$j_z = \sum_{\alpha} j_z^{\alpha}$$

is the total primary current flowing along the axis perpendicular to the sandwich plane (4).

Thus, we obtained expressions for the Hall resistivity of a magnetic sandwich with a tunnel junction across a dielectric spacer for both configurations of the magnetization. These expressions are given in the Appendix.

3. DISCUSSION OF RESULTS

For greater clarity of the results, the expressions for the Hall resistivity of the magnetic sandwich are normalized to the reduced magnetic moment and the Hall resistivity of a bulk sample of the same ferromagnetic material:

$$\rho_0 = \frac{D}{2B} \lambda \frac{m}{\pi^2 \hbar^2} \frac{l_{\uparrow}^2 k_{F\uparrow}^5 - l_{\downarrow}^2 k_{F\downarrow}^5}{(l_{\uparrow} k_{F\uparrow}^2 + l_{\downarrow} k_{F\downarrow}^2)^2},$$

where B and D are constants, $D = x(1-x)(1-2x)\delta^2 a_0^4$, $B = e^2\hbar^3/8\pi^2 m$.

This made it possible to study different types of effects associated with the introduction of a tunnel junction into the magnetic structure.

First, it should be noted that a tunnel junction across a dielectric spacer creates a considerably greater obstacle for the electron motion compared with scattering at impurities and crystal-lattice defects inside the metal layers since the coefficient of electron transmission decreases exponentially with increasing dielectric thickness. Thus, we can assume that if the potential difference applied to the two magnetic layers of the sandwich is constant, the main voltage drop takes place at the dielectric spacer. This assumption holds as far as very large metal thicknesses (of the order of meters for a dielectric thickness of around 10 Å) when its resistivity is of the same order of magnitude of that of the tun-

nel junction. If the thickness of the metal layers is not so large, the accelerating component E_z^α of the effective field in each ferromagnetic layer does not depend on its thickness but depends on the transmission coefficient of the potential barrier for electrons having the spin projection α on the magnetization axis. This situation differs fundamentally from the case of a bulk sample where the field is simply equal to the ratio of the potential difference to its thickness and is the same for any direction of electron spin.

If the current strength is fixed perpendicular to the plane of the layers, the ratio of the resistivities of the two current channels with a different direction of spin for the entire sandwich will not be the same as that for the homogeneous metal and thus, the ratio of the effective fields will differ. Consequently the distribution of the effective electric field in a magnetic sandwich with a tunnel junction where the current flows perpendicular to the layers will differ fundamentally from the field in a metal film or a bulk sample. This strongly influences the Hall effect and in particular explains why the normalized Hall resistivity does not tend to unity as the thickness of the ferromagnetic layers increases, i.e., in addition to various types of size effects, this structure also exhibits another important field redistribution effect as a result of the presence of a tunnel junction. This may be called the "geometric size effect" because it only appears in a geometry where the current flows perpendicular to the layers of a sandwich with a tunnel junction and does not appear when the current flows in the plane of the layers. Hence, the Hall resistivity of this structure cannot be calculated simply as the sum of the resistivities of two independent ferromagnetic films connected in series.

The geometric size effect is responsible for the fact that when the thickness of the magnetic layers is fairly large so that other size effects can be excluded but not large enough for the resistivity of the metal sections to be comparable with that of the dielectric, the reduced Hall resistivity of a sandwich with antiparallel orientation of the layer magnetization tends to

$$\frac{M_0}{M_s}\frac{\rho^{\uparrow\downarrow}}{\rho_0} = \frac{(l_\uparrow k_{F\uparrow}^3 - l_\downarrow k_{F\downarrow}^3)(l_\uparrow k_{F\uparrow}^2 + l_\downarrow k_{F\downarrow}^2)}{2(l_\uparrow^2 k_{F\uparrow}^5 - l_\downarrow^2 k_{F\downarrow}^5)} \sim \frac{1}{2},$$

whereas for parallel orientation it tends to

$$\frac{\rho^{\uparrow\uparrow}}{\rho_0} = \frac{(I_2^\uparrow l_\uparrow k_{F\uparrow}^3 - I_2^\downarrow l_\downarrow k_{F\downarrow}^3)(l_\uparrow k_{F\uparrow}^2 + l_\downarrow k_{F\downarrow}^2)}{(I_2^\uparrow + I_2^\downarrow)(l_\uparrow^2 k_{F\uparrow}^5 - l_\downarrow^2 k_{F\downarrow}^5)} \sim 1.$$

Here M_0/M_s is the ratio of the magnetic moment of the bulk sample to the magnetic moment of an antiparallel-magnetized sandwich of the same material and

$$\frac{M_0}{M_s} \propto \frac{|a_1 - a_3|}{a_1 + a_3},$$

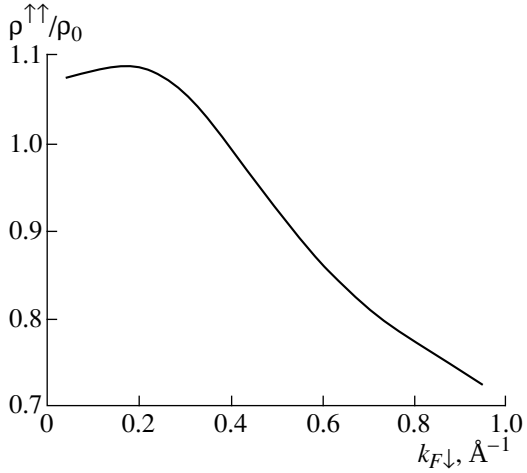


Fig. 1. Dependence of the normalized Hall conductivity of a magnetic sandwich with a tunnel junction and a parallel configuration of layer magnetization on the Fermi wave vector $k_{F\downarrow}$ of spin $\downarrow$ electrons. The thickness of the layers is $a_1 = a_3 = 20$ Å, $a_2 = 10$ Å, the mean free path is $l_\uparrow = l_\downarrow = 60$ Å, $k_{F\uparrow} = 1$ Å^{-1}, and $Q = 1$ Å^{-1}.

$I_2^{\uparrow,\downarrow}$ are integrals proportional to the transparency of the potential barrier for electrons having spin $\uparrow$ and $\downarrow$, respectively (see Appendix). This effect is stronger for an antiparallel configuration. Reflecting the field redistribution mechanism described, the relationships for the Hall resistivity given in the Appendix also have a structure in which we can identify three contributions describing the influence on the Hall effect in a magnetic sandwich with a dielectric spacer:

(1) Skew scattering in the bulk of the ferromagnetic layers with a typical field distribution for this structure;

(2) A classical size effect associated with the geometric dimensions of the ferromagnetic layers and proportional to the ratio of the electron mean free path to the layer thickness;

(3) An effect associated with the influence of the strong field in the dielectric on the electron motion in the adjacent magnetic layer (this is a new size effect obtained by us).

If the sandwich layers are parallel magnetized, the second and third effects yield contributions to the size effect of the same order of magnitude but of opposite sign, which is thus not very well defined and depends weakly on the *d*-band splitting. Only the average Hall resistivity varies (Fig. 1).

In contrast, for an antiparallel magnetization configuration (Fig. 2), the influence of the field in the dielectric on the skew scattering of band electrons near the insulating layer is strongest because whereas in a parallel geometry the motion of electrons having spins $\uparrow$ and $\downarrow$ is completely independent and the accelerating field of the dielectric is mainly related to the transmission coefficient of the potential barrier for spin $\uparrow$ electrons, in an

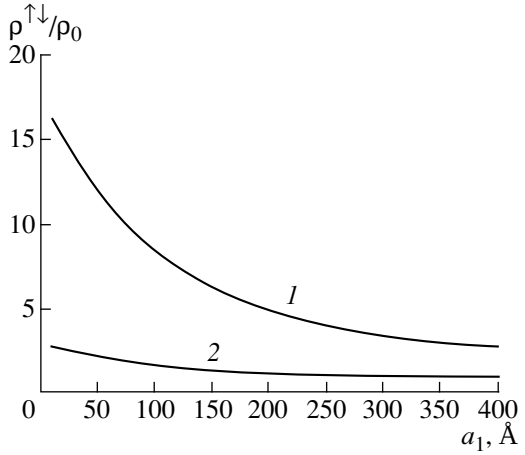


Fig. 2. Dependence of the normalized Hall conductivity of a magnetic sandwich with a tunnel junction and an antiparallel configuration of layer magnetization on the magnetic layer thickness a_1 for values of the Fermi wave vector $k_{F\downarrow}$ = 0.1 Å^{-1} (curve *1*), and $k_{F\downarrow}$ = 0.2 Å^{-1} (curve *2*); $k_{F\uparrow}$ = 1 Å^{-1}. The thickness of the other layers is a_2 = 10 Å, a_3 = 15 Å, mean free path $l_\uparrow$ = 100 Å, $l_\downarrow$ = 60 Å, and Q = 1 Å^{-1}.

antiparallel geometry an electron tunneling from one magnetic layer to another is converted to a different subband. Then the transmission coefficient for both directions of spin is the same and is related to the density of states of spin $\uparrow$ electrons and to the density of states of spin $\downarrow$ electrons, and thus depends more strongly on $k_{F\downarrow}$ and has a lower value. In this case, the effective field in the magnet differs more strongly from the field in the dielectric layer since their ratio is proportional to the transmission coefficient of the potential barrier. This leads to an increase in the size effect. We can postulate that the reasons for the appearance of an extraordinarily high Hall resistivity in granular magnetic alloys [21] are associated with the effects described.

It has thus been shown that in a multilayer magnetic structure with a tunnel junction, the anomalous Hall effect not only includes the classical size effect but another two new effects: a geometric size effect associated with the distribution of the effective electric field when the current flows perpendicular to the structure layers and the influence of the strong field of the dielectric layer.

APPENDIX

The expression for the Hall resistivity of a sandwich with antiparallel magnetized magnetic layers is given by

$$\rho^{\uparrow\downarrow} = \frac{D}{4B}\lambda\frac{m}{\pi^2\hbar^2}\frac{1}{(a_1+a_3+a_2)(l_\uparrow k_{F\uparrow}^2 + l_\downarrow k_{F\downarrow}^2)}$$

$$\times\left\{(a_1-a_3)(l_\uparrow k_{F\uparrow}^3 - l_\downarrow k_{F\downarrow}^3) + \frac{9}{4}\left[\frac{l_\uparrow^2}{k_{F\uparrow}^4}(J_\uparrow(a_1)-J_\uparrow(a_3)) - \frac{l_\downarrow^2}{k_{F\downarrow}^4}(J_\downarrow(a_1)-J_\downarrow(a_3))\right] + \frac{3}{4I_2^{\uparrow\downarrow}}\right.$$

$$\left.\times\left[\frac{l_\uparrow^2}{k_{F\uparrow}}(I_1^{\uparrow\downarrow}(a_1)-I_1^{\uparrow\downarrow}(a_3)) - \frac{l_\downarrow^2}{k_{F\downarrow}}(I_3^{\uparrow\downarrow}(a_1)-I_3^{\uparrow\downarrow}(a_3))\right]\right\},$$

where

$$J_\uparrow(a_{1,3}) = \int_0^{k_{F\uparrow}}\frac{k^3dk}{c_\uparrow}\int_0^{k_{F\uparrow}}\tilde{k}d\tilde{k}\tilde{c}_\uparrow^2$$

$$\times\left[\frac{\tilde{c}_\uparrow^2\exp(-2\tilde{d}_\uparrow a_{1,3}) - c_\uparrow^2\exp(-2d_\uparrow a_{1,3})}{\tilde{c}_\uparrow^2 - c_\uparrow^2}\right],$$

$$J_\downarrow(a_{1,3}) = \int_0^{k_{F\downarrow}}\frac{k^3dk}{c_\downarrow}\int_0^{k_{F\downarrow}}\tilde{k}d\tilde{k}\tilde{c}_\downarrow^2$$

$$\times\left[\frac{\tilde{c}_\downarrow^2\exp(-2\tilde{d}_\downarrow a_{1,3}) - c_\downarrow^2\exp(-2d_\downarrow a_{1,3})}{\tilde{c}_\downarrow^2 - c_\downarrow^2}\right],$$

$$I_1^{\uparrow\downarrow}(a_{1,3}) = \int_0^{k_{F\uparrow}}\frac{k^3dk}{c_\uparrow}\int_0^{k_{F\uparrow}}\tilde{k}d\tilde{k}\tilde{q}^2\exp(-2\tilde{q}a)\tilde{c}_\uparrow^2\tilde{c}_\downarrow$$

$$\times\left[1+\frac{\tilde{c}_\uparrow^2\exp(-2\tilde{d}_\uparrow a_{1,3}) - c_\uparrow^2\exp(-2d_\uparrow a_{1,3})}{c_\uparrow^2 - \tilde{c}_\uparrow^2}\right],$$

$$I_3^{\uparrow\downarrow}(a_{1,3}) = \int_0^{k_{F\downarrow}}\frac{k^3dk}{c_\downarrow}\int_0^{k_{F\downarrow}}\tilde{k}d\tilde{k}\tilde{q}^2\exp(-2\tilde{q}a)\tilde{c}_\downarrow^2\tilde{c}_\uparrow$$

$$\times\left[1+\frac{\tilde{c}_\downarrow^2\exp(-2\tilde{d}_\downarrow a_{1,3}) - c_\downarrow^2\exp(-2d_\downarrow a_{1,3})}{c_\downarrow^2 - \tilde{c}_\downarrow^2}\right],$$

$$I_2^{\uparrow\downarrow} = \int_0^{k_{F\downarrow}}kdkq^2c_\uparrow c_\downarrow\exp(-2qa_2).$$

The Hall resistivity for a parallel magnetization configuration is given by:

$$\rho^{\uparrow\uparrow} = \frac{D}{4B}\lambda\frac{m}{\pi^2\hbar^2}$$

$$\times\frac{1}{(a_1+a_3+a_2)(l_\uparrow k_{F\uparrow}^2 + l_\downarrow k_{F\downarrow}^2)(I_2^\uparrow + I_2^\downarrow)}$$

$$\times\left\{2(a_1+a_3)(I_2^\uparrow l_\uparrow k_{F\uparrow}^3 - I_2^\downarrow l_\downarrow k_{F\downarrow}^3) - \frac{3}{2}(I_2^\uparrow I_\uparrow^2 k_{F\uparrow}^3 - I_2^\downarrow l_\downarrow^2 k_{F\downarrow}^3)\right.$$

$$\left.+\frac{9}{2}\left[\frac{I_2^\uparrow l_\uparrow^2}{k_{F\uparrow}^4}(J_\uparrow(a_1)+J_\uparrow(a_3)) - \frac{I_2^\downarrow l_\downarrow^2}{k_{F\downarrow}^4}(J_\downarrow(a_1)+J_\downarrow(a_3))\right]\right.$$

$$+\frac{3}{2}\left[\frac{l_\uparrow^2}{k_{F\uparrow}(Q^2+k_{F\uparrow}^2)^2}(I''_\uparrow(a_1)+I''_\uparrow(a_3))\right.$$

$$\left.-\frac{l_\downarrow^2}{k_{F\downarrow}(Q^2+k_{F\downarrow}^2)^2}(I''_\downarrow(a_1)+I''_\downarrow(a_3))\right]\Bigg\},$$

where

$$I''_\uparrow(a_{1,3}) = \int_0^{k_{F\uparrow}}\frac{k^3dk}{c_\uparrow}\int_0^{k_{F\uparrow}}\tilde{k}d\tilde{k}\tilde{q}^2\exp(-2\tilde{q}a_2)\tilde{c}_\uparrow^3$$

$$\times\left[1+\frac{\tilde{c}_\uparrow^2\exp(-2\tilde{d}_\uparrow a_{1,3})-c_\uparrow^2\exp(-2d_\uparrow a_{1,3})}{c_\uparrow^2-\tilde{c}_\uparrow^2}\right],$$

$$I''_\downarrow(a_{1,3}) = \int_0^{k_{F\downarrow}}\frac{k^3dk}{c_\downarrow}\int_0^{k_{F\downarrow}}\tilde{k}d\tilde{k}\tilde{q}^2\exp(-2\tilde{q}a_2)\tilde{c}_\downarrow^3$$

$$\times\left[1+\frac{\tilde{c}_\downarrow^2\exp(-2\tilde{d}_\downarrow a_{1,3})-c_\downarrow^2\exp(-2d_\downarrow a_{1,3})}{c_\downarrow^2-\tilde{c}_\downarrow^2}\right],$$

$$I_2^\uparrow = \frac{1}{(Q^2+k_{F\uparrow}^2)^2}\int_0^{k_{F\uparrow}}kdkq^2c_\uparrow^2\exp(-2qa_2),$$

$$I_2^\downarrow = \frac{1}{(Q^2+k_{F\downarrow}^2)^2}\int_0^{k_{F\downarrow}}kdkq^2c_\downarrow^2\exp(-2qa_2).$$

REFERENCES

1. P. Grünberg, R. Schreiber, Y. Pang, *et al.*, Phys. Rev. Lett. **57**, 2442 (1986).
2. C. Carbone and S. F. Alvarado, Phys. Rev. B **36**, 2433 (1987).
3. M. N. Baibich, J. M. Broto, A. Fert, *et al.*, Phys. Rev. Lett. **61**, 2472 (1988).
4. A. Barthelemy, J. Childress, V. Cros, *et al.*, Phys. Low-Dimens. Struct. **3**, 17 (1994).
5. W. H. Butler, X.-G. Zhang, D. M. Nicholson, *et al.*, Phys. Rev. B **52**, 13399 (1995).
6. J. Mathon, M. Villeret, and H. Itoh, Phys. Rev. B **52**, R6983 (1995).
7. A. V. Vedyaev, N. G. Karachanskaya, O. A. Kotel'nikova, *et al.*, Vestn. Mosk. Univ. **2**, 51 (1993).
8. A. Crepieux, C. Lacroix, N. Ryzhanova, *et al.*, Phys. Lett. A **229**, 401 (1997).
9. S. Zhang, Phys. Rev. B **51**, 3632 (1995).
10. A. Crepieux, C. Lacroix, N. Ryzhanova, *et al.*, Phys. Rev. B **57**, 2943 (1998).
11. J. S. Moodera, L. R. Kinder, T. M. Wong, *et al.*, Phys. Rev. Lett. **74**, 3273 (1995).
12. Y. Suezawa and Y. Gondo, J. Magn. Magn. Mater. **109**, 79 (1992).
13. T. Miyazaki and N. Tezuka, J. Magn. Magn. Mater. **139**, L231 (1995).
14. N. Tezuka and T. Miyazaki, J. Appl. Phys. **79**, 6262 (1996).
15. T. S. Plaskett, P. P. Freitas, N. P. Barradas, *et al.*, J. Appl. Phys. **76**, 6104 (1994).
16. C. Platt, B. Dieny, and A. Berkowitz, Appl. Phys. Lett. **69**, 2291 (1996).
17. M. Julliere, Phys. Lett. A **54**, 225 (1975).
18. C. Slonszewski, Phys. Rev. B **39**, 6995 (1989).
19. A. Vedyaev, N. Ryzhanova, C. Lacroix, *et al.*, Europhys. Lett. **39**, 219 (1997).
20. A. Vedyayev, B. Dieny, and N. Ryzhanova, Europhys. Lett. **19**, 329 (1992).
21. A. B. Pakhomov, X. Yan, and Y. Xu, J. Appl. Phys. **79**, 6140 (1996).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 1042–1049.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1198–1206.

SOLIDS
Electronic Properties

High-Field Magnetoresistance and Hall Effect in $Bi_2Sr_2CuO_x$ Single Crystals¶

S. I. Vedeneev*, **, A. G. M. Jansen, and P. Wyder****

**Lebedev Physical Institute, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 117924 Russia*
**e-mail: vedeneev@sci.lebedev.ru*

***Grenoble High Magnetic Field Laboratory, Max-Planck-Institut für Festkörperforschung and Centre National de la Recherche Scientifique, B. P. 166, F-38042 Grenoble Cedex 9, France*

Received December 9, 1999

Abstract—We investigated the in-plane magnetoresistance and the Hall effect of high-quality $Bi_2Sr_2CuO_x$ single crystals with T_c (midpoint) = 3.7–9.6 K in dc magnetic fields up to 23 T. For $T < 10$ K, the crystals show the classical positive magnetoresistance. Starting at $T \approx 14$ K, an anomalous negative magnetoresistance appears at low magnetic fields; for $T \geq 40$ K, the magnetoresistance is negative in the whole studied range of magnetic fields. Temperature and magnetic field dependences of the negative-magnetoresistance single crystals are qualitatively consistent with the electron interaction theory developed for simple semiconductors and disordered metals. As is observed in other cuprate superconductors, the Hall resistivity is negative in the mixed state and changes its sign with increasing field. The linear T-dependence of $\cot\theta_H$ for the Hall angle in the normal state closely resembles that of the normal-state resistivity as expected for a Fermi liquid picture. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The magnetic properties of high-T_c superconductors (HTSC), especially of the highly anisotropic Bi- and Tl-layered superconductors, are characterized by anomalous quasi-two-dimensional (2D) states, which have been very extensively studied in recent years [1]. In theoretical and experimental investigations of high-T_c superconductivity, great interest has been focused on the analysis of normal-state properties, because there is a widespread viewpoint that the knowledge of these properties can help to understand the exotic superconducting and magnetic properties of HTSC. However, the normal-state properties of HTSC also have several unusual features. One of these is the coexistence of the metallic in-plane resistivity ρ_{ab} and the "semiconducting" out-of-plane resistivity ρ_c (see, e.g., [2, 3]). Recently, this behavior of ρ_{ab} and ρ_c was measured by Y. Ando *et al.* [4] in La-doped $Bi_2Sr_2CuO_y$ ($T_c \simeq 13$ K) down to temperatures as low as $T/T_c \sim 0.04$. This implies a 2D confinement and is incompatible with Fermi liquid behavior. The temperature dependence of the normal-state Hall coefficient R_H is too strong to be easily understood. In addition, near T_c, there is a sign change of the Hall effect as a function of the magnetic field or temperature. Moreover, the normal-state properties of HTSC essentially depend on the carrier concentration or doping, and an additional quantum critical point in the HTSC phase diagram has been proposed to explain this dependence. It should describe a transition between the non-Fermi-liquid corresponding to hole doping around and below the optimal doping and the Fermi liquid in the overdoped regime.

A number of theories have been suggested to describe the unusual transport properties of HTSC. Anderson [5] has suggested that anomalous features of magnetotransport data can be explained by a spin-charge separation in CuO_2 planes resulting in two different relaxation rates, the transport relaxation rate and the Hall relaxation rate. The first gives the linear-T resistivity ρ_{xx}, while the second gives the Hall angle dependence

$$\cot\theta_H = \rho_{xx}/\rho_{xy} = aT^2.$$

The Hall coefficient

$$R_H = \rho_{xy}/B$$

is to have a $1/T$-dependence. The different temperature dependences of ρ_{xx} and θ_H provide evidence for a non-Fermi-liquid transport. Both the experimental evidence and a deviation from this suggestion were found for $YBa_2Cu_3O_7$, $La_{2-x}Sr_xCuO_4$, and $Bi_2Sr_2CaCu_2O_8$ for different doping. There are more conventional models that are based on an anisotropy of the Fermi surface and the scattering rate. In the latter case, the magnitude and temperature dependence of the scattering rate are different on different parts of the Fermi surface [6]. At present, no single theory is able to account for all the anomalies found in the normal-state properties of HTSC and it is unclear whether these properties can be described by the conventional Fermi liquid theory. Although many experiments have been done with the aim of understanding the normal-state properties, sev-

¶This article was submitted by the authors in English.

eral basic measurements, such as dc transport, are not well understood. Previous studies of dc magnetotransport in HTSC [7–11] also showed contradictory results. The mechanism responsible for the anomalous Hall effect and its temperature dependences are still unclear.

The low-T_c phase $Bi_2Sr_2CuO_6$ (Bi2201) of Bi-based HTSC is an ideal material for measuring the magnetotransport properties in cuprates at low temperatures, because it has a very low T_c, while its structure and properties are closely related to HTSC. At the same time, pure Bi2201 single crystals are difficult objects for investigation, because they are nonstoichiometric and, as a rule, imperfect. For this reason, the greater part of the measurements of physical properties have been carried out on La-doped Bi2201 samples. The introduction of La allows one to obtain single-phase samples and modify the hole concentration. Unfortunately, this doping deteriorates the crystal quality. Details of nondoped Bi2201 single crystals remain unexplored up to now. Recalling the debate about normal-state properties of cuprates, a study of magnetotransport in pure Bi2201 single crystals is clearly desirable. In this paper, we present magnetotransport measurements in high-quality nondoped Bi2201 single crystals in the normal and mixed states under magnetic fields up to 23 T.

2. EXPERIMENTAL

Sr-deficient $Bi_2Sr_2CuO_x$ single crystals were grown in a gaseous phase in big closed cavities of a KCl solution-melt [12]. The number of crystals reached several tens in the cavity, and their size was around 0.5–2.5 mm × 0.4–2 mm × 1.5–5 μm. These freely grown single crystals are free of the appearance of strains during the growth process and when cooling to room temperature. Properties of the crystals grown inside the same cavity were closely similar. The high quality of the crystals was verified by measuring the dc resistance and ac susceptibility and by X-ray diffraction and scanning electron microscopy. The composition of the crystals was studied using a Philips CM-30 electron microscope with a Link analytical AN-95S energy dispersion X-ray spectrometer at the Lebedev Institute and at the Laboratory of Solid State Physics, University of Groningen, The Netherlands. The half-width of the main reflections in the X-ray rocking curves for single crystals consisting of two or three blocks did not exceed 0.3°, whereas for crystals consisting of one block only (with the dimensions 0.3 × 0.3 mm), it was less than 0.1°. Because this value is close to the resolution limit of the diffractometer, we analyzed the crystal perfection using the supercell parameters that are more sensitive to structural imperfections. Crystals with higher T_c contain no low-angle boundaries. Well-separated $K_{\alpha 1} - K_{\alpha 2}$ lines of the supercell reflections in the X-ray diffraction profiles and very narrow rocking curves imply structural perfection of the crystals [12]. As far as we know, this is the best among the HTSC cuprates. Our single crystals have a linear relationship between the T_c value and the monoclinic superlattice angle value [12], which is directly related to the concentration of the carriers. The values of the lattice parameters were in the range a = 5.360–5.385 Å and c = 24.60–24.638 Å according to the magnitude range of T_c [13]. Low-resistance contacts were made on the samples using evaporated and fired-on gold films. A six-probe contact configuration with a symmetric position of contacts on the *ab*-surface of the sample was used for measuring the in-plane (R_{xx}) and Hall (R_{xy}) resistances. Measurements of R_{xx} and R_{xy} on different pairs of voltage contacts were simultaneously performed by turning the crystal through 180°. To transform the measured resistances into specific resistivities ρ_{xx} and ρ_{xy}, the sample dimensions were determined by a high-resolution optical microscope.

The investigated crystals had critical temperatures T_c = 4–9.5 K (midpoint transitions) with ΔT_c = 0.7–1.5 K (10–90% points of the transition). The T_c value of the crystals formed by our free-growth method ranges up to 13 K, but the crystals with $T_c \le 9.5$ K have better structural perfection. It should be noted that the onset temperature of superconducting transitions and the transition widths for the dc resistance and the ac susceptibility were very close. In the dc method employed for the resistance measurements, the current was about 50–100 μA. The crystals were studied with the magnetic field **B** applied parallel to the **c**-axis. A configuration with the in-plane transport current **J** perpendicular to **B** was used. High-magnetic-field measurements were carried out in continuous magnets at the High Magnetic Field Laboratory in Grenoble.

3. RESULTS AND DISCUSSION

3.1. Resistivity and Transverse Magnetoresistance

The zero-field temperature dependence of the in-plane resistivity ρ_{ab} of our Bi2201 single crystals with T_c < 5 K and T_c = 8–9.5 K was described in considerable detail earlier in [14, 15], respectively. As an example, we display in Fig. 1 the typical temperature dependences of the in-plane (ρ_{ab}) and out-of-plane (ρ_c) resistivities for two single crystals with T_c (midpoint) = 4.6 K (*A*) and 9 K (*B*) under zero magnetic field. The inset in Fig. 1 shows an expanded scale of the low-temperature $\rho_{ab}(T)$ data near T_c. The crystals with T_c = 8–9.5 K (type *B*) showed a nearly linear temperature dependence $\rho_{ab}(T)$, whereas the crystals with T_c = 4–6 K (type *A*) had a linear $\rho_{ab}(T)$ at high temperatures that saturated to a residual resistivity ρ_0 below 20–40 K. The slope $\Delta\rho_{ab}/\Delta T$ of 0.5–1.5 μΩ cm/K was obtained at high temperatures. In the present measurements, only the crystals with the residual resistivities ρ_0 = 50–140 μΩ cm were investigated. In samples with higher T_c, smaller ρ_0 and larger $\Delta\rho_{ab}/\Delta T$ values were observed. As the T_c value of the crystal decreases, the temperature dependence of the resistivity changes from

$$\rho_0 + AT + BT^2$$

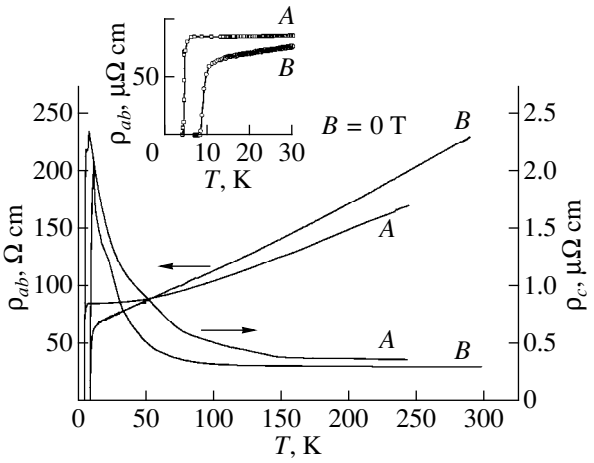


Fig. 1. Temperature dependences of the in-plane ρ_{ab} and out-of-plane ρ_c resistivities for two single crystals with T_c (midpoint) = 4.6 (*A*) and 9 K (*B*) under zero magnetic field. The inset shows an expanded scale of the low-temperature $\rho_{ab}(T)$ data near T_c.

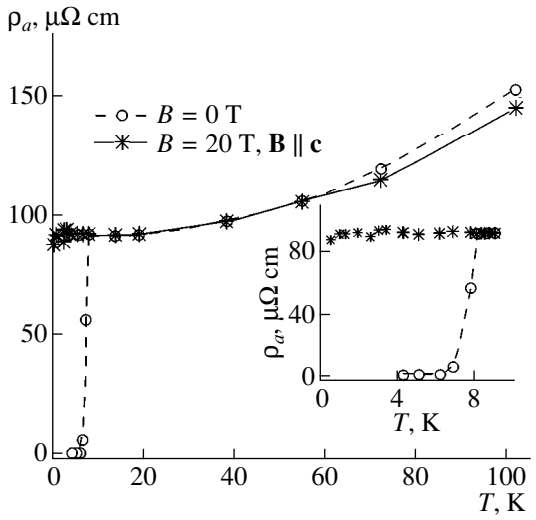


Fig. 2. Main panel: the zero-field resistivity vs temperature for one of the crystals (open circles). The stars show the resistivity data at B = 20 T taken from the resistivity saturation region. The inset shows an expanded scale of the low-temperature data.

with the strong linear term for the type *B* crystals to

$$\rho_0 + AT + BT^2 + CT^3$$

for the type *A* samples. Recently [16], it was also shown for $Tl_2Ba_2CuO_{6+\delta}$ single crystals that as the doping increases, the temperature dependence of ρ_{ab} changes from linear in the optimally doped samples to $\rho_{ab}(T)$ with a considerably larger curvature in overdoped samples with reduced T_c. In our single crystals, there is a linear relationship between the T_c value and a monoclinic superlattice angle value [12] that is directly related to the concentration of carriers [17]. The out-of-plane resistivity $\rho_c(T)$ of Bi2201 single crystals varies as $T^{-\alpha}$ with α = 0.7–1.6 over the temperature region T = 4–300 K. The largest anisotropy ratio ρ_c/ρ_{ab} is 5.3×10^4 at T = 0.4 K.

In the two-dimensional model, from the effective sheet resistance per CuO_2 layer, the residual resistivity ρ_0, and the interlayer spacing (c/2), we determined the disorder parameter values [18]

$$(k_F l)_{ab} = (c/2)h/e^2\rho_0 \approx 60\text{–}20$$

for two limiting values ρ_0 = 50 and 140 μΩ cm, respectively (k_F and l represent the Fermi wave vector and the elastic scattering length in the *ab*-plane). We measured the normal-state Hall coefficient R_H in the temperature region 4.2–50 K (see below) and determined the carrier density in our crystals $n = (4.8\text{–}6.3) \times 10^{21}$ cm^{-3}. The carrier density in the lower T_c samples was larger than that in samples with T_c = 9 K. According to the data obtained by Tsvetkov *et al.* [19], in optical experiments with our Bi2201 single crystals, the value of the carrier density equals $n = 5.1 \times 10^{21}$ cm^{-3} at 10 K and the effective mass in the *ab*-plane is $m^* = 3m_0$. Based on the Fermi liquid theory and the assumption of a cylindrically shaped Fermi surface with a highly anisotropic dispersion relation, Kresin and Wolf [20] proposed a model to analyze the normal-state properties of the cuprates. If we take the Sommerfeld constant $\gamma \approx 9.2$ mJ mol^{-1} K^{-2} from the heat-capacity data for Bi2201 [21] and $E_F = (\pi^2 k_B^2/3)n/\gamma$ [20], we obtain with the averaged value $n \approx 5.6 \times 10^{21}$ cm^{-3}, the Fermi velocity $\overline{v_F} \approx 1.7 \times 10^7$ cm/s, $\overline{k_F}$ = 0.45 Å^{-1}, and $l \approx$ 130–50 Å for $(k_F l)_{ab}$ = 60–20. It should be noted that $\rho_{ab}(0)$ and $\Delta\rho_{ab}/\Delta T$ were obtained using the Montgomery method. As shown in [22], however, they can be twice the values obtained by means of the Van der Pauw 4-probe methods. It is known that this discrepancy is due to a strong anisotropy of the layered cuprates. It has been shown [23] that in $YBa_2Cu_3O_7$ with $\rho_c \gg \rho_{ab}$, the absolute value of ρ_{ab} measured by the contacts on one side of the single crystal is roughly two times greater than that obtained with the current and voltage contacts on opposite sides of the sample. A similar phenomenon was also observed in our specimens [15]. This uncertainty may far exceed the 10% error that is due to measurements of the crystal thickness.

Some of the superconducting properties and the magnetic phase diagram of our Bi2201 crystals were described in detail earlier [15]. In the study of the superconducting resistive transition of the crystals in magnetic fields, we observed that all transition curves in the magnetic field saturated at temperatures down to 0.4 K. In Fig. 2 (main panel), the zero-field resistivity is plotted as a function of the temperature (open circles) for one of the studied crystals. The stars represent the data points for $\rho_{ab}(T)$ measured in the magnetic field

$B = 20$ T. At temperatures $T < T_c$, the resistivity data were determined from the resistive transition as a function of the magnetic field in the saturated portion of the $\rho_{ab}(B)$ curves. This presented no special problems, because the resistive transitions in the normal state for a given sample are completed at 20 T and all the $\rho_{ab}(B)$ curves saturate to the same value, even at the lowest experimental temperature (see Fig. 5). The inset in Fig. 2 shows an expanded scale of the low-temperature data. The high-magnetic-field measurements testify that the *ab*-plane resistivity in the normal state shows the ordinary behavior for a metal down to 0.4 K. Analogous results were obtained on the low-T_c phases of Tl2201 [24] and La-doped Bi2201 [4]. The question of the coexistence of metallic in-plane resistivity with "semiconducting" out-of-plane resistivity down to low temperatures has been attracting much attention recently, and several models have been proposed to explain it.

In Fig. 3, we have plotted the magnetoresistance data

$$\Delta\rho/\rho(0) = [\rho(B) - \rho(0)]/\rho(0)$$

for four samples at various temperatures that were extracted from the experimental curves ρ_{ab} vs. B. The zero-field transition temperature regions are equal to 3–4.5 K, 3.8–4.3 K, 6–7.5 K, and 6.8–8 K for the respective samples nos. 8, 3, 10, and 11. At relatively low temperatures just above T_c, we observed a positive magnetoresistance with a quadratic dependence up to about 8 T. At higher fields, the magnetoresistance saturates (for lower T_c samples) or goes over to a linear dependence (for higher T_c samples, Fig. 3). The magnetoresistance magnitude in the low-T_c samples with larger ρ_0 was nearly three times greater than that in the samples with $T_c = 8$–9 K and lower ρ_0. A simple estimate shows that the weak-field regime is realized in magnetic fields up to 20 T even at low temperatures. Considering the quadratic dependence of $\Delta\rho/\rho(0)$, it can then be believed that in the *ab*-plane at low temperatures in the magnetic fields **B** || **c**, Bi2201 behaves like a conventional many-band metal with closed electron trajectories in **k** space. "Saturation" of the magnetoresistance in low-T_c samples is most likely caused by a magnetic breakdown between semiclassical trajectories closely disposed in **k** space. On the other hand, it is quite possible that for the low-T_c samples, the intermediate-magnetic-field regime is realized and the magnetoresistance in Fig. 3 actually saturates in accordance with the models based on the anisotropy of the Fermi surface, the effective mass, and the scattering rate [6]. The physical origin of the weak- and intermediate-field magnetoresistance in HTSC has been studied in detail in $Tl_2Ba_2CuO_{6+\delta}$ single crystals [25]. Our results obtained for the lower T_c samples are in qualitative agreement with these experiments. The obvious source of the positive magnetoresistance at low temperatures is the ordinary suppression of superconducting fluctuations. However, this effect is probably very insignificant, because the magnetoresistance shown in Fig. 3 continues to increase under magnetic fields that are larger than the upper critical fields of Bi2201 [15].

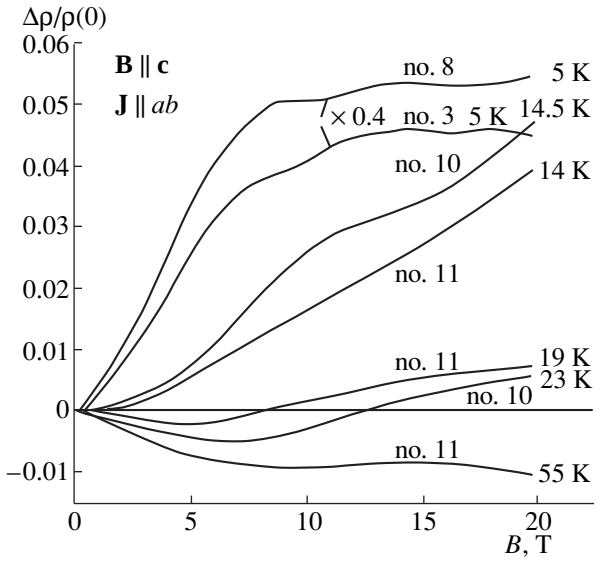


Fig. 3. Transverse (**B** || **c**) magnetoresistance data for four samples at various temperatures just near and above T_c. The zero field T_c regions are equal to 3–4.5 K, 3.8–4.3 K, 6–7.5 K, and 6.8–8 K for the respective samples nos. 8, 3, 10, and 11.

At $T < 40$ K, the magnetoresistance decreases rapidly with increasing temperature in all the crystals studied. Starting at $T \approx 14$ K (Fig. 3), an anomalous negative magnetoresistance appears at low magnetic fields; for $T \geq 40$ K, the magnetoresistance is negative in the whole studied range of magnetic fields. The absolute magnitude of the magnetoresistance increases with temperature but decreases again at higher temperatures (≥100 K). In the temperature range 40–135 K, the negative magnetoresistance varies as B^2 for low fields, is proportional to $\sqrt{B}$ at intermediate fields, and is better described by the $\ln B$-dependence for high fields. At very high magnetic fields, the magnetoresistance tends gradually towards saturation.

Converse results were obtained in earlier measurements of the transverse magnetoresistance in nonsuperconducting $Bi_2Sr_2CuO_6$ crystals with the field normal to the *ab*-plane [26]. In [26], at magnetic fields of 0–8 T, the magnetoresistance was negative in the temperature range 0.5–20 K and became positive with increasing temperature above 20 K. To account for these results, the authors invoked the localization theory [18], which describes the low-temperature negative magnetoresistance in conventional metals in a weak-localization regime. Since all wave functions are localized in disordered electron systems, the zero-field conductivity decreases with decreasing temperature and vanishes at zero temperature. The negative magnetoresistance resulted from magnetic field suppression of localiza-

tion effects. The zero-field temperature dependence of the *ab*-plane resistivity in the studied samples [26] had a "semiconducting" form at T below 20 K, and localization effects should be very important, especially at low temperatures. In our crystals in the zero magnetic field, on the other hand, we observed a nearly linear temperature dependence $\rho_{ab}(T)$ (samples with T_c = 8–9.5 K) or a weak departure from it at low temperatures (samples with T_c = 4–6 K). In fact, localization theory [18] predicts a linear relationship between $\Delta\sigma$ and $\sqrt{B}$ at relatively high fields, with the temperature-independent slope σ-vs-$\sqrt{B}$. The $\Delta\sigma$ data are shown in Fig. 4a as a function of $\sqrt{B}$ for various temperatures. One can see that in the region where the $\sqrt{B}$-dependence of $\Delta\sigma$ is close to linear, the $\Delta\sigma$-vs-$\sqrt{B}$ slope is different at different temperatures. In addition, a simple estimation shows that the values of $\Delta\sigma$ are much larger than those predicted by the theory [18]. It might be assumed, therefore, that localization effects play a minor role in the conductivity of our samples.

On the other hand, an effort can be made to explain, at least qualitatively, the behavior of the negative magnetoresistance in our samples at high fields for $T > 15$ K on the basis of the interaction effects [18]. Altshuler *et al.* [27] considered the Coulomb electron–electron interaction effect on the magnetic field dependence of the conductivity in disordered systems. Electron interaction theories, as well as the theory related to localization, lead to similar B-dependences of the magnetoresistance. In the three-dimensional (3D) case, the conductivity change with the magnetic field is expected to be linearly related to B^2 for low fields and to $\sqrt{B}$ for high fields. In the 2D case, the $\sqrt{B}$-dependence is replaced by the $\ln B$-dependence. Recalling that we are dealing with a quasi-two-dimensional system, we can explain the observed $\ln B$-dependence of the magnetoresistance. Localization leads to a negative magnetoresistance, whereas the electron gas interaction due to spin splitting and orbital effects leads to a positive magnetoresistance [18]. The sign of the interaction magnetoresistance caused by the influence of the magnetic field on the kinetic coefficients is determined by the interaction constant sign, and negative magnetoresistance indicates an attractive electron interaction [27]. This magnetoresistance is already essential at classical small fields and can be observed even in relatively clean metals at high temperatures such that $\hbar/\tau \gg k_B T$, where τ is the electron relaxation time [27]. It follows from the interaction theories that $\Delta\sigma$ is also proportional to $\sqrt{B}$ but $\sigma(B, T)$ has a $\sqrt{T}$-dependence at a sufficiently high fixed field [18, 27]. The data in Fig. 4b show a clear-cut linear $\sqrt{T}$-dependence of $\sigma(B, T)$. The interaction theory predicts that in high fields, the magnetoresistance depends on the electron–electron interaction constant only and that it saturates at a B value such that

$$\omega_c \tau \approx k_B T / E_F,$$

where ω_c is the cyclotron frequency

$$\omega_c = eB/m^*,$$

that is, at

$$B \approx 2k_B T / v_F^2 e\tau.$$

We determined the electron relaxation time

$$\tau(T) = m^*/ne^2\rho_{ab}(T)$$

from the data for the carrier density n and resistivity ρ_{ab} in this sample (no. 11). Taking into account that

$$\Delta\sigma/\sigma_0 \approx \Delta\rho/\rho_0,$$

we next found that in the temperature region 35–100 K, the saturation in the magnetoresistance is observed at

$$B = (0.2\text{–}0.3)T,$$

where B is measured in T and T in K. The values of B estimated from this relation are 8, 13, 19, and 30 T at the respective temperatures T = 38, 55, 72, and 102 K. These are in good agreement with the observed experimental values of $\Delta\sigma(B, T)$ in Fig. 4.

We cannot perform a quantitative analysis of the interaction correction to the conductivity $\Delta\sigma_{int}(T)$, because the interaction theory [27] has been developed for simple semiconductors and disordered metals, whereas Bi2201 is a high-anisotropic layered compound. However, the qualitative behavior of negative-magnetoresistance Bi2201 single crystals is consistent with the attractive electron interaction considered in [18, 27]. This is indicative of a tendency to electron attraction in the temperature region where the high-T_c superconducting phases Bi2212 and Bi2223 exist, in spite of the lack of these phases in our Bi2201 single crystals.

3.2. Mixed- and Normal-State Hall Effect

As pointed out above, previous experiments with $YBa_2Cu_3O_7$ [28, 29], $Tl_2Ba_2CaCu_2O_{8+\delta}$ [30], and $HgBa_2CaCu_2O_6$ [31] have shown a sign-reversal anomaly of the Hall coefficient R_H at T near T_c. A large number of models have been proposed to describe this phenomenon, and a consensus regarding their origin has evidently been attained. Dorsey and Kopnin *et al.* [32] have found that a flux-flow Hall conductivity is the sum of quasiparticle and vortex parts

$$\sigma_H = \sigma_H^n + \sigma_H^f,$$

with σ_H^f having the same sign for any vortex direction [32] and σ_H^n being independent of the pinning. The sign change during measurements of the Hall resistivity

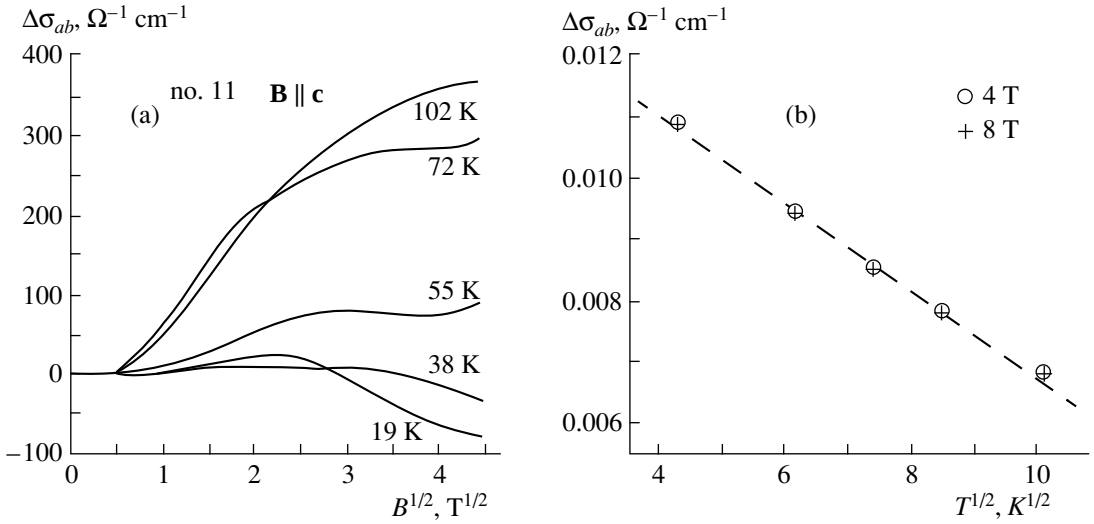


Fig. 4. (a) Transverse (**B** ‖ **c**) magnetoconductivity data $\Delta\sigma$ as a function of $\sqrt{B}$ for sample no. 11 at various temperatures and (b) the linear $\sqrt{T}$-dependence of σ at 4 and 8 T.

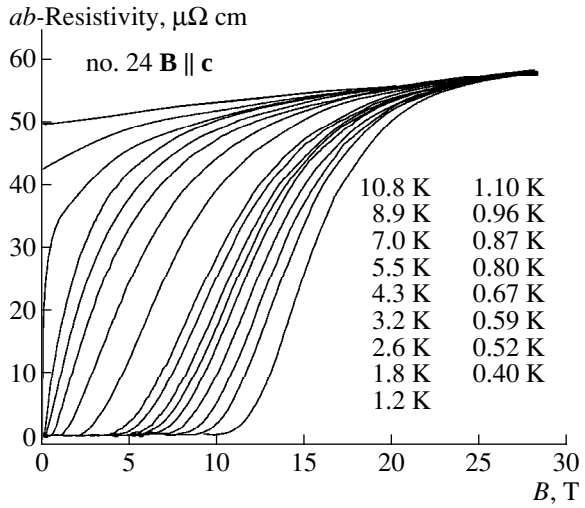


Fig. 5. Resistive transitions of crystal no. 24 in a magnetic field **B** ‖ **c** at different temperatures.

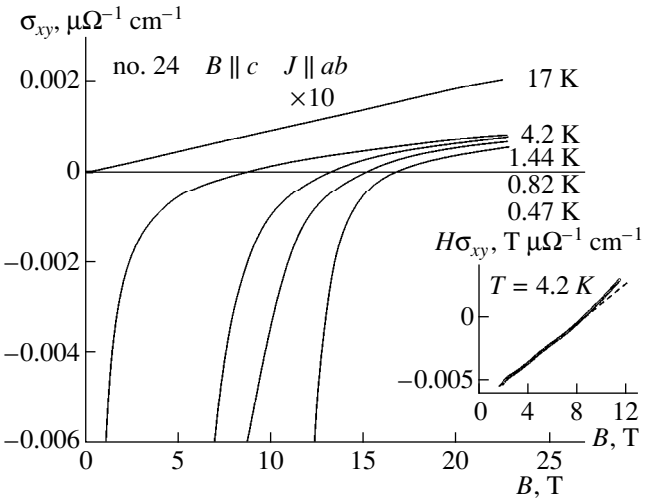


Fig. 6. The Hall conductivity σ_{xy} vs. **B** ‖ **c** at various temperatures below the zero field T_c (crystal no. 24). The inset shows the magnetic field dependence of the product $B\sigma_{xy}$ at $T = 4.2$ K along with a linear data extrapolation (crystal no. 24).

$\rho_{xy}(B)$ is related to exceeding positive σ_H^n relative to negative σ_H^f.

We have measured the field dependence of the mixed-state Hall effect on one of our Bi2201 single crystals (no. 24) with $T_c = 8.7$–9.5 K at various temperatures down to 0.4 K. In Fig. 5, we report the resistive transitions of this crystal in a magnetic field directed perpendicular to the *ab*-plane at different temperatures. In spite of a strong broadening of the magnetic transitions, one can see that the resistive transitions in the normal state are completed at 25 T, even at the lowest experimental temperature. We find that the Hall resistivity $\rho_{xy}(B)$ at $T < T_c$ is negative with a broad minimum at low fields and is hole-like at high fields in the normal state.

The theories developed for viscous flux motion [32–34] have shown that the low-field Hall conductivity

$$\sigma_{xy} = \rho_{xy}/\rho_{xx}^2$$

(at $\rho_{xy} \ll \rho_{xx}$) is negative in the mixed state and inversely proportional to B, while the normal-state σ_{xy} is positive and proportional to B. We converted ρ_{xy} into the Hall conductivity, and as can be seen from Fig. 6 showing the upper parts of the $\sigma_{xy}(B)$ curves, our results

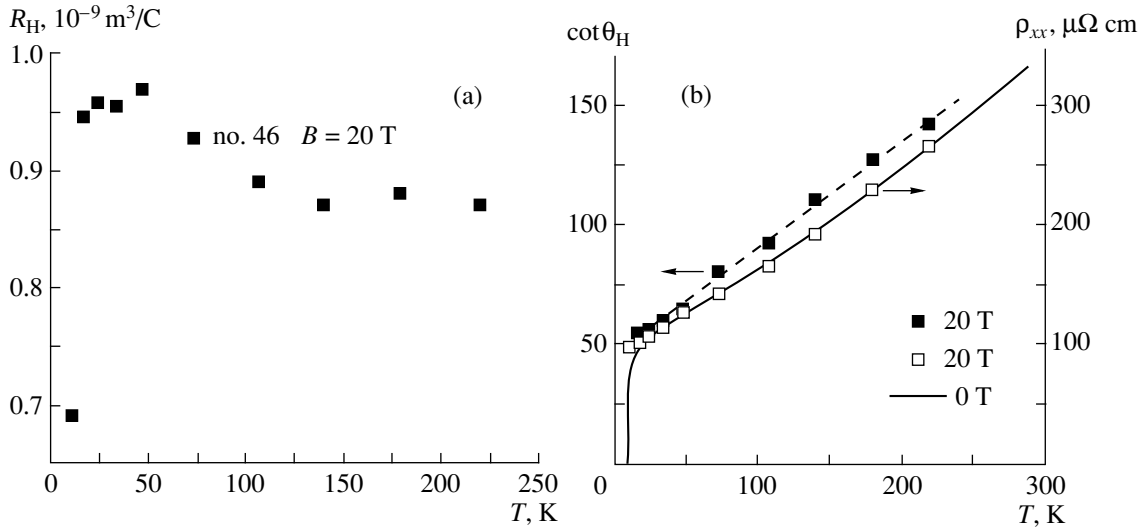


Fig. 7. Temperature dependence of the normal-state Hall constant R_H at 20 T (a) and the temperature dependence of the Hall angle $\cot\theta_H$ at 20 T (solid symbols), together with ρ_{xx} at 0 and 20 T (b) for single crystal no. 46 (T_c = 8.7–10.5 K).

are in agreement with the above-mentioned theories for viscous flux flow. In the inset, the magnetic field dependence of the product $B\sigma_{xy}$ at T = 4.2 K is shown along with a linear extrapolation. These results are also in complete agreement with similar experimental data obtained in untwinned $YBa_2Cu_3O_7$ crystals [29] and $Tl_2Ba_2CaCu_2O_{8+\delta}$ thin films [30].

Several features of the normal-state Hall effect were already mentioned in the Introduction. To these must be added a nonregular change of the temperature dependence of the Hall coefficient $R_H(T)$ at different carrier concentrations that leads to considerable difficulties in interpreting the experimental data [16]. Anderson [5] has shown that plotting $\cot\theta_H = \rho_{xx}/\rho_{xy}$ vs. T is a better way to describe the transport properties of highly correlated systems. By separating relaxation rates of carriers between their motion, Anderson's theory predicts that $\cot\theta_H$ can be expressed as

$$\cot\theta_H = a + bT^2,$$

where a is independent of the doping level and b depends on it. The evidence for this behavior has been obtained in experiments with many cuprates. At the same time, a significant deviation from the T^2-dependence for $\cot\theta_H$ was observed at temperatures below 25 K in $Tl_2Ba_2CuO_{6+\delta}$ [16], below 100 K in overdoped $La_{2-x}Sr_xCuO_4$ [35], and in La-doped Bi2201 [36].

Figure 7a shows the temperature dependence of the normal-state Hall constant $R_H = \rho_{xy}/B$ for single crystal no. 46 (T_c = 8.7–10.5 K) at 20 T. At higher temperatures, R_H is nearly independent of the temperature in the normal state, except for a slight broad maximum in the region 25–100 K. In spite of a considerable body of theoretical works devoted to the temperature dependence of R_H in HTSC, the mechanism responsible for this maximum, which is characteristic of cuprates [37], remains to be explored. The decrease in R_H with temperature, which is in agreement with other reports, is most likely caused by the proximity of the superconducting transition and superconducting fluctuations. The magnitude and the temperature dependence of R_H obtained here agree closely with the experimental data for La-doped Bi2201 [35] and nondoped Bi2201 [37] single crystals.

In Fig. 7b, we report the temperature dependence of the Hall angle $\cot\theta_H$ at 20 T showing the linear dependence (dashed line), together with ρ_{xx} at 0 (solid line) and 20 T (symbols). A linear temperature dependence is found for $\cot\theta_H$; it is the same as the temperature dependence of the resistivity. This behavior is expected for a metallic system in the low-field limit for uniform scattering on the Fermi surface.

One can see that in optimally La-doped Bi2201 polycrystal samples [13] and films [38] with the maximum value $T_c \simeq 25$ K for Bi2201, a carrier density near $n = 3 \times 10^{21}$ cm^{-3} was obtained. Although the dependence of T_c on the carrier density in nondoped Bi2201 is different from that in La-doped Bi2201, it should be assumed that our crystals are overdoped because of a significant excess of Bi. On the other hand, as reported in [35], the carrier concentration in underdoped $Bi_2Sr_{2-x}La_xCuO_{6+\delta}$ single crystals with T_c = 13 K is similar. Because the largest value of T_c in our pure Bi2201 single crystals approximately equals 13 K, the samples studied here with T_c = 8.7–10.5 K are most likely to be near the optimal doping. Taking a relatively large value of the elastic scattering length into account, this assumption seems plausible. Moreover, the complementary measurements of our Bi2201 single crystal composition performed at the Material Science Center, Uni-

versity of Groningen (The Netherlands) have shown that our crystals are slightly underdoped by the oxygen depletion. We presented here the Hall effect data from two samples with the highest T_c, but all basic features were also observed in other samples.

In summary, we have studied both the in-plane magnetoresistance and the Hall effect in several high-quality $Bi_2Sr_2CuO_x$ single crystals with the critical temperature T_c (midpoint) = 3.7–9.6 K in dc magnetic fields up to 23 T. We found that the crystals show the classical positive magnetoresistance for temperatures $T < 10$ K. Above 14 K, an anomalous negative magnetoresistance appears that is qualitatively discussed in view of the electron–electron interaction effects in charge transport. As observed for other cuprate superconductors, the Hall resistivity is negative in the mixed state and changes its sign with increasing field. The linear T-dependence of $\cot\theta_H$ for the Hall angle in the normal state closely resembles that of the normal-state resistivity, as expected for a Fermi liquid picture.

ACKNOWLEDGMENTS

The authors are grateful to V.F. Gantmakher of the Institute of Solid State Physics, Russian Academy of Sciences, for stimulating discussions. We would like to thank V.P. Martovitskiĭ for the careful X-ray studies of the single crystals. One of us (S. I. V.) was partially supported by the Russian Ministry of Science and Technical Policy within the Program "Actual Problems of Condensed Matter Physics" (project no. 96001) and by the Russian Foundation for Basic Research (project no. 99-02-17877).

REFERENCES

1. Wu Liu, T. W. Clinton, A. W. Smith, and C. J. Lobb, Phys. Rev. B **55**, 11802 (1997), and references cited therein.
2. S. Martin, A. T. Fiory, R. M. Fleming, *et al.*, Phys. Rev. B **41**, 846 (1990).
3. G. Briceo, M. F. Crommie, and A. Zettl, Phys. Rev. Lett. **66**, 2164 (1991).
4. Y. Ando, G. S. Boebinger, A. Passner, *et al.*, Phys. Rev. Lett. **77**, 2065 (1996).
5. P. W. Anderson, Phys. Rev. Lett. **67**, 2092 (1991).
6. B. P. Stojkovic and D. Pines, Phys. Rev. Lett. **76**, 811 (1996); Phys. Rev. B **55**, 8576 (1997), and references cited therein.
7. A. T. Fiory, S. Martin, R. M. Fleming, *et al.*, Phys. Rev. B **41**, 2627 (1990).
8. M. Suzuki and M. Hikita, Phys. Rev. B **44**, 249 (1991).
9. N. W. Preyer, M. A. Kastner, C. Y. Chen, *et al.*, Phys. Rev. B **44**, 407 (1991).
10. H. Raffy, S. Labdi, O. Laborde, and P. Monceau, Phys. Rev. Lett. **66**, 2515 (1991).
11. M. S. Osofsky, V. M. Browning, M. M. Miller, and R. J. Soulen, Bull. Am. Phys. Soc. **38** (1), 223 (1993).
12. J. I. Gorina, G. A. Kaljushnaia, V. I. Ktitorov, *et al.*, Solid State Commun. **91**, 615 (1994); V. P. Martovitsky, J. I. Gorina, and G. A. Kaljushnaia, Solid State Commun. **96**, 893 (1995); J. I. Gorina, G. A. Kaljushnaia, V. P. Martovitsky, *et al.*, Solid State Commun. **108**, 275 (1998).
13. A. Maeda, M. Hase, I. Tsukada, *et al.*, Phys. Rev. B **41**, 6418 (1990).
14. S. I. Vedeneev, A. G. M. Jansen, A. A. Tsvetkov, and P. Wyder, Phys. Rev. B **51**, 16380 (1995).
15. S. I. Vedeneev, A. G. M. Jansen, E. Haanappel, and P. Wyder, Phys. Rev. B **60**, 12467 (1999).
16. A. P. Mackenzie, S. R. Julian, D. C. Sinclair, and R. S. Lin, Phys. Rev. B **53**, 5848 (1996).
17. H. Zhang and H. Sato, Phys. Rev. Lett. **70**, 1697 (1993).
18. P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985).
19. A. A. Tsvetkov, J. Schützmann, J. I. Gorina, *et al.*, Phys. Rev. B **55**, 14152 (1997).
20. V. Z. Kresin and S. A. Wolf, Phys. Rev. B **41**, 4278 (1990).
21. S. J. Collocott, R. Driver, and C. Andrikidis, Physica C (Amsterdam) **173**, 117 (1991).
22. T. Manako, Y. Kubo, and Y. Shimakawa, Phys. Rev. B **46**, 11019 (1992).
23. H. Safar, P. L. Gammel, D. A. Huse, *et al.*, Phys. Rev. Lett. **72**, 1272 (1994).
24. A. P. Mackenzie, S. R. Julian, G. G. Lonzarich, *et al.*, Phys. Rev. Lett. **71**, 1238 (1993).
25. A. W. Tyler, Y. Ando, F. F. Balakirev, *et al.*, Phys. Rev. B **57**, R728 (1998).
26. T. W. Jing, N. P. Ong, T. V. Ramakrishnan, *et al.*, Phys. Rev. Lett. **67**, 761 (1991).
27. B. L. Al'tshuler, A. G. Aronov, A. I. Larkin, and D. E. Khmel'nitskiĭ, Zh. Éksp. Teor. Fiz. **81**, 768 (1981) [Sov. Phys. JETP **54**, 411 (1981)]; B. L. Al'tshuler and A. G. Aronov, Zh. Éksp. Teor. Fiz. **77**, 2028 (1979) [Sov. Phys. JETP **50**, 968 (1979)].
28. J. M. Harris, Y. F. Yan, O. K. C. Tsui, *et al.*, Phys. Rev. Lett. **73**, 1711 (1994).
29. J. M. Harris, N. P. Ong, P. Matl, *et al.*, Phys. Rev. B **51**, 12053 (1995), and references cited therein.
30. A. V. Samoilov, Z. G. Ivanov, and L.-G. Johansson, Phys. Rev. B **49**, 3667 (1994).
31. W. N. Kang, S. H. Yun, J. Z. Wu, and D. H. Kim, Phys. Rev. B **55**, 621 (1997).
32. A. Dorsey, Phys. Rev. B **46**, 8376 (1992); N. V. Kopnin, B. I. Ivlev, and V. A. Kalatsky, J. Low Temp. Phys. **90**, 1 (1993).
33. V. B. Geshkenbein and A. I. Larkin, Phys. Rev. Lett. **73**, 609 (1994).
34. V. M. Vinokur, V. B. Geshkenbein, M. V. Feigelman, and G. Blatter, Phys. Rev. Lett. **71**, 1242 (1993).
35. Y. Ando, G. S. Boebinger, A. Passner, *et al.*, Phys. Rev. B **56**, 8530 (1997).
36. Y. Ando and T. Murayama, cond-mat/9812334.
37. A. P. Mackenzie, S. D. Hughes, J. R. Cooper, *et al.*, Phys. Rev. B **45**, 527 (1992).
38. J. Hejtmanek, M. Nevriva, E. Pollert, D. Sedmidubsky, and P. Vasek, Physica C (Amsterdam) **235–240**, 1389 (1994).

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 1050–1057.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1207–1215.

SOLIDS
Electronic Properties

Electronic Thermal Properties of the Interface between a Normal Metal and a High-Temperature Superconducting Material

I. A. Devyatov[a, *], M. Yu. Kupriyanov[a], L. S. Kuzmin[a], A. A. Golubov[b], and M. Willander[c]

[a]*Institute of Nuclear Physics, Moscow State University, Moscow, 119899 Russia*
**e-mail: idev@pn.npi.msu.ru*

[b]*Department of Applied Physics, University of Twente, P.O. Box 217, 7500AE, Enschede, the Netherlands*

[c]*Chalmers University of Technology, S-412 96 Göteborg, Sweden*

Received December 20, 1999

Abstract—The electron thermal flux and electron thermal conductivity of the interface between a normal metal and a high-temperature superconductor (HTSC) are calculated using quasiclassical equations. Calculations are made for various values of the interface transparency and various orientations of the axes of a HTSC crystal. It was shown that compared with an interface between a normal metal and an "ordinary" superconductor (*s*-type symmetry, isotropic order parameter), the thermal conductivity of an HTSC–normal-metal interface is substantially higher and has a nonactivation dependence. The thermal properties were calculated for various interface models, including mirror and diffuse, and also for various potential barrier profiles. An analysis is made of the possibility of using devices based on normal-metal–HTSC interfaces for bolometric and microrefrigerator applications. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Recent interest in thermoelectric effects in structures containing interfaces between normal (N) and superconducting (S) materials has been associated with the development of a new type of electromagnetic radiation detector, a hot electron bolometer with Andreev reflection of electrons in superconducting banks,[1–12] developed for the Russian segment of the International Space Station [4]. This bolometer is a thin film of N metal sandwiched between superconducting electrodes which supply the signal current. The temperature of the electron gas in the film is monitored by additional S–I–N tunnel junctions positioned in the central part of the N film.

Two types of thermoelectronic effects are used in this device. The first takes place at the S–I–N interface and consists in the generation of a thermal flux induced by the electric current flowing across the interface. We shall subsequently take a positive thermal flux to imply the direction which leads to cooling of the N metal and a negative flux to imply a flux which results in heating. This effect is used for preliminary cooling of the electron gas in the N film [1, 5, 13] and is similar [14] to the Peltier effect in metal–semiconductor junctions. As a result of the presence of an energy gap Δ in superconductors, electrons having energies ε higher than the gap are removed more effectively (for a corresponding bias across the junction) from the N metal than electrons having energies below the gap. This reduces the effective electron temperature in the N metal. However, unlike semiconducting structures, S–I–N junctions have an additional channel for propagation of electrons of energy $|\varepsilon| < \Delta$ across the interface as a result of coherent Andreev reflection [15]. The efficiency of Andreev reflection is proportional to D^2 where D is the transparency of the interface. As D increases, the cooling effect is rapidly suppressed until the sign of the thermal flux changes, i.e., overheating of the electron gas occurs for completely transparent interfaces.

This second effect, involving substantial overheating of the electron gas in the N film of an S–N–S structure with transparent S–N interfaces, is used in the new types of bolometers. The heat dissipated as a result of the signal current flowing across the N film cannot be transferred to the substrate because of the low efficiency of the phonon mechanism of heat transfer in the operating temperature range ($T < 1$ K) nor can it be transferred to the electrons because of the low thermal conductivity of the transmitting S–N interface. The overheating is set by one of the S–I–N tunnel junctions while the other S–I–N structures [5, 13] are used for preliminary cooling of the electron gas in the N film.

Electron heat transport in ordinary isotropic superconductors with *s*-type pairing was analyzed theoretically in [14, 15]. It was shown that in the tunnel limit ($D \ll 1$) the maximum cooling power is proportional to Δ whereas the thermal conductivity of a pure S–N interface is proportional to $\exp(-\Delta/T)$ where T is the operating temperature of the junction. Thus, both thermoelec-

tronic effects become more significant as the superconductor energy gap increases.

The aim of the present study is to make a theoretical analysis of these thermoelectronic effects in structures with an HTSC electrode. The solution of this problem is of both practical and fundamental interest. This is because in HTSC materials the energy gap is not only considerably greater than that in ordinary superconductors but is essentially anisotropic and has *d*-type symmetry.

The existence of this anisotropy should lead to the discovery of additional channels for heat transport across the interface between a N metal and a *d*-type superconductor (N–I–D and N–D interfaces) via subgap [16, 17] and Andreev bound states [18]. In addition, a reduction in the coefficient of specular reflection accompanying electron scattering by the interface leads to *s*-type nongap superconductivity being induced in its vicinity [19] which should be accompanied by substantial suppression of thermoelectric effects.

It should also be noted that electron heat transport across an N–I–D structure depends strongly on the profile of the potential barrier at the interface. For "irregular" barriers comprising a large number of "punctures" simulated by δ-functional scatterers of transparency D in an almost nontransparent interface [20, 21], the thermal effects should be strongly influenced by the difference between the order parameters in the directions corresponding to the angles of electron incidence and reflection [17]. However, for extended barriers satisfying the condition $\kappa_0 d \gg 1$ ($\kappa_0 = \sqrt{2m(V-\mu)}$, d is the barrier half-width, V is its height, and μ is the chemical potential), transport is only significant in the narrow cone of angles normal to the interface.

The aim of the present study is to analyze electron thermal effects in N–I–D structures taking all these factors into account.

2. MODEL OF AN N–I–D JUNCTION

We shall assume that an N–I–D junction has the form of a constriction with geometric dimensions much shorter than the coherence length of the superconductor and also the elastic and inelastic electron scattering lengths in the metal. This assumption allows us to neglect any suppression of the order parameter in the superconductor as a result of the closeness to the N metal for any transparency of the N–D interface. We shall also assume that the Fermi surface of the HTSC material is cylindrical (we shall analyze transport in the *ab* plane). We express the transport of quasiparticles across the junction in the form of a sum of several decoupled transverse modes. The current in each mode is described by the probabilities of Andreev $A(\varepsilon, \theta_+)$ and normal $B(\varepsilon, \theta_-)$ reflections from the surface of the superconductor [14, 22]:

$$A(\varepsilon, \theta_+) = \frac{D^2|a_+(\varepsilon)|^2}{|1 - Ra_+(\varepsilon)a_-(\varepsilon)|^2},$$
$$B(\varepsilon, \theta_-) = \frac{R|1 - a_+(\varepsilon)a_-(\varepsilon)|^2}{|1 - Ra_+(\varepsilon)a_-(\varepsilon)|^2}. \quad (1)$$

Here $\theta_+ = \theta$ is the angle of incidence of an electron on the superconductor surface, $\theta_- = \pi - \theta$ corresponds to a specularly reflected electron, and $R = 1 - D$ is the electron reflection coefficient. The functions $a_\pm(\varepsilon)$ in (1) are the coefficients of Andreev reflection from a pure interface having the transparency $D = 1$, corresponding to the angles $\theta_\pm$. The angular and energy dependence of the coefficients $a_\pm(\varepsilon)$ is determined by the choice of interface model and will be calculated below.

Using the reflection coefficients (1), we can write the balance equations for the distribution functions of electrons moving away from and toward the interface at an angle θ to the normal. For electrons moving from the N metal toward the interface the distribution function is the same as the Fermi distribution shifted by eV: $f^+(\varepsilon) = f(\varepsilon - eV)$ where V is the voltage drop at the interface.

Electrons moving away from the interface toward the N metal at the angle θ occur as a result of three processes [22]:

(1) Holes moving away from the N metal at the angle θ undergo Andreev reflection to form electrons with the probability $A(\varepsilon, \theta_+)$:

(2) Electrons moving from the N metal at the angle $\pi - \theta$ are specularly reflected with the probability $B(\varepsilon, \theta_-)$;

(3) Quasiparticles approaching the interface from the superconductor enter the N metal with the probability $1 - A(\varepsilon, \theta_+) - B(\varepsilon, \theta_-)$.

Thus, the distribution function $f^-(\varepsilon, \theta)$ of electrons moving toward the N metal at the angle θ has the form

$$f^-(\varepsilon, \theta) = A(\varepsilon, \theta_+)[1 - f^+(-\varepsilon)] + B(\varepsilon, \theta_-) f^+(\varepsilon) + [1 - B(\varepsilon, \theta_-) - A(\varepsilon, \theta_+)] f(\varepsilon). \quad (2)$$

Then, following the method put forward in [14] and bearing in mind the angular dependence of the distribution functions in (2) of fundamental importance for HTSC, we obtain the thermal flux across the N–S interface

$$j = \frac{k_0}{2\pi^2\hbar}\int d\varepsilon(\varepsilon - eV) \times \int_{-\pi/2}^{\pi/2} d\theta \cos\theta [f^+(\varepsilon) - f^-(\varepsilon, \theta)], \quad (3)$$

where k_0 is the absolute value of the wave vector at the Fermi surface.

In addition to determining the thermal flux (cooling power), for bolometric applications it is also interesting to calculate the thermal conductivity $\kappa = j/\delta T$ of an ideal interface (having the transparency $D = 1$) where δT is the small difference between the temperatures of the N metal and the superconductor. The expression for the thermal conductivity is obtained from (3)

$$\kappa = \frac{k_0 k_B T^2}{2\pi^2 \hbar}\int \frac{\varepsilon^2 e^{\varepsilon} d\varepsilon}{[1+e^{\varepsilon}]^2}\int_{-\pi/2}^{\pi/2} d\theta \cos\theta (1 - |a_+(\varepsilon)|^2) \quad (4)$$

and is the same as the expression for an ordinary anisotropic superconductor since (4) only depends on a single angle θ. A similar situation occurred in calculations of the current–voltage dependence of a HTSC–normal-metal junction having the interface transparency $D = 1$ [17].

Formulas (1)–(4) express the thermal flux and thermal conductivity as a function of the voltage, temperature, angle of orientation α of the HTSC crystal, and interface transparency D and reduce the problem of finding them to determining the energy and angular dependences of the coefficients of Andreev reflection $a_{\pm}(\varepsilon, \theta)$.

3. ANDREEV REFLECTION COEFFICIENTS

For the case of an atomically abrupt mirror N–D interface the coefficients of Andreev reflection $a_{\pm}(\varepsilon)$ depend on the angle of orientation α of one of the crystal directions in the *ab* plane relative to the normal to this plane. For $\alpha = 0$ no suppression of the order parameter occurs near the N–D interface and the coefficients $a_{\pm}(\varepsilon)$ have the normal form deduced from BCS theory (see, for example [14, 17]):

$$a_{\pm} = \begin{cases} \dfrac{\varepsilon - \mathrm{sgn}(\varepsilon)\sqrt{\varepsilon^2 - |\Delta_{\pm}|^2}}{\Delta_{\pm}}, & |\varepsilon| > |\Delta_{\pm}|, \\ \dfrac{\varepsilon - i\sqrt{|\Delta_{\pm}|^2 - \varepsilon^2}}{\Delta_{\pm}}, & |\varepsilon| < |\Delta_{\pm}|, \end{cases} \quad (5)$$

where the order parameter should be taken to be the values $\Delta_{\pm} = \Delta(\theta_{\pm}) = \Delta_0(T)\cos[2(\theta \mp \alpha)]$ which allow for its anisotropy.

For arbitrary α the coefficients $a_{\pm}(\varepsilon)$ were calculated numerically in two stages [23, 24]. At the first stage by solving the quasiclassical Eilenberger equations

$$v\cos\theta \frac{dp}{dx} + (2\omega + \Delta p)p - \Delta = 0, \quad (6)$$

$$v\cos\theta \frac{dq}{dx} - (2\omega + \Delta p)p + \Delta = 0, \quad (7)$$

$$\Delta \ln\frac{T}{T_c} + 2\pi T\sum_{\omega}\left[\frac{\Delta}{\omega} - \sqrt{2}\cos(2(\theta - \alpha)) \times \left\langle \sqrt{2}\cos(2(\theta' - \alpha))\frac{p+q}{1+pq}\right\rangle\right] = 0, \quad (8)$$

where $\langle\ldots\rangle = (1/2\pi)\int_0^{2\pi}(\ldots)\,d\theta'$, we determine the spatial dependence of the order parameter $\Delta(x)$ for various values of the angle of orientation α. The functions p and q were introduced in [23] to parametrize the Eilenberger equations. They are related to the Eilenberger functions f, f_+, and g by

$$f = \frac{2p}{1+pq}, \quad f^+ = \frac{2q}{1+pq}, \quad g = \frac{1-pq}{1+pq}. \quad (9)$$

Equations (6)–(8) were closed by boundary conditions determined from the condition for obtaining a spatially homogeneous solution inside the superconductor:

$$q(\infty, \theta) = p(\infty, \pi + \theta) = \frac{\sqrt{\omega^2 + \Delta^2(\infty)} - \omega}{\Delta(\infty)}. \quad (10)$$

The interface of the N–I–D structure was simulated by a layer of thickness $\tilde{d}$ in which strong diffuse scattering of electrons occurred. It was shown in [19] that at this interface the functions $p(0, \theta)$ and $q(0, \theta)$ should be related by

$$q(0, \theta) = p(0, -\theta)\frac{1 + (F - G)\tanh(\tilde{k}\tilde{d})}{[Fp(0, -\theta) + G]\tanh(\tilde{k}\tilde{d}) + 1}, \quad (11)$$

where

$$\tilde{k} = \frac{\sqrt{\langle g\rangle^2 + \langle f\rangle^2}}{l\cos\theta}, \quad F = \frac{\langle f\rangle}{\sqrt{\langle g\rangle^2 + \langle f\rangle^2}},$$

$$G = \frac{\langle g\rangle}{\sqrt{\langle g\rangle^2 + \langle f\rangle^2}},$$

and l is the electron mean free path in the diffuse layer. The first term in (11) describes the correlation between an electron approaching the interface in the direction characterized by the angle –θ and an electron reflected from the interface in the direction θ. For $\tilde{d} \longrightarrow 0$ this terms yields a boundary condition valid for completely specular electron reflection:

$$q(0, \theta) = p(0, -\theta). \quad (12)$$

The second term in (11) is responsible for the average contribution to the electron flux moving away from the interface in the direction θ made by particles approaching the interface at all other angles. For $\tilde{d} \gg l$ the second term in (11) is the determining factor since as a result of strong multiple scattering in the diffuse layer,

the incoming and outgoing electrons are completely uncorrelated:

$$q(0, \theta) = \frac{F}{G+1}. \quad (13)$$

The numerical solution procedure involved first integrating along the trajectory of equation (6) for $p(0, -\theta)$ between the values determined by condition (10) and the interface for a given initial approximation for $\Delta(x)$. Then, the values of $q(0, \theta)$ were determined using (11) and equation (7) was integrated along the trajectory extending from the interface to infinity. The functions $q(x, \theta)$ and $p(x, \theta)$ thus determined were then used to refine the values of $\Delta(x)$ and the values of F and G appearing in (11) at the next iteration step.

At the second stage the Eilenberger equations were continued analytically by substituting $\omega \longrightarrow i\varepsilon$ and this system of equations was solved numerically using the values of $\Delta(x)$ determined at the first stage. It is easy to see that the boundary condition (10) is transformed as a result of the analytic continuation to a form exactly the same as the expression for the Andreev reflection coefficient (5) apart from imaginary unity. Thus, the values of $-ip(0, \theta, \varepsilon)$ give the coefficient of Andreev reflection of electrons approaching the interface in the direction of the angle θ to form holes. Similarly, the functions $iq(0, \theta, \varepsilon)$ determine the coefficient of scattering of hole excitations to form electrons at the interface.

The energy dependence of the Andreev scattering coefficient $|a_+(\varepsilon)|$ calculated numerically for the angle of orientation $\alpha = 10°$ is plotted in Fig. 1. For a specular interface and $\alpha = 0$ the energy dependence of $|a_\pm(\varepsilon)|$is the same as that predicted by BCS theory with the order parameter depending on the angle θ, $\Delta \propto \cos 2\theta$. For angles $\alpha \neq 0$ the energy gap is suppressed near the interface (see Fig. 1). This leads to the formation of Andreev bound states [18] for quasiparticles trapped by the potential well between the interface and the point having the coordinate L determined by the equation $\varepsilon = \Delta(L)$, where ε is the quasiparticle energy. The closer the angle θ to $\pi/2$, the larger the width L of the potential well for the same Δ and the larger the number of Andreev bound states. This is observed as an increase in the number of peaks on the dependence $|a_+(\varepsilon)|$ with increasing θ and can be seen clearly in Fig. 1.

In the vicinity of a diffuse interface nongap superconductivity is induced which leads to a linear dependence of the Andreev reflection coefficient $|a_+(\varepsilon)|\alpha\varepsilon$ at low energies (Fig. 2). Note that in this case the maximum value of $|a_+(\varepsilon)|$ is substantially lower than the values obtained for a specular interface.

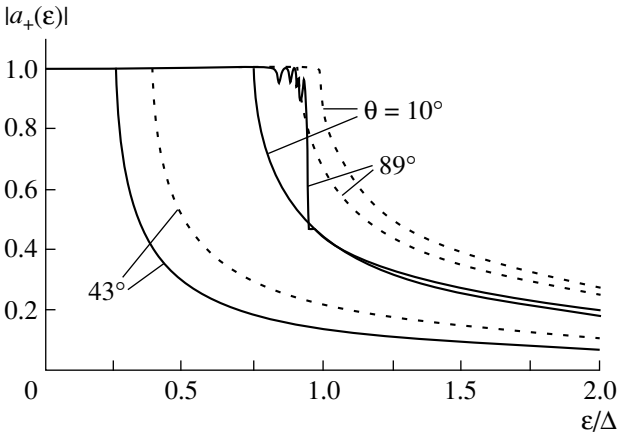


Fig. 1. Coefficients of Andreev reflection for a specular N–D interface at temperature $T = 0.1T_c$ for the angle of orientation $\alpha = 10°$ and $\theta = 10°$, 43°, 89°. The solid curves give the results of self-consistent calculations and the dashed curves give the results of BCS theory for an infinite superconductor.

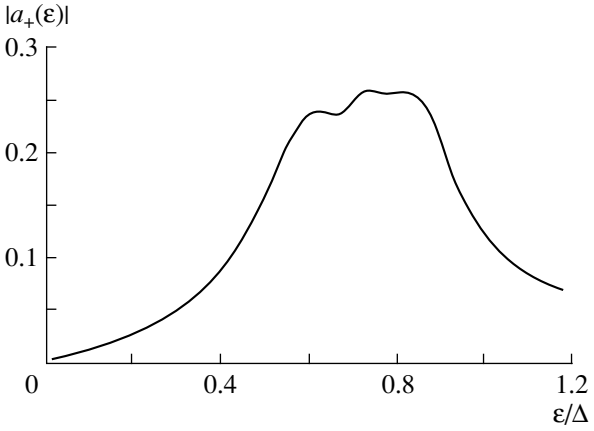


Fig. 2. Coefficients of Andreev reflection for a diffuse N–D interface ($T/T_c = 0.1$).

4. N–I–D JUNCTION WITH SPECULAR INTERFACES

4.1. Thermal Flux across an N–I–D Junction with Specular Interfaces and δ-Functional Scatterers

The properties of an N–I–D junction with specular interfaces depend strongly on the angle of orientation α and also on the profile of the potential barrier. We shall first consider the situation most commonly encountered in practice comprising irregular barriers with numerous "punctures" in an intermediate layer, simulated by δ-functional scatterers [21]. In this case the coefficient of electron transparency D in formula (1) is characterized by the parameter $z = mH/\hbar^2 k_0$:

$$D = \frac{1}{1 + z^2/\cos^2\theta}. \quad (14)$$

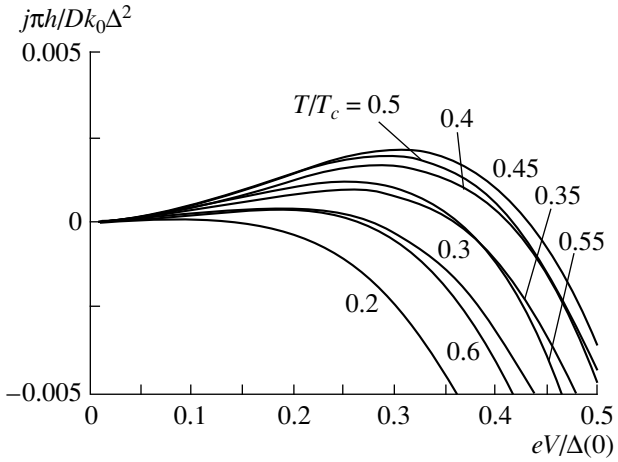


Fig. 3. Thermal flux j for an N–I–D junction with specular interfaces and δ-functional scatterer as a function of junction voltage V calculated at various temperatures T for the angle of inclination $\alpha = 0$ and low (tunnel) transparency $D = 10^{-4}$.

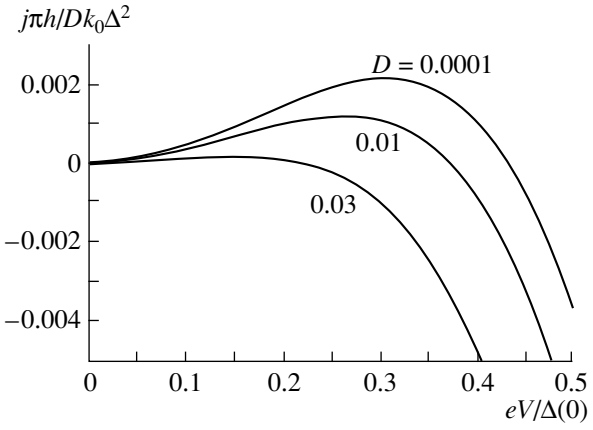


Fig. 4. Thermal flux j for an N–I–D junction with specular interfaces and δ-functional scatterer as a function of voltage V for interface transparencies $D = 10^{-4}$ (tunnel limit), 0.01, and 0.03, optimum temperature $T/T_c = 0.45$, and angle of orientation $\alpha = 0$.

Here H is the coefficient of a δ-functional barrier at the interface. Substituting into (1)–(3) the numerically calculated values of $a_{\pm}(\varepsilon)$ and using (14), we made numerical calculations of the thermal flux j as a function of the voltage across the junction V for various temperatures T for $\alpha = 0$ and low (tunneling) transparency $D = 10^{-4}$. The results of the calculations are plotted in Fig. 3. It can be seen that for each temperature there is an optimum voltage at which the thermal flux has a maximum. In this case, the highest value of the thermal flux is achieved at temperature $T/T_c \approx 0.45$. This temperature is close to the similar optimum temperature of a superconductor junction with ordinary *s*-type symmetry [14]. However, the amplitude of the thermal flux j for superconductors with *d*-type symmetry is more than an order of magnitude lower. This is because of the strong anisotropy of HTSC materials.

An increase in the transparency D of the interface suppresses the cooling power of the N–I–D junction because of the appearance of a channel for coherent Andreev current transport across the interface. A similar effect was also observed for a junction with an *s*-type superconductor. Figure 4 shows how the cooling power decreases with increasing transparency for an optimum set of parameters $\alpha = 0$, $T/T_c = 0.45$ for an N–I–D junction.

Figure 5 gives the dependence $j(V)$ for various values of α and the optimum temperature $T/T_c = 0.45$ for a low (tunneling) interface transparency. It can be seen from Fig. 5 that an increase in the angle of orientation α leads to suppression of the cooling power. This is attributed to the possible appearance of coherent current transport via subgap states [17, 18] and to suppression of the order parameter near the interface for $\alpha \neq 0$.

Figure 6 gives the maximum cooling power (the maximum in terms of V in Fig. 4) of an N–I–D junction as a function of the interface transparency D. For small D the cooling power increases as a function of D as for a junction with an isotropic *s*-type superconductor [14]. Then, as the transparency increases, the thermal flux begins to decrease as a result of the increasing contribution of coherent Andreev reflection to the current transport. It is important to note that the maximum thermal flux for a junction with a *d*-type superconductor is more than two orders of magnitude lower than the similar (normalized) value achieved for junctions with isotropic *s*-type superconductors [14].

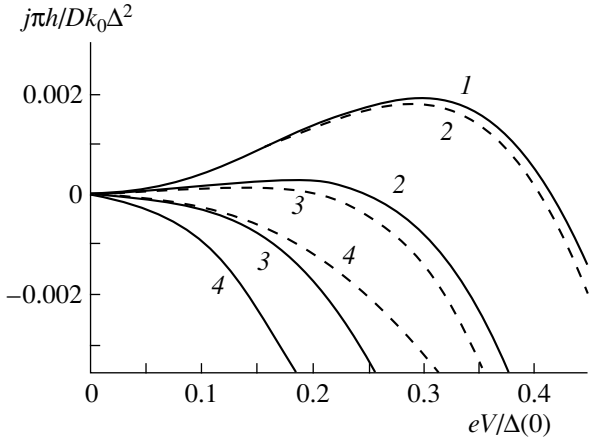


Fig. 5. Thermal flux j for an N–I–D junction with a specular interface and δ-functional scatterer as a function of voltage V for various angles of orientation: α = (*1*) 0°, (*2*) 5°, (*3*) 20°, and (*4*) 45° for tunnel transparency of the interface $D = 10^{-4}$ and optimum temperature $T/T_c = 0.45$. The dashed curves give values of the thermal flux calculated using formula (3) neglecting suppression of the superconductor order parameter near the interface.

4.2. Thermal Flux across an N–I–D Junction with Specular Interfaces and a Regular Long Barrier

The electron transport of heat across an N–I–D junction changes qualitatively when we assume that a high-quality insulator layer (I layer) is formed at the HTSC–normal-metal interface, simulated by a homogeneous rectangular barrier having the transparency

$$D = D(k, \kappa) = \frac{16k^2\kappa^2}{(k^2+\kappa^2)}e^{-4\kappa d}, \tag{15}$$

where $k = \sqrt{k_0^2 - k_\perp^2}$, $\kappa = \sqrt{\kappa_0^2 + k_\perp^2}$, $k_\perp$ is the wave vector perpendicular to the normal to the interface, and d is the thickness of the I layer. For $\kappa_0 d \gg 1$, when $D \ll 1$ only small angles of electron incidence are significant

$$\theta \simeq \frac{\kappa_0}{k_0}\frac{1}{\sqrt{\kappa_0 d}} \ll 1, \tag{16}$$

which ensures that the condition $a_+(\varepsilon, \theta) \simeq a_-(\varepsilon, \theta) = a(\varepsilon, 0)$ is satisfied in formulas (1)–(3). In this case, formula (3) for $\alpha = 0$ reduces to

$$j_{\kappa_0 d \gg 1} = \frac{k_0}{2\pi^2\hbar}\langle D\rangle_\theta \int d\varepsilon(\varepsilon - eV) \times \frac{1-|a|^4}{|1-a^2|^2}[f(\varepsilon - eV) - f(\varepsilon)], \tag{17}$$

where

$$\langle D\rangle_\theta = \sqrt{\frac{\pi}{2}}\frac{\kappa_0}{k_0}\frac{1}{\sqrt{\kappa_0 d}}D(k_0, \kappa_0)$$

is the transparency (15) of a long rectangular barrier averaged over the angle θ. Using (1), (17), and the values of the Andreev reflection coefficients obtained in Section 3, we can easily calculate the thermal flux j as a function of the junction voltage V for various temperatures T. Results of the calculations for $\alpha = 0$ are plotted in Fig. 7. A comparison of the curves in Figs. 3 and 7 shows that the dependences $j(V)$ in these figures are qualitatively the same whereas the maximum of $j(V)$ in Fig. 7 is more than an order of magnitude greater than that in Fig. 3 and is comparable (in relative units) to the values obtained for s-type superconductors [14]. This is because condition (16) is satisfied for angles of incidence of quasiparticles on the interface so that the strong anisotropy of the HTSC materials can be neglected and also because no suppression of the order parameter occurs for $\alpha = 0$.

Figure 8 gives dependences $j(V)$ for various values of α and the optimum temperature $T/T_c = 0.45$ for a long barrier $\kappa_0 d \gg 1$. It can be seen that an increase in the angle of orientation α leads to significant suppression of the cooling power. However, unlike similar curves in Fig. 5 plotted for a δ-functional scatterer, suppression of the thermal flux with increasing angle of

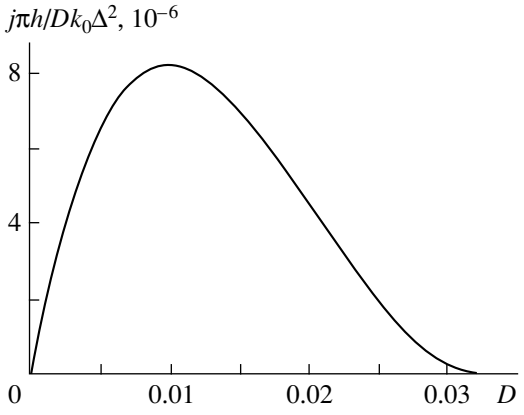


Fig. 6. Maximum values of the thermal flux j for an N–I–D junction as a function of the transparency D for angle of orientation $\alpha = 0$ and temperature $T = 0.45T_c$.

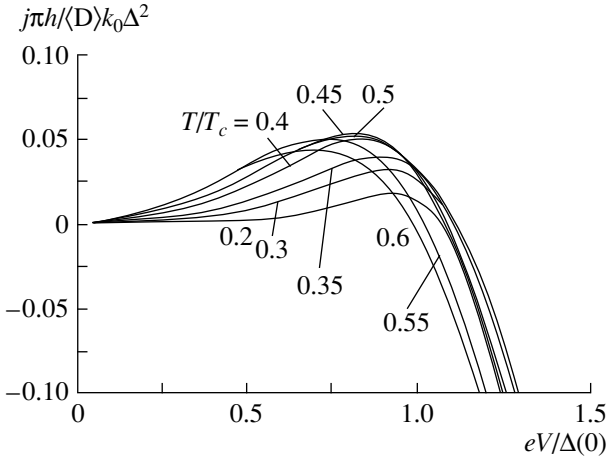


Fig. 7. Thermal flux j for an N–I–D junction with specular interfaces and a long rectangular barrier as a function of the junction voltage V calculated for various temperatures T and angle of orientation $\alpha = 0$.

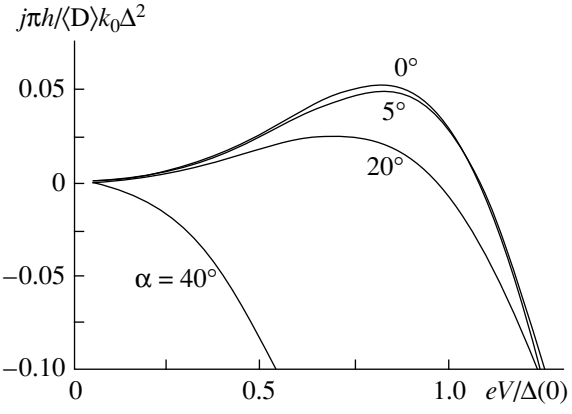


Fig. 8. Thermal flux j for an N–I–D junction with a long rectangular barrier and specular interface as a function of voltage V for $\alpha = 0°$, 5°, 20°, and 40° at temperature $T/T_c = 0.45$.

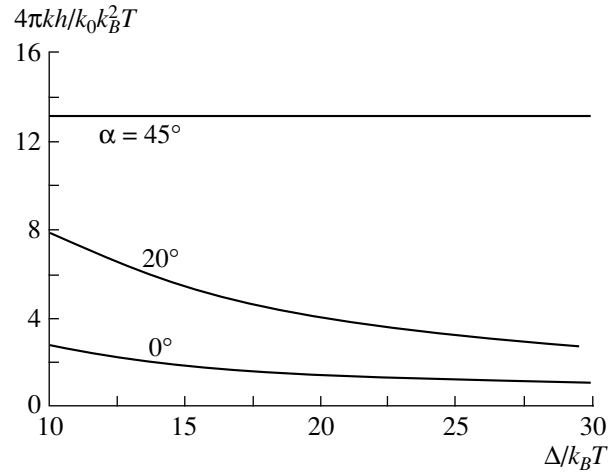


Fig. 9. Thermal conductivity of a pure N–D interface as a function of Δ/k_BT for various angles of orientation α.

orientation α is observed from considerably higher values of the thermal flux for α = 0.

4.3. Thermal Conductivity of an N–D Junction with Specular Interfaces

The thermal conductivity κ calculated using formula (4) as a function of the ratio Δ/k_BT for α = 0°, 20°, and 45° is plotted in Fig. 9. It can be seen that the thermal conductivity of the junction is not exponentially small as for a junction based on an isotropic two-dimensional *s*-type superconductor [5]:

$$\kappa_s = \frac{\sqrt{2}}{\pi^{3/2}}\frac{k_0}{\hbar}\Delta k_B\sqrt{\frac{\Delta}{k_BT}}\exp\left(-\frac{\Delta}{k_BT}\right), \quad \frac{\Delta}{k_BT} \gg 1. \quad (18)$$

Estimates made using formula (18) and results of numerical calculations plotted in Fig. 9 show that for any values of the angle of orientation α the thermal conductivity of the N–D junction is more than two orders of magnitude higher than that of an N–S junction. This is a consequence of the strong anisotropy of *d*-type superconductors and the suppression of the order parameter near the interface for nonzero values of the angle of orientation. For α = 45° the order parameter is suppressed to zero so that the thermal conductivity of this structure is the same as that of a two-dimensional N–N structure:

$$\kappa \approx 4\frac{k_0k_B^2T}{\pi h}\int_0^\infty dx\frac{x^2}{\cosh^2 x} \approx 13.159\frac{k_0k_B^2T}{\pi h}. \quad (19)$$

For α ≠ 45° the suppression of the order parameter is weaker. Nevertheless, it is possible to have an effective leakage of heat because of the strong anisotropy of the order parameter of *d*-type superconductors. Thus, the thermal conductivity of an N–D junction forα ≠ 45° is lower than that for α = 45° but substantially higher than that for N–S junctions (18).

5. N–I–D JUNCTIONS WITH A DIFFUSE INTERFACE

In the model of N–I–D junctions proposed in [19] the order parameter at the interface is similar to that obtained in the case of isotropic nongap superconductivity. In this case, the coefficients of Andreev reflection (Fig. 2) are substantially lower than those for specular interfaces (Fig. 1).

Numerical calculations of the thermal flux across this type of interface based on formulas (1)–(3) and the values of the Andreev reflection coefficients plotted in Fig. 2 yield negative values for any voltages and interface transparencies. In this case, the numerical values of the thermal flux are similar (to within a few percent) to those for a similar N–I–N junction. The thermal conductivity of an amorphous interface with transparency $D = 1$ is also similar to the value given by expression (19) for a two-dimensional N–N structure.

6. CONCLUSIONS

This analysis has shown that the cooling effect in N–I–D junctions is most significant in junctions with atomically abrupt specular interfaces and a high-quality insulator layer which can be simulated by a rectangular potential barrier. For this we need to have a homogeneous HTSC crystal with angle of orientation α ≈ 0 when the suppression of the order parameter at the interface is negligible. The most effective cooling takes place at temperature $T = 0.45T_c$. The cooling power of electronic refrigerators based on HTSC materials will be approximately two orders of magnitude higher than that of the best refrigerators using low-temperature superconductors (at lower temperatures). However, in the temperature range of practical interest for bolometric applications T = 100–300 mK the cooling power of N–S refrigerators becomes negative for HTSC materials because of their high order parameters which eliminates the possible application of these materials in these low-temperature devices.

The cooling effect in N–I–D junctions with specular interfaces and an irregular barrier simulated in the form of constrictions with δ-functional scatterers yields values of the cooling power approximately two orders of magnitude lower [in a corresponding normalization, see (2)] than those in N–I–S structures even for α = 0. This implies that in this case, the advantage of high values of the order parameter in HTSC materials is canceled out by their strong pairing anisotropy.

The exponentially strong suppression of the thermal conductivity characteristic of N–S structures does not occur in N–D structures because of the strong anisotropy of the order parameter and because of the suppression of the order parameter at the interface for nonzero

values of the angle of orientation α. Thus, pure N–D interfaces cannot be used for bolometric applications.

For the case of a diffuse N–D interface the nongap superconductivity generated at the interface leads to extremely low values of the coefficients of Andreev reflection (see Fig. 2) which makes these structures similar (in the sense of the electronic thermal properties) to N–N structures in which no cooling and no suppression of the thermal conductivity typical of N–I–S structures occurs.

ACKNOWLEDGMENTS

This work was supported by a joint project of the Russian Foundation for Basic Research and INTAS (grant no. 97-0731) and the program "Topical Problems in Physics of the Condensed State".

REFERENCES

1. M. Nahum, P. L. Richards, and C. A. Mears, IEEE Trans. Appl. Supercond. **3**, 2124 (1993).
2. M. Nahum and J. Martinis, Appl. Phys. Lett. **63**, 3075 (1993).
3. M. Nahum, T. M. Eiles, and J. Martinis, Appl. Phys. Lett. **65**, 3123 (1994).
4. L. Kuzmin, N. Kardashev, V. Kurt, *et al.*, in *Proceedings of the 2nd European Symposium on the Utilization of the International Space Station, ESTEC, The Netherlands, Noordwijk, 1998*, p. 127.
5. L. S. Kuzmin, I. A. Devyatov, and D. Golubev, in *Proceedings of the 4th International Conference on mm and submm Waves and Applications, San-Diego, 1998*, Vol. 3465, p. 193.
6. D. Chouvaev, L. Kuzmin, and M. Tarasov, in *Proceedings of ISEC-99, Claremont Resort Berkeley, California, 1999*, p. 447.
7. A. Vystavkin, D. Chouvaev, L. Kuzmin, *et al.*, in *Proceedings of the 4th International Conference on mm and submm Waves and Applications, San-Diego, 1998*, Vol. 3465, p. 441.
8. D. Chouvaev, L. Kuzmin, M. Tarasov, *et al.*, in *Proceedings of the IX International Symposium on Space Terahertz Technology, Pasadena, 1998*, p. 331.
9. A. Vystavkin, D. Chuvaev, L. Kuzmin, *et al.*, Zh. Éksp. Teor. Fiz. **115**, 1085 (1999) [JETP **88**, 598 (1999)].
10. A. Vystavkin, D. Chouvaev, T. Claeson, *et al.*, in *Proceedings of the X International Symposium on Space Terahertz Technology, Charlottesville, 1999*, p. 372.
11. D. Chouvaev, D. Golubev, M. Tarasov, *et al.*, in *Proceedings of the X International Symposium on Space Terahertz Technology, Charlottesville, 1999*, p. 552.
12. L. Kuzmin, D. Chouvaev, M. Tarasov, *et al.*, IEEE Trans. Appl. Supercond. **9**, 3186 (1999).
13. M. M. Leivo, J. P. Pekola, and D. Averin, Appl. Phys. Lett. **68**, 1996 (1996).
14. A. Bardas and D. Averin, Phys. Rev. B **52**, 12873 (1995).
15. A. F. Andreev, Zh. Éksp. Teor. Fiz. 46, 1823 (1964) [Sov. Phys. JETP **19**, 1228 (1964)].
16. C. R. Hu, Phys. Rev. Lett. **72**, 1526 (1994).
17. Y. Tanaka and S. Kashiwaya, Phys. Rev. Lett. **74**, 3451 (1995).
18. Yu. S. Barash, A. A. Svidzinsky, and H. Burkhurdt, Phys. Rev. B **55**, 15282 (1997).
19. A. A. Golubov and M. Yu. Kupriyanov, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 478 (1998) [JETP Lett. **67**, 501 (1998)].
20. A. A. Golubov, V. M. Krasnov, and M. Yu. Kupriyanov, IEEE Trans. Appl. Supercond. **7**, 3204 (1997).
21. M. Yu. Kupriyanov, J. Low Temp. Phys. **106**, 149 (1997).
22. G. E. Blonder, M. Tinkham, and T. M. Klapwijk, Phys. Rev. B **25**, 4515 (1982).
23. N. Schopohl, cond-mat/9804064 (1998).
24. A. A. Golubov and M. Yu. Kupriyanov, JETP Lett. **69**, 262 (1999).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 1058–1062.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1216–1220.

SOLIDS
Electronic Properties

Conductivity of Two-Component Systems in a Magnetic Field

Yu. N. Ovchinnikov

Centre de Recherches sur les Très Basses Températures associé à l'Université Joseph Fourier, CNRS BP166, 38042 Grenoble, Cedex 9, France

Landau Institute for Theoretical Physics, Russian Academy of Sciences, ul. Kosygina 2, Moscow, 117940 Russia
e-mail: ovc@itp.ac.ru

Received January 13, 2000

Abstract—The influence of a magnetic field on the effective conductivity of a system with a chessboard structure is studied under conditions of the Hall effect. It is shown that in this case a new physical effect occurs, involving an oscillatory dependence of the charge density at the interfaces on passage through the bifurcation point. The system possesses considerable magnetoresistance. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Although studies of the conductivity of multicomponent systems with ordered or random distributions of components have a fairly long history [1–3], this problem cannot be considered to be completely resolved. Dykhne [1] showed that a two-component system in which the conductivities of the components are $\sigma_{1,2}$ and the components (1, 2) have a symmetric random distribution, has the effective conductivity $\sigma_{\text{eff}} = (\sigma_1\sigma_2)^{1/2}$. We obtained an exact solution for the conductivity and electric field distribution for a two-component ordered system with a chessboard structure [4] and confirmed the Dykhne hypothesis that $\sigma_{\text{eff}} = (\sigma_1\sigma_2)^{1/2}$. It is interesting to determine whether this simple expression for σ_{eff} is related to the chessboard structure or whether it also holds for other ordered systems. In addition, the point $\sigma_2/\sigma_1 \longrightarrow 0$ is the bifurcation point for the equation for the scalar potential. The physical values of the conductivity are limited to the range $\sigma_{1,2} > 0$ and the bifurcation point is unattainable. The existence of finite Hall conductivity changes the situation.

An analysis will be made of the influence of the Hall conductivity on an ordered two-component system with a chessboard structure. In this case, the bifurcation point is attainable and passage through this point leads to oscillations of the charge density along the edges of the cells. These results are valid for the two-dimensional case and for samples formed from cylinders. It is assumed that the cross section of the sample has the structure indicated above.

2. CONDUCTIVITY OF A TWO-COMPONENT SYSTEM IN A MAGNETIC FIELD

We shall determine the effective conductivity of a two-component system having a chessboard structure under conditions when each component also possesses Hall conductivity. We express the conductivity tensor $\hat{\sigma}_{1,2}$ of each component in the form

$$\hat{\sigma}_{1,2} = \sigma_{1,2}\delta_{ik} + \hat{\sigma}^{\text{H}}_{1,2}, \tag{1}$$

where $\hat{\sigma}^{\text{H}}_{1,2}E = \sigma^{\text{H}}_{1,2}[H, E]$ is the Hall conductivity. The condition for continuity of the current across the interface may be written in the form

$$(\hat{\sigma}E)^{(1)}_{\mathbf{n}} = (\hat{\sigma}E)^{(2)}_{\mathbf{n}}, \tag{2}$$

where $\mathbf{n}$ is the vector of the normal to the surface and E is the electric field at the interface. Equation (2) can be conveniently rewritten for the following analysis in the form

$$\begin{aligned}\sigma_1\mathbf{n}\left(\frac{\partial\varphi}{\partial x}, \frac{\partial\varphi}{\partial y}\right)_1 - \sigma^{\text{H}}_1 H\mathbf{n}\left(\frac{\partial\varphi}{\partial y}, -\frac{\partial\varphi}{\partial x}\right)_1 \\ = \sigma_2\mathbf{n}\left(\frac{\partial\varphi}{\partial x}, \frac{\partial\varphi}{\partial y}\right)_2 - \sigma^{\text{H}}_2 H\mathbf{n}\left(\frac{\partial\varphi}{\partial y}, -\frac{\partial\varphi}{\partial x}\right)_2,\end{aligned} \tag{3}$$

where φ is the scalar potential.

A jump in the normal component of the electric field leads to the appearance of the charge density ρ at the interface

$$(E)^{(1)}_{\mathbf{n}} - (E)^{(2)}_{\mathbf{n}} = 4\pi\rho. \tag{4}$$

In order to find the tensor of the effective conductivity, it is sufficient to analyze the orientation of the electric field shown in the figure.

We shall consider a periodic structure with unit cell vectors $\sqrt{2a}\,(1, 0)$ and $\sqrt{2a}\,(0, 1)$. Thus, only the four densities ρ shown in the figure can be independent. It is easy to check that two relationships exist between them:

$$\rho_4 = -\rho_1, \quad \rho_3 = -\rho_2. \tag{5}$$

As before, the equation for the scalar potential φ is given by

$$\begin{aligned}\varphi = E_0 y - 4\pi\int d^2\mathbf{r}_1 G(\mathbf{r} - \mathbf{r}_1)\rho(\mathbf{r}_1), \\ G = \frac{1}{2\pi}\ln|\mathbf{r} - \mathbf{r}_1|.\end{aligned} \tag{6}$$

1063-7761/00/9006-1058$20.00

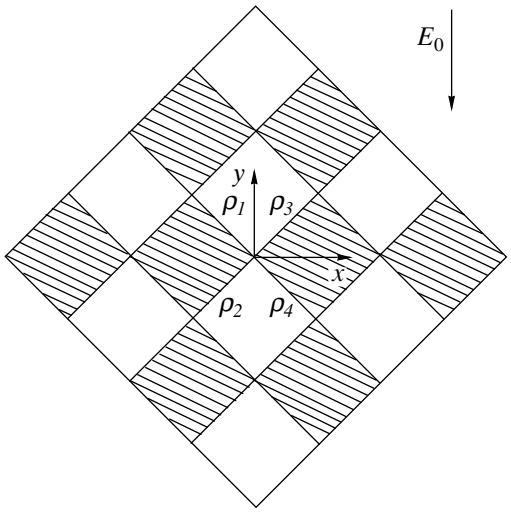
Converting to the dimensionless variables

$$t \longrightarrow \tilde{t}a, \quad \rho_{1,2} \longrightarrow E\tilde{\rho}_{1,2}, \tag{7}$$

we rewrite the equation of continuity (3) in the form

$$\frac{Z}{2\pi}\left\{\frac{1}{\sqrt{2}} - 2\pi\int_0^1 dt\tilde{\rho}_2(t)\sum_N \tan\left(\frac{\pi}{2}(t+N)\right)\right.$$
$$\times\left[\sinh^2\left(\frac{\pi}{2}(t'+N)\right) + \cosh^2\left(\frac{\pi}{2}(t'+N)\right)\right.$$
$$\left.\left.\times\tan^2\left(\frac{\pi}{2}(t+N)\right)\right]^{-1}\right\} - \frac{Z^H}{2\pi}\left\{\frac{1}{\sqrt{2}} - \pi\oint_0^1 dt\tilde{\rho}_1(t)\right. \tag{8}$$
$$\left.\times\sum_N\left[\frac{\tan\left(\frac{\pi}{2}(t'-t+N)\right)}{\sinh^2\frac{\pi N}{2} + \cosh^2\frac{\pi N}{2}\tan^2\left(\frac{\pi}{2}(t'-t+N)\right)}\right.\right.$$
$$\left.\left.-\frac{\tan\left(\frac{\pi}{2}(t'+t+N)\right)}{\sinh^2\frac{\pi N}{2} + \cosh^2\frac{\pi N}{2}\tan^2\left(\frac{\pi}{2}(t'+t+N)\right)}\right]\right\} = \tilde{\rho}_1(t),$$

$$\frac{Z}{2\pi}\left\{\frac{1}{\sqrt{2}} + 2\pi\int_0^1 dt\tilde{\rho}_1(t)\sum_N \tan\left(\frac{\pi}{2}(t+N)\right)\right.$$
$$\times\left[\sinh^2\left(\frac{\pi}{2}(t'+N)\right) + \cosh^2\left(\frac{\pi}{2}(t'+N)\right)\right.$$
$$\left.\left.\times\tan^2\left(\frac{\pi}{2}(t+N)\right)\right]^{-1}\right\} + \frac{Z^H}{2\pi}\left\{\frac{1}{\sqrt{2}} + \pi\oint_0^1 dt\tilde{\rho}_2(t)\right. \tag{9}$$
$$\left.\times\sum_N\left[\frac{\tan\left(\frac{\pi}{2}(t'-t+N)\right)}{\sinh^2\frac{\pi N}{2} + \cosh^2\frac{\pi N}{2}\tan^2\left(\frac{\pi}{2}(t'-t+N)\right)}\right.\right.$$
$$\left.\left.-\frac{\tan\left(\frac{\pi}{2}(t'+t+N)\right)}{\sinh^2\frac{\pi N}{2} + \cosh^2\frac{\pi N}{2}\tan^2\left(\frac{\pi}{2}(t'+t+N)\right)}\right]\right\} = -\tilde{\rho}_2(t'),$$



Two-component system with a chessboard structure.

where

$$Z = (\sigma_1 - \sigma_2)/(\sigma_1 + \sigma_2),$$
$$Z^H = H(\sigma_1^H - \sigma_2^H)/(\sigma_1 + \sigma_2).$$

The integral ($\oint$) in equations (8) and (9) indicates integration in the sense of the principal value. It follows from the system of equations (8) and (9) that the functions $\tilde{\rho}_{1,2}$ are analytic doubly periodic functions of the variable t with periods $T_{1,2}$ given by

$$T_{1,2} = (1 \pm i). \tag{10}$$

The points (0, 1) and those equivalent to them are branching points of the functions $\tilde{\rho}_{1,2}$.

Near zero, we find

$$\rho_1(t) \simeq \frac{A_1}{t^{2\kappa}}, \quad \rho_2(t) = \frac{A_2}{t^{2\kappa}}, \tag{11}$$

where $A_{1,2}$ and κ are certain constants. Taking into account the relationships

$$\int_0^\infty \frac{dt}{(t+1)t^{2\kappa}} = \frac{\pi}{\sin 2\pi\kappa},$$
$$\oint_0^\infty \frac{dt}{(1-t)t^{2\kappa}} = \pi\cot 2\pi\kappa, \tag{12}$$

we obtain correlations between the coefficients $A_{1,2}$ derived from the system of equations (8) and (9)

$$-\frac{A_2 Z}{\sin\pi\kappa} - A_1 Z^H \tan\pi\kappa = A_1,$$
$$-\frac{A_1 Z}{\sin\pi\kappa} + A_2 Z^H \tan\pi\kappa = A_2. \tag{13}$$

The condition for solubility of equations (13) determines the value of the parameter κ

$$1-(Z^H)^2\tan^2\pi\kappa = \frac{Z^2}{\sin^2\pi\kappa}. \tag{14}$$

Solving equation (14), we find

$$\sin^2\pi\kappa = \frac{1+Z^2-\sqrt{(1-Z^2)^2-4Z^2(Z^H)^2}}{1+(Z^H)^2}.$$

Equation (14) for the parameter κ differs qualitatively from the corresponding equation for zero Hall conductivity. In equation (14), the bifurcation point lies in the physical region and after this has been reached, the value of κ becomes complex, resulting in oscillations of the charge density.

Near the point $t = 1$, the functions $\rho_{1,2}$ have singularities of the type

$$\rho_1 = \frac{\tilde{A}_1}{t^{2\kappa}}, \quad \rho_2 = \frac{\tilde{A}_2}{t^{2\kappa}}, \tag{15}$$

where the coefficients $\tilde{A}_{1,2}$ are related by

$$\tilde{A}_2 = \tilde{A}_1\frac{Z}{(1+Z^H\tan\pi\kappa)\sin\pi\kappa}. \tag{16}$$

The solution of the system of equations (8) and (9) may be expressed in terms of the Weierstrass function $\wp(t)$ for which the parameters $g_{2,3}$ have the following values:

$$g_3 = 0, \quad g_2 = -4K^4(1/\sqrt{2}). \tag{17}$$

This is simply related to the Jacobi elliptic function cn [5]:

$$\wp(t) = K^2\frac{1+\mathrm{cn}(2Kt, 1/\sqrt{2})}{1-\mathrm{cn}(2Kt, 1/\sqrt{2})}, \quad K \equiv K\left(\frac{1}{\sqrt{2}}\right), \tag{18}$$

where

$$\mathrm{cn}(2Kt, 1/\sqrt{2}) = \frac{2\sqrt{2}\pi}{K(1/\sqrt{2})} \times \sum_{n=1}^{\infty}\frac{e^{-\pi(n-1/2)}}{1+e^{-\pi(2n-1)}}\cos(\pi t(2n-1)). \tag{19}$$

Taking into account formulas (13) and (16), we find

$$\tilde{\rho}_1 = A\left(\frac{\wp(t)}{K^2(1/\sqrt{2})}\right)^\kappa + B\left(\frac{\wp(1-t)}{K^2(1/\sqrt{2})}\right)^\kappa,$$

$$\tilde{\rho}_2 = -\frac{AZ}{(1-Z^H\tan\pi\kappa)\sin\pi\kappa}\left(\frac{\wp(t)}{K^2(1/\sqrt{2})}\right)^\kappa + \frac{BZ}{(1+Z^H\tan\pi\kappa)\sin\pi\kappa}\left(\frac{\wp(1-t)}{K^2(1/\sqrt{2})}\right)^\kappa. \tag{20}$$

We introduce the two functions $G_{1,2}$ using the relationships

$$G_1 = \int_0^1 dt\left(\frac{\wp(t)}{K^2(1/\sqrt{2})}\right)^\kappa,$$

$$G_2 = \int_0^1 dt'\int_0^1 dt\left(\frac{\wp(t)}{K^2(1/\sqrt{2})}\right)^\kappa\sum_N\left[\tan\left(\frac{\pi}{2}(t+N)\right)\right] \times\left\{\sinh^2\left(\frac{\pi}{2}(t'+N)\right)+\cosh^2\left(\frac{\pi}{2}(t'+N)\right)\times\tan^2\left(\frac{\pi}{2}(t+N)\right)\right\}. \tag{21}$$

Both these functions were analyzed in an earlier study [4] and we merely give their values here

$$G_1 = \frac{1}{4\sqrt{\pi}K(1/\sqrt{2})}\Gamma\left(\frac{1}{4}-\frac{\kappa}{2}\right)\Gamma\left(\frac{1}{4}+\frac{\kappa}{2}\right), \quad G_2 = G_1\tan\frac{\pi\kappa}{2}. \tag{22}$$

Integrating the system of equations (8) and (9) with respect to t' over the interval (0, 1), we obtain allowing for the formulas (21)

$$\frac{Z}{2\pi}\left\{\frac{1}{\sqrt{2}}+\frac{2\pi G_1 Z}{\sin\pi\kappa}\tan\left(\frac{\pi\kappa}{2}\right)\left(\frac{A}{1-Z^H\tan\pi\kappa}+\frac{B}{1+Z^H\tan\pi\kappa}\right)\right\}-\frac{Z^H}{2\sqrt{2}\pi} = G_1(A+B),$$

$$\frac{Z}{2\pi}\left\{\frac{1}{\sqrt{2}}+2\pi G_1(A-B)\tan\frac{\pi\kappa}{2}\right\}+\frac{Z^H}{2\sqrt{2}\pi} = \frac{ZG_1}{\sin\pi\kappa}\left(\frac{A}{1-Z^H\tan\pi\kappa}-\frac{B}{1+Z^H\tan\pi\kappa}\right). \tag{23}$$

Solving the system of equations (23) yields the following value for the coefficients A and B:

$$A+B = \frac{1}{2\sqrt{2}\pi DG_1} \times\left\{Z\left[\frac{\sin\pi\kappa}{Z}-Z\tan\frac{\pi\kappa}{2}+Z^H\sin\pi\kappa\tan\pi\kappa\tan\frac{\pi\kappa}{2}\right]\right.$$

$$+ Z^H\left[-\frac{\sin\pi\kappa}{Z} + Z\tan\frac{\pi\kappa}{2} + Z^H\sin\pi\kappa\tan\pi\kappa\tan\frac{\pi\kappa}{2}\right]\Bigg\}, \quad (24)$$

$$A - B = \frac{1}{2\sqrt{2}\pi D G_1}$$

$$\times\Bigg\{Z\left[-\frac{Z^H}{Z}\sin\pi\kappa\tan\pi\kappa + 1 - \sin\pi\kappa\tan\frac{\pi\kappa}{2}\right]$$

$$+ Z^H\left[\frac{Z^H}{Z}\sin\pi\kappa\tan\pi\kappa + 1 - \sin\pi\kappa\tan\frac{\pi\kappa}{2}\right]\Bigg\},$$

where

$$D = \left(\frac{1}{Z} + Z\tan^2\frac{\pi\kappa}{2}\right)\sin\pi\kappa - 2Z\tan\frac{\pi\kappa}{2}. \quad (25)$$

The effective conductivity is related to the charge density $\rho_{1,2}$ by

$$\sigma_{\text{eff}} = \frac{4\pi\sigma_1\sigma_2}{\sqrt{2}(\sigma_1 - \sigma_2)}\int_0^1 dt(\rho_1(t) - \rho_2(t)). \quad (26)$$

Substituting into formula (26) the explicit expression (20) for the charge densities $\rho_{1,2}$ and taking into account (24), we obtain a relatively simple expression for the effective conductivity σ_{eff} in a magnetic field:

$$\sigma_{\text{eff}} = \frac{2\sigma_1\sigma_2}{\sigma_1 + \sigma_2}$$

$$\times\frac{\sin\pi\kappa - Z^2\tan\frac{\pi\kappa}{2} + (Z^H)^2\sin\pi\kappa\tan\pi\kappa\tan\frac{\pi\kappa}{2}}{\sin\pi\kappa\left(1 + Z^2\tan^2\frac{\pi\kappa}{2}\right) - 2Z^2\tan\frac{\pi\kappa}{2}}. \quad (27)$$

The effective Hall conductivity is expressed in terms of the charge density and, consequently, in terms of the coefficients A and B using

$$H\sigma_{\text{eff}}^{\text{H}} = H\sigma_2^{\text{H}} - \frac{2\pi G_1\sigma_2}{\sqrt{2}}\Bigg\{(A + B) - \frac{Z}{\sin\pi\kappa}$$

$$\times\left(\frac{A}{1 - Z^H\tan\pi\kappa} - \frac{B}{1 + Z^H\tan\pi\kappa}\right) - \Bigg[(A - B) \quad (28)$$

$$- \frac{Z}{\sin\pi\kappa}\left(\frac{A}{1 - Z^H\tan\pi\kappa} + \frac{B}{1 + Z^H\tan\pi\kappa}\right)\Bigg]\tan\frac{\pi\kappa}{2}\Bigg\}.$$

Using equation (24), expression (28) for the Hall conductivity can be simplified substantially and reduced to the form

$$H\sigma_{\text{eff}}^{\text{H}} = \frac{H(\sigma_1\sigma_2^{\text{H}} + \sigma_2\sigma_1^{\text{H}})}{\sigma_1 + \sigma_2} - \frac{4\pi\sigma_1\sigma_2}{\sqrt{2}Z(\sigma_1 + \sigma_2)}$$

$$\times\Bigg\{G_1(A + B) - G_1Z(A - B)\tan\frac{\pi\kappa}{2} - \frac{Z}{2\sqrt{2}\pi} + \frac{Z^H}{2\sqrt{2}\pi}\Bigg\}. \quad (29)$$

Substituting into formula (29) expressions (24) for the coefficients A and B, we obtain the final expression for the Hall component of the effective conductivity

$$\sigma_{\text{eff}}^{\text{H}} = \frac{\sigma_1\sigma_2^{\text{H}} + \sigma_2\sigma_1^{\text{H}}}{\sigma_1 + \sigma_2} - \frac{2\sigma_1\sigma_2Z^2(\sigma_1^{\text{H}} - \sigma_2^{\text{H}})}{\sigma_1^2 - \sigma_2^2}$$

$$\times\left[\left(\tan^2\frac{\pi\kappa}{2} + \tan\frac{\pi\kappa}{2}\tan\pi\kappa\right)\sin\pi\kappa - \tan\frac{\pi\kappa}{2}\right] \quad (30)$$

$$\times\left[\left(1 + Z^2\tan^2\frac{\pi\kappa}{2}\right)\sin\pi\kappa - 2Z^2\tan\frac{\pi\kappa}{2}\right]^{-1}.$$

Equation (14) contains a bifurcation point after which the parameter κ becomes complex. The bifurcation point is determined from the condition that the derivative of κ with respect to the parameter $(Z^H)^2$ goes to infinity. From this condition, we find the bifurcation point

$$Z^2 + 2|ZZ^H| - 1 = 0. \quad (31)$$

On account of the linearity of the initial equations, above the bifurcation point in equations (27) and (30) for the effective conductivities σ_{eff} and $\sigma_{\text{eff}}^{\text{H}}$ the real part should be taken from the right-hand sides. The appearance of an imaginary part of the parameter κ leads to oscillations of the charge density at the interfaces. If the Hall conductivities of components 1 and 2 differ, the effective conductivity will have a high magnetoresistance.

3. CONCLUSIONS

If the Hall conductivity of at least one of the components is nonzero under the condition $\sigma_1^{\text{H}} \neq \sigma_2^{\text{H}}$, the effective conductivity possesses strong magnetoresistance. In this case, there is a bifurcation point above which the charge density oscillates along the interfaces. For a fixed value of the parameter $Z = (\sigma_1 - \sigma_2)/(\sigma_1 + \sigma_2)$, the bifurcation point is realized with increasing magnetic field. This property of a two-component system is promising from the point of view of experimental confirmation.

ACKNOWLEDGMENTS

The author thanks P. Monceau for supporting the work at CRTBT (Grenoble).

REFERENCES

1. A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 110 (1970) [Sov. Phys. JETP **32**, 63 (1971)].
2. A. L. Efros and B. I. Shklovskii, Phys. Status Solidi B **76**, 475 (1976).
3. A. M. Dykhne and I. M. Kaganova, Phys. Rep. **288**, 263 (1997).
4. Yu. N. Ovchinnikov and A. M. Dyugaev, Zh. Éksp. Teor. Fiz. **117**, 1013 (2000) [JETP **90**, 881 (2000)].
5. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Nauka, Moscow, 1971; Academic, New York, 1980).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 1063–1070.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1221–1229.

SOLIDS
Electronic Properties

Method of Envelope Functions and Intervalley Γ–X_z Interaction of States in (001) III–V Semiconductor Heterostructures

É. E. Takhtamirov* and V. A. Volkov**

Institute of Radio Engineering and Electronics, Russian Academy of Sciences, Moscow, 103907 Russia

**e-mail: gibbon@royal.net*

***e-mail: VoVA@mail.cplire.ru*

Received January 11, 2000

Abstract—The **kp** method is used to analyze the problem of intervalley Γ–X_z interaction of conduction-band states in the (001) lattice-matched III–V semiconductor heterostructures. A convenient basis for expansion of the wave function is systematically selected and a multiband system of equations is derived for the envelope functions which is then reduced to a system of three equations for three valleys (Γ_1, X_1, and X_3) by using a unitary transformation. Intervalley Γ–X_z mixing is described by short-range potentials localized at heterojunctions. The expressions for the parameters determining the Γ–X_z mixing strength explicitly contain the chemical-composition profile of the structure since mixing is naturally stronger for abrupt heterojunctions than for structures with a continuously varying chemical composition. It is shown that direct Γ_1–X_1 interaction of comparable strength to the Γ_1–X_3 interaction exists. This must be taken into account when interpreting tunnel and optical experiments since the X_1 valley is substantially lower in energy than the X_3 valley. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Heterostructures with quantum wells and barriers using GaAs/AlGaAs are popular objects for studying the physical processes accompanying resonant tunneling and optical transitions. These structures are particularly interesting because in GaAs the minimum of the conduction band is situated at the center of the first Brillouin zone (Γ_1 symmetry) whereas the minimum of the AlAs conduction band lies near the X point where the bands having X_1 and X_3 symmetry are close together, the X_1 band being the principal one (the X_3 band is 350 meV higher energy). On each cubic axis, there is a pair of X valleys, one from the X_1 band and one from the X_3 band. Layers with predominantly Ga content act as barriers for X electrons while regions with Al predominating act as barriers for Γ electrons. Under certain conditions, resonant tunneling interaction can take place between energetically close states of the Γ and X valleys and this is even observed on the current–voltage characteristics of single-barrier structures (see, for example, [1, 2]) and also in optical experiments [3, 4]. The conditions for the existence of this interaction generally involve the presence of structural defects, impurities, heterojunction roughness, and/or interaction with short-wavelength phonons. However, in the (001) heterostructures, interaction of a Γ valley with X_z valleys (i.e., X valleys lying on the z axis directed along the normal to the surface) is caused by the potential of the structure itself which leads to nonconservation of the quasimomentum component k_z perpendicular to the junction. As a result of size quantization for fairly thin (less than 50 Å) AlGaAs layers, X_z-valley states are situated below all X valleys so that an analysis of Γ–X_z interaction at the heterojunction is important for an accurate description of Γ–X junctions [5].

This type of intervalley mixing has been studied theoretically both phenomenologically [6] and using a tight-binding model [7, 8], and also using a pseudopotential method [9]. Nevertheless, our level of understanding of the processes leading to Γ–X_z mixing of electronic states is far from satisfactory. For instance, the results of [8, 9] indicate that direct Γ_1–X_1 interaction is extremely weak and Γ–X_z mixing is merely attributable to Γ_1–X_3 interaction, whereas, according to [7], Γ_1–X_1 interaction is the determining factor (however, the parameters of the tight-binding model used in [7] are such that the X_3 band is very high in terms of energy and is in fact eliminated from the analysis). Studies [10, 11] in which the form of the Γ–X_z interaction potential was determined by direct calculations of the matrix elements of the model heterointerface potential using the complete wave functions of the states also give a contradictory answer to the question of the strength of the Γ_1–X_1 interaction. According to the results of [10], this interaction is weak whereas the results of [11] suggest the opposite: Γ_1–X_1 mixing can be comparable with

Γ_1–X_3 mixing. The solution of this problem is important for interpreting tunneling and optical experiments since Γ_1–X_1 junctions are usually observed experimentally.

In the present study Γ–X_z interaction (mixing) is analyzed using the method of envelope functions. This is a fairly explicit technique which does not have the disadvantage of the tight-binding model in which the heterointerface is oversimplified, while calculations using the empirical pseudopotential method are much more cumbersome than those using the envelope function method. Note that so far attempts to derive a system of equations for the envelope functions jointly describing the Γ and X states have not produced satisfactory results. This is mainly because an atomically abrupt change in the crystal potential at the heterojunctions produces Γ –X_z mixing [12] and correct allowance for such abrupt changes in the potential is outside the scope of the usual method of Luttinger–Kohn envelope functions. In [10, 11] in which the problems were analyzed using the envelope function method, Γ–X interaction was analyzed using perturbation theory but the selected basis functions were a generally overfull set corresponding to the set of Kohn–Luttinger functions for the Γ and X states. An overfull (and nonorthogonal) basis can, in principle, give an erroneous result. In addition, this approach cannot be applied directly to describe the states of the continuous spectrum which is important for the Γ–X_z tunneling problem. Nevertheless, the results of the present study agree qualitatively with the conclusions of [11].

The essential features of the problem of adequately describing intervalley mixing of states in heterostructures is similar to the intervalley splitting of impurity states in multivalley semiconductors (for example, in Si and Ge). We know [13] that allowance for the short-range part of the impurity potential ("the correction to the central cell") not only yields a chemical shift of the impurity-state energy but also lifts the valley degeneracy. In [14], we proposed a fairly simple method of analyzing heterojunctions with an atomically abrupt change in chemical composition. This method involves isolating the "smooth" component of the heterostructure potential and the "abrupt" component which is only nonzero near the heterojunction. The smooth component is "processed" by a standard technique (the Kohn–Luttinger method) while the abrupt component is considered as a correction to the central cell. In [14], we only considered states near the Γ point in the Brillouin zone and in the present study we develop the method further to describe the interaction of states near different points in $\mathbf{k}$ space. These results were first presented at the III All-Russia Conference on the Physics of Semiconductors, see [15] and also [16].

2. FORMALISM OF THE ENVELOPE FUNCTION METHOD

2.1. Formulation of the Problem

We shall consider the electron states in the (001) III–V heterostructures formed from related lattice-matched semiconductors having zinc blende symmetry. By related structural materials we understand a fairly small band offset so that in the energy range of interest the conduction-band states near the point Γ can be described using a single-band variant and the states near the point X can be described using two-band (for the X_1 and X_3 bands) variants of the envelope function method. We shall also assume that the energy gap between the states of interest to us in the Γ and X valleys is smaller than or of the order of the band offset. This means that we can consider the direct interaction of the Γ_1, X_1, and X_3 states exactly and interaction via all other bands can be taken into account using perturbation theory. Since the desired Hamiltonian of the equation for the envelope functions should depend explicitly on the number of monoatomic layers of each material forming the structure [7, 17], we shall consider a structure with two symmetric heterojunctions as the simplest case to obtain this dependence. For simplicity, we shall neglect spin–orbit interaction and also external smooth potentials. The single-electron Schrödinger equation then has the following form:

$$\left(\frac{\mathbf{p}^2}{2m_0} + U\right)\Psi(\mathbf{r}) = \epsilon\Psi(\mathbf{r}). \tag{1}$$

Here, m_0 is the free electron mass, $\mathbf{p}$ is the momentum operator, and $U \equiv U(\mathbf{r})$ is the crystal potential of the heterostructure. We shall first use the following model for this potential (a more realistic situation will be discussed in Section 3):

$$U = U_1 + P(z)[U_2 - U_1] \equiv U_1 + P(z)\delta U, \tag{2}$$

and $U_1 \equiv U_1(\mathbf{r})$ and $U_2 \equiv U_2(\mathbf{r})$ are periodic (continued to all space) potentials of the two heterostructure materials, the z-axis is perpendicular to the heterojunction plane; the form factor $P(z)$ of a heterostructure having the heterointerfaces $z = 0$ and $z = L$ is defined so that

$$P(z) = \begin{cases} 0, & z < -d \\ 1, & d < z < L - d \\ 0, & z > L + d. \end{cases} \tag{3}$$

The behavior of the function $P(z)$ in the transition regions (of width $2d$ near the heterointerfaces) may be fairly arbitrary. Here, the symmetry of the structure implies that $P(z) = P(L - z)$, $L > d$. We shall assume that a layer of width L contains an integer number of monolayers: $L = \mathcal{N}a/2$, where $\mathcal{N}$ is a natural number.

2.2. Choice of Complete Orthonormalized Set of Basis Functions

As in [14], the potential $P(z)\delta U$ is analyzed in terms of perturbation theory and the complete and orthonormalized set of functions used to expand the complete wave function $\Psi(\mathbf{r})$ is constructed of Bloch functions of the base semiconductor (having the crystal potential U_1). In our case, the most natural basis is a mixed basis of Kohn–Luttinger functions for the points Γ and X_z. In order to make this set complete and orthonormalized, it should be constructed as follows. We first expand $\Psi(\mathbf{r})$ in terms of the Bloch functions $u_{n\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}$ of the base crystal which correspond to the energy eigenvalues $\epsilon_{n\mathbf{k}}$, where n and $\mathbf{k}$ are the band index and the quasiwave vector, respectively:

$$\Psi(\mathbf{r}) = \sum_{n'}\int \mathscr{A}_{n'}(\mathbf{k}')e^{i\mathbf{k}'\cdot\mathbf{r}}u_{n'\mathbf{k}'}d^3k'. \tag{4}$$

Summation in (4) is performed over all bands and integration is performed over the region Λ_0 of nonequivalent $\mathbf{k}$. In order not to consider two equivalent X_z points having the coordinates $(0, 0, 2\pi/a)$ and $(0, 0, -2\pi/a)$ in $\mathbf{k}$ space, we shall not operate in the first Brillouin zone constructed as a Wigner–Seitz cell but we shall define a region Λ_0 such that points Γ and are contained in this region, for example, $\mathbf{q} = (0, 0, 2\pi/a)$ with their vicinities. Following [13, §7–3] we divide Λ_0 into two subregions Λ_Γ and $\tilde{\Lambda}_X$ containing the points Γ and X_z with their vicinities where $\Lambda_\Gamma \cup \tilde{\Lambda}_X = \Lambda_0$ and $\Lambda_\Gamma \cap \tilde{\Lambda}_X = 0$ (see comments on this method of division in Section 2.4). Now, following [18], we use series expansions of the periodic function $u_{n\mathbf{k}}$:

$$u_{n\mathbf{k}} = \sum_{m'} b_{m'n}(\mathbf{k})u_{m'0}, \quad u_{n\mathbf{k}} = \sum_{m'} c_{m'n}(\mathbf{k})u_{m'\mathbf{q}}.$$

Then (4) can be rewritten in the following form:

$$\Psi(\mathbf{r}) = \sum_{n', m'} \int_{\mathbf{k}' \in \Lambda_\Gamma} \mathscr{A}_{n'}(\mathbf{k}')b_{m'n'}(\mathbf{k}')e^{i\mathbf{k}'\cdot\mathbf{r}}u_{m'0}d^3k' \tag{5}$$

$$+ \sum_{n', m'} \int_{\mathbf{k}'+\mathbf{q} \in \tilde{\Lambda}_X} \mathscr{A}_{n'}(\mathbf{k}'+\mathbf{q})c_{m'n'}(\mathbf{k}'+\mathbf{q})e^{i\mathbf{k}'\cdot\mathbf{r}}e^{i\mathbf{q}\cdot\mathbf{r}}u_{m'\mathbf{q}}d^3k'.$$

We then define the functions

$$\mathscr{F}^{(\Gamma)}_{m'}(\mathbf{k}') = \sum_{n'} \mathscr{A}_{n'}(\mathbf{k}')b_{m'n'}(\mathbf{k}'), \tag{6}$$

$$\mathscr{F}^{(X)}_{m'}(\mathbf{k}') = \sum_{n'} \mathscr{A}_{n'}(\mathbf{k}'+\mathbf{q})c_{m'n'}(\mathbf{k}'+\mathbf{q}), \tag{7}$$

which will specifically comprise the envelopes of the functions of the Γ and X states in the $\mathbf{k}$ representation and we shall define the region Λ_X such that the condition $\mathbf{k} \in \Lambda_X$ (shift of the origin in $\mathbf{k}$ space) is satisfied for all $\mathbf{k} + \mathbf{q} \in \tilde{\Lambda}_X$. Now (5) can have the form of an expansion of $\Psi(\mathbf{r})$ in terms of a complete set of Kohn–Luttinger functions near the points Γ and X_z:

$$\Psi(\mathbf{r}) = \sum_{m'} \int_{\Lambda_\Gamma} \mathscr{F}^{(\Gamma)}_{m'}(\mathbf{k}')\chi^{(\Gamma)}_{m'\mathbf{k}'}d^3k' + \sum_{m'} \int_{\Lambda_X} \mathscr{F}^{(X)}_{m'}\chi^{(X)}_{m'\mathbf{k}'}d^3k' \tag{8}$$

$$= \sum_{\nu' = \Gamma, X} \sum_{m'} \int_{\Lambda_{\nu'}} \mathscr{F}^{(\nu')}_{m'}(\mathbf{k}')\chi^{(\nu')}_{m'\mathbf{k}'}d^3k',$$

where the Kohn–Luttinger functions are

$$\chi^{\Gamma}_{m\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{r}}u_{m0} \equiv e^{i\mathbf{k}\cdot\mathbf{r}}\phi^{(\Gamma)}_m,$$
$$\chi^{(X)}_{m\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{r}}e^{i\mathbf{q}\cdot\mathbf{r}}u_{m\mathbf{q}} \equiv e^{i\mathbf{k}\cdot\mathbf{r}}\phi^{(X)}_m.$$

Thus, the Fourier transforms of the envelope functions constructed above only differ from the usual ones [18] in terms of the domain of definition. In our case, these are the regions Λ_Γ (for states near the center of the Brillouin zone) and Λ_X (for states near the X_z point) rather than the complete first Brillouin zone.

We shall also assume that $\phi^{(\nu)}_m$ are real. Taking the following orthonormalization relationship for the Bloch functions:

$$\int_{\substack{\text{all}\\ \text{spase}}} u^*_{n'\mathbf{k}'}e^{-i\mathbf{k}'\cdot\mathbf{r}}u_{n\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{r}}d^3r = \delta_{nn'}\delta(\mathbf{k}-\mathbf{k}'), \tag{9}$$

we obtain the required orthonormalization relationship for the basis functions [18]:

$$\int_{\substack{\text{all}\\ \text{spase}}} (\chi^{(\nu')}_{n'\mathbf{k}'})^*\chi^{(\nu)}_{n\mathbf{k}}d^3r = \delta_{\nu\nu'}\delta_{nn'}\delta(\mathbf{k}-\mathbf{k}'), \tag{10}$$

*2.3. Multiband System of **kp** Equations*

In the basis specified above the procedure for obtaining the **kp** system of equations is trivial, see [18]. Using the expansion (8) in (1), multiplying both sides of the equation by $(\chi^{(\nu')}_{n'\mathbf{k}'})^*$, and integrating over all $\mathbf{r}$ space, we obtain the following system of equations:

$$\left(\epsilon_m^{(\nu)} + \frac{\hbar^2\mathbf{k}^2}{2m_0}\right)\mathcal{F}_m^{(\nu)}(\mathbf{k}) + \sum_{m'}\frac{\hbar\mathbf{p}_{mm'}^{(\nu)}\mathbf{k}}{m_0}\mathcal{F}_{m'}^{(\nu)}(\mathbf{k}) \tag{11}$$

$$+ \sum_{\nu' = \Gamma, X}\sum_{m'}\int_{\Lambda_{\nu'}}\mathcal{M}_{mm'}^{(\nu\nu')}(\mathbf{k}, \mathbf{k}')\mathcal{F}_{m'}^{(\nu')}(\mathbf{k}')d^3k' = \epsilon\mathcal{F}_m^{(\nu)}(\mathbf{k}).$$

Here, $\epsilon_m^{(\Gamma)} = \epsilon_{m0}$ and $\epsilon_m^{(X)} = \epsilon_{m\mathbf{q}}$,

$$\mathbf{p}_{mm'}^{(\nu)} = \langle m, \nu|\mathbf{p}|m', \nu\rangle \equiv \frac{(2\pi)^3}{\Omega}\int_{\text{cell}}\phi_m^{(\nu)}\mathbf{p}\phi_{m'}^{(\nu)}d^3r,$$

$\langle m, \nu|\mathbf{p}|m', \nu'\rangle = 0$ for $\nu \neq \nu'$, Ω is the unit cell volume, and

$$\mathcal{M}_{mm'}^{(\nu\nu')}(\mathbf{k}, \mathbf{k}') = \int_{\substack{\text{all}\\ \text{spase}}} e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}}\phi_m^{(\nu)}P(z)\delta U\phi_{m'}^{(\nu')}d^3r.$$

The matrix elements $\mathcal{M}_{mm'}^{(\nu\nu)}(\mathbf{k}, \mathbf{k}')$ were analyzed in [14]. It was shown that the contribution of the perturbation potential can be divided into smooth and abrupt components (the latter is exponentially small for smooth perturbations on the scale a). The contribution of the abrupt component is a correction to that of the smooth component (for the case when the width of the heterostructure layers is much greater than a) and it can be written in the form of converging series in powers of $(k_z - k_z')$. We shall use the effective-mass approximation with spatially independent effective-mass parameters and we shall only allow for the abruptness of the heterojunctions in the first order in terms of the parameter $a\bar{k}_z$ [14, 19], where $\bar{k}_z$ is the characteristic quasimomentum of the state. We shall consider the intervalley elements $\mathcal{M}_{mm'}^{(\nu\nu')}(\mathbf{k}, \mathbf{k}')$, $\nu \neq \nu'$, in greater detail and for $\mathcal{M}_{mm'}^{(\nu\nu')}(\mathbf{k}, \mathbf{k}')$, following [14, 19], we obtain

$$\mathcal{M}_{mm'}^{(\nu\nu)}(\mathbf{k}, \mathbf{k}') = \left[\mathcal{P}(k_z - k_z')\delta U_{mm'}^{(\nu)}\right.$$

$$\left. + \frac{1}{2\pi}\sum_{j\neq 0}\frac{\langle m, \nu|\delta U e^{iK_jz}|m', \nu\rangle}{iK_j}\right. \tag{12}$$

$$\left.\times\left(\int_{-d}^{d}P'(z)e^{-iK_jz}dz - e^{-i(k_z-k_z')L}\int_{-d}^{d}P'(z)e^{iK_jz}dz\right)\right]\delta(k_\parallel - k_\parallel').$$

Here we introduce the notation: $\mathcal{P}(k_z)$ is the Fourier transform of the function $P(z)$; the matrix element is $\delta U_{mm'}^{(\nu)} = \langle m, \nu|\delta U|m', \nu\rangle$; $K_j = (4\pi/a)j$, j is an integer; $P'(z) = dP(z)/dz$, and $\mathbf{k}_\parallel = (k_x, k_y, 0)$. We used the symmetry $P(z) = P(L - z)$ and also the fact that $e^{iK_jL} = 1$.

Here, we give all the matrix elements (12) required subsequently. We assign the indices **w**, **u**, and **v** to the conduction-band states Γ_1, X_1, and X_3, respectively, and omit the valley indices since the band number in our approximation now uniquely defines the state. Using symmetry concepts we obtain

$$\mathcal{M}_{ss}(\mathbf{k}, \mathbf{k}') = \left[\mathcal{P}(k_z - k_z')\delta U_{ss} + \frac{1}{2\pi}d_{ss}\left(1 + e^{-i(k_z-k_z')L}\right)\right]$$

$$\times\delta(\mathbf{k}_\parallel - \mathbf{k}_\parallel'),$$

where s = **w**, **u**, **v**; and the parameters d_{ss} are determined as follows:

$$d_{ss} = -\sum_{j\neq 0}\frac{\langle s|\delta U\cos(K_jz)|s\rangle}{K_j}\int_{-d}^{d}P'(z)\sin(K_jz)dz.$$

The potential also makes a contribution to the direct interaction of the X_1 and X_3 states:

$$\mathcal{M}_{\mathbf{uv}}(\mathbf{k}, \mathbf{k}') = \frac{1}{2\pi}d_{\mathbf{uv}}\left(1 - e^{-i(k_z-k_z')L}\right)\delta(\mathbf{k}_\parallel - \mathbf{k}_\parallel'),$$

where

$$d_{\mathbf{uv}} = \sum_{j\neq 0}\frac{\langle \mathbf{u}|\delta U\sin(K_jz)|\mathbf{v}\rangle}{K_j}\int_{-d}^{d}P'(z)\cos(K_jz)dz.$$

We shall now consider the most interesting intervalley matrix elements $\mathcal{M}_{mm'}^{(\nu\nu')}$, $\nu \neq \nu'$:

$$\mathcal{M}_{mm'}^{(\Gamma X)}(\mathbf{k}, \mathbf{k}') = \int_{\substack{\text{all}\\ \text{spase}}} P(z)e^{-i(\mathbf{k}-\mathbf{k}'-\mathbf{q})\cdot\mathbf{r}}u_{m0}\delta U u_{m'\mathbf{q}}d^3r.$$

We shall analyze this matrix element using the same method which yielded (12). We shall use an expansion of the periodic function $u_{m0}\delta U u_{m'\mathbf{q}}$ as a Fourier series which gives

$$\mathcal{M}_{mm'}^{(\Gamma X)}(\mathbf{k}, \mathbf{k}') = \sum_l C_l^{m(\Gamma)m'(X)}\mathcal{P}(k_z - k_z' - q_z + K_{zl}) \tag{13}$$

$$\times\delta(\mathbf{k}_\parallel - \mathbf{k}_\parallel' + \mathbf{K}_{\parallel l}),$$

where $\mathbf{K}_l \equiv (K_{zl}, \mathbf{K}_{\parallel l})$ are the vectors of the reciprocal lattice, and

$$C_l^{m(\Gamma)m'(X)} = \frac{(2\pi)^3}{\Omega}\int_{\text{cell}}u_{m0}\delta U e^{i\mathbf{K}_l\cdot\mathbf{r}}u_{m'\mathbf{q}}d^3r$$

$$\equiv\left\langle m, \Gamma\left|\delta U e^{i(\mathbf{K}_l-\mathbf{q})\cdot\mathbf{r}}\right|m', X\right\rangle.$$

As was shown in [19], for the region where $|k_x| + |k_y| < \pi/a$ whose size is fairly large for our purposes [we obviously used this constraint in the derivation of (12)],

only the vectors of the reciprocal lattice with $\mathbf{K}_{\|l} = 0$ will contribute to (13):

$$\begin{aligned}\mathcal{M}^{(\Gamma X)}_{mm'}(\mathbf{k}, \mathbf{k}') &= \sum_j \left\langle m, \Gamma \middle| \delta U e^{i(K_j - q_z)z} \middle| m', X \right\rangle \\ &\times \mathcal{P}(k_z - k_z' + K_j - q_z)\delta(\mathbf{k}_\| - \mathbf{k}_\|') \\ &= \sum_{j = \pm 1, \pm 3, \pm 5, \ldots} \left\langle m, \Gamma \middle| \delta U e^{i(2\pi/a)jz} \middle| m', X \right\rangle \\ &\times \mathcal{P}\left(k_z - k_z' + \frac{2\pi}{a} j\right)\delta(\mathbf{k}_\| - \mathbf{k}_\|').\end{aligned} \tag{14}$$

We shall analyze the functions $\mathcal{P}(k_z - k_z' + (2\pi/a)j)$ contained in (14) in greater detail:

$$\begin{aligned}\mathcal{P}\left(k_z - k_z' + \frac{2\pi}{a} j\right) &= \frac{1}{2\pi i}\frac{1}{k_z - k_z' + 2\pi j/a} \\ &\times \int_{-\infty}^{\infty} P'(z)\exp\left\{-i\left(k_z - k_z' + \frac{2\pi}{a} j\right)z\right\}dz \\ &= \frac{1}{2\pi i}\frac{1}{k_z - k_z' + 2\pi j/a} \\ &\times \left(\int_{-d}^{d} P'(z)\exp\left\{-i\frac{2\pi}{a} jz\right\}\exp\{-i(k_z - k_z')z\}dz\right. \\ &- \exp\{-i(k_z - k_z')L\}\exp\left\{-i\frac{2\pi}{a} jL\right\} \\ &\left.\times \int_{-d}^{d} P'(z)\exp\left\{i\frac{2\pi}{a} jz\right\}\exp\{i(k_z - k_z')z\}dz\right),\end{aligned}$$

where we again used the property $P(z) = P(L - z)$. We now expand $(k_z - k_z' + (2\pi/a)j)^{-1}$, $\exp(-i(k_z - k_z')z)$, and $\exp(i(k_z - k_z')z)$ as series in powers of $(k_z - k_z')$ where the convergence of the first series is ensured by the property $|k_z - k_z'| < 2\pi/a$ since $\mathbf{k} \in \Lambda_\Gamma$ and $\mathbf{k}' \in \Lambda_X$. We only retain the first terms of the expansions. This gives a good approximation for $\bar{k}_z \ll 2\pi/a$ and $\bar{k}_z \ll 1/(2d)$, where $\bar{k}_z$ is the characteristic quasimomentum of the state or the reciprocal characteristic length of variation of the envelope functions (6) and (7) in the **r** representation (i.e., the envelope functions should vary continuously on scales of order a and on scales of the order of the widths of the interface regions). Now, bearing in mind that for all values of the summation index j in (14) $\exp(-2\pi ijL/a) = (-1)^{\mathcal{N}}$, we obtain

$$\begin{aligned}\mathcal{P}\left(k_z - k_z' + \frac{2\pi}{a} j\right) &= \frac{1}{2\pi i}\frac{a}{2\pi j}\left(\int_{-d}^{d} P'(z)\exp\left(-i\frac{2\pi}{a} jz\right)dz\right. \\ &\left.- (-1)^{\mathcal{N}}\exp\{-i(k_z - k_z')L\}\int_{-d}^{d} P'(z)\exp\left(i\frac{2\pi}{a} jz\right)dz\right).\end{aligned}$$

Here, there is a dependence of the effective potential on the number of monoatomic layers $\mathcal{N}$, which mixes the valley states assigned to different points in **k** space. We now write the matrix elements we require (we again drop the valley index):

$$\mathcal{M}_{\mathbf{wu}}(\mathbf{k}, \mathbf{k}') = \frac{1}{2\pi} d_{\mathbf{wu}}\left(1 + (-1)^{\mathcal{N}} e^{-i(k_z - k_z')L}\right)\delta(\mathbf{k}_\| - \mathbf{k}_\|'),$$

where

$$\begin{aligned}d_{\mathbf{wu}} &= -\sum_{j = \pm 1, \pm 3, \pm 5, \ldots} \frac{a}{2\pi j}\left\langle \mathbf{w} \middle| \delta U \cos\left(\frac{2\pi}{a} jz\right) \middle| \mathbf{u} \right\rangle \\ &\times \int_{-d}^{d} P'(z)\sin\left(\frac{2\pi}{a} jz\right)dz\end{aligned} \tag{15}$$

and

$$\mathcal{M}_{\mathbf{wv}}(\mathbf{k}, \mathbf{k}') = \frac{1}{2\pi} d_{\mathbf{wv}}\left(1 - (-1)^{\mathcal{N}} e^{-i(k_z - k_z')L}\right)\delta(\mathbf{k}_\| - \mathbf{k}_\|'),$$

where

$$\begin{aligned}d_{\mathbf{wv}} &= \sum_{j = \pm 1, \pm 3, \pm 5, \ldots} \frac{a}{2\pi j}\left\langle \mathbf{w} \middle| \delta U \sin\left(\frac{2\pi}{a} jz\right) \middle| \mathbf{v} \right\rangle \\ &\times \int_{-d}^{d} P'(z)\cos\left(\frac{2\pi}{a} jz\right)dz.\end{aligned} \tag{16}$$

It can be seen that the abrupt potential of the heterointerfaces not only ensures Γ_1–X_3 interaction [8, 17, 10] but also Γ_1–X_1 interaction [11].

We have obtained a multiband **kp** system of equations and explicit expressions for all the elements of this system required subsequently. We shall now obtain a system of 3×3 equations for the strongly interacting states Γ_1, X_1, and X_3.

2.4. Elimination of Far-Field Zones and Transition to **r** *Space*

In the effective-mass approximation with spatially independent effective-mass parameters, the approximate unitary transformation procedure which eliminates the influence of far-field zones in the required order of perturbation theory is performed by a standard method [18] and is not given here. We obtain an integral system of equations in the **k** representation. The aim of the present study is to simplify the final results as far as

possible which may be achieved if the corresponding equations are differential. The problem of the accuracy of the envelope function method which occurs on transition from integral to differential equations and also when the far-field zones are eliminated has already been discussed in [14, 19]. We shall summarize the constraints imposed on the accuracy of the envelope function method for this case.

First, we need to determine the effective radius of the region in **k** space for which the system of equations for the envelope functions is correct (we shall call this **k** region the fundamental region). Whereas the multiband system of equations (11) is valid for all $\mathbf{k} \in \Lambda_\nu$ and $\mathbf{k}' \in \Lambda_{\nu'}$, the unitary transformation of this system which eliminates the far-field zones can, in principle, reduce the dimensions of the fundamental **k** regions. This is easily understood from the following. The spectrum of states of a bulk semiconductor $\epsilon_n^{(\mathbf{k}_0)}(\mathbf{k})$ near the point $\mathbf{k}_0$ in a band numbered n may be represented as a series in powers of **k** (for degenerate states the spectrum is determined by diagonalizing the matrix whose elements are these series). The series has a finite radius of convergence R_0 which is determined by the strength of the **kp** interaction with the far-field zones. This radius can be estimated as $R_0 = m_0 \bar{E}_g / 2\hbar \bar{P}$, where $\bar{E}_g$ and $\bar{P}$ are the characteristic values of the interband energy at point $\mathbf{k}_0$ and the interband matrix element of the momentum. States having quasimomenta which do not belong to the fundamental region cannot be correctly taken into account in the transformed equation and should be neglected. We denote the corresponding radii of the fundamental **k** regions as $R_0^{(\Gamma)}$ and $R_0^{(X)}$ for conduction-band states near the points Γ and X, respectively.

Secondly, on changing from the **k** to the **r** representation, integration is performed over regions of **k** space of finite dimensions which makes it difficult to obtain differential equations directly in **r** space. The local approximation formula involves replacing the finite regions of **k** space by infinite ones. Since the heterostructure potential is not smooth and the envelope functions obtained (or their derivatives) can vary appreciably on scales of the order of a, this procedure does not give an exponentially small error as in the case of smooth perturbations but an error which is only small in terms of power (for this analysis it is convenient to consider the limiting case of a mathematically abrupt potential and then the envelope functions or their derivatives obtained as a result of the local approximation may have a discontinuity). Since the perturbation-theory series used for the unitary transformation is a power series, it is important to avoid the inclusion of extra-accuracy terms. For the simple single-valley case (Γ states) analyzed in detail in [19], the error of the method is of the order of $(\bar{k}_z^{(\Gamma)}/R_0^{(\Gamma)})^M$ where $1/\bar{k}_z^{(\Gamma)}$ is the characteristic length of variation of the corresponding envelope function and the exponent M is a measure of the smoothness of the latter (for an isolated heterojunction or a fairly wide quantum well $M = 3$). In this case, we need to determine two characteristic quasimomenta $\bar{k}_z^{(\Gamma)}$ and $\bar{k}_z^{(X)}$ for the Γ and X states. In the zero-order approximation which is acceptable for estimating the accuracy of the method, the Γ and X states do not interact and the constraint associated with the transition to differential equations is determined by the error $(\bar{k}_z^{(\Gamma)}/R_0^{(\Gamma)})^3$ for the Γ states (we shall assume that the layer width L is sufficiently large so that $\bar{k}_z^{(\Gamma)} L \gtrsim 1$). For X states, the situation is slightly more complex: the system of equations for these contains both second derivatives of the envelope functions with respect to z and first derivatives as a result of **kp** interaction between the X_1 and X_3 bands, where the effect of this interaction may be comparable with the contribution of terms which are quadratic with respect to the momentum operator. This means that in the "worst" case, the accuracy of the local approximation for the X states is limited by the error $(\bar{k}_z^{(\Gamma)}/R_0^{(\Gamma)})^2$. All these factors allow us to consider the corrections which appear as a result of the abruptness of the change in the heterointerface potential which are small with respect to the parameters $a\bar{k}_z^{(\Gamma)}$ and $a\bar{k}_z^{(X)}$. We also note that splitting the region Λ_0 into subregions Λ_Γ and $\tilde{\Lambda}_X$ should be performed so that $R_0^{(\Gamma)}$ and $R_0^{(X)}$ are not larger than the radii of the regions Λ_Γ and Λ_X, respectively, and then $R_0^{(\Gamma)}$ and $R_0^{(X)}$, not the radii Λ_Γ and Λ_X, appear in the expressions to estimate the accuracy of the local approximation (this was implied above).

As a result, the required system of differential equations for the transformed envelope functions $\tilde{F}_m(\mathbf{r})$, where $m = \mathbf{w}, \mathbf{u}, \mathbf{v}$, has the following form for the strongly interacting Γ_1, X_1, and X_3 states:

$$\sum_{m' = \mathbf{w}, \mathbf{u}, \mathbf{v}} (T_{mm'} + V_{mm'}(z))\tilde{F}_{m'}(\mathbf{r}) = \epsilon \tilde{F}_m(\mathbf{r}). \qquad (17)$$

Here, $\mathbf{T}$ and $\mathbf{V}$ are the (3×3) matrices of the kinetic and potential energies. The form of the matrix of the effective kinetic energy operator is known [20]:

$$\mathbf{T} = \begin{pmatrix} \dfrac{\mathbf{p}^2}{2m_{\mathbf{w}}} & 0 & 0 \\ 0 & \dfrac{\mathbf{p}_\parallel^2}{2m_{\mathbf{u}\parallel}} + \dfrac{p_z^2}{2m_{\mathbf{u}\perp}} & \dfrac{(p_z)_{\mathbf{uv}} p_z}{m_0} + \gamma p_x p_y \\ 0 & \dfrac{(p_z)_{\mathbf{vu}} p_z}{m_0} + \gamma p_x p_y & \dfrac{\mathbf{p}_\parallel^2}{2m_{\mathbf{v}\parallel}} + \dfrac{p_z^2}{2m_{\mathbf{v}\perp}} \end{pmatrix},$$

where $m_{\mathbf{w}}$ is the effective mass for the Γ conduction-band states, $m_{l\parallel}$ and $m_{l\perp}$ are the longitudinal and transverse effective masses for the l band ($l = \mathbf{u}, \mathbf{v}$) at point X;

and the bulk parameter γ, in particular, determines the magnitude of the linear photogalvanic effect [20]. Before giving the form of the matrix of the potential energy operator, we go over from the function $P(z)$ to the function $\Theta(z) - \Theta(z - L)$ merely for reasons of convenience [19]:

$$P(z) \approx \Theta(z) - \Theta(z - L) + \rho_0(\delta(z) + \delta(z - L)), \quad (18)$$

where

$$\rho_0 = \int_{-d}^{d} P(z)dz - d.$$

Now the matrix of the effective potential energy operator $\mathbf{V}$ may be expressed as the sum of three matrices: $\mathbf{V}_1 + \mathbf{V}_2 + \mathbf{V}_3$. The diagonal matrix $\mathbf{V}_1$ corresponds to the standard (bulk) effective-mass approximation [18]:

$$(V_1)_{ss'} = \{\epsilon_{ss'} + \delta U_{ss'}[\Theta(z) - \Theta(z - L)]\}\delta_{ss'}.$$

As a result of the abruptness of the heterointerface potential, the matrix $\mathbf{V}_2$ contains both intravalley contributions and contributions which mix the X_1 and X_3 states (all the parameters $d_{mm'}$ are real):

$$\mathbf{V}_2 = \begin{pmatrix} \tilde{d}_{\mathbf{ww}}(\delta(z) + \delta(z - L)) & 0 & 0 \\ 0 & \tilde{d}_{\mathbf{uu}}(\delta(z) + \delta(z - L)) & d_{\mathbf{uv}}(\delta(z) - \delta(z - L)) \\ 0 & d_{\mathbf{uv}}(\delta(z) - \delta(z - L)) & \tilde{d}_{\mathbf{vv}}(\delta(z) + \delta(z - L)) \end{pmatrix}.$$

Here, the parameters $\tilde{d}_{ss}$ are related to d_{ss} as follows:

$$\tilde{d}_{ss} = d_{ss} + \delta U_{ss}\rho_0.$$

Finally, the matrix $\mathbf{V}_3$ contains contributions which mix the Γ and X_z states (also as a result of the abruptness of the heterointerface potential). We give the nonzero elements of $\mathbf{V}_3$:

$$(V_3)_{\mathbf{wu}} = (V_3)_{\mathbf{uw}} = d_{\mathbf{wu}}(\delta(z) + (-1)^{\mathcal{N}}\delta(z - L)), \quad (19)$$

$$(V_3)_{\mathbf{wv}} = (V_3)_{\mathbf{vw}} = d_{\mathbf{wv}}(\delta(z) - (-1)^{\mathcal{N}}\delta(z - L)). \quad (20)$$

In our approximation, the expression linking the envelope functions and the complete wave function has the usual form (we again drop the valley index):

$$\Psi(\mathbf{r}) = \sum_{m = \mathbf{w}, \mathbf{u}, \mathbf{v}} \left[\tilde{F}_m(\mathbf{r})\phi_m + \sum_n{}' \frac{\hbar \mathbf{p}_{nm}(\nabla \tilde{F}_m(\mathbf{r}))}{im_0(\epsilon_m - \epsilon_n)}\phi_n\right].$$

For an arbitrary number of heterojunctions $z = z_j$ numbered by the index j, the nonzero elements of $\mathbf{V}_3$ may be written in the following form:

$$(V_3)_{\mathbf{wu}} = (V_3)_{\mathbf{uw}} = \sum_j e^{iqz_j} d^{(j)}_{\mathbf{wu}}\delta(z - z_j),$$

$$(V_3)_{\mathbf{wv}} = (V_3)_{\mathbf{vw}} = \sum_j e^{iqz_j} d^{(j)}_{\mathbf{wv}}\delta(z - z_j),$$

where q is the distance in $\mathbf{k}$ space between the centers of the valleys under consideration (in our case, $q = (0, 0, 2\pi/a)$). Since the parameters $d^{(j)}_{\mathbf{wu}}$ and $d^{(j)}_{\mathbf{wv}}$ depend not only on the heterojunction material but also on the microscopic structure of the interface, these must generally be determined for each interface separately.

3. DISCUSSION OF RESULTS AND CONCLUSIONS

Using the **kp** formalism, we have constructed a generalization of the method of envelope functions suitable to describe the interaction of the Γ and X_z states in the (001) III–V nanostructures formed from similar lattice-matched semiconductors. In the derived system of equations (17), mixing of the states of different valleys is determined by the heterosurface effective potentials (19) and (20) similar to those introduced phenomenologically in [6]. The system (17) contains information on the number $\mathcal{N}$ of monoatomic layers of the structure. An oscillatory dependence of the effective intervalley mixing strength on $\mathcal{N}$ was obtained, in particular, in [7] and was also introduced from symmetry concepts (in terms of the boundary conditions for the envelope functions) in [17]. However, in addition to this, there are some difference between the results of the present study and the results of other authors. The most important of these is the appearance of direct interaction between Γ_1 and X_1 states whose strength is determined by the parameter $d_{\mathbf{wu}}$. This was predicted recently in [11, 15, 16]. The strength of the Γ_1–X_1 interaction, i.e., the value of the constant $d_{\mathbf{wu}}$, depends strongly on the structure of the heterointerface on atomic scales. In simplified models, such as the simplest variants of the tight-binding method, this interaction may be absent. It is clear from (15) that in the hypothetical case of mathematically abrupt heterojunctions when $P(z) = \Theta(z) - \Theta(z - L)$, we in fact find $d_{\mathbf{wu}} = 0$. The conclusion reached in [10] that $d_{\mathbf{wu}} \ll d_{\mathbf{wv}}$ was specifically a consequence of the selected heterojunction model for which

the simplest case is a mathematically abrupt jump (see also [11]). In general, there is no basis for assuming that d_{wu} differs substantially from d_{wv} and since the X_1 band is lower in energy than the X_3 band, the existence of Γ_1–X_1 interaction may be very significant for interpreting experiments.

In [19], we considered a more complex model of the potential of an ideal heterostructure and, specifically, took into account a periodic coordinate dependence of the heterostructure form factor in the (001) plane. We showed that this complication merely leads to renormalization of some parameters obtained using (2). However, whereas it was found using the simple form factor model that the mixing strength of heavy and light holes at the center of the 2D Brillouin zone was higher for abrupt heterojunctions than for structures with a continuously varying composition, this conclusion was not obtained using the more complex form-factor model. We can merely conclude that the strength of the lh–hh mixing depends on the structure of the transition region of the heterointerfaces. Having made a similar analysis for our case, we can easily show that the Γ–X mixing strength in fact depends strongly on how abruptly (on scales of the order a) the chemical composition at the heterojunction varies. This conclusion which follows directly from expressions (15) and (16) for the parameters d_{wu} and d_{wv} is quite natural since junctions with such a large (of the order of the dimensions of the Brillouin zone) change in the quasimomentum k_z can only be achieved by electron scattering at an atomically abrupt heterointerface (in formal terms the Fourier components of the heterointerface potential with the wave vector $k_z = 2\pi/a$ are responsible for the Γ–X junctions).

ACKNOWLEDGMENTS

This work was supported financially by the Russian Foundation for Basic Research (project no. 99-02-17592), INTAS (grant no. 97-11475), and under the programs “Physics of Solid-State Nanostructures” (project no. 99-1124) and “Surface Atomic Structures” (project no. 3.1.99).

REFERENCES

1. R. Teissier, J. J. Finley, M. S. Skolnick, *et al.*, Phys. Rev. B **54**, R8329 (1996).
2. Yu. N. Khanin, E. E. Vdovin, Yu. V. Dubrovskiĭ, *et al.*, Pis’ma Zh. Éksp. Teor. Fiz. **67**, 814 (1998) [JETP Lett. **67**, 863 (1998)].
3. M. Nakayama, K. Imazawa, K. Suyama, *et al.*, Phys. Rev. B **49**, 13564 (1994).
4. N. Ohtani, M. Hosoda, H. Mimura, *et al.*, Jpn. J. Appl. Phys. **36**, 1884 (1997).
5. O. E. Raichev, Phys. Rev. B **49**, 5448 (1994).
6. H. C. Liu, Appl. Phys. Lett. **51**, 1019 (1987).
7. Yan-Ten Lu and L. J. Sham, Phys. Rev. B **40**, 5567 (1989).
8. T. Ando and H. Akera, Phys. Rev. B **40**, 11619 (1989).
9. A. Franceschetti and A. Zunger, Phys. Rev. B **52**, 14664 (1995).
10. B. A. Foreman, Phys. Rev. Lett. **81**, 425 (1998).
11. P. C. Klipstein, in *Proceedings of the 24th International Conference on the Physics of Semiconductors, Jerusalem, 1998*, Ed. by M. Heiblum and E. Cohen (World Sci., Singapore, 1999).
12. J. N. Schulman, J. Vac. Sci. Technol. B **1**, 644 (1983).
13. F. Bassani and G. Pastori Parravicini, *Electronic States and Optical Transitions in Solids* (Pergamon, New York, 1975; Nauka, Moscow, 1982).
14. V. A. Volkov and É. E. Takhtamirov, Usp. Fiz. Nauk **167**, 1123 (1997) [Phys. Usp. **40**, 1071 (1997)].
15. V. A. Volkov and É. E. Takhtamirov, in *Proceedings of the III All-Russia Conference on the Physics of Semiconductors, “Poluprovodniki’97”, 1997* (Fiz. Inst. Akad. Nauk, Moscow, 1997), p. 127.
16. É. E. Takhtamirov, Candidate’s Dissertation (Moscow, 1998).
17. Y. Fu, M. Willander, E. L. Ivchenko, and A. A. Kiselev, Phys. Rev. B **47**, 13498 (1993).
18. J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).
19. É. E. Takhtamirov and V. A. Volkov, Zh. Éksp. Teor. Fiz. **116**, 1843 (1999) [JETP **89**, 1000 (1999)].
20. E. L. Ivchenko and G. E. Pikus, Fiz. Tekh. Poluprovodn. (Leningrad) **13**, 992 (1979) [Sov. Phys. Semicond. **13**, 579 (1979)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 1071–1080.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1230–1241.

SOLIDS
Electronic Properties

Quantization of the Motion and Cherenkov Structure of a Josephson Vortex

V. P. Silin* and A. V. Studenov

Lebedev Physical Institute, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow 117924, Russia
**e-mail: silin@sci.lebedev.ru*

Received January 20, 2000

Abstract—A generalization of the Sakai–Tateno–Pedersen model is proposed. This model is used to establish the existence of a discrete eigenvalue spectrum of the free-motion velocities of a Josephson vortex, or 2π kink, produced by the Cherenkov structure of extraordinary vortex-trapped Swihart waves having the spectrum (3.9). The dependence of the current across the Josephson junction on the vortex velocity was obtained, and, at comparatively high velocities, this was characterized by equidistant dips attributable to the Cherenkov resonant interaction of the vortex with Swihart waves. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A theory of multikink structures of Josephson vortices in a nondissipative long Josephson junction was constructed in [1–3], and an effect involving Cherenkov trapping of Swihart waves by a moving Josephson vortex was established theoretically. The coherent structure formed as a result of this trapping corresponds to the gluing of Josephson vortices by trapped waves to form a multikink vortex. The coherence of the trapped waves leads to a discrete difference in the family of $2\pi n$ kinks. This discreteness is responsible, for example, for a discrete number of trapped wavelengths. It is of prime importance that, this new property is observed in the discrete ("quantized") velocity spectrum of the free motion of Josephson vortices, or multikinks. The theory constructed in [1–3] is based on the Aubry–Volkov model [4–7] (see also [8]). Although this model could describe freely moving multivortex structures, the free motion of an isolated vortex, or 2π kink, carrying a single magnetic-flux quantum was forbidden in this model. Apart from this result, the absence of any other exact statements related to the free motion of 2π kinks under conditions of possible Cherenkov interaction of Josephson vortices with Swihart waves naturally raised the question as to whether the exclusion of the free motion of 2π kinks in the Aubry–Volkov model is merely a property of this model or whether it is more general. In this study, we answer this question on the basis of an analytic description of Josephson vortices using a generalized Sakai–Tateno–Pedersen model [9, 10]. Our generalization put forward in Section 2 involves making additional allowance for the fourth spatial derivatives in this model. This generalization is required at Josephson vortex velocities close to the Swihart velocity. At the same time, this generalization can describe the Cherenkov interaction of Josephson vortices with Swihart waves and is put forward in Section 3. Note that, in [11], allowance for the fourth derivatives in a generalization of the ordinary sine-Gordon equation was used to obtain a numerical description of a 4π kink. Using our nondissipative generalization of the Sakai–Tateno–Pedersen model, we subsequently show that Cherenkov interaction of a Josephson vortex with waves does not generally forbid the free motion of a 2π kink. In other words, the exclusion for the motion of a 2π kink as a result of the Cherenkov effect in the Aubry–Volkov model of Josephson vortices is a specific result of this model. By allowing the free motion of an isolated Josephson vortex in the Sakai–Tateno–Pedersen model, we are able to establish, for the first time, the internal coherent structure of a 2π kink produced by Swihart waves Cherenkov-trapped by a moving Josephson vortex. As in the case of multivortices in the Aubry–Volkov model, our generalization of the Sakai–Tateno–Pedersen model allows us to predict theoretically for the first time that the velocities of free motion of a 2π kink (the simplest Josephson vortex) are discrete, and we establish a law determining the eigenvalue spectrum of these velocities. We stress that, here, we have a qualitative difference between this family of Josephson vortices and the Josephson vortex having a continuous spectrum of free-motion velocities usually considered in the sine-Gordon model [12]. The results of this theory are presented in Section 4. In Section 5, we use a generalized Sakai–Tateno–Pedersen model to take into account current from an external source. This model can be used to analyze the forced uniform motion of Josephson vortices with an accompanying Cherenkov field of trapped waves and a trail of Cherenkov radiation following the vortex. In particular, we can obtain a unique dependence of the current on the vortex velocity which can be observed in the current–voltage characteristic of Josephson junctions. This uniqueness is observed as a oscillating dependence of the current–voltage characteristic on the velocity of the Josephson vortex which

1063-7761/00/9006-1071$20.00

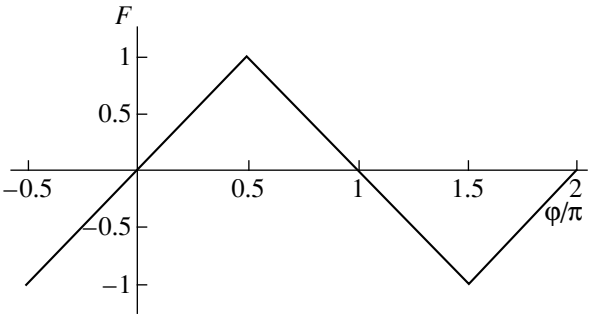


Fig. 1. The function $F[\varphi]$ corresponding to the nonlinearity in the Sakai–Tateno–Pedersen model in accordance with formula (2.2).

corresponds to the appearance of a spectrum of characteristic velocities of the freely moving Josephson vortex and, hence, to a dependence of the current on the voltage. This last result differs qualitatively from that obtained in [13] in which the influence of Cherenkov radiation on the current–voltage characteristic of a Josephson junction was described. This qualitative difference is quite natural since the study [13] is based on the assumption that the velocities of a 2π kink have a continuous eigenvalue spectrum, whereas the discreteness of the velocity spectrum of a 2π kink is established for the first time in the present study. Finally, in Section 6, we allow for ordinary dissipation, discuss the results, and put forward the concept of Cherenkov gluing of a Josephson vortex from elementary structures comprising 1π kinks which correspond to half a magnetic-flux quantum.

2. ORIGINAL MODEL

Instead of the usual model of sine nonlinearity of the Josephson current, the authors of [9, 10] proposed a different model that can give a comparatively simple exact analytic description under conditions when generalizations of the ordinary sine-Gordon equation do not allow an exact description. In the Sakai–Tateno–Pedersen model [9, 10] the following equation is proposed for the phase difference φ of superconducting pairs on different sides of the Josephson junction:

$$\frac{1}{\omega_j^2}\frac{\partial^2\varphi}{\partial t^2} - \lambda_j^2\frac{\partial^2\varphi}{\partial z^2} + F[\varphi] = -\frac{j}{j_c} - \frac{\beta}{\omega_j^2}\frac{\partial\varphi}{\partial t} + \frac{\eta\lambda_j^2}{\omega_j^2}\frac{\partial^3\varphi}{\partial t\partial z^2}. \quad (2.1)$$

Here, ω_j and λ_j are the Josephson frequency and length, j is the current density of the external source, j_c is the critical Josephson current density, β characterizes the dissipation of the junction material separating the superconductors, and η characterizes the so-called surface dissipation determined by the normal superconductor electrons. All the terms in equation (2.1), apart from the last term on the left-hand side, are the same as the usual ones for local Josephson electrodynamics, which is based on generalizations of the sine-Gordon equation. A unique characteristic of the Sakai–Tateno–Pedersen model is that, instead of the ordinary nonlinearity $F[\varphi] = \sin\varphi$, it uses a different nonlinearity that, to some extent, simulates sine nonlinearity. The nonlinearity of this model in the region $-\pi/2 < \varphi < 2\pi$ sufficient for studying the 2π kinks of interest to us has the form

$$F[\varphi] = \begin{cases} (2/\pi)\varphi, & -\pi/2 < \varphi < \pi/2 \quad (2.2a) \\ (2/\pi)(\pi - \varphi), & \pi/2 < \varphi < 3\pi/2 \quad (2.2b) \\ (2/\pi)(\varphi - 2\pi), & 3\pi/2 < \varphi < 2\pi. \quad (2.2c) \end{cases}$$

This dependence, illustrated in Fig. 1, is an approximation, which is qualitatively similar to a sine one and has already yielded various important dependences characterizing Josephson vortices [9, 10].

For comparison with the further analysis, we also give the solution of equation (2.1) for the free motion of a Josephson vortex when

$$\varphi(z, t) = \psi(s = z - Vt) \quad (2.3)$$

and when the right-hand side of equation (2.1) is neglected. In this case, we have

$$\psi(s) = 2\pi - \frac{\pi}{2}\exp\left[-sk_j + \frac{\pi}{4}\right], \quad s > \frac{\pi}{4k_j}, \quad (2.4a)$$

$$\psi(s) = \pi + \frac{\pi}{\sqrt{2}}\sin k_j s, \quad -\frac{\pi}{4k_j} < s < \frac{\pi}{4k_j}, \quad (2.4b)$$

$$\psi(s) = \frac{\pi}{2}\exp\left[k_j s + \frac{\pi}{4}\right], \quad s < \frac{\pi}{4k_j}. \quad (2.4c)$$

Here, we have

$$k_j = \sqrt{\frac{2}{\pi}}\frac{1}{\lambda_j(1 - V^2/V_s^2)^{1/2}}, \quad (2.5)$$

where $V_s = \omega_j l_j$ is the Swihart velocity. Qualitatively, the dependences (2.4) are similar to those obtained from the sine-Gordon equation for the free motion of a Josephson vortex [12]:

$$\psi(s) = 4\arctan\left(\exp\left[-\frac{s}{\lambda_j(1 - V^2/V_s^2)^{1/2}}\right]\right). \quad (2.6)$$

In the sine-Gordon model and in the Sakai–Tateno–Pedersen model, the characteristic scale of the Josephson-vortex dimension is reduced, becoming much smaller than the Josephson length, as the vortex velocity approaches the Swihart velocity. This reduction in the spatial scale indicates that, in order to describe Josephson vortices moving at velocities close to the Swihart velocity, appropriate a more equation than the ordinary sine-Gordon equation or equation (2.1) of the Sakai–Tateno–

Pedersen model should be used. In order to formulate such an equation, in the general derivation [14] of the corresponding nonlocal equation for the phase difference φ of superconducting pairs on different sides of the Josephson junction, we allow for a small correction term which corresponds to the influence of a small London length and leads to the appearance of the fourth spatial derivative. As a result, we obtain (see [11] in the sine-nonlinearity model)

$$\frac{1}{\omega_j^2}\frac{\partial^2\varphi}{\partial t^2} - \lambda_j^2\frac{\partial^2\varphi}{\partial z^2} - \frac{1}{2}\lambda_j^2\lambda^2\frac{\partial^4\varphi}{\partial z^4} + F[\varphi] = -\frac{j}{j_c} - \frac{\beta}{\omega_j^2}\frac{\partial\varphi}{\partial t} + \frac{\eta\lambda_j^2}{\omega_j^2}\frac{\partial^3\varphi}{\partial t\partial z^2}. \quad (2.7)$$

The third term on the left-hand side of this equation distinguishes the differential operator obtained here from the ordinary differential operator of the sine-Gordon equation. Equation (2.7) corresponds to the generalized Sakai–Tateno–Pedersen model used by us in the following analysis.

3. CHERENKOV RESONANCE CONDITIONS

We shall first discuss the wave perturbations described by equations (2.2) and (2.7). In this section, we shall completely neglect the terms on the right-hand side of equation (2.7). Then in regions (2.2a) and (2.2c) for small $\varphi = \delta\varphi$ and a small difference of φ from 2π, respectively, i.e., $(\varphi - 2\pi) = \delta\varphi$, we have the same equation:

$$\frac{1}{2}\lambda^2\lambda_j^2\frac{\partial^4\delta\varphi}{\partial z^4} + \lambda_j^2\frac{\partial^2\delta\varphi}{\partial z^2} - \frac{1}{\omega_j^2}\frac{\partial^2\delta\varphi}{\partial t^2} - \frac{2}{\pi}\delta\varphi = 0. \quad (3.1)$$

For the wave dependence

$$\delta\varphi \approx \exp(-i\omega t + ikz), \quad (3.2)$$

(3.1) yields the dispersion equation

$$\omega^2 = \frac{2}{\pi}\omega_j^2 + V_s^2k^2 - \frac{1}{2}V_s^2\lambda^2k^4, \quad (3.3)$$

corresponding to ordinary Swihart waves. If we now write the Cherenkov condition for resonant interaction of a source moving at velocity V with the wave perturbation,

$$k^2V^2 = \omega^2, \quad (3.4)$$

in accordance with formula (3.3) we obtain the following expression for the resonant wave vector:

$$k_t^2 = \frac{1}{\lambda^2}\left(1 - \frac{V^2}{V_s^2}\right)\left\{1 + \sqrt{1 + \frac{4\lambda^2}{\pi\lambda_j^2(1 - V^2/V_s^2)^2}}\right\} \equiv \frac{2}{\varepsilon^2}k_j^2(1 + \sqrt{1 + \varepsilon^2}), \quad (3.5)$$

where the following notation is used:

$$\varepsilon^2 = \frac{4\lambda^2}{\pi\lambda_j^2(1 - V^2/V_s^2)^2}. \quad (3.6)$$

In addition to the purely real solution of equation (3.4) described by formula (3.5), which corresponds to a Cherenkov-excited wave perturbation, equation (3.4) for the spectrum (3.3) also has a purely imaginary solution $k = ik_e$, where

$$k_e^2 = \frac{1}{\lambda^2}\left(1 - \frac{V^2}{V_s^2}\right)\left\{-1 + \sqrt{1 + \frac{4\lambda^2}{\pi\lambda_j^2(1 - V^2/V_s^2)}}\right\} \equiv \frac{2k_j^2}{\varepsilon^2}(-1 + \sqrt{1 + \varepsilon^2}). \quad (3.7)$$

A slightly different picture emerges for region (2.2b), where, for a small phase difference from π, i.e., for $\delta\varphi = \varphi - \pi$, we have

$$\frac{1}{2}\lambda^2\lambda_j^2\frac{\partial^4\delta\varphi}{\partial z^4} + \lambda_j^2\frac{\partial^2\delta\varphi}{\partial z^2} - \frac{1}{\omega_j^2}\frac{\partial^2\delta\varphi}{\partial t^2} + \frac{2}{\pi}\delta\varphi = 0. \quad (3.8)$$

For the wave perturbations (3.2), we then have the following dispersion equation:

$$\omega^2 = -\frac{2}{\pi}\omega_j^2 + V_s^2k^2 - \frac{1}{2}V_s^2\lambda^2k^4. \quad (3.9)$$

This equation differs from (3.2) in respect of the sign of the first term on the right-hand side. For small values of the wave vector, the frequency corresponding to the dispersion equation (3.9) is purely imaginary, which is consistent with the usual property of the sine-Gordon equation describing instability of the state near $\varphi = \pi$. Perturbations having the spectrum (3.9) may also be Cherenkov-excited in accordance with the resonance condition (3.4). For the resonant wave vectors, in accordance with (3.4) and (3.9) we then have

$$k_0 = \left\{\frac{1}{\lambda^2}\left(1 - \frac{V^2}{V_s^2}\right)\left[1 + \sqrt{1 - \frac{4\lambda^2}{\pi\lambda_j^2(1 - V^2/V_s^2)^2}}\right]\right\}^{1/2} = \left\{\frac{2k_j^2}{\varepsilon^2}(1 + \sqrt{1 - \varepsilon^2})\right\}^{1/2}, \quad (3.10)$$

$$k_1 = \left\{\frac{1}{\lambda^2}\left(1 - \frac{V^2}{V_s^2}\right)\left[1 - \sqrt{1 - \frac{4\lambda^2}{\pi\lambda_j^2(1 - V^2/V_s^2)^2}}\right]\right\}^{1/2} = \left\{\frac{2k_j^2}{\varepsilon^2}(1 - \sqrt{1 - \varepsilon^2})\right\}^{1/2}. \quad (3.11)$$

Usually the Cherenkov excitation of waves having the spectrum (3.9) is of no interest. However, it might seem

unexpected, but we shall see that waves having the spectrum (3.9) will play a key role in our analysis.

In accordance with relations (3.5), (3.8), (3.10), and (3.11), the influence of the term containing the fourth spatial derivative in equation (2.7) is significant, provided that

$$\left|1 - V^2/V_s^2\right| \ll 1. \tag{3.12}$$

Moreover, when this inequality is not satisfied, it is simply not justified to confine ourselves to the term containing the fourth derivative and neglect higher terms because, when inequality (3.12) is not satisfied, perturbations having a spatial scale comparable with the London length λ are Cherenkov-excited. However, because of the smallness of the ratio

$$\lambda/\lambda_j \ll 1, \tag{3.13}$$

in the Josephson junctions usually studied we can specify the velocity range

$$\frac{\lambda}{\lambda_j} \ll \left|1 - \frac{V^2}{V_s^2}\right| \ll 1 \tag{3.14}$$

where $\varepsilon^2 \ll 1$ and where formulas (3.5), (3.7), (3.10), and (3.11) take the following simple form:

$$k_t^2 \simeq k_0^2 \simeq 4k_j^2/\varepsilon^2, \tag{3.15}$$

$$k_e^2 \simeq k_1^2 \simeq k_j^2. \tag{3.16}$$

The expressions (3.16) describe a comparatively smooth spatial variation on the scale k_j^{-1}, which corresponds to the spatial scale of variation of the 2π kinks in local Josephson electrodynamics [12]. In contrast, formula (3.15) describes small-scale perturbations of Cherenkov-excited Swihart waves of wavelength

$$\lambda_t \simeq \frac{\varepsilon}{2k_j} = \frac{\lambda}{\sqrt{2}(1 - V^2/V_s^2)^{1/2}}. \tag{3.17}$$

It should be stressed that, if condition (3.14) is satisfied, in accordance with (3.16) we have $k_e \simeq k_1$; however, if, k_1 corresponds to an oscillating wave perturbation, k_e corresponds to an exponentially varying spatial variation $\delta\varphi$. The relations given in this section and describing the laws governing Cherenkov excitation in Josephson electrodynamics with the fourth spatial derivative are used as the basis for our following analysis.

4. CHERENKOV TRAPPING OF WAVES BY A 2π KINK AND EIGENVALUE SPECTRUM OF VELOCITIES OF A MOVING JOSEPHSON VORTEX

We now apply equations (2.7) and (2.2) to the motion of a vortex (2π kink) at constant velocity V when $\varphi(z, t) = \psi(z - Vt)$. As in the preceding section, we shall completely neglect all terms on the right-hand side of equation (2.7). Then equation (2.7) gives

$$\frac{d^4\psi}{ds^4} + \frac{2}{\lambda^2}\left(1 - \frac{V^2}{V_s^2}\right)\frac{d^2\psi}{ds^2} - \frac{2}{\lambda^2\lambda_j^2}F[\psi] = 0, \tag{4.1}$$

$$s = z - Vt.$$

The fact that we are interested in a solution of equation (4.1) in the form of a 2π kink implies that the real axis of the variable s is divided into three regions. The first region is $s < -s_0$, where $0 < \psi < \pi/2$ and where the solution of equation (4.1) has the form

$$\psi(s) = \frac{\pi}{2}\exp\{k_e(s + s_0)\}. \tag{4.2}$$

The second region is $-s_0 < s < s_0$, where $\pi/2 < \psi < 3\pi/2$ and where the solution of equation (4.1) has the form

$$\psi(s) = \pi + B\sin k_0 s + D\sin k_1 s. \tag{4.3}$$

Finally, the third region is $s_0 < s$, where $3\pi/2 < \psi < 2\pi$ and where the solution of equation (4.1) has the form

$$\psi(s) = 2\pi - \frac{\pi}{2}\exp\{-k_0(s - s_0)\}. \tag{4.4}$$

Expressions (4.2) and (4.3) and expressions (4.3) and (4.4) should be continuous at contiguous points of their regions of application, and their first, second, and third derivatives should also be continuous. The corresponding conditions of continuity first yield expressions for the coefficients of formula (4.3):

$$B = -\frac{\pi}{2\sin(k_0 s_0)}\left(\frac{k_1^2 - k_e^2}{k_0^2 - k_1^2}\right), \tag{4.5}$$

$$D = \frac{\pi}{2\sin(k_1 s_0)}\left(\frac{k_0^2 - k_e^2}{k_0^2 - k_1^2}\right), \tag{4.6}$$

and second yield two key equations for the following analysis:

$$M_1 = 1 - \frac{k_e}{k_1}\cot(k_1 s_0) = 0, \tag{4.7}$$

$$M_0 = 1 - \frac{k_e}{k_0}\cot(k_0 s_0) = 0. \tag{4.8}$$

A solution of equation (4.1) describing a moving 2π kink is thus completely determined. We stress that the system of equations (4.7) and (4.8) first determines the width $2s_0$ of the region in which the phase changes from $\pi/2$ to $3\pi/2$ and second determines the velocity of a 2π kink with allowance for formulas (3.7), (3.10), and (3.11). We also stress that, according to these formulas, the vortex velocity is lower than the Swihart velocity,

$$V^2 < V_s^2(1 - \sqrt{4\lambda^2/\pi\lambda_j^2}), \tag{4.9}$$

and, unlike the local theory which neglects the fourth derivative in (4.1), cannot be extremely close to the Swihart velocity. Of course, the width of this forbidden region is extremely small since $\lambda \ll \lambda_j$. To prevent the analysis from becoming too cumbersome, we shall subsequently confine ourselves to Josephson vortex velocities not too close to the Swihart velocity when condition (3.14) is satisfied (for the general case, see Appendix 1). Then, using the small parameter

$$\varepsilon \ll 1, \tag{4.10}$$

we can write equations (4.7) and (4.8) in the following approximate form:

$$\tan\left(\frac{2}{3}k_j s_0\right) = \frac{\varepsilon}{2}, \tag{4.11}$$

$$\tan(k_j s_0) = 1. \tag{4.12}$$

Thus, for the coefficients (4.5) and (4.6) we have

$$B = -\frac{\pi}{8}\varepsilon^3, \tag{4.13}$$

$$D = \frac{\pi}{\sqrt{2}}\left(1 + \frac{\varepsilon^4}{8}\right). \tag{4.14}$$

In formulas (4.11) and (4.12), we use expression (2.5).

If condition (4.10) is satisfied when

$$k_e \simeq k_1 \simeq k_j, \tag{4.15}$$

expression (2.5) describes an exponential change in the phase difference in the external regions of the vortex:

$$\begin{aligned}\psi(s) &= \frac{\pi}{2}\exp\{k_j(s + s_0)\}, \quad s < -s_0,\\ \psi(s) &= 2\pi - \frac{\pi}{2}\exp\{-k_j(s - s_0)\}, \quad s > s_0.\end{aligned} \tag{4.16}$$

On account of (4.19) (see below) these approximate dependences are the same as those describing the spatial variation of a 2π kink in local electrodynamics (2.4a) and (2.4c). In the internal region of the 2π kink, in the approximation (4.11) we can write:

$$\psi(s) = \pi + \frac{\pi}{\sqrt{2}}\left(1 + \frac{\varepsilon^4}{8}\right)\sin k_j s - \frac{\pi}{8}\varepsilon^3 \sin\left(\frac{2}{\varepsilon}k_j s\right), \quad -s_0 < s < s_0. \tag{4.17}$$

One of the terms in (4.17) varies on the same scale as expression (4.16). The last term varies on a substantially smaller scale and corresponds to the wave vector

$$k_0 \simeq \frac{2}{3}k_j = \frac{\sqrt{2}}{\lambda}\left(1 - \frac{V^2}{V_s^2}\right)^{1/2} \tag{4.18}$$

of the Swihart waves that are Cherenkov-trapped by the vortex and move with it. We shall now determine the characteristic velocities of the free motion of a 2π kink. First we obtain from equation (4.12)

$$k_j s_0 = \pi/4 \simeq 0.785\ldots. \tag{4.19}$$

On the right-hand side of this formula, we could add πm in accordance with (4.12). However, for $m \neq 0$, the condition $\pi/2 < \varphi < 3\pi/2$ is not satisfied in the internal region of the kink. Thus, only the solution (4.19) remains. It therefore follows from (4.19) that the size of the internal region of the vortex, $2s_0$, has the same scale k_j^{-1} as the external, front and back regions. This size is the same as the size of the middle region of the vortex described by local Josephson electrodynamics in the Sakai–Tateno–Pedersen model [see (2.4)]. This agreement is quite natural because (4.19) corresponds to an approximation in which the small parameter ε^2 characterizing the difference between our model and the Sakai–Tateno–Pedersen model is negligible. Now going over to equation (4.11), we note that a small right-hand side of equation (4.11) is required to ensure that the solution (4.17) satisfies the matching conditions. At the same time, in order to search for an approximate expression for the velocity spectrum, the right-hand side of (3.16) can be neglected. Then, allowing for (4.19), we obtain

$$\frac{2}{\varepsilon}k_j s_0 = \frac{\pi}{2\varepsilon} = \pi n. \tag{4.20}$$

The required discrete velocity spectrum for a Josephson vortex then follows directly:

$$V = V_s\left(1 - \frac{2\lambda}{\lambda_j\sqrt{\pi}}n\right). \tag{4.21}$$

This equidistant velocity spectrum is obtained when

$$1 \ll n \ll (\lambda_j/\lambda). \tag{4.22}$$

In accordance with (4.20) n is the number of lengths of the trapped Swihart waves having the extraordinary spectrum (3.9) in the internal region of the vortex $2s_0$. With reference to the completeness of the description of the velocity spectrum using formula (4.21), we note that, for small mode numbers n, a description of the spectrum is given in Appendix 1. For larger mode numbers n which do not satisfy the right-hand side of inequality (4.22), it should be noted that, first, in this case our differential generalization of the Sakai–Tateno–Pedersen model is unsuitable. Second, an analysis of the range of multikink velocities much lower than the Swihart velocity reported in [1–3] indicates that the velocity spectrum has a condensation point for $V \longrightarrow 0$. We can therefore assume that the denumerable set of velocities of the free motion of Josephson vortices also has a similar condensation point in our particular case of 2π kinks.

5. FORCED MOTION OF JOSEPHSON VORTICES: CHERENKOV LOSS BALANCE

In this section, we consider the uniform motion of Josephson vortices under the influence of a current j that is established as a result of the balance between the action of the current and the Cherenkov losses caused by emission of waves by the vortex. In this formulation of the problem, we retain the current j on the right-hand side of equation (2.7) and drop the dissipative terms. Below, we give an analytic description of a moving vortex and establish a relation between its velocity and the current density j. We shall assume that the "head" of the vortex is situated in the region, $s > s_j$, where $3\pi/2 < \psi < 2\pi$. In this region, we have the equation

$$\frac{d^4\psi}{ds^4} + \frac{2}{\lambda^2}\left(1 - \frac{V^2}{V_s^2}\right)\frac{d^2\psi}{ds^2} - \frac{4}{\pi\lambda^2\lambda_j^2}(\psi - 2\pi) = \frac{2j}{j_c\lambda_j^2\lambda^2}. \tag{5.1}$$

We determine the integration constants assuming that no Cherenkov radiation field occurs ahead of the vortex and from the condition $\psi(s_j) = 3\pi/2$, and we obtain a solution of equation (5.1) in the following form:

$$\psi(s) = \psi_h(s) \equiv 2\pi - \frac{\pi j}{2j_c} - \frac{\pi}{2}\left[1 - \frac{j}{j_c}\right]\exp[-k_e(s - s_j)]. \tag{5.2}$$

In the middle region of the Josephson vortex, $-s_j < s < s_j$, where $\pi/2 < \psi < 3\pi/2$, we have

$$\frac{d^4\psi}{ds^4} + \frac{2}{\lambda^2}\left(1 - \frac{V^2}{V_s^2}\right)\frac{d^2\psi}{ds^2} - \frac{4}{\pi\lambda^2\lambda_j^2}(\psi - \pi) = \frac{2j}{j_c\lambda_j^2\lambda^2}. \tag{5.3}$$

The solution of this equation has the form

$$\psi(s) = \psi_m(s) \equiv \pi + \frac{\pi j}{2j_c} + a_m\cos k_0 s + b_m\sin k_0 s + c_m\cos k_1 s + d_m\sin k_1 s. \tag{5.4}$$

Matching the solutions (5.2) and (5.4) and also their first three derivatives, we obtain

$$a_m = -\left(1 - \frac{j}{j_c}\right)\frac{B_j M_{0j}\sin 2k_0 s_j}{2(1 - M_{0j})}, \tag{5.5}$$

$$b_m = \left(1 - \frac{j}{j_c}\right)\frac{B_j(1 - M_{0j}\sin^2 k_0 s_j)}{1 - M_{0j}}, \tag{5.6}$$

$$c_m = -\left(1 - \frac{j}{j_c}\right)\frac{D_j M_{1j}}{2(1 - M_{1j})}\sin 2k_1 s_j, \tag{5.7}$$

$$d_m = \left(1 - \frac{j}{j_c}\right)\frac{D_j(1 - M_{0j}\sin^2 k_1 s_j)}{1 - M_{1j}}, \tag{5.8}$$

Here, B_j and D_j are determined by formulas (4.5) and (4.6) with s_0 replaced by s_j, and the functions M_{0j} and M_{1j} are determined by formulas (4.7) and (4.8) with the same substitution. Obviously, in the limit $j = 0$ when the functions M_0 and M_1 given by formulas (4.7) and (4.8) go to zero, we obtain $a_m = c_m = 0$, $b_m = B$, and $d_m = D$, i.e., the solution (5.4) becomes (4.3)

Finally, in the tail region of the vortex, $s < -s_j$, where $\psi < \pi/2$, we have

$$\frac{d^4\psi}{ds^4} + \frac{2}{\lambda^2}\left(1 - \frac{V^2}{V_s^2}\right)\frac{d^2\psi}{ds^2} - \frac{4}{\lambda^2\lambda_j^2}\psi = \frac{2j}{j_c\lambda_j^2\lambda^2}, \tag{5.9}$$

whose solution has the form

$$\psi(s) = \psi_t(s) \equiv -\frac{\pi j}{2j_c} + a_t\exp[k_e(s + s_j)] + b_t\sin k_t(s + s_j) + c_t\cos k_t(s + s_j). \tag{5.10}$$

The matching conditions of the solutions (5.4) and (5.10) and also their three derivatives for $s = -s_j$ yield the following four expressions:

$$a_t = \pi\left(1 - \frac{j}{j_c}\right)\times\left\{\frac{1}{2} + A_0 M_{0j}\sin^2 k_0 s_j - A_1 M_{1j}\sin^2 k_1 s_j\right\}, \tag{5.11}$$

$$a_t = \frac{2\pi k_t^2}{k_e^2 + k_t^2}\frac{j}{j_c} + \frac{\pi}{2}\left(1 - \frac{j}{j_c}\right)\times\left\{1 + \frac{k_e}{k_0}A_0 M_{0j}\sin 2k_0 s_j - \frac{k_e}{k_1}A_1 M_{1j}\sin 2k_1 s_j\right\}, \tag{5.12}$$

$$b_t = \pi\left(1 - \frac{j}{j_c}\right)\frac{k_e}{k_t} \times C\{M_{0j}\sin^2 k_0 s_j - M_{1j}\sin^2 k_1 s_j\}, \tag{5.13}$$

$$c_t = \frac{2\pi k_t^2}{k_e^2 + k_t^2}\frac{j}{j_c} + \frac{\pi}{2}\left(1 - \frac{j}{j_c}\right)\times C\left\{\frac{k_e}{k_0}M_{0j}\sin 2k_0 s_j - \frac{k_e}{k_1}M_{1j}\sin 2k_1 s_j\right\}, \tag{5.14}$$

where the following notation is used:

$$A_0 = \frac{(k_t^2 - k_0^2)(k_1^2 + k_e^2)}{(k_0^2 - k_1^2)(k_t^2 + k_e^2)} = \frac{1}{2}\left[\frac{1}{\sqrt{1 - \varepsilon^4}} - 1\right], \tag{5.15}$$

$$A_1 = \frac{(k_t^2 - k_1^2)(k_0^2 + k_e^2)}{(k_0^2 - k_1^2)(k_t^2 + k_e^2)} = \frac{1}{2}\left[\frac{1}{\sqrt{1-\varepsilon^4}} + 1\right], \quad (5.16)$$

$$C = \frac{(k_0^2 + k_e^2)(k_1^2 + k_e^2)}{(k_0^2 - k_1^2)(k_t^2 + k_e^2)} = \frac{\varepsilon^2}{2\sqrt{1-\varepsilon^4}}. \quad (5.17)$$

Equating expressions (5.11) and (5.12) gives

$$\frac{j}{j_c} = \left(1 - \frac{j}{j_c}\right)\frac{1}{2(\sqrt{1-\varepsilon^2} + \sqrt{1-\varepsilon^4})} \times ([1 - \sqrt{1-\varepsilon^4}]\sin^2(k_0 s_j)M_0^2 - [1 + \sqrt{1-\varepsilon^4}]\sin^2(k_0 s_j)M_1^2). \quad (5.18)$$

This equation relates three quantities: the current density j; the size of the middle (internal) region of the Josephson vortex, $2s_j$; and the vortex velocity V. The second equation for these quantities is obtained from the condition $\psi_m(-s_j) = \pi/2$ and has the following form:

$$\frac{j}{j_c} = \frac{1}{4}\left(1 - \frac{j}{j_c}\right)\left(\left[\sqrt{\frac{1+\varepsilon^2}{1-\varepsilon^2}} + 1\right]\frac{k_e}{k_1}M_1 \sin 2k_1 s_j - \left[\sqrt{\frac{1+\varepsilon^2}{1-\varepsilon^2}} - 1\right]\frac{k_e}{k_0}M_0 \sin 2k_0 s_j\right). \quad (5.19)$$

These expressions are obtained under the condition $j < j_c$, which is satisfied over a wide range of manifestation of the discrete spectrum of velocities of a Josephson vortex.

Quite clearly, in the limit $j = 0$, equations (5.18) and (5.19) reduce to equations (4.7) and (4.8) characterizing the free motion of a Josephson vortex for discrete values of the velocities of a 2π kink. This set corresponds to the discrete points where the current density goes to zero on the curve of j as a function of the vortex velocity, as given by equations (4.7) and (4.8). At the same time, expressions (5.13) and (5.14) go to zero for these velocities. This implies that, at velocities corresponding to the eigenvalues of the free-motion velocity of a Josephson vortex, no Cherenkov radiation field appears in the vortex tail corresponding to the oscillating term in formula (5.10).

We shall now describe some simple consequences of this description, as in the preceding section, for the case when the Josephson vortex velocities are not too close to the Swihart velocity when condition (3.14) is satisfied. The solution of equations (5.18) and (5.19) is described in Appendix 2. According to formulas (A.2.3) and (A.2.7), the size of the middle region of the Josephson vortex can be accurately given by $s_j \simeq s_0$, and we can use the approximate relationship (4.19) corresponding to the approximation of the local model. In addition, in accordance with formulas (A.2.6) and (A.2.8) we use the following approximate relation for our discussions:

$$\frac{j}{j_c} = \frac{\varepsilon^4}{8}\left(\sin\frac{\pi}{2\varepsilon} - \frac{\varepsilon}{2}\cos\frac{\pi}{2\varepsilon}\right)^2. \quad (5.20)$$

The right-hand side of formula (5.20) is a function of velocity as given by (3.6), which characterizes the dependence of ε on the velocity of a Josephson vortex. Consequently, in the approximation used here, relationship (5.20) characterizes the dependence of the current density j on the vortex velocity. In the approximation (3.14), we have

$$D_j \simeq \frac{\pi}{\sqrt{2}}, \quad B_j \simeq -\frac{\pi\varepsilon^4}{16\sin k_0 s_j}, \quad M_{0j} \simeq 1 - \frac{\varepsilon}{2}\cot k_0 s_j, \quad M_{1j} = \frac{\varepsilon^3}{2}\left(\sin\frac{\pi}{2\varepsilon} - \frac{\varepsilon}{2}\cos\frac{\pi}{2\varepsilon}\right)\cos\frac{\pi}{2\varepsilon}. \quad (5.21)$$

Thus, formulas (5.5)–(5.8) have the following simple form:

$$a_m \simeq \frac{\pi\varepsilon^3}{8}\left(\sin\frac{\pi}{2\varepsilon} - \frac{\varepsilon}{2}\cos\frac{\pi}{2\varepsilon}\right), \quad b_m \simeq -\frac{\pi\varepsilon^3}{8}\left(\cos\frac{\pi}{2\varepsilon} + \frac{\varepsilon}{2}\cos\frac{\pi}{2\varepsilon}\right), \quad c_m \simeq -\frac{\pi\varepsilon^3}{4\sqrt{2}}\left(\sin\frac{\pi}{2\varepsilon} - \frac{\varepsilon}{2}\cos\frac{\pi}{2\varepsilon}\right)\cos\frac{\pi}{2\varepsilon}, \quad d_m \simeq \frac{\pi}{\sqrt{2}}. \quad (5.22)$$

The field of Cherenkov-trapped waves having the extraordinary spectrum (3.9) is described in the solution (5.4) by terms having small coefficients a_m and b_m. These small expressions must be taken into account to obtain the law (5.20). However, in the zeroth approximation with respect to the parameter ε, the solution (5.4) has the form

$$\psi_m(s) = \pi + \frac{\pi}{\sqrt{2}}\sin k_j s, \quad (5.23)$$

which is consistent with the description of the free motion of a Josephson vortex in the usual local approximation of the Sakai–Tateno–Pedersen model (2.4b).

Also in the approximation (4.10), formulas (5.11), (5.13), and (5.14) characterizing the tail of the Josephson vortex may be written in the following simple form:

$$a_t \simeq \frac{\pi}{2} - \frac{\pi\varepsilon^3}{4}\left[\sin\frac{\pi}{2\varepsilon} - \frac{\varepsilon}{2}\cos\frac{\pi}{2\varepsilon}\right]\cos\frac{\pi}{2\varepsilon},$$

$$b_t \simeq \frac{\pi\varepsilon^3}{4}\left[\sin\frac{\pi}{2\varepsilon} - \frac{\varepsilon}{2}\cos\frac{\pi}{2\varepsilon}\right]\sin\frac{\pi}{2\varepsilon}, \qquad (5.24)$$

$$c_t \simeq \frac{\pi\varepsilon^3}{4}\left[\sin\frac{\pi}{2\varepsilon} - \frac{\varepsilon}{2}\cos\frac{\pi}{2\varepsilon}\right]\cos\frac{\pi}{2\varepsilon}.$$

The small coefficients b_t and c_t characterize the Cherenkov radiation field of waves having the ordinary spectrum (3.3) in the vortex tail. The coefficient a_t differs from the corresponding solution for free motion by a small correction of order ε^3, which goes to zero for those discrete values of the velocity that correspond to the eigenvalues of the free-motion velocity of a Josephson vortex. At these velocities, the coefficients b_t and c_t also go to zero, which indicates that no Cherenkov radiation field exists in the vortex tail at these velocities. In Fig. 2, the dotted curve corresponds to the right-hand side of equation (5.20). The zeroes in this curve correspond to the eigenvalues of the velocity of a Josephson vortex.

6. DISCUSSION

To gain deeper insight into the problem under consideration, the dissipation should be taken into account in the law linking the current and velocity of a Josephson vortex. Since, according to formula (5.20), the current density causing the motion of 2π kinks is low compared with the critical Josephson value, the corresponding dissipation contribution to the law generalizing formula (5.20) may be considered additive. This means that the role of dissipation may be analyzed directly in the Sakai–Tateno–Pedersen model. Here, we indicate the form of the solution of equation (2.1) for a uniformly moving Josephson vortex:

$$\psi(s) = -\frac{\pi j}{2j_c} + \frac{\pi}{2}\left(1 + \frac{j}{j_c}\right)\exp[(s + s_0)k_j(1 - \alpha_-)], \quad s < -s_0, \qquad (6.1a)$$

$$\psi(s) = \pi + \frac{\pi j}{2j_c} + \frac{\pi}{\sqrt{2}}\sin(k_j s)\exp(-sk_j\alpha_+) - \frac{\pi^2\beta V k_j}{4\sqrt{2}\omega_j^2}\cos(k_j s)\exp(-sk_j\alpha_+), \quad -s_0 < s < s_0, \qquad (6.1b)$$

$$\psi(s) = 2\pi - \frac{\pi j}{2j_c} - \frac{\pi}{2}\left(1 - \frac{j}{j_c}\right)\exp[-(s - s_0)k_j(1 + \alpha_-)], \quad s > s_0. \qquad (6.1c)$$

Here, $s_0 = \pi/4k_j$, and the following notation is used

$$\alpha_\pm = \frac{\pi}{4}[\beta \pm \eta\lambda_j^2 k_j^2]\frac{k_j V}{\omega_j^2}. \qquad (6.2)$$

Formula (6.1) then yields the law

$$\frac{j}{j_c} = \frac{\pi^2}{16}\left[\beta\left(\frac{4}{\pi} + 1\right) + \eta\lambda_j^2 k_j^2\right]\frac{k_j V}{\omega_j^2}. \qquad (6.3)$$

In the simplest case of low temperatures, when the contribution of surface losses is negligible because of the freezing of normal electrons, the dependence (6.3) is given by the dashed curve in Fig. 2.

The combined contribution of ordinary dissipation and Cherenkov radiation losses yields the following equation:

$$\frac{j}{j_c} = \frac{\pi^{3/2}}{8\sqrt{2}}\left[\frac{\beta}{\omega_j}\left(\frac{4}{\pi} + 1\right) + \frac{2\eta}{\omega_j\pi(1 - V^2/V_s^2)}\right] \times \frac{V}{V_s(1 - V^2/V_s^2)^{1/2}} + \frac{2\lambda^4}{\pi^2\lambda_j^4(1 - V^2/V_s^2)^4}\left\{\sin\left[\frac{\pi^{3/2}\lambda_j}{4\lambda}\left(1 - \frac{V^2}{V_s^2}\right)\right] - \frac{\lambda}{\sqrt{\pi}\lambda_j(1 - V^2/V_s^2)}\cos\left[\frac{\pi^{3/2}\lambda_j}{4\lambda}\left(1 - \frac{V^2}{V_s^2}\right)\right]\right\}^2. \qquad (6.4)$$

In order to reveal the manifestation of the Cherenkov effect more clearly, we write formula (6.4) in terms of the variable ν that, for integer values, characterizes resonant values of the velocity (4.21) and is given by

$$V = V_s[1 - \nu(2\lambda/\sqrt{\pi}\lambda_j)].$$

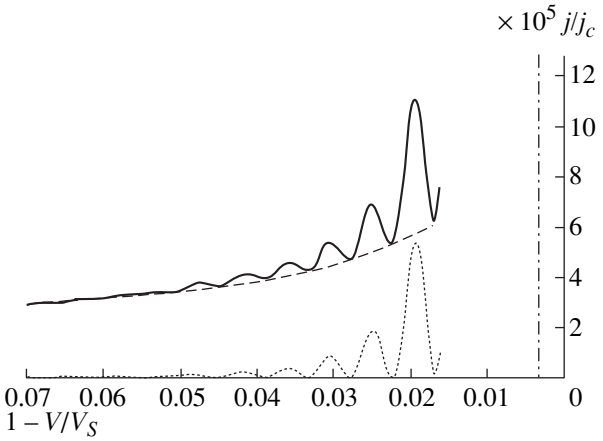


Fig. 2. Dependence of the current density on the velocity of a 2π kink near the Swihart velocity for $\beta/\omega_j = 10^{-5}$, $\lambda/\lambda_j = 5 \times 10^{-3}$, and $\eta = 0$: (solid curve) equation (6.4), (dashed curve) equation (6.3), and (dotted curve) equation (5.20); the dot-and-dash vertical line defines the velocity region for which our analysis using formula (4.9) is valid.

Then (6.4) has the following form:

$$\frac{j}{j_c} = \frac{\pi^{7/4}V}{16\sqrt{2}V_s}\left[\frac{\beta}{\omega_j}\left(\frac{4}{\pi}+1\right)+\frac{\eta\lambda_j}{2\sqrt{\pi}\lambda\omega_j\nu}\right]\sqrt{\frac{\lambda_j}{\lambda\nu}} + \frac{1}{128\nu^4}\left(\sin\pi\nu-\frac{1}{4\nu}\cos\pi\nu\right)^2.$$

The first term in this expression, which is determined by normal dissipation describes a monotonic increase in j as the Josephson vortex velocity approaches the Swihart velocity. The second Cherenkov term on the right-hand side of (6.4) represent pronounced oscillating dependence. This dependence may be seen in the experimental results [15]. Using the experimental parameters given in [15], we can suppose that the observed equidistance of the dips on the current–voltage characteristic in the range of Josephson vortex velocities close to the Swihart velocity may be associated with our law (6.4). Note that the "descending" sections of the dependence of j on V given by equation (6.4), which correspond to "descending" sections on the current–voltage characteristic are usually unstable, which is consistent with their absence on the detailed curve of the current–voltage characteristic given in [15] for sample B.

To sum up, we can state that we have established for the first time that the velocities of the free motion of the simplest Josephson vortex carrying a single magnetic-flux quantum, known as a 2π kink, are discrete (quantized). Moreover, like the results of [1–3], we established that this quantization of the characteristic velocities is produced by the number of wavelengths of those Swihart waves trapped by a Josephson vortex as a result of resonant interaction. A qualitative difference between the results of [1–3], which use the Aubry–Volkov model [4–6], is that, whereas, according to [1–3], Cherenkov gluing of 2π kinks to form multikinks takes place as a result of the trapping of ordinary Swihart waves having the spectrum (3.3), in our case, waves having the extraordinary spectrum (3.9) play the role of glue. This, is because the Aubry–Volkov model does not allow these extraordinary waves. Thus, our use of the Sakai–Tateno–Pedersen model makes it possible to identify the gluing of a 2π kink from its two halves. This suggests that the elementary constituents from which complex vortices may be constructed are 1π kinks corresponding to half a magnetic-flux quantum and described by formula (4.2) [or the corresponding formula (4.4)] and also by formula (4.3) in the range $-s_0 < s < 0$ (or $0 < s < s_0$, respectively). The dependence of the current on the Josephson vortex velocity obtained using these concepts demonstrates that the spectrum of velocities of a Cherenkov-glued 2π kink is discrete. In addition to being of physical interest, the results of the present study may well be of mathematical interest, which will help to stimulate studies of local and nonlocal generalizations of the sine-Gordon equation.

ACKNOWLEDGMENTS

This work was supported financially by the interdisciplinary project "Studies of Characteristics of High-Temperature and Other Superconductors with High Critical Parameters", section "Study of the Nature and Fundamental Properties of Superconductors" of the Scientific Council on Superconductivity, the Russian Foundation for Basic Research (project no. 00-02-16076), and under the State Support for Leading Scientific Schools.

APPENDIX 1

In this Appendix, we give results of a general description suitable for a Josephson vortex under conditions where inequality (4.10) [or (4.22)] is not satisfied, i.e., for a small number of modes n when the velocity of the Josephson vortex is very close to the Swihart velocity. In this general case, the solution of equation (4.1) may be expressed in the form

$$\psi(s) = \frac{\pi}{2}\exp\left[\sqrt{(\sqrt{2-\delta^2}-1)(1-V^2/V_s^2)}(s+s_0)/\lambda\right],$$

$$s < -s_0,$$

$$\psi(s) = \pi + \{\pi[2+\delta-\sqrt{2-\delta^2}]$$

$$\times \sin[\sqrt{(1-\delta)(1-V^2/V_s^2)}(s/\lambda)]\}$$

$$\times\{4\delta\sin[\sqrt{(1-\delta)(1-V^2/V_s^2)}(s_0/\lambda)]\}^{-1} \quad \text{[A.1.1]}$$

$$-\{\pi[2-\delta-\sqrt{2-\delta^2}]\sin[\sqrt{(1+\delta)(1-V^2/V_s^2)}(s/\lambda)]\}$$

$$\times\{4\delta\sin[\sqrt{(1+\delta)(1-V^2/V_s^2)}(s/\lambda)]\}^{-1},$$

$$-s_0 < s < s_0,$$

$$\psi(s) = 2\pi$$

$$-\frac{\pi}{2}\exp\left[-\sqrt{(\sqrt{2-\delta^2}-1)(1-V^2/V_s^2)}(s-s_0)/\lambda\right],$$

$$s > s_0,$$

where $\delta^2 = 1 - \varepsilon^2$. Thus, the velocity eigenvalues are given by

$$V^2 = V_s^2\left[1-\frac{2\lambda}{\lambda_j\sqrt{\pi(1-\delta^2)}}\right], \quad \text{(A.1.2)}$$

where the discrete values correspond to solutions of the equation

$$\sqrt{\frac{1-\delta}{1+\delta}} = \frac{\arctan P(\delta)}{\arctan Q(\delta)+\pi n}, \quad \text{(A.1.3)}$$

where

$$P(\delta) = \sqrt{\frac{\sqrt{2-\delta^2}-1}{1-\delta}}, \qquad Q(\delta) = \sqrt{\frac{\sqrt{2-\delta^2}-1}{1+\delta}}. \tag{A.1.4}$$

Here, we give the first three solutions of equation (A.1.3): $n = 1$, $\delta_1 = 0.901$, and $k_{j1}s_{01} = 0.762$; $n = 2$, $\delta_2 = 0.971$, and $k_{j2}s_{02} = 0.778$; and $n = 3$, $\delta_3 = 0.986$, and $k_{j3}s_{03} = 0.782$, where k_{jn} ($n = 1, 2, 3, \ldots$) denote values of k_1 (3.11) for various values of δ. Quite clearly, the approximation (4.19) is poorly satisfied even for the third mode.

APPENDIX 2

We shall briefly give the solutions of equations (5.18) and (5.19) for Josephson vortex velocities not too close to but also not too far from the Swihart velocity when condition (3.14) is satisfied and when we can use the small parameter ε. This allows us to write the system (5.18) and (5.19) in the form

$$\frac{j}{j_c} \simeq \left(1-\frac{j}{j_c}\right)\left(\frac{\varepsilon^4}{8}\left[\sin k_0 s_j - \frac{\varepsilon}{2}\cos k_0 s_j\right]^2 - \frac{1}{2}\sin^2 k_1 s_j\left[1+\frac{\varepsilon^2}{4}+\frac{\varepsilon^4}{8}\right]\left[1-\left(1-\frac{\varepsilon^2}{4}+\frac{\varepsilon^4}{32}\right)\cot k_1 s_j\right]^2\right), \tag{A.2.1}$$

$$\frac{j}{j_c} \simeq \left(1-\frac{j}{j_c}\right)\left(-\frac{\varepsilon^3}{8}\left[1-\frac{\varepsilon}{2}\cot k_0 s_j\right]\sin 2k_0 s_j + \frac{1}{2}\sin 2k_1 s_j\left[1+\frac{\varepsilon^2}{4}+\frac{5\varepsilon^4}{32}\right]\left[1-\left(1-\frac{\varepsilon^2}{4}+\frac{\varepsilon^4}{32}\right)\cot k_1 s_j\right]\right). \tag{A.2.2}$$

The method of solving this system of equations becomes obvious after substituting into these equations

$$\cot k_1 s_j = 1 + \varepsilon^2/4 - 2\Delta, \tag{A.2.3}$$

where $\Delta = o(\varepsilon^2)$. Equations (A.2.1) and (A.2.2) then have the simple form

$$\frac{j}{j_c} \simeq \left(1-\frac{j}{j_c}\right)\left\{\frac{\varepsilon^4}{8}\left(\sin k_0 s_j - \frac{\varepsilon}{2}\cos k_0 s_j\right)^2 - \frac{1}{2}\sin^2 k_1 s_j\left(1+\frac{\varepsilon^2}{4}+\frac{\varepsilon^4}{8}\right)\left[\frac{\varepsilon^4}{32}+2\Delta\right]^2\right\}, \tag{A.2.4}$$

$$\frac{j}{j_c} = \left(1-\frac{j}{j_c}\right)\left\{\frac{1}{2}\sin 2k_1 s_j\left[\frac{\varepsilon^4}{32}+2\Delta\right] - \frac{\varepsilon^3}{4}\left[\sin k_0 s_j - \frac{\varepsilon}{2}\cos k_0 s_j\right]\cos k_0 s_j\right\}. \tag{A.2.5}$$

Now, it is easy to see that equation (A.2.4) yields the following simple relationship:

$$\frac{j}{j_c} = \frac{\varepsilon^4}{8}\left(\sin k_0 s_j - \frac{\varepsilon}{2}\cos k_0 s_j\right)^2, \tag{A.2.6}$$

and equation (A.2.5) gives

$$\Delta = \frac{\varepsilon^3}{4}\cos k_0 s_j\left(\sin k_0 s_j - \frac{\varepsilon}{2}\cos k_0 s_j\right) - \frac{\varepsilon^4}{64}. \tag{A.2.7}$$

Formulas (A.2.3) and (A.2.7) distinguish s_j from the approximate value of s_0 used in Section 4 which is the same as the result of the original Sakai–Tateno–Pedersen model. This difference is very small. However, if this difference is neglected, it is not possible to obtain the correct law (A.2.6) linking the current density j and the Josephson vortex velocity. In this law, the zeroth approximation (4.19) with respect to ε can be used as s_j. This gives

$$k_0 s_j \simeq \frac{2}{\varepsilon}k_j s_0 \simeq \frac{\pi}{2\varepsilon}. \tag{A.2.8}$$

The approximate equation (5.20) used in the main text corresponds to relation (A.2.6) in which (A.2.8) is used.

REFERENCES

1. A. S. Malishevskiĭ, V. P. Silin, and S. A. Uryupin, Phys. Lett. A **253**, 333 (1999).
2. A. S. Malishevskiĭ, V. P. Silin, and S. A. Uryupin, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 318 (1999) [JETP Lett. **69**, 348 (1999)].
3. A. S. Malishevskiĭ, V. P. Silin, and S. A. Uryupin, Zh. Éksp. Teor. Fiz. **117**, 771 (2000) [JETP **90**, 671 (2000)].
4. S. Aubry and P. J. Le Daeron, Physica D (Amsterdam) **7**, 240 (1983).
5. S. Aubry, J. Phys. C **16**, 2497 (1983).
6. A. F. Volkov, Physica C (Amsterdam) **183**, 177 (1991).
7. A. F. Volkov, Physica C (Amsterdam) **192**, 306 (1992).
8. S. N. Artemenko and S. V. Remizov, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 811 (1997) [JETP Lett. **66**, 853 (1997)].
9. S. Sakai and H. Tateno, Jpn. J. Appl. Phys. **22**, 1374 (1983).
10. S. Sakai and N. F. Pedersen, Phys. Rev. B **34**, 3506 (1986).
11. G. L. Alfimov, V. M. Eleonsky, N. E. Kulagin, *et al.*, Chaos **3**, 405 (1993).
12. A. Barone and G. Paterno, *Physics and Applications of the Josephson Effect* (Wiley, New York, 1982; Mir, Moscow, 1984).
13. R. G. Mintz and I. B. Shapiro, Phys. Rev. B **52**, 969 (1995).
14. Yu. M. Aliev and V. P. Silin, Zh. Éksp. Teor. Fiz. **104**, 2526 (1993) [JETP **77**, 142 (1993)].
15. M. Martuccello, G. Mygind, V. P. Koshelets, *et al.*, Phys. Rev. B **57**, 5444 (1998).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 1081–1088.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1242–1250.

SOLIDS
Electronic Properties

Influence of Silver on the Galvanomagnetic Properties and Energy Spectrum of Mixed $(Bi_{1-x}Sb_x)_2Te_3$ Crystals

V. A. Kul'bachinskiĭ[a, *], A. Yu. Kaminskiĭ[a], V. G. Kytin[a], P. Lošťak[b], Č. Drašar[b], and A. de Visser[c]

Moscow State University, Moscow, 119899 Russia
**e-mail: kulb@mig.phys.msu.su*
[b]*Pardubice University, Pardubice, Czech Republic*
[c]*Van der Waals Institute, University of Amsterdam, the Netherlands*

Received January 25, 2000

Abstract—An investigation is made of the temperature dependences of the resistivity in the range 4.2–300 K, the Hall effect, and the Shubnikov–de Haas effect in magnetic fields up to 40 T in $(Bi_{1-x}Sb_x)_2Te_3Ag_y$ single crystals ($0 \le x \le 0.75$). Doping $(Bi_{1-x}Sb_x)_2Te_3$ crystals with silver showed that in Sb_2Te_3 and $(Bi_{1-x}Sb_x)_2Te_3$ crystals unlike Bi_2Te_3 silver exhibits acceptor properties. The angular and concentration dependences of the Shubnikov–de Haas effect were studied in $(Bi_{1-x}Sb_x)_2Te_3Ag_y$. It was established that the anisotropy of the ellipsoids of the upper valence band in $Bi_{0.5}Sb_{1.5}Te_3$ remains unchanged as a result of silver doping. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Semiconductors of the Sb_2Te_3 type are layered crystals having a rhombohedral structure and symmetry space group R_{3m}–D_{3d}^5 with the twofold and threefold axes of symmetry C_2 and C_3. The crystal lattice is formed by periodically ordered layers lying in the plane perpendicular to the C_3 symmetry axis. Each layer consists of five atomic planes (quintets) forming the following sequence: Te^1–Sb–Te^2–Sb–Te^1. Here Te^1 and Te^2 denote Te atoms in various positions. In each individual layer the atoms are similar and form a plane hexagonal lattice. The atoms of each successive layer are positioned above the centers of the triangles formed by atoms of the preceding layer (hexagonal close packing), i.e., the Te^1 and Sb atoms occupy octahedral positions in a tetradymite structure. The chemical bond within the quintets is covalent-ionic. Between the quintets there is a comparatively large spacing and a weak bond produced by van der Waals forces. This is responsible for the anisotropy of the electrophysical properties of single crystals [1].

Crystals of Sb_2Te_3 have a very high concentration of holes because of the presence of a large number of charged point defects. Single crystals of Sb_2Te_3 typically have the following types of defects: Sb and Te vacancies, atoms of these elements in interstitial sites, antistructural Sb_{Te} (an Sb atom in a Te position) and Te_{Sb} defects (a Te atom in an Sb position), impurity antistructural defects, impurity atoms in interstitial sites, and so on. Since antistructural defects are negatively charged, Sb_2Te_3 grown under stoichiometric conditions always possesses *p*-type conductivity and has a fairly high hole concentration. Similar defects may be found in antimony telluride solid solutions.

Antistructural defects in Sb_2Te_3 semiconductors are formed under the influence of the bond polarity. The weak polarity of the Sb–Te bonds is conducive to the formation of antistructural defects. A change in the polarity of the bonds caused by doping or a shift of the stoichiometry leads to a change in the concentration of antistructural defects (see, for example, [2, 3]). Assuming that the charge of the antistructural defects is compensated by holes, any change in the concentration of antistructural defects will correspond to a change in the hole concentration. Dopants in layered crystals such as antimony telluride have a strong influence on the concentration of point defects and therefore on the carrier concentration. Consequently doping with elements in particular groups of the periodic table may have a donor or acceptor effect not depending on the number of the group but as a result of the influence of the incorporated element on the bond polarity. As an example we can quote the Group III element indium which has a donor effect in Sb_2Te_3 [4].

Mixed crystals, i.e., crystals of the type $(Bi_{1-x}Sb_x)_2Te_3$ or $Sb_2Te_{3-y}Se_y$ are of particular interest because the highest thermoefficiencies *Z* are observed in these. Different combinations of the elements Bi, Sb, As and Te, Se, S are possible. Despite intensive studies, very little data are currently available on the influence of doping with Group I elements on the energy spectrum of mixed $(Bi_{1-x}Sb_x)_2Te_3$ crystals.

1063-7761/00/9006-1081$20.00

Table 1

Samples	y determined by AAS	c, 10^{19} atoms/cm^3
$Sb_2Te_3Ag_y$	0.0035	2.2
	0.0093	5.9
	0.0100	6.2
	0.0152	9.6
$Bi_{0.5}Sb_{1.5}Te_3Ag_y$	0.0014	0.8
	0.0030	1.9
	0.0048	3.0

The Shubnikov–de Haas effect is an effective method of studying semiconductors since the parameters of the energy spectrum can be determined directly, i.e., the effective masses, Fermi surface cross sections, carrier concentration, and so on. However, in antimony telluride and mixed crystals such measurements are difficult since quantum oscillations of the magnetoresistance are initiated in magnetic fields having an induction higher than 10 T because of the fairly low hole mobility.

In the present study we investigated the Shubnikov–de Haas effect in magnetic fields up to 40 T which allowed us to observe oscillations of the magnetoresistance in all the samples studied. We also studied the influence of silver doping on the galvanomagnetic properties and energy spectrum of mixed $(Bi_{1-x}Sb_x)_2Te_3$ single crystals.

2. SAMPLES AND METHOD OF MEASUREMENT

We investigated *p*-type single-crystal samples of silver-doped antimony telluride Sb_2Te_3 and mixed $(Bi_{1-x}Sb_x)_2Te_3$ crystals grown by the Bridgman method from polycrystalline materials. First we synthesized polycrystalline samples from 99.999% pure elements in stoichiometric ratio in a quartz ampoule. We then added silver to the stoichiometric polycrystal for doping so that the sample compositions will subsequently be given in the form $Sb_2Te_3Ag_y$ and $(Bi_{1-x}Sb_x)_2Te_3Ag_y$. After preparing a polycrystal of the required composition we grew the single crystals. The grown single crystals were released from the quartz ampoule and cleaved perpendicular to the C_3 axis which, in the as-grown ingots, was always perpendicular to the longitudinal axis of the sample. The samples for the measurements, in the form of a parallelepiped having average dimensions of $1 \times 1 \times 5$ mm^3 with the long axis directed along the C_2 axis, were cut using an electric-arc machine. Current and potential leads made of 30 μm diameter copper wire were soldered using Bi + 4%Sb alloy.

The Ag content in the samples was determined by atomic absorption spectroscopy (AAS) for the specific sample used for all the electrophysical measurements. Table 1 gives the silver content y determined by AAS for the investigated samples and the experimentally determined silver concentration c in the measured samples. Attempts to obtain single crystals having an even higher silver content resulted in its segregation.

The Hall effect was investigated at temperatures of 4.2 K, 77 K, and 300 K. The dc current was directed along the C_2 axis and the magnetic field induction vector **B** was directed along the C_3 axis. The magnetic field at 4.2 K was generated by two methods. For measurements in fields not exceeding 11 T it was generated using a superconducting solenoid immersed in liquid helium. The measurements were made for two directions of the magnetic field induction vector and the Hall voltage was taken as half the difference between these values. In addition, the samples selected for the measurements had symmetrically positioned Hall contacts, i.e. in the absence of a magnetic field the potential from the Hall contacts was zero. For measurements in strong magnetic fields up to 40 T we used a pulsed magnetic field. In one of the $Bi_{0.5}Sb_{1.5}Te_3Ag_{0.003}$ samples the oscillations of the magnetoresistance were investigated by rotating the magnetic field in the C_3C_1 plane (see below).

3. RESULTS OF MEASUREMENTS

3.1. Galvanomagnetic Properties of $(Bi_{1-x}Sb_x)_2Te_3Ag_y$

Tables 2 and 3 give the resistivity ρ, Hall coefficients R, and Hall mobility μ of $(Bi_{1-x}Sb_x)_2Te_3Ag_y$ crystals at various temperatures. It should be noted that R does not depend on the magnetic field for all samples. Figure 1 gives the dependences $\rho(T)$ for various samples. For all samples the value of ρ decreases with decreasing temperature and saturates at low temperatures. In the temperature range 77–300 K the resistivity depends on temperature as $\rho \propto T^m$ where $m \approx 1.1$ for $Sb_2Te_3Ag_y$ and for $Bi_{0.5}Sb_{1.5}Te_3Ag_y$ samples. The small difference between this value of m and that of $m = 1.5$ typical of acoustic phonon scattering may be attributed to the additional scattering on ionized impurities and a possible temperature dependence of the effective mass when the current is directed along the C_2 axis. For a magnetic field parallel to the C_3 axis direct measurements have established that the cyclotron mass does not depend on temperature [5].

The Hall coefficients R are positive for all samples and depend on temperature: as the temperature decreases from room temperature to liquid-nitrogen temperature, the value of R decreases slightly and then remains almost constant (see Table 2 for $Sb_2Te_3Ag_y$ and Table 3 for $Bi_{0.5}Sb_{1.5}Te_3Ag_y$. This $R(T)$ dependence is typical of Sb_2Te_3 single crystals and may be explained quantitatively by the existence of two groups of holes belonging to the upper and lower valence bands, having different masses and probabilities (see, for example [6]). In the silver-doped samples R decreases monotonically with increasing Ag content, which indicates an

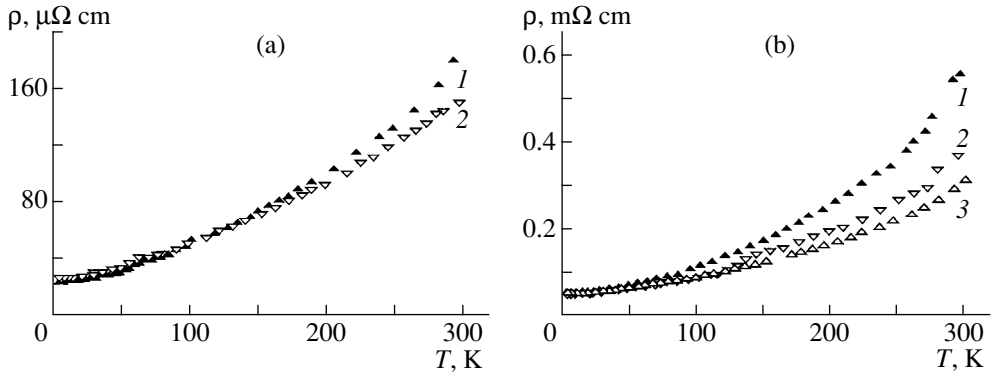


Fig. 1. Temperature dependences of the resistivity $\rho(T)$ for various $Sb_2Te_3Ag_y$ samples (a): (*1*) $Sb_2Te_3Ag_{0.0035}$; (*2*) $Sb_2Te_3Ag_{0.015}$ and $(Bi_{1-x}Sb_x)_2Te_3Ag_y$ (b): (*1*) $(Bi_{1-x}Sb_x)_2Te_3$, (*2*) $(Bi_{1-x}Sb_x)_2Te_3Ag_{0.0014}$, (*3*) $(Bi_{1-x}Sb_x)_2Te_3Ag_{0.0030}$.

increase in the hole concentration. The Hall coefficients do not depend on the magnetic field for all samples. The Hall mobility μ in the silver-doped samples decreases with increasing silver content at all temperatures, the decrease in μ being greater for $(Bi_{0.25}Sb_{0.75})_2Te_3Ag_y$ than for $Sb_2Te_3Ag_y$, as is illustrated in Fig. 2.

It can be seen from the doping results that the average efficiency of silver in $(Bi_{0.25}Sb_{0.75})_2Te_3$ crystals is lower than that in Sb_2Te_3 crystals, i.e., the number of additional holes per silver atom is lower in $(Bi_{0.25}Sb_{0.75})_2Te_3$. This result may be explained by the fact that in Sb_2Te_3 silver mainly enters the Sb sublattice and forms negatively charged substitutional point defects whereas in Bi_2Te_3 silver forms interstitial atoms. This is because the Bi–Te atomic bonds in Bi_2Te_3 crystals are more ionic than the Sb–Te bonds in Sb_2Te_3 crystals. The negative charge of the Te^1 and Te^2 atoms in the Bi_2Te_3 lattice is approximately an order of magnitude higher than the negative charge of the same atoms in the Sb_2Te_3 lattice. Similarly the positive charge of the Bi atoms in Bi_2Te_3 is an order of magnitude higher than the positive charge of Sb in Sb_2Te_3. Thus, very few point defects where Bi is substituted by Ag are formed in the Bi_2Te_3 lattice. At the same time, the incorporation of Ag suppresses the formation of antistructural negatively charged Bi_{Te}

Table 2. $Sb_2Te_3Ag_y$

y	$\rho^{4.2}$, mΩ cm	ρ^{77}, mΩ cm	ρ^{300}, mΩ cm	$R^{4.2}$, cm³/C	R^{77}, cm³/C	R^{300}, cm³/C	$\mu_H^{4.2}$, m²/V s	μ_H^{77}, m²/V s	μ_H^{300}, m²/V s
0	0.031	–	0.250	0.059	–	0.086	0.190	–	0.034
0.0035	0.023	0.050	0.180	0.028	0.020	0.035	0.122	0.040	0.019
0.0093	0.021	0.044	0.149	0.022	0.021	0.030	0.105	0.048	0.020
0.010	0.027	0.042	0.146	0.012	0.016	0.020	0.044	0.038	0.014
0.015	0.026	0.038	0.145	0.009	0.012	0.013	0.035	0.031	0.009

Table 3. $Bi_{0.5}Sb_{1.5}Te_3Ag_y$

y	$\rho^{4.2}$, mΩ cm	ρ^{77}, mΩ cm	ρ^{300}, mΩ cm	$R^{4.2}$, cm³/C	R^{77}, cm³/C	R^{300}, cm³/C	$\mu_H^{4.2}$, m²/V s	μ_H^{77}, m²/V s	μ_H^{300}, m²/V s
0	0.052	0.092	0.557	0.091	0.078	0.110	0.175	0.085	0.020
0.0014	0.047	0.077	0.366	0.037	0.030	0.050	0.079	0.039	0.014
0.0030	0.049	0.078	0.309	0.031	0.028	0.046	0.063	0.036	0.015
0.0048	0.044	0.073	0.333	0.016	0.025	0.044	0.036	0.034	0.013

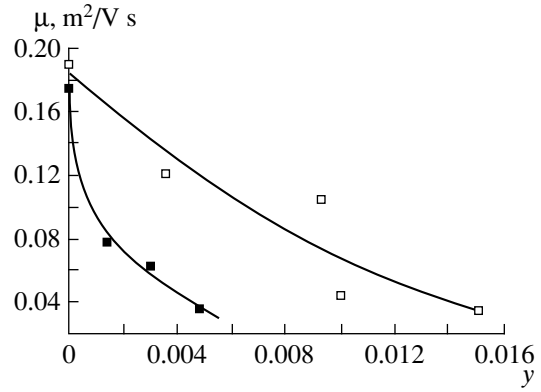


Fig. 2. Dependences of the Hall mobility μ at $T = 4.2$ K on the silver content y in $Sb_2Te_3Ag_y$ (open squares) and $(Bi_{1-x}Sb_x)_2Te_3Ag_y$ (filled squares).

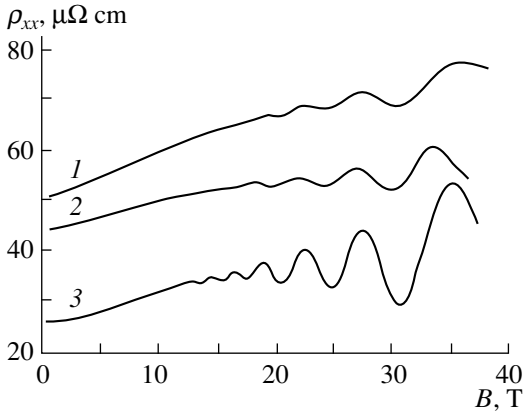


Fig. 3. Oscillations of the transverse magnetoresistivity ρ_{xx} of $Sb_2Te_3Ag_y$ samples at $T = 4.2$ K for samples having y contents: (*1*) 0.0093; (*2*) 0.010; (*3*) 0.015. Curves *1* and *2* are shifted upward.

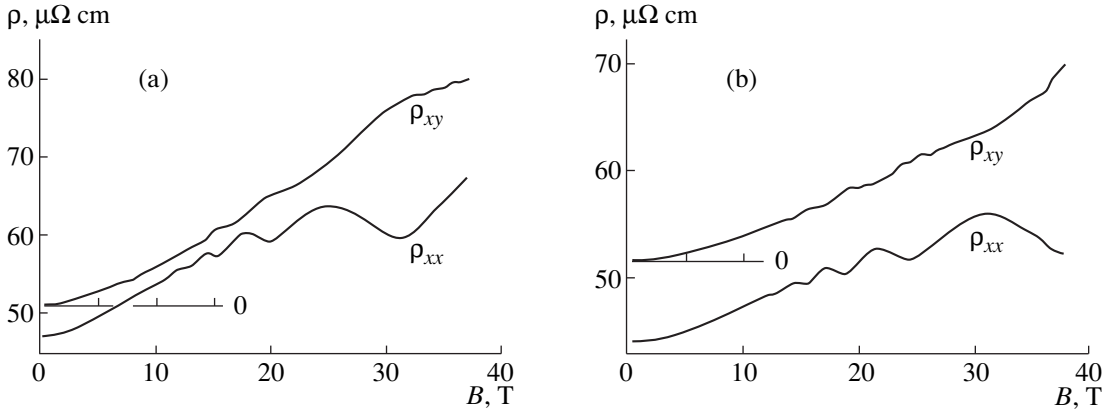


Fig. 4. Oscillations of the transverse magnetoresistivity ρ_{xx} and the Hall resistivity ρ_{xy} for (a) $Bi_{0.5}Sb_{1.5}Te_3Ag_{0.0014}$ and (b) $Bi_{0.5}Sb_{1.5}Te_3Ag_{0.0048}$ samples at $T = 4.2$ K. The dependences $\rho_{xy}(B)$ are measured from zero.

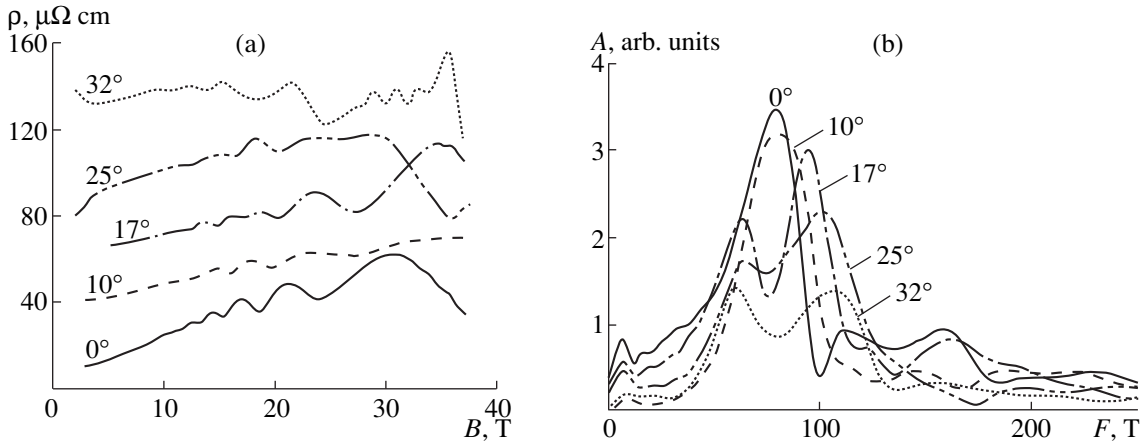


Fig. 5. Angular dependences of oscillations of the transverse magnetoresistivity ρ_{xx} in a $Bi_{0.5}Sb_{1.5}Te_3Ag_{0.0003}$ crystal and their Fourier spectra.

defects. All this gives rise to a different type of defect and ultimately to different behavior of Ag: in Sb_2Te_3 silver behaves as an acceptor whereas in Bi_2Te_3 silver doping induces a donor effect [7]. It is then easy to understand why the efficiency of silver as an acceptor is lower in a mixed $(Bi_{0.25}Sb_{0.75})_2Te_3Ag_y$ crystal.

3.2. Shubnikov–de Haas Effect in $(Bi_{1-x}Sb_x)_2Te_3Ag_y$ Crystals

Shubnikov–de Haas oscillations in $(Bi_{1-x}Sb_x)_2Te_3Ag_y$ crystals were investigated at liquid helium temperature with the magnetic field vector was directed along the C_3 axis. For this orientation of the vector **B** the cross sections of all six ellipsoids of the upper valence band of the Fermi surface are the same. Oscillations of the transverse magnetoresistivity ρ_{xx} are shown in Fig. 3 for some $Sb_2Te_3Ag_y$ samples. A single oscillation frequency is observed in these crystals although the lower valence band is filled at this high hole concentration. The absence of oscillations from the lower valence band is attributable to the higher values of the effective mass in this band and is normal for Sb_2Te_3 (see, for example [4, 6]).

Shubnikov–de Haas oscillations in $(Bi_{1-x}Sb_x)_2Te_3Ag_y$ samples are shown in Fig. 4. For $Bi_{0.5}Sb_{1.5}Te_3Ag_{0.003}$ the Shubnikov–de Haas effect was also studied when the magnetic field vector **B** was rotated in the plane C_3C_1, see Fig. 5. The area of the extreme cross section S_H of the Fermi surface can be determined from the frequency of the Shubnikov–de Haas oscillations F: $S_H = 2\pi e[F/\hbar]$. Values of the oscillation frequency F are given in Tables 4 and 5.

We now turn our attention to the oscillations of the Hall resistivity ρ_{xy} as a function of the magnetic field (see Fig. 4). When the temperature falls below 4.2 K, these oscillations have the form of horizontal plateaus as in the quantum Hall effect. The oscillations typical exhibit phase singularities: the beginning of the downward deflection of $\rho_{xy}(B)$ corresponds to the resistivity minimum. These oscillations are caused by the existence of a lower valence band having a high density of states, which is filled in these samples, and by redistribution of holes between the upper [from which the oscillations of $\rho_{xx}(B)$ are observed] and the lower valence bands. This effect is discussed in further detail in [8].

4. DISCUSSION OF RESULTS

The first Brillouin zone of Sb_2Te_3 is similar to the Brillouin zone of an fcc lattice but is highly compressed along the z-axis. The energy spectrum of the Sb_2Te_3 crystal has two valence bands: a light hole band or upper valence band (UVB) and a heavy hole band or lower valence band (LVB), see Fig. 6. As can be seen from Fig. 6, there are also two conduction bands, an upper (UCB) and a lower (LCB) conduction band.

Table 4. $Sb_2Te_3Ag_y$

y	F, T	E_F, meV	p_{SdH}, 10^{19} cm^{-3}	$1/eR$, 10^{19} cm^{-3}
0	52.0	98	2.20	10.5
0.0035	76.5	144	5.64	22.3
0.0093	78.0	147	5.80	28.4
0.0100	115.0	217	10.40	52.1
0.0152	120.8	228	11.20	69.4

Table 5. $Bi_{0.5}Sb_{1.5}Te_3Ag_y$

y	F, T	E_F, meV	p_{SdH}, 10^{19} cm^{-3}	$1/eR$, 10^{19} cm^{-3}
0	56.8	100	3.28	6.9
0.0014	64.6	114	3.97	16.9
0.0030	82.6	146	5.76	20.2
0.0048	92.0	163	6.74	39.0

The Fermi surface of the upper valence band of Sb_2Te_3 is described by the six-ellipsoid Drabble–Wolfe model [9]. Both valence bands and the lower conduction band each have six Fermi-surface ellipsoids [5]. The six Fermi-surface ellipsoids are positioned as shown in Fig. 7a. One of the axes of the ellipsoid centered in the xz plane is parallel to the y (C_2)-axis. The major axes of the ellipsoids are inclined in the mirror plane $xz(C_1C_3)$ with respect to the crystallographic axes by the angle θ (Fig. 7b) which is determined as follows:

$$\tan 2\theta = 2\alpha_{23}/(\alpha_{22} - \alpha_{33}), \qquad (1)$$

where $\alpha_{ij} = m_0/m_j$ are the components of the tensor of the reciprocal effective masses which depend on energy

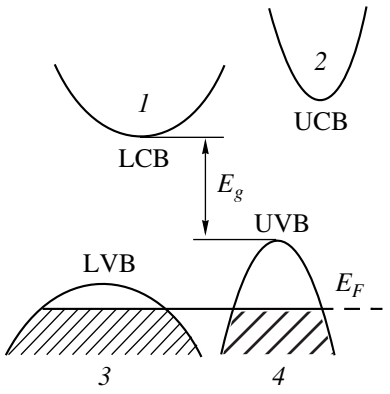


Fig. 6. Band structure of antimony telluride: (*1*, *2*) are the two extrema of the conduction band; (*3*, *4*) are the two extrema of the valence band, E_g is the indirect band gap, and E_F are the Fermi levels in p-type samples.

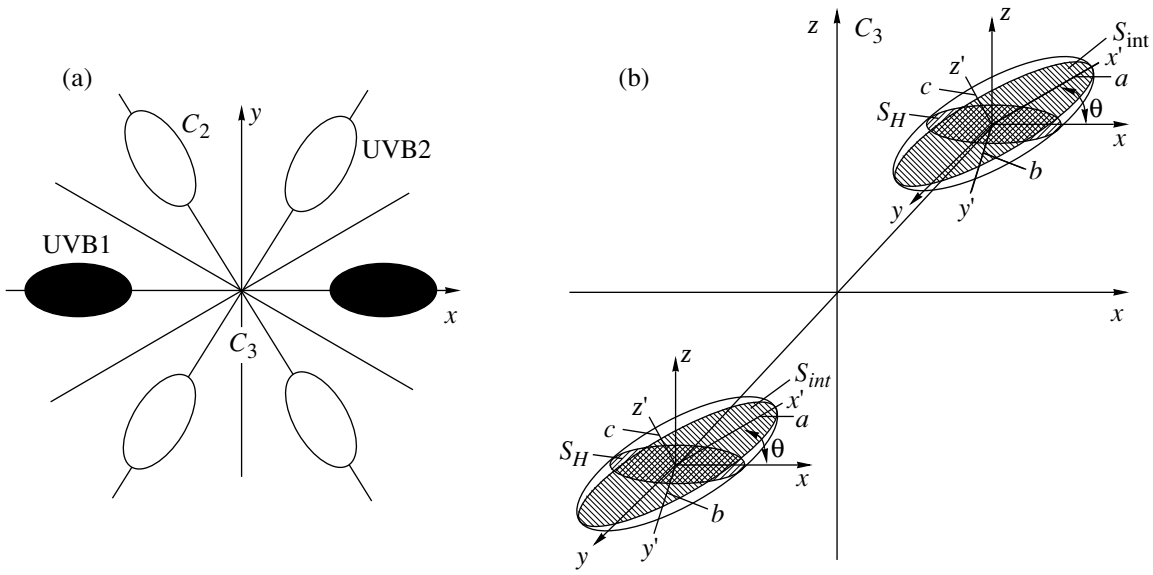


Fig. 7. Six-ellipsoid Fermi surface of the upper valence band of antimony telluride: (a) view in the *xy* plane perpendicular to the C_3-axis and (b) position of ellipsoids in the *xz* plane; C_2 is the twofold axis, θ is the angle of inclination of the ellipsoids to the basal plane, *x*', *y*', *z*' are the canonical coordinates of an ellipsoid having the semiaxes *a*, *b*, *c*.

because of some nonparabolicity of the spectrum. The angle θ in $(Bi_{1-x}Sb_x)_2Te_3$ crystals is only known as far as $x = 0.6$ [10] and is approximately 42° for $(Bi_{0.4}Sb_{0.6})_2Te_3$. For the mixed $Bi_{0.5}Sb_{1.5}Te_3$ crystals studied here the angle $\theta \approx 46°$ was determined by extrapolation.

In accordance with the ellipsoidal nonparabolic model, the general form of the Fermi surface in the upper valence band may be described by the following expression:

$$\alpha_{11}k_x^2 + \alpha_{22}k_y^2 + \alpha_{33}k_z^2 + \alpha_{23}k_yk_z = 2m_0E/\hbar^2, \quad (2)$$

where $k_{x,y,z}$ are the components of the wave vector, k_x is normal to the mirror plane in which the C_2- and C_3-axes lie, k_y is parallel to the C_2-axis, and k_z is parallel to the C_3-axis. In principle, the values of α_{ij} depend on energy but the Fermi surface remains ellipsoidal. It is interesting to study the dependence of the cross-sectional area S_H of each ellipsoid with its plane perpendicular to the direction of the magnetic field vector on the angle of inclination φ of this vector relative to the *z*-axis (C_3). In the experiments the direction of the field was varied in the *zx* (C_3C_1 plane). In this case, the six ellipsoids can be arbitrarily divided into two groups: two lying in this plane (UVB1) and the other four positioned outside this plane and being symmetric to it (UVB2) (Fig. 7a). By virtue of this symmetry, the ellipsoids in each group are equivalent in the sense that their cross-sectional areas S_H are the same for any φ. For a nonzero angle φ the cross sections of the ellipsoids in the same group differ from those of the ellipsoids in the other group. Thus, the ellipsoids in each group should be considered separately.

We introduce the following notation: *a*, *b*, and *c* are the principal axes of the ellipsoids; S_{min}, S_{max}, and S_{int} are the extreme cross sections of the ellipsoid. By S_H we denote the ellipsoid cross section with its plane passing through the center of the ellipsoid and perpendicular to the magnetic field vector, which is directed along the *z* (C_3)-axis in Fig. 7b. One of the axes of each of the two ellipsoids centered in the *xz* plane is parallel to the *y* (C_2) coordinate axis. The major axes of the ellipsoids are inclined in the C_1C_3 planes by the angle θ with respect to the crystallographic axes, which is determined by formula (1). The ellipsoid semiaxes may be expressed as:

$$a = \frac{1}{\hbar}\left(\frac{2m_0E_F}{\alpha'_{11}}\right)^{1/2}, \quad b = \frac{1}{\hbar}\left(\frac{2m_0E_F}{\alpha'_{22}}\right)^{1/2},$$
$$c = \frac{1}{\hbar}\left(\frac{2m_0E_F}{\alpha'_{33}}\right)^{1/2}, \quad (3)$$

where

$$\alpha'_{22} + \alpha'_{33} = \alpha_{22} + \alpha_{33},$$
$$\alpha'_{22}\alpha'_{33} = \alpha_{22}\alpha_{33} - (\alpha_{23})^2. \quad (4)$$

Using simple transformations, we obtain expressions for the extreme cross sections of the ellipsoid:

$$S_{\min} = \pi cb = \frac{2m_0E_F}{(\alpha'_{22}\alpha'_{33})^{1/2}\hbar^2},$$

$$S_{\max} = \pi ac = \frac{2m_0E_F}{(\alpha'_{11}\alpha'_{33})^{1/2}\hbar^2}, \quad (5)$$

$$S_{\text{int}} = \pi ab = \frac{2m_0E_F}{(\alpha'_{11}\alpha'_{22})^{1/2}\hbar^2}.$$

Simple transformations then yield the following dependence of the extreme cross sections of the Fermi surface for the two UVB1 ellipsoids:

$$S_{H1} = 2\pi m_0E_F/\hbar^2(\alpha'_{22}\alpha'_{33}\sin^2\theta + \alpha'_{11}\alpha'_{22}\cos^2\theta)^{1/2}, \quad (6)$$

where φ is the experimental value of the angle between the direction of the magnetic field vector and the *z* axis, and θ is the angle of inclination of the ellipsoid axes with respect to the crystallographic axes.

For the four UVB2 ellipsoids we obtain

$$S_{H2} = 4\pi m_0E_F/\hbar^2$$
$$\times [4(\alpha'_{22}\alpha'_{33}\sin^2\theta + \alpha'_{11}\alpha'_{22}\cos^2\theta)\cos^2\varphi$$
$$+ (\alpha'_{11}\alpha'_{22}\sin^2\theta + 3\alpha'_{11}\alpha'_{33} + \alpha'_{22}\alpha'_{33}\cos^2\theta)\sin^2\varphi \quad (7)$$
$$+ \alpha'_{22}(\alpha'_{11} - \alpha'_{33})\sin 2\theta \sin 2\varphi]^{1/2}.$$

The hole concentration in the six ellipsoids is given by

$$p = \frac{6 \times 2V}{(2\pi\hbar)^3}, \quad (8)$$

where the volume of a single ellipsoid is $V = 4\pi abc/3$.

Using the data obtained from a study of the Shubnikov–de Haas effect in $(Bi_{1-x}Sb_x)_2Te_3Ag_y$ crystals, we calculated the hole concentrations in the upper valence band for the six-ellipsoid Fermi surface and the Fermi energy and the results are given in Tables 4 and 5. We used the following band parameters: $\alpha'_{11} = 2.26$, $\alpha'_{22} = 32.5$, and $\alpha'_{33} = 11.6$ [6].

Figure 8 gives theoretical (curves) and experimental (symbols) dependences of the areas of the extreme cross sections of the Fermi-surface ellipsoids on the angle of inclination of the magnetic field vector relative to the C_3-axis in the C_1C_3 plane for UVB1 and UVB2 ellipsoids. The solid curves give the theoretical angular dependences of the cross-sectional areas of ellipsoids for which the anisotropy of the extreme cross sections is $S_{\max}/S_{\min} = 3.8$. It can be seen that for this anisotropy the experimental points are a good fit to the theoretical curves. Also plotted are similar curves for ellipsoids with different band parameters. The dotted curves refer to ellipsoids with 1.4 times lower anisotropy ($S_{\max}/S_{\min} = 2.8$) and the dot-dash curves refer to ellipsoids having anisotropy 1.4 times higher than 3.8 ($S_{\max}/S_{\min} = 5.4$). It can be seen that these curves differ from the experimental points (the size of the point approximately corresponds to the error in determining the extreme cross section of the Fermi surface) from which it follows that values of $S_{\max}/S_{\min} \approx 3.8$ may be considered suitable for samples of this particular composition.

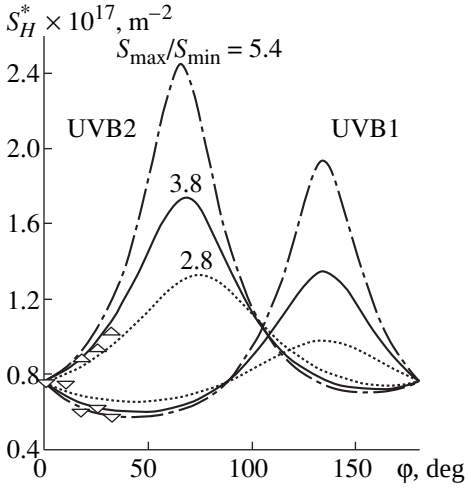


Fig. 8. Theoretical (curves) and experimental (symbols) dependences of the areas of the extreme surface cross sections S_H on the angle φ between the C_3 crystal axis and the vector **B** when the magnetic field is rotated in the C_1C_3 plane. The theoretical dependences are plotted for various values of the anisotropy $S_{\max}/S_{\min}$ of the Fermi surface (indicated in the figure).

The hole concentrations p_{SdH} in the upper valence band calculated using the Shubnikov–de Haas effect were compared with the corresponding Hall concentrations $1/eR$. All these values are given in Tables 4 and 5 which show that the values of p_{SdH} are always lower than the corresponding values of $1/eR$, confirming that the second lower valence band (LVB) is filled at these high hole concentrations.

Thus, a study of the resistivity, Hall effect at different temperatures, and Shubnikov–de Haas effect in $(Bi_{1-x}Sb_x)_2Te_3Ag_y$ single crystals has shown that unlike Bi_2Te_3, in Sb_2Te_3 and $Bi_{0.5}Sb_{1.5}Te_3$ crystals silver exhibits acceptor properties: the hole concentration increases with increasing silver content in the crystal.

An investigation of the angular and concentration dependences of the Shubnikov–de Haas effect with silver doping showed that the anisotropy of the upper valence band ellipsoids for $(Bi_{1-x}Sb_x)_2Te_3Ag_y$ remains the same as $(Bi_{1-x}Sb_x)_2Te_3$.

REFERENCES

1. B. M. Gol'tsman, V. A. Kudinov, and I. A. Smirnov, *Semiconducting Thermoelectric Materials Based on Bi_2Te_3* (Nauka, Moscow, 1972).

2. J. Horak, P. Lostak, and L. Benes, Philos. Mag. B **50**, 665 (1984).
3. P. Lostak, J. Horak, and L. Koudelka, Phys. Status Solidi A **76**, K71 (1983).
4. V. A. Kulbachinskiĭ, Z. M. Dashevskiĭ, M. Inoue, *et al.*, Phys. Rev. B **52**, 10915 (1995).
5. V. A. Kulbachinskiĭ, N. Miura, H. Arimoto, *et al.*, J. Phys. Soc. Jpn. **68**, 3328 (1999).
6. V. A. Kulbachinskiĭ, N. Miura, H. Nakagawa, *et al.*, J. Phys.: Condens. Matter **11**, 5273 (1999).
7. J. Navratil, I. Klichova, S. Karamazov, *et al.*, J. Solid State Chem. **140**, 29 (1998).
8. V. A. Kul'bachinskiĭ, A. Yu. Kaminskiĭ, N. Miyajima, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 754 (1999) [JETP Lett. **70**, 767 (1999)].
9. J. R. Drabble and R. Wolfe, Proc. Phys. Soc. London, Sect. B **69**, 1101 (1956).
10. H. Kohler and A. Freudenberger, Phys. Status Solidi B **84**, 195 (1977).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 1089–1092.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1251–1254.

SOLIDS
Electronic Properties

Properties of a Classical Electron Gas at the Surface of Condensed Media

A. M. Dyugaev*, **, P. D. Grigor'ev, and Yu. N. Ovchinnikov*****

**Landau Institute for Theoretical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*
***Grenoble High Magnetic Field Laboratory, MPI-FRF and CNR, BP166, F-38042 Grenoble, France*
****Centre de Recherches sur les Très Basses Températures associé à l'Université Joseph Fourier, CNRS BP166, F-38042 Grenoble, France*
****e-mail: ovc@itp.ac.ru*

Received February 14, 2000

Abstract—An analysis is made of the problem of the density distribution of a classical electron gas in a charged capacitor whose electrodes are coated with dielectric films (liquid or solid). At high density the electrons form thin layers at each of the capacitor electrodes. The electric field distribution is determined both inside and outside these layers. © *2000 MAIK "Nauka/Interperiodica".*

1. It was shown in [1, 2] that the surface of condensed media, such as liquid helium, can retain electrons by means of electrostatic image forces. The population of the lower electron levels is low even at low temperatures because electrons drift into the bulk above the surface [3]. Hence, experiments to study the properties of a two-dimensional electron gas are carried out in an electric field. A typical experiment is shown schematically in the figure. The capacitor electrodes are coated with insulator layers of thickness h and d having permittivities ε_h and ε_d. An electron source is switched on in the vacuum gap $0 < z < z_0$. A voltage $V = V_1 - V_2$ is applied to the capacitor electrodes, pressing the electrons toward the lower dielectric layer on which an electron layer forms. As the surface density n_e increases, the electrons can almost completely compensate for the electric field in the bulk of the capacitor and in [4] it was assumed that there is a maximum electron density n_s such that when $n_e > n_s$, electron transport to the helium surface ceases. However, in a recent study [5] qualitative reasoning was put forward to suggest that a helium surface may be filled with electrons at a density higher than n_s since, in the same experimental setup, electrons may form a two-dimensional system on the upper capacitor plate coated with a thin helium film.

2. In the present study, an exact solution of the problem of the electron distribution $n_e = n_e(z)$ and the electric field distribution $E = E(z)$ is given in the classical limit for the capacitor shown schematically in the figure. At high temperatures, the mean field approximation can be used and the potential $\varphi(z)$ acting on electrons of density $n_e(z)$ can be obtained from a solution of the Laplace equation [4]

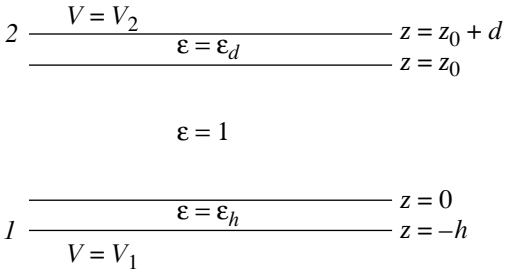
$$\varphi''(z) = 4\pi e n_e(z), \quad n_e(z) = n_0 \exp(e\varphi(z)/T). \quad (1)$$

In dimensionless variables (1) has the form

$$y''(x) = \lambda^2 e^{y(x)}, \quad x \equiv \frac{z}{z_0},$$
$$y = \frac{e\varphi}{T}, \quad \lambda^2 = \frac{4\pi e^2 n_0 z_0^2}{T}. \quad (2)$$

By substituting, we can confirm that solution (2) has the form

$$y = -2\ln(C_1 e^{\gamma x} + C_2 e^{-\gamma x}), \quad \lambda^2 = -8C_1 C_2 \gamma^2. \quad (3)$$



The electrodes *1* and *2* have potentials V_1 and V_2. The insulator layers have thicknesses h and d and permittivities ε_h and ε_d. The vacuum gap $0 < z < z_0$ is accessible to the electrons.

1063-7761/00/9006-1089$20.00

It follows from the theory of linear differential equations [6] that (3) is a general solution of (2). Of the three constants C_1, C_2, and γ, one is determined using the surface electron density n_e:

$$n_e \equiv \int_0^{z_0} n_e(z)dz = n_0 z_0 \int_0^1 e^{y(z)}dx,$$
$$n_e = \frac{z_0 n_0}{2\gamma C_1}\left[\frac{1}{C_1 + C_2} - \frac{1}{C_1 e^{2\gamma} + C_2}\right]. \quad (4)$$

The other two constants are determined by the constraints for continuity of the potential $\varphi(z)$ and the induction $D = \varepsilon E$ at the dielectric–vacuum interfaces where $z = 0$ and $z = z_0$ (see figure)

$$\frac{eV_1}{T} = 2\gamma\frac{h_*(C_1 - C_2)}{z_0(C_1 + C_2)} - 2\ln(C_1 + C_2),$$
$$\frac{eV_2}{T} = -2\gamma\frac{d_*(C_1 e^{\gamma} - C_2 e^{-\gamma})}{Z_0(C_1 e^{\gamma} + C_2 e^{-\gamma})} - 2\ln(C_1 e^{\gamma} + C_2 e^{-\gamma}). \quad (5)$$

The parameters h_* and d_* are the effective thicknesses of the capacitor dielectric layers:

$$h_* \equiv h/\varepsilon_h, \quad d_* \equiv d/\varepsilon_d.$$

Since the physical effect gives the potential difference $V = V_1 - V_2$ and not V_1 and V_2 separately, in (3) it is convenient to introduce new constants instead of C_1 and C_2:

$$C_1 = \frac{C_+ + C_-}{2}, \quad C_2 = \frac{C_+ - C_-}{2},$$
$$y = -\ln(C_+\cosh\gamma + C_-\sinh\gamma). \quad (6)$$

From (4) and (5), we then obtain two relationships to determine the two parameters γ and C_-/C_+, and C_+ is dropped from the expressions for the observable quantities:

$$\frac{2\pi n_e e^2 z_0}{T} = \left(\frac{C_-^2}{C_+^2} - 1\right)\frac{\gamma\sinh\gamma}{\cosh\gamma + (C_-/C_+)\sinh\gamma}, \quad (7)$$

$$\frac{eV}{2T} = \gamma\frac{h_*}{z_0}\frac{C_-}{C_+} + \gamma\frac{d_*}{z_0}\frac{\sinh\gamma + (C_-/C_+)\cosh\gamma}{\cosh\gamma + (C_-/C_+)\sinh\gamma} + \ln\left(\cosh\gamma + \frac{C_-}{C_+}\sinh\gamma\right). \quad (8)$$

Relationships (7) and (8) are exact. Of physical interest however is the limit of the high electron density n_e when the parameter C_-/C_+ is large. To achieve this it is sufficient for one of the parameters A or B given by the following expressions to be large:

$$A \equiv \frac{2n_e\pi e^2 z_0}{T}, \quad B \equiv \frac{eVz_0}{2Th_*}. \quad (9)$$

Expressions (7)–(9) yield simple equations for the two quantities C_* and φ:

$$C_* \equiv \frac{C_-}{C_+}\gamma, \quad \varphi \equiv \gamma\coth\gamma, \quad A = \frac{C_*^2}{C_* + \varphi},$$
$$B = C_* + \frac{d_*}{h_*}\frac{C_*\varphi}{(\varphi + C_*)}. \quad (10)$$

In (8), we neglected the last logarithmic term compared with eV/T. From solution (10), we obtain final expressions for $y = y(x)$, C_*, and $\varphi(\gamma)$:

$$y = -2\ln\left(\cosh\gamma x + C_*\frac{\sinh\gamma x}{\gamma}\right),$$
$$C_* = B\frac{(h_* + d_* n_e/n_s)}{(h_* + d_*)},$$
$$Bh_*\frac{(n_s - n_e)}{n_e}\frac{(h_* + d_* n_e/n_s)}{(h_* + d_*)^2} = \varphi(\gamma). \quad (11)$$

The electron density n_s was determined in [4]:

$$4\pi e^2 n_s = \frac{eV}{h_*} \text{ or } B = \frac{2\pi n_s e^2 z_0}{T}. \quad (12)$$

3. We shall now analyze solution (10) and (11). For the electron density $n_e < n_s$, a solution with a real value of γ exists. If $n_e > n$, complex solution (11) is obtained with $\gamma = i\delta$, where the phase δ varies in the range $0 < \delta < \pi$:

$$y = -\ln\left(\cos\delta x + C_*\frac{\sin\delta x}{\delta}\right). \quad (13)$$

For the electron density distribution $n(x)$, we obtain from (1) and (13)

$$n(x) = \frac{\nu}{(\kappa\cos\delta x + (\sin\delta x)/\delta)^2},$$
$$\nu \equiv \frac{T}{2\pi e^2 z_0^2}, \quad \kappa \equiv \frac{1}{C_*}. \quad (14)$$

The value $\delta = 0$ corresponds to the solution obtained in [4]. Since the parameter B determined in (9) is large,

the phase δ rapidly approaches the value $\delta = \pi$ as the density n_e increases if

$$n_e - n_s \gg n_s/B.$$

In this case, the electron density $n(x)$ has two maxima: one at $x = 0$ and the other at $x = 1$ which corresponds to two thin electron layers at the lower and upper capacitor plates:

$$\begin{aligned} n(x) &= \frac{\nu}{(\kappa_1 + x)^2} \quad \text{at} \quad x \ll 1, \\ \kappa_1 &= \frac{1}{C_*}; \\ n(x) &= \frac{\nu}{(\kappa_2 + 1 - x)^2} \quad \text{at} \quad 1 - x \ll 1, \\ \kappa_2 &= \frac{n_s(h_* + d_*)}{Bh_*(n_e - n_s)}. \end{aligned} \tag{15}$$

The total electron density is $n_e = n_h + n_d$, where n_h and n_d are the densities of the electron layers near the lower and upper capacitor plates:

$$n_h = \frac{n_s h_* + n_e d_*}{h_* + d_*}, \quad n_d = \frac{(n_e - n_s)h_*}{h_* + d_*}. \tag{16}$$

In the bulk of the capacitor outside the electron layers, we can assume in (14) $\kappa = 0$, $\delta = \pi$, and then

$$n(x) = \frac{\nu\pi^2}{(\sin\pi x)^2}, \quad \kappa_1 \ll x, \quad 1 - x \gg \kappa_2. \tag{17}$$

From the expression for y (13), we can also obtain the electric field distribution $E(x)$:

$$\theta(x) = \frac{2T}{ez_0}\frac{(C_*\cos\delta x - \delta\sin\delta x)}{(C_*(\sin\delta x)/\delta + \cos\delta x)}. \tag{18}$$

In the region outside the electron layers, we have

$$E(x) = \frac{2T}{ez_0}\cot\pi x, \quad \kappa_1 \ll x, \quad 1 - x \gg \kappa_2. \tag{19}$$

The electric field $E(x)$ only goes to zero at a single point $x = 1/2$ where it changes sign. Thus, complete screening of the field never occurs in the bulk of the capacitor and electrons can be transported to the dielectric layers for any electron density n_e. For liquid helium, the only constraint on n_e is associated with loss of stability of its surface (7).

4. We shall now consider the case of low density n_s, i.e., a low value of the parameter B in (9) and (12) and we identically rewrite (11) in the form

$$\begin{aligned} C_* &= A\frac{(d_* + h_* n_s/n_e)}{h_* + d_*}, \\ \varphi &= Ah_*\left(\frac{n_s}{n_e} - 1\right)\frac{(d_* + h_* n_s/n_e)}{(h_* + d_*)^2}. \end{aligned} \tag{20}$$

It follows from (20) that even for $n_s = 0$ when the voltage applied to the capacitor electrodes is switched off, our solutions (14) and (18) obtained for $n(x)$ and $E(x)$ still hold if the electron density n_e is high and $A \gg 1$ [see (9)]. In this case, the electron layers are narrow: $\kappa_1, \kappa_2 \ll 1$ [see (15)]. If the total electron density n_e is fixed and the voltage V varied, according to (9), (12), and (16), the electron density will be redistributed between the layers on the capacitor dielectric films. This redistribution takes the form of a tunneling current since the electron potential energy $\vartheta = -e\varphi$ has a maximum at $x = 1/2$ [see (13)]:

$$\vartheta = 2T\ln\left[C_*\frac{\sin\pi x}{\pi}\right], \quad x \gg \kappa_1, \quad 1 - x \ll \kappa_2. \tag{21}$$

The range of validity of the classical equation (1) for helium was determined in [4]. The principal approximation involves a loss of discreteness of the electron levels: their populations should be low. This is not the case for a thin layer of helium on the upper electrode of a capacitor [5]. However, the range of validity of the "quantum" approach [5] involves neglecting the contribution of the bulk electron levels. Even in a strong electric field, the quantization of the levels becomes insignificant as they become filled with electrons. In cases where the electric field is almost completely screened when n_e approaches n_s, the discrete electron levels are squeezed into a continuous spectrum. An effect involving divergence of the electron partition functions comes into play [3] which inevitably leads to a reduction in the population of the discrete levels and the range of validity of equation (1) becomes broader. Interestingly, for densities $n_e \sim 10^6$ cm^2 and temperatures $T > 1$ K, the classical equation (1) can be applied for the lower electron layer on a thick helium film [4], but, on the upper thin helium film, the electrons fill a single surface level [5].

These results are also fully applicable to a capacitor whose dielectric layers are solid hydrogen or neon. As is seen from (16), for $d \gg h$ (see figure) electron transport mainly enriches the lower electron layer that suggest the possible way for designing an experiment. The tunneling current between the electron layers can be measured using a well-developed technique for helium [8]. The drift of "hot" electrons from the surface of hydrogen or neon can be observed experimentally from the cyclotron resonance [9].

ACKNOWLEDGMENTS

The authors thank P. Wyder, I. Wagner, and P. Monceau for their hospitality at the LCMI and CRTBT Laboratories in Grenoble where this work was carried out.

REFERENCES

1. M. W. Cole and M. H. Cohen, Phys. Rev. Lett. **23**, 1238 (1969).
2. V. B. Shikin, Zh. Éksp. Teor. Fiz. **58**, 1748 (1970) [Sov. Phys. JETP **31**, 936 (1970)].
3. R. S. Grandall and R. Williams, Phys. Rev. A **5**, 2183 (1972).
4. D. M. Chernikova, Zh. Éksp. Teor. Fiz. **68**, 249 (1975) [Sov. Phys. JETP **41**, 121 (1975)].
5. V. B. Shikin, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 274 (1999) [JETP Lett. **70**, 283 (1999)].
6. E. Kamke, *Gewöhnliche Differentialgleichungen* (Acad. Verlag, Leipzig, 1959; Nauka, Moscow, 1976).
7. L. P. Gor'kov and D. M. Chernikova, Pis'ma Zh. Éksp. Teor. Fiz. **18**, 119 (1973) [JETP Lett. **18**, 68 (1973)].
8. A. P. Volodin, M. S. Khaĭkin, and V. S. Édel'man, Pis'ma Zh. Éksp. Teor. Fiz. **23**, 524 (1976) [JETP Lett. **23**, 478 (1976)].
9. V. S. Édel'man, Pis'ma Zh. Éksp. Teor. Fiz. **25**, 422 (1977) [JETP Lett. **25**, 394 (1977)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 1093–1104.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1255–1269.

SOLIDS
Electronic Properties

Interwell Excitons in GaAs/AlGaAs Double Quantum Wells and Their Collective Properties

A. V. Larionov*, V. B. Timofeev*, **, J. Hvam*, and K. Soerensen*****

**Institute of Solid State Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432, Russia*
***e-mail: timofeev@issp.ac.ru*
****Microelectronic Centre, DK 2800 Lyngby, Denmark*

Received February 17, 2000

Abstract—Luminescence spectra of interwell excitons in GaAs/AlGaAs double quantum wells with electric-field-tilted bands (*n–i–n*) structures were studied. In these structures the electron and the hole in the interwell exciton are spatially separated between neighboring quantum wells by a narrow AlAs barrier. Under resonant excitation by circularly polarized light the luminescence line of the interwell excitons exhibited appreciable narrowing as their concentration increased and the degree of circular polarization of the photoluminescence increased substantially. Under resonant excitation by linearly polarized light the alignment of the interwell excitons increased as a threshold process with increasing optical pumping. By analyzing time-resolved spectra and the kinetics of the photoluminescence intensity under pulsed excitation it was established that under these conditions the rate of radiative recombination increases substantially. The observed effect occurs at below-critical temperatures and is interpreted in terms of the collective behavior of the interwell excitons. Studies of the luminescence spectra in a magnetic field showed that the collective exciton phase is dielectric and in this phase the interwell excitons retain their individual properties. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Quasi-two-dimensional systems, double quantum wells, and superlattices are attracting interest in particular because of the fundamental possibility of spatially separating photoexcited electron and hole carriers between neighboring quantum wells [1–15]. In double quantum wells with an applied electric field which tilts the bands, excitons can be excited whose electron and hole are situated in different quantum wells separated by a barrier transparent to tunneling. These excitons are called spatially indirect (I) or interwell excitons in contrast to the direct intrawell excitons (D) where the electron and the hole in the exciton are located in the same quantum wells. Interwell excitons are long-lived compared with intrawell excitons so that they can easily build up, and a gas consisting of these excitons can be cooled to fairly low temperatures. As a result of destroyed inversion symmetry, interwell excitons have a dipole moment even in the ground state. The theory predicts various possible scenarios for collective behavior in a fairly dense system of spatially separated electrons and holes [1, 9–14]. It was shown in [14] that despite the dipole–dipole repulsion of interwell excitons, for certain critical parameters, i.e., the dipole moment of the interwell excitons, their density, and temperature, the stable state in the *e–h* system may be the liquid dielectric phase of these excitons. In an earlier study [11] it was noted that the condensed dielectric exciton phase can only occur in the presence of lateral confinement (random or artificially produced) in the quantum-well plane. Under these confinement conditions and the associated external compression, it is easier for excitons to build up to high critical densities sufficient for the appearance of collective exciton interaction effects.

It should be borne in mind that in real semiconducting-heterostructure tunnel-coupled quantum systems there is always a random potential as a result of various structural defects, i.e., residual impurities, charged and neutral, fluctuations of the barrier width and the widths of the quantum wells themselves, and so on. These fluctuations create a random potential relief in the quantum-well planes so that photoexcited electrons and holes spatially separated between neighboring wells, and also interwell excitons are highly localized at these fluctuations at fairly low temperatures. This strong localization effect in coupled quantum systems is manifest in particular in lateral thermoactivated carrier tunneling and is observed in experiments to study spectral narrowing of the luminescence line with increasing temperature, which corresponds to interwell radiative recombination [7].

In the present study we investigate the photoluminescence of interwell excitons in double quantum wells with a barrier containing four AlAs monolayers separating the quantum wells. With such narrow barriers the interwell excitons are fairly strongly coupled. Under these conditions the interwell excitons are localized at lateral fluctuations of the random potential without any significant changes in their internal structure if the linear scales of the fluctuations are $l > a_B$ (Bohr radius of an exciton $a_B \leq 100$ Å) and the fluctuation amplitudes are $\Delta > kT$. We know that in structures with narrow AlAs barriers the fluctuations of the barrier width are

large-scale and consequently the corresponding fluctuations of the lateral potential relief will also be large-scale. It is naturally predicted that at fairly low temperatures interwell excitons will be located in these large-scale random lateral potential wells. It is interesting to know whether a system of interwell excitons under conditions of such random lateral confinement will demonstrate critical behavior with increasing density and at fairly low temperatures. In the present paper we attempt to answer these questions.

The paper is constructed as follows. After describing the double-quantum-well heterostructures being studied and the experimental technique in Section 2, in Section 3 we describe the radiative recombination properties of localized and delocalized interwell excitons under conditions of resonant excitation by circularly polarized light when the pump power, electrical bias voltage, and temperature are varied. The time evolution of the spectra and the decay kinetics of the luminescence intensity of interwell excitons under conditions of pulsed laser excitation are discussed in Section 4. The diamagnetic properties of interwell excitons, their Zeeman splitting, and *g*-factor are presented in Section 5. In Section 6 we describe experiments using interwell excitons which accumulated in a lateral potential well of deformation origin far from the photoexcitation region. Finally in the concluding section 7 various properties of interwell excitons observed in their luminescence spectra and their critical behavior as a function of the optical pumping and temperature are interpreted in terms of collective exciton behavior.

2. EXPERIMENTAL TECHNIQUE AND STRUCTURES

We investigated GaAs/AlGaAs *n–i–n* heterostructures with a GaAs/AlAs/GaAs double quantum well and narrow AlAs tunnel barrier (4 ML) between the wells (the width of the GaAs quantum wells was approximately 120 Å and the AlAs barrier was approximately 11 Å). The entire structure was grown by molecular beam epitaxy on a (001)-oriented *n*-type doped GaAs substrate having an Si dopant concentration of 10^{18} cm^{-3}. First a 0.5 μm thick Si-doped (10^{18} cm^{-3}) GaAs buffer layer was grown on the substrate, followed by a 0.15 μm AlGaAs isolating layer ($x = 0.33$) and GaAs/AlAs/GaAs double quantum wells.

The heterojunction of each GaAs quantum well with the isolating AlGaAs layer was also separated by a narrow (4 ML) AlAs barrier. The narrow AlAs barriers were grown using a stop growth regime. In this growth technique the fluctuations of the AlAs barrier widths are large-scale. The double quantum wells were followed by a 0.15 μm thick isolating AlGaAs layer then a 0.1 μm thick Si-doped (10^{18} cm^{-3}) GaAs layer. Single broad GaAs quantum wells (of width ≈300 Å) were located in the isolating AlGaAs barriers near the doped regions. The luminescence from these quantum wells was used to assess the *e–h* excitations "percolating" toward the contact, doped regions of the structure and also to monitor qualitatively the density of interwell excitons building up in the double quantum wells under resonant excitation. The upper part of the structure was covered with a 100 Å thick GaAs layer. Mesas having dimensions of 1×1 mm were fabricated on the as-grown structure by a lithographic technique. Metal contacts of Au + Ge + Pt alloy were deposited as a frame on the upper part of the mesa and also the doped buffer layer.

The luminescence spectra were investigated under conditions of cw resonant excitation of intrawell direct heavy-hole excitons using a tunable Ti-sapphire laser. Circularly polarized resonant excitation was used to achieve optical orientation of the angular momentum in the exciton. The luminescence signal was analyzed using a circular analyzer. The time evolution of the luminescence spectra and their intensity kinetics were studied under pulsed excitation by a picosecond laser (wavelength 6200 Å, pulse duration 30 ps, repetition frequency 0.8 MHz). Pulsed measurements of the spectra and luminescence kinetics were made using a time-correlated photon counting system.

In order to detect the luminescence spectra we projected the optical excitation spot on the mesa onto crossed slits which were used to monitor the uniformity of the excitation and to select suitable regions of the structure for detection. Under cw excitation the luminescence spectra of the interwell excitons were also studied in the presence of a magnetic field perpendicular to the quantum-well planes (Faraday geometry). The diamagnetic shift of the excitons and their Zeeman splitting (σ^+, σ^- components of the spectra) were analyzed under these conditions.

3. LUMINESCENCE SPECTRA OF INTERWELL EXCITONS UNDER RESONANT PHOTOEXCITATION

In this section we shall analyze the behavior of the luminescence spectra of interwell excitons when intrawell spatially direct heavy-hole excitons (1*s*HH excitons) are resonantly excited by circularly polarized light. We are interested in how these spectra behave (their profile, intensity, and degree of circular polarization) when the excitation power density, applied electric field, and temperature vary.

3.1. Figure 1 shows luminescence spectra of intrawell (D) and interwell (I) excitons measured under resonant excitation and various applied electric fields. The optical transitions studied are illustrated schematically in Fig. 2. In the intrawell luminescence range at zero electric bias two lines can be identified, free, 1*s*HH, and bound excitons [15]. At low temperatures the direct exciton (line *D*) is observed as a weak "shoulder" on the violet wing of the exciton complex line. At negative electric fields from –0.3 V an interwell radiative recombination line appears which shifts almost linearly toward lower energies as the applied voltage increases in accor-

dance with the linear Stark shift of the size-quantization levels in the quantum wells (see inset to Fig. 1). Then only the line of the charged exciton complex remains significant in the intrawell recombination spectrum [15]. At higher negative fields $U < -0.4$ V under cw excitation only the photoluminescence line of the interwell excitons dominates in the spectra while the luminescence of the direct intrawell excitons and exciton complexes is of considerably lower intensity under these conditions.

The intensity of the interwell exciton line behaves nonmonotonically as a function of the applied bias. It appears in the luminescence spectrum at voltages when the Stark shift exceeds the difference between the binding energies of the intrawell and interwell excitons $eFz \geq E_D - E_I$.

At $U \sim -0.7$ V the intensity of this line has a maximum and then decreases with increasing electric field (see Fig. 1). This behavior is easily understood if we bear in mind that as the field increases, the effective dipole moment of the interwell exciton increases in the growth direction of the structure (z-axis) and the overlap of the electron and hole wave functions in the exciton decreases monotonically.

3.2. At fairly low temperatures (T = 2 K) and low pumping the luminescence line of the interwell excitons has a large width (FWHM = 4–5 meV) and the line profile itself is asymmetric with a fairly extended long-wavelength tail and relatively steep violet edge (see Figs. 1, 3). These characteristics of the photoluminescence line of the interwell excitons are a consequence of their strong localization at fluctuations of the random potential [6]. In this case the line width reflects the statistical distribution of the amplitudes of the random potential. The pumping levels are so low that the average density of the spatially separated electrons and holes is $n_{e-h} < 10^9$ cm^{-2}. At these concentrations the average statistical filling of lateral random potential wells having linear scales $l < 1$ μm by interwell excitons is less than unity and the inhomogeneous width of the photoluminescence spectrum of the interwell excitons is fairly large (Fig. 1).

The luminescence intensity, profile, and line width of the interwell excitons varies substantially as the power of the resonant excitation of direct intrawell excitons 1sHH by circularly polarized light increases (see Fig. 3). Line I narrows to 1.3 meV with increasing pumping, i.e., it becomes almost four times narrower. Its intensity at the maximum increases superlinearly while the line profile becomes almost symmetric or homogeneously broadened in contrast to the clearly inhomogeneously broadened profile of the photoluminescence spectrum of the interwell excitons at low excitation densities. In the strongly narrowing region the interwell exciton line is shifted by up to 1 meV toward lower energies as the pumping increases. Only at negligible pump powers $P > 6$ W/cm^2 does this line shift toward higher energies and become broader. A line shift toward higher energies indicates that the applied

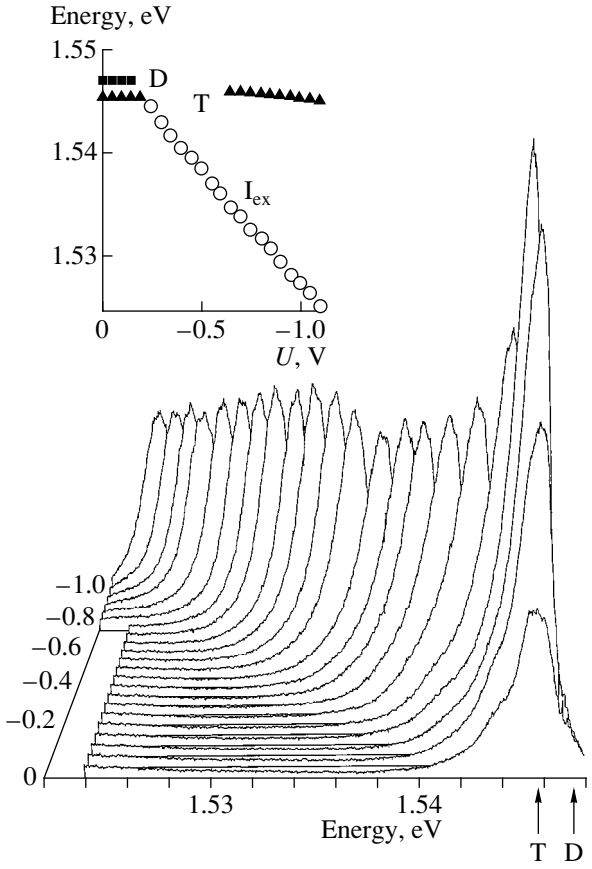


Fig. 1. Behavior of the luminescence spectra of interwell excitons as a function of applied voltage; T and D are the intrawell luminescence lines of an exciton complex and a direct 1sHH exciton at T = 2 K. The interwell exciton line (I) is constructed as a linear function of the applied electric field (see also inset to figure). The numbers to the left of the measured spectra correspond to the applied electric field. These photoluminescence spectra were measured at various applied voltages between 0 and –1.05 V at intervals of 0.05 V.

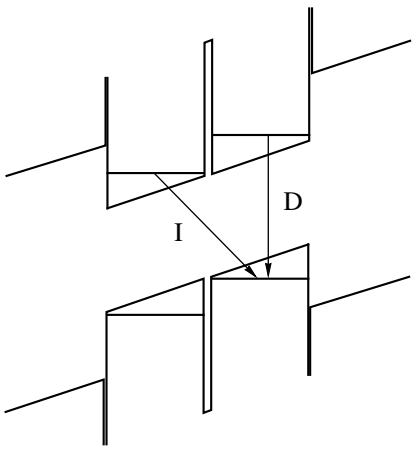


Fig. 2. Schematic of optical transitions on application of an electrical voltage which shifts the size-quantization levels in the double quantum wells. The arrows show the optical transitions corresponding to the intrawell (direct) D and interwell (spatially indirect) I excitons.

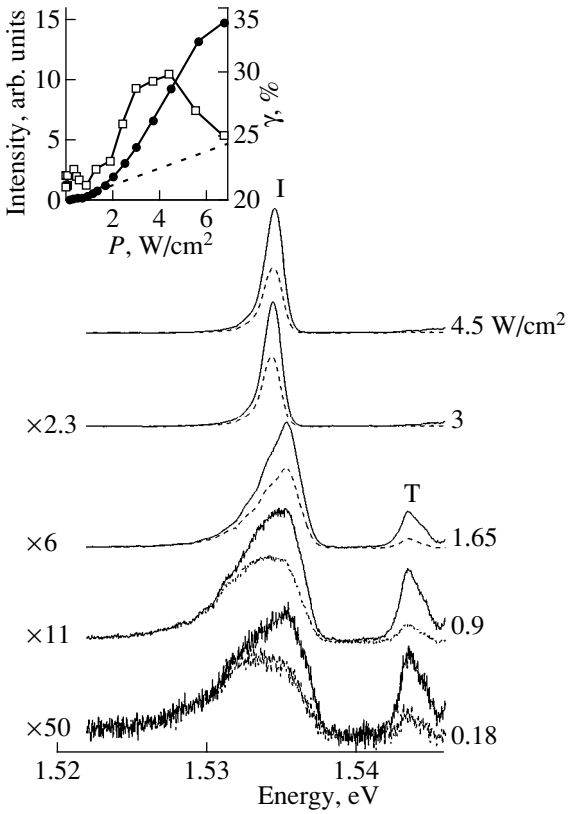


Fig. 3. Luminescence spectra and degree of circular polarization of interwell excitons for various optical pump powers under conditions of resonant excitation of heavy-hole intrawell excitons and $T = 1.8$ K: T and I are the lines of the intrawell exciton complex and an interwell exciton, the solid curves give the photoluminescence spectra measured for σ^+ polarization and the dashed curves give the spectra for the σ^- polarization. The numbers on the left of the spectra give the intensity scale factors and those on the right give the excitation power densities. The inset gives the photoluminescence intensity of line I (filled circles) and its degree of circular polarization (open squares) as a function of the power density. The dashed curve gives the linear approximation of the photoluminescence intensity of line I.

electric field is screened when the density of the interwell excitons becomes sufficiently high. Then, using the Gauss formula we can obtain an upper estimate of the interwell exciton density from the spectral shift. This estimate gives $n = 3 \times 10^{10}$ cm^{-2} for the concentration of interwell excitons when the line width becomes minimal. We observed appreciable narrowing of the luminescence line of the interwell excitons for various negative bias voltages between –0.5 V and –1.2 V. At high negative voltages similar narrowing of the interwell-exciton luminescence line occurred at significantly lower pump powers.

The strong narrowing of the interwell-exciton photoluminescence line at low temperatures suggests that at high excitation densities the lateral fluctuations of the random potential begin to be significantly screened. As a result of this screening of the random potential relief at fairly high pumping levels the interwell excitons lie above the percolation threshold (or the mobility threshold associated with the strong localization effect) and are delocalized. As the density of these delocalized interwell excitons increases, the narrow photoluminescence line is shifted toward lower energies (see Fig. 3). From this observation it follows that the ground-state energy of the interacting interwell excitons decreases as their density increases despite dipole–dipole repulsion. This behavior is typical of a dense system of Bose particles as their concentration increases at fairly low temperature.

3.3. The degree of circular polarization of the interwell-exciton luminescence line exhibits interesting behavior in the region where its intensity increases superlinearly with increasing resonant excitation power (see Fig. 3). In our experiments using circularly polarized exciting light, direct, completely spin-oriented 1*s*HH excitons were created for which the heavy-hole angular momentum is $J_h = +3/2$ and the electron spin $S_z = -1/2$. As a result of carrier tunneling and binding to form interwell excitons, and also as a result of spin–lattice relaxation and strong spin–orbit interaction for the holes, the spin "memory" of the interwell excitons is partially lost but still remains appreciable and is almost 5–10% at a low excitation density, although the interwell excitons are localized under these conditions and the corresponding photoluminescence line is inhomogeneously broadened. At constant pumping the degree of circular polarization of the interwell-exciton photoluminescence decreases monotonically with increasing bias voltage. As the power density of the resonant photoexcitation increases when the interwell-exciton photoluminescence line exhibits substantial narrowing, the degree of circular polarization of the corresponding line increases severalfold as a threshold process. Assuming that the rate of spin relaxation varies little with increasing pumping (most likely it only increases), this increase in the degree of circular polarization is naturally attributed to a reduction in the lifetime of the interwell excitons. This is deduced from a simple kinetic expression linking the degree of circular polarization with the lifetimes and spin relaxation [16]:

$$\gamma = \gamma_0/(1 + \tau_d/\tau_s),$$

where γ_0 and γ are the degrees of polarization of the intrawell and interwell excitons, and τ_d and τ_s are the radiative recombination and spin relaxation times of the interwell excitons, respectively. Assuming that τ_s is barely sensitive to the pumping, using this expression we can easily conclude that the experimentally observed trebling of the degree of circular polarization of the interwell excitons with increasing excitation power is a consequence of an at least fivefold increase in their rate of radiative annihilation. In the next section (Section 4) we discuss the lifetimes of interwell excitons determined directly using pulsed measurements.

3.4. When direct 1*s*HH excitons (polarized parallel to the layer plane) were excited resonantly by linearly

polarized light, as the pump power increased we observed a threshold increase in the linear polarization of the narrow photoluminescence line (alignment of interwell excitons) in the region of superlinearly increasing intensity. When the excitation densities were low and the interwell excitons were strongly localized at fluctuations of the random potential, their spectrum remained weakly polarized under the same conditions. The results of these experiments are illustrated in Fig. 4 which clearly shows an abrupt increase in the linear polarization of the interwell-exciton photoluminescence in a narrow range of pumping. This alignment of the lateral dipole moment of the interwell excitons reached a maximum and then decreased as the power density of the resonant excitation increased further, when screening of the applied electric voltage became appreciable (the pump range in which the photoluminescence line began to show appreciable broadening and shift toward higher energies).

3.5. Thus, as the concentration of interwell excitons increases, the intensity of the corresponding photoluminescence line increases superlinearly and the line exhibits strong narrowing while its degree of polarization increases, which indirectly indicates that the lifetime of the interwell excitons is reduced. These effects were very sensitive to temperature. We observed that when the temperature increased above critical values at high constant pump power, the line width of the interwell excitons increased abruptly and the degree of circular polarization dropped to its previous level. The temperature behavior of the degree of circular polarization and the photoluminescence line width of the interwell excitons are illustrated in Fig. 5. It can be seen that the critical temperature at which such dramatic spectral changes occurred in this case is $T_c \le 6$ K ($\Delta T = \pm 1$ K).

4. KINETICS OF THE PHOTOLUMINESCENCE SPECTRA

In this section we shall discuss the time evolution of the luminescence spectra of intrawell and interwell excitons and also the kinetics of the intensities of the corresponding photoluminescence spectra under conditions of pulsed excitation using a picosecond laser.

Under these pulsed excitation conditions at the time of action of the laser pulse hot photoexcited electrons and holes are generated in each quantum well with equal density and not spatially separated. The spatial separation of the carriers between neighboring tunnel-coupled quantum wells is the result of complex kinetic processes involving intrawell relaxation and recombination of carriers and also carrier tunneling through the interwell barrier. Figure 6 shows the time evolution of the photoluminescence spectra under pulsed excitation measured for different delays relative to the exciting laser pulse at $T = 1.8$ K and applied voltage $U = -0.7$ V. For zero delays and integration of the signal with 1 ns time gates the spectra only reveal a region of direct intrawell luminescence. The photoluminescence spectrum of the interwell excitons only begins to form at

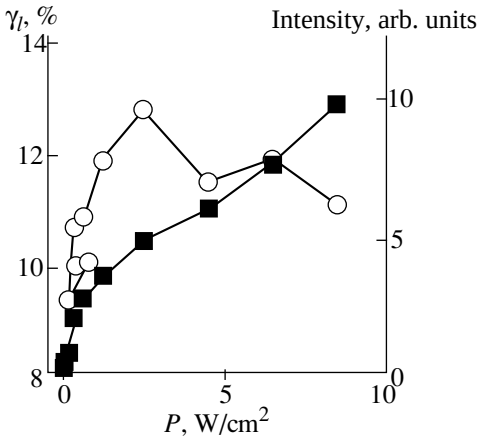


Fig. 4. Degree of linear polarization of interwell excitons (open circles) and photoluminescence intensity (filled squares) under resonant excitation of intrawell excitons by linearly polarized light as a function of optical pumping at $T = 1.5$ K.

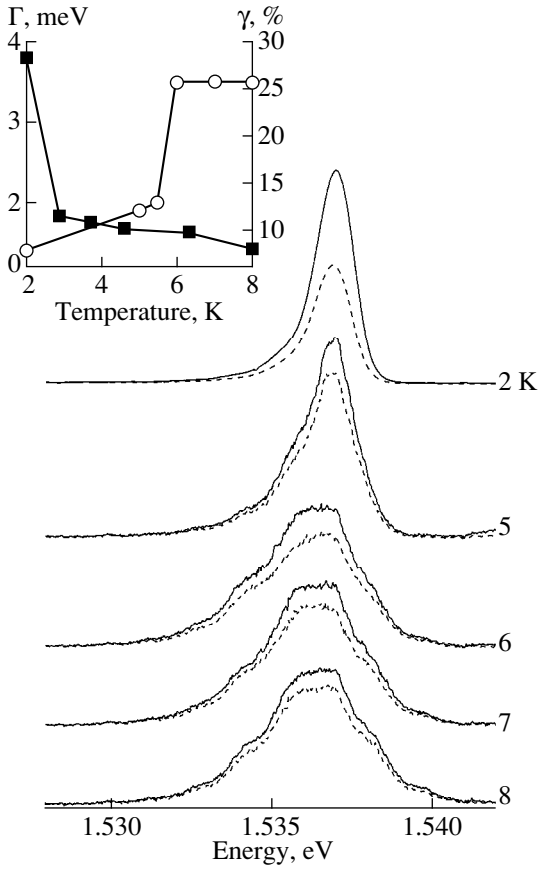


Fig. 5. Photoluminescence spectra of interwell excitons under resonant excitation by circularly polarized light, measured at various temperatures (numbers on the right of the spectra give the temperature) and optical pump power 3 W/cm^2. The solid curves give the photoluminescence spectra measured for the σ^+ polarization and the dashed curves give those for the σ^- polarization. The temperature dependences of the luminescence line width of the interwell excitons Γ and the degree of circular polarization γ are shown in the inset.

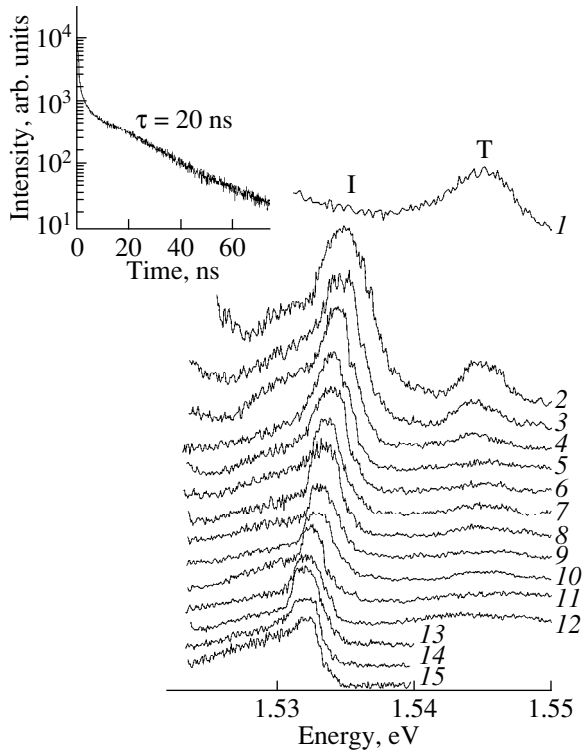


Fig. 6. Time evolution of the luminescence spectra of interwell excitons and intensity kinetics of the narrow line (see inset) under conditions of pulsed excitation and $T = 2$ K at –0.75 V. Spectrum *1* was measured with zero delay and an integration time of 1 ns; spectra *2–9* were measured with delays of 2, 3, 4, 5, 6, 7, 8, and 10 ns and an integration time of 2 ns; spectra *10–13* correspond to delays of 12, 16, 25, and 25 ns and integration times of 4, 4, 5, and 6 ns, respectively; spectra *14* and *15* correspond to delays of 30 and 40 ns and integration times of 10 and 12 ns, respectively.

time delays $\tau \geq 2$ ns. This delay is caused by carrier tunneling through the barrier (mainly electrons for which the effective mass in the vertical direction is much lower than the hole mass), accompanied by spatial separation of carriers between quantum wells, intrawell energy relaxation (thermalization) and simultaneous recombination. It can be seen from Fig. 6 that for small delays ($\tau = 2$–4 ns) the width of the interwell photoluminescence spectrum is large (3–4 meV). However, as the time delay increases, a fairly narrow line begins to form at the violet edge of the interwell photoluminescence spectrum. The width of this line is 1.5 meV for an 8 ns delay, i.e., the width is reduced almost threefold compared with that for the initial delays. The intensity of this line decays far more rapidly with time than the structureless interwell luminescence spectrum below it. For delays greater than 20 ns this line can no longer be resolved and merges with the structureless part of the spectrum whose profile remains almost unchanged and can be observed for delays greater than 50 ns. This behavior can be seen very clearly in the intensity kinetics measured directly at the spectral position of the narrow line and the structureless continuum below it (see inset to Fig. 6).

It is important to note that the narrow line of interwell-exciton photoluminescence characterized by short-lived intensity decay kinetics under pulsed excitation conditions is only observed at fairly low temperatures, as under cw excitation. For example, for a 12 ns delay the narrow line begins to become broader with increasing temperature and merges with the structureless background below it at $T \geq 6$ K. For shorter delays and therefore higher interwell exciton density this line disappears in the spectrum at significantly higher temperatures. For example, for a 7 ns delay the narrow line disappears in the spectra at $T \geq 9$ K. The interwell exciton lifetime can be measured directly from the intensity decay kinetics. In accordance with Fig. 6 (see inset), this time measured for a narrow photoluminescence line of delocalized interwell excitons at $T = 1.8$ K is 20 ns whereas the decay time of the broad photoluminescence band corresponding to localized exciton states is almost three times greater at 70 ns. This explains at least qualitatively why the degree of circular polarization of the narrow photoluminescence line of the delocalized excitons increases. Direct measurements of the intensity decay kinetics can be used to estimate the spin relaxation times in an interwell exciton. Using the formula given above for the degree of circular polarization and the measured lifetime of an interwell exciton at $T = 1.8$ K, we obtain for the spin relaxation time $\tau_s \sim 2 \times 10^{-8}$ s. The intensity decay kinetics of the narrow photoluminescence line are also sensitive to temperature and the corresponding decay time increases monotonically with decreasing temperature. For example, the decay time is $\tau = 20$ ns at $T = 1.8$ K whereas at $T = 5$ K we have $\tau = 10$ ns. The same qualitative behavior of the time-resolved spectra and the interwell photoluminescence kinetics was observed for applied voltages between –0.4 V and –0.9 V.

5. LUMINESCENCE SPECTRA OF INTERWELL EXCITONS IN A MAGNETIC FIELD

In this section we are interested in the diamagnetic properties of interwell excitons and their Zeeman splitting. These investigations were carried out at low excitation densities when the interwell excitons are strongly localized and their spectrum inhomogeneously broadened and at high pump densities when the photoluminescence line of the interwell recombination is substantially narrower. In these experiments we analyzed the Zeeman components σ^+ and σ^- of the interwell photoluminescence spectrum measured in a magnetic field perpendicular to the planes of the quantum wells (Faraday geometry).

Figure 7 shows the interwell photoluminescence spectra (σ^- component) measured when direct 1*s*HH excitons were resonantly excited at high power density in a magnetic field between 0 and 2 T at 0.1 T intervals.

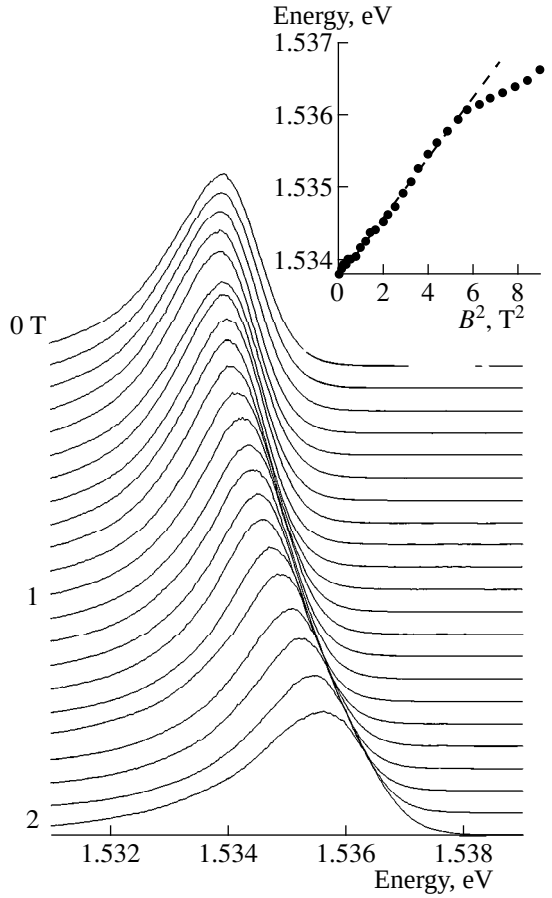


Fig. 7. Photoluminescence spectra of interwell excitons in a magnetic field perpendicular to the quantum-well plane and $T = 2$ K. These spectra were measured in the range 0–2 T at 0.1 T intervals. The spectra were measured using the σ^- polarization and pump power density 3 W/cm^2. The inset shows the diamagnetic shift of the delocalized interwell excitons.

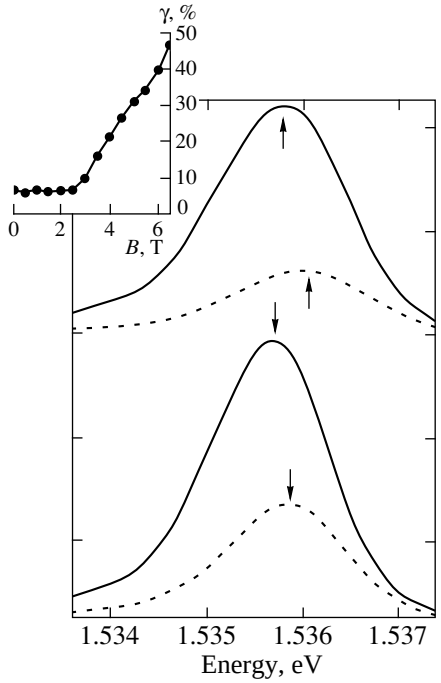


Fig. 8. Zeeman splitting of the photoluminescence line of delocalized interwell excitons at $B = 6.5$ T and $T = 2$ K. The solid and dashed curves correspond to the σ^- and σ^+ circular polarizations, respectively. The inset shows the intensity ratio of the Zeeman components as a function of the magnetic field.

The figure clearly shows a superlinear shift of the narrow photoluminescence line in the region of weak magnetic fields as B increases. The inset gives the spectral position of the maximum of the narrow interwell photoluminescence line as a function of the square of the magnetic field. The quadratic diamagnetic shift of this line is clearly satisfied in fields $B < 2$ T whereas in strong fields a linear contribution to the magnetic susceptibility begins to become significant [17]. Using a correction quadratic in B to the ground-state energy of an interwell exciton, we can estimate its Bohr radius in the quantum-well plane using the well-known formula:

$$\Delta E = \frac{e^2}{8\mu c^2} B^2 \langle a \rangle^2,$$

where $\langle a \rangle^2 = \pi a_B^2$ and μ is the reduced exciton mass ($m_e = 0.067 m_0$ and $m_h = 0.2 m_0$). The Bohr radius of an interwell exciton thus determined was $a_B = 170$ Å. From this we can obtain a lower estimate of the binding energy of an interwell exciton $E_{exc} = e^2/\epsilon a_B$ which was $E_{exc} \approx 3$ meV.

The narrow interwell-exciton line in a transverse magnetic field is split into a doublet, as in the case of an intrawell heavy-hole direct exciton. By way of example Fig. 8 shows the Zeeman splitting (σ^+ and σ^- components) of this line at $B = 6.5$ T. In this case unpolarized light was used for excitation (see lower part of Fig. 8). Figure 9 is a diagram of the allowed optical transitions predicted in Faraday geometry. The Zeeman splitting, which is 0.2 meV ($B = 6.5$ T) can be used to determine the effective g-factor of an interwell exciton $g_{exc} = 0.53$. The magnitude of the Zeeman splitting and the related effective exciton g-factor can also be determined using the intensity ratio of the corresponding σ^+ and σ^- components assuming that the spin states are uniformly populated and the electron temperature given. The independently determined exciton g-factor agreed to within 10% with the value obtained directly from the Zeeman splitting. We thus conclude that the distribution between the split spin states is quasi-equilibrium and also the temperature of the electron system is 0.2 K higher than the temperature of the helium bath at $T = 2$ K and pump power density 6 W/cm^2. Hence the optical pumping used experimentally does not strongly overheat the electron (exciton) system relative to the helium bath temperature and the lattice temperature.

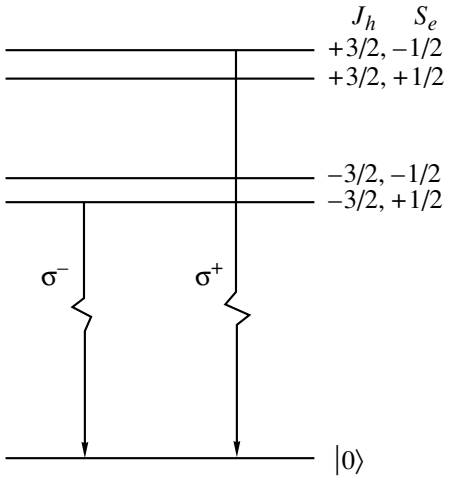


Fig. 9. Diagram of optical transitions under conditions of Zeeman splitting of the ground state of an interwell exciton.

The photoluminescence spectra shown in the upper part of Fig. 8 were measured under resonant excitation by circularly polarized light (σ^+ component). In this case, as can be seen from Fig. 8, the degree of circular polarization between the split Zeeman components increased more than 1.1 times compared with the photoluminescence spectra measured in the same magnetic field but excited by resonant unpolarized light. This experimental observation indicates that in a magnetic field under conditions of resonant photoexcitation by circularly polarized light the spin splitting in the exciton increases. In particular in a magnetic field B = 6.5 T this splitting increases 1.5 times and is 0.3 meV. The increase in the spin splitting in this case is naturally attributed to amplification of the effective electron g-factor:

$$\Delta E_{sp} = (|g_e^{\text{eff}}| + |g_h|)\mu_B B,$$

where the effective amplified electron g-factor is $g_e^{\text{eff}} = g_e^0 + (\Delta g)^{\text{eff}}$, and g_e^0 is the unperturbed ("bar") electron g-factor. The value thus determined in $(\Delta g)^{\text{eff}} = 0.27$ and the published value $g_e^0 = -0.44$. We attribute the amplification of the effective electron g-factor to alignment of the spins of the nuclear subsystem which occurs as a result of contact interaction between spin-oriented electrons excited by the circularly polarized light and nuclei of the dominant heterostructure material (in particular Ga nuclei, see for example [16]). In the presence of an external magnetic field this effect is significant since the field defines the preferential orientation of the nuclear spins. In the absence of an external magnetic field the orientations of the nuclear spins are random and the contributions of the various components compensate for each other.

We also investigated the photoluminescence spectra of interwell radiative recombination in a magnetic field at low excitation density when the interwell excitons are strongly localized at fluctuations of the random potential and the corresponding line in the spectrum is broad. In this case, the diamagnetic correction to the ground-state energy was determined from the dependence of the first moment M_1 of the photoluminescence spectrum of the interwell excitons on the square of the magnetic field ($M_1 = \int E\, I(E)dE / \int I(E)dE$). It was found that the diamagnetic correction to the energy for localized excitons is approximately 1.5 times this correction for delocalized excitons. The diamagnetic shift can be used to estimate the linear scales of lateral confinement associated with the random potential at which the interwell excitons are localized at low densities and low temperatures. This linear scale of localization was ~400 Å. The broad photoluminescence line of the localized interwell excitons in a magnetic field is also split into a Zeeman doublet. The doublet splitting and the exciton g-factor were estimated most accurately using the intensity ratio of the σ^+ and σ^- components. The value obtained for the g-factor was close to that for delocalized excitons.

6. BUILDUP OF INTERWELL EXCITONS IN A LATERAL POTENTIAL WELL FAR FROM THE PHOTOEXCITATION REGION

In previous sections we have discussed the spectra of interwell excitons measured under conditions where the luminescence was detected directly in the region of excitation on the mesa. In this section we discuss experiments in which interwell excitons accumulated in a lateral potential well far from the photoexcitation point. The interwell excitons were located in this potential well as a result of natural drift from the region of photoexcitation under the action of forces associated with the gradient of the strain potential. It was established that a potential well of strain origin appears in the quantum-well plane if a narrow (100 μm wide) metal (Au) stripe around 1 μ thick is deposited on the mesa surface of the structure. A substantial difference between the coefficients of thermal expansion of the metal stripe and the base material of the GaAs mesa gave rise to strain forces, which at liquid helium temperatures lead to nonuniform compression of the structure perpendicular to the quantum well plane. The largest strain occurred directly beneath the metal stripe and propagated inside the mesa. This strain was observed in particular in the fact that the luminescence line of the intrawell excitons measured directly beneath the metal stripe was spectrally shifted toward lower energies by around 1.5–2 meV as a result of mechanical compressive strain.

Experiments taking this into account were carried out as follows. An exciting laser spot smaller than 100 μm was focused near one of the sides of the metal stripe on the mesa surface. The mesa surface was projected onto the plane of crossed slits which could be used to monitor the "sampling" of the photoluminescence both directly from the region of excitation and at the opposite edge of the metal stripe, i.e., approximately 100 μ

from the exciting spot [regions (1) and (2) in Fig. 10 which shows the photoexcitation and detection conditions used experimentally]. These experiments were carried out using a GaAs/AlGaAs heterostructure (p–i–n structure) with a double quantum well (the width of the quantum well was 80 Å and the width of the AlGaAs barrier 40 Å). Figure 11 illustrates the behavior of the luminescence spectra at various pump powers when detected from region (2) some distance from the region of excitation. The luminescence from region (2) associated with interwell radiative recombination only became appreciable at high optical pump powers when the applied electric bias was strongly screened directly in the excitation zone (1) Thus, it can be seen from Fig. 11 that initially at low pump powers the interwell exciton line is strongly inhomogeneously broadened (its half-width is 4–5 meV) which is typical of low interwell exciton densities when these are localized at fluctuations of the random potential. With increasing pumping a narrow line begins to form at the violet edge of the photoluminescence spectrum. The intensity of this line increases superlinearly with increasing excitation power (see inset to Fig. 11) and the line itself narrows substantially, to a minimum width of 1.3 meV. The maximum of this narrow line is initially shifted toward lower energies (around 1–1.5 meV) and only at high pump powers does this line begin to broaden and shift into the violet as a result of screening of the applied electric field. This narrowing of the narrow photoluminescence line of the interwell excitons and superlinear increase in its intensity is observed when detected from region (2) at various applied bias voltages. Figure 12 illustrates the linear Stark shift of this line when the electric field is varied, which irrefutably indicates its interwell exciton nature. Direct measurements of the decay kinetics of the photoluminescence intensity made under pulsed excitation at various temperatures demonstrated (see Fig. 13) that the lifetime of delocalized interwell excitons at low temperatures, $T \leq 6$ K, i.e., when the corresponding photoluminescence line is narrow, is several times shorter than the lifetime of the localized interwell excitons (low pump powers or temperatures $T \geq 6$K).

These experiments have therefore demonstrated that the strong narrowing of the interwell radiative recombination line as the interwell exciton density increases is of a general nature in the presence of lateral confinement, either random (as a result of fluctuations of the random potential) or artificially created (accumulation of interwell excitons in a strain potential well far from the region of photoexcitation).

7. DISCUSSION OF EXPERIMENTAL RESULTS AND CONCLUSIONS

When we attempt to explain the complete set of experimental results presented above in terms of a simple single-particle picture of radiative annihilation of interwell excitons localized at fluctuations of the random potential, we encounter major difficulties. If the

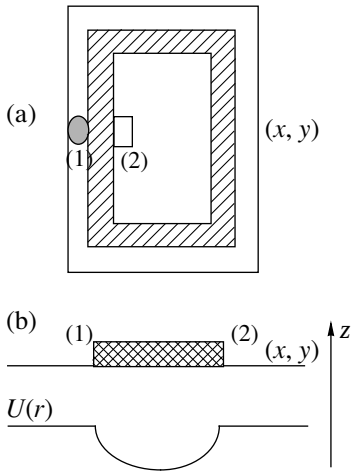


Fig. 10. Diagram of experiments to detect photoluminescence of interwell excitons far from the region of excitation. (a) Top view, metal stripe on mesa shown hatched, width of metal stripe 100 μm, (*1*) pump region, (*2*) region from which photoluminescence detected. (b) Qualitative picture of the deformation potential $U(r)$ beneath metal frame. The z direction is perpendicular to the quantum well plane (x, y).

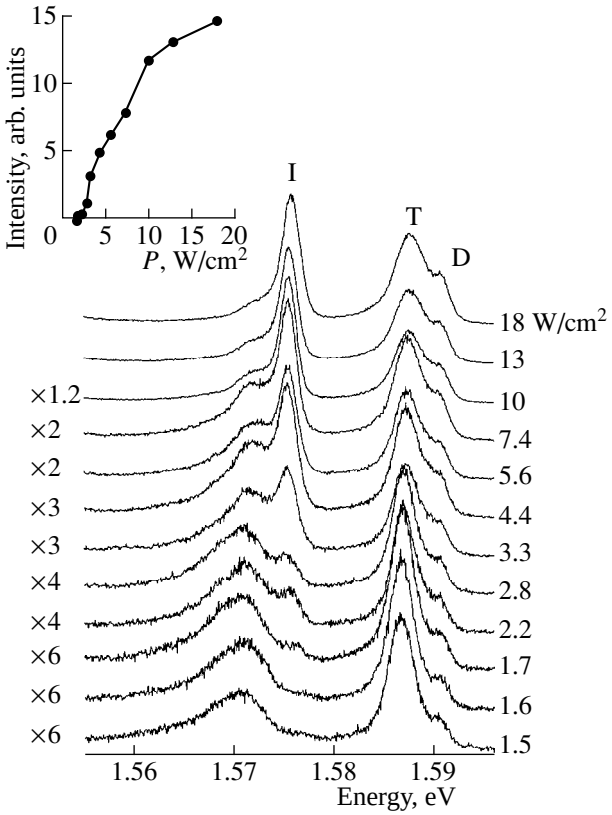


Fig. 11. Photoluminescence spectra of interwell I and intrawell D and T excitons at various pump powers, detected from region (2) on the mesa as shown in the experimental setup in Fig. 10. The numbers on the right of the spectra give the excitation power density and those on the left give the scale factors for the intensities of the corresponding spectra $T = 2$ K. The inset gives the intensity of the narrow interwell-exciton line as a function of the pump power density.

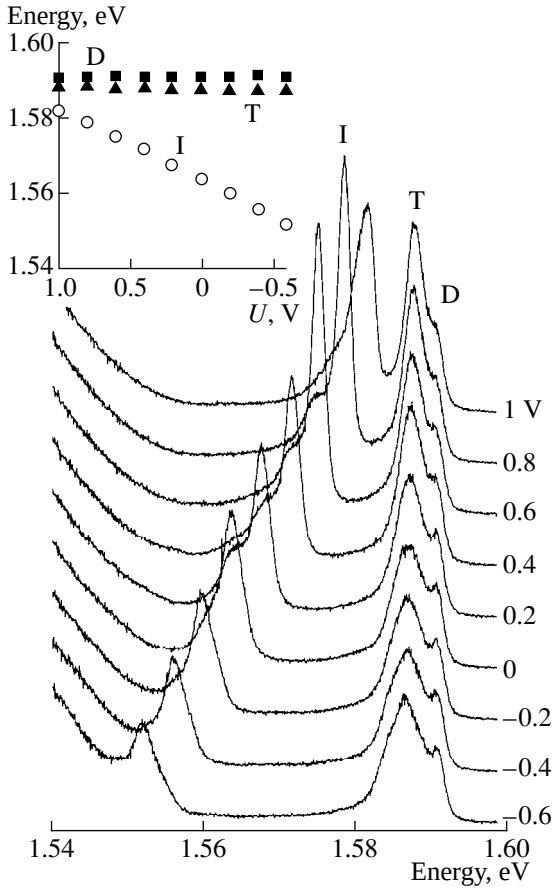


Fig. 12. Photoluminescence spectra of interwell excitons at various voltages: I, D, and T are the photoluminescence lines of the interwell excitons, direct intrawell excitons, and exciton charged complexes, respectively. The spectral position of the photoluminescence lines as a function of the electric field is illustrated in the inset top left.

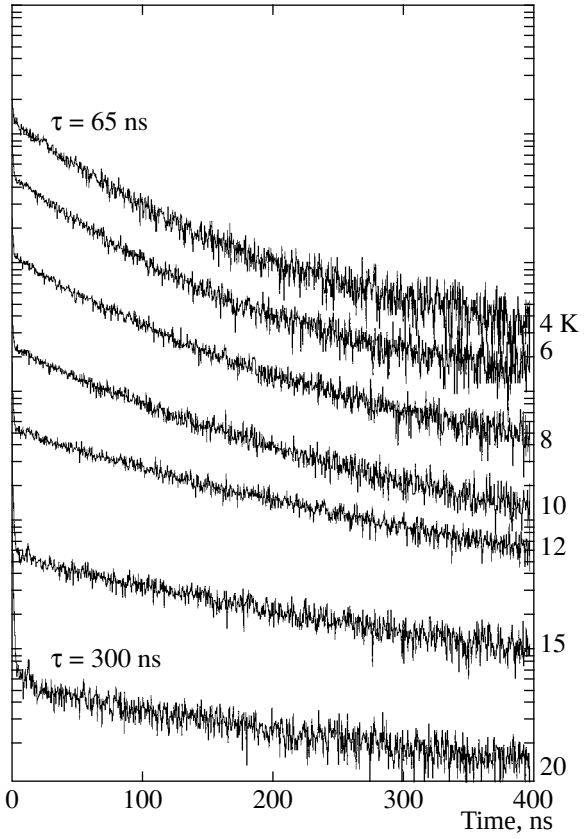


Fig. 13. Photoluminescence kinetics of the narrow interwell-exciton line at various temperatures (numbers on the right of the spectra).

appearance of a narrow line in the photoluminescence spectra is attributed to interwell delocalized excitons, which can appear above the percolation threshold as a result of screening of the random potential, it is difficult to explain why this effect is so temperature-critical and does not occur when $T > T_c$. The threshold increase in the degree of circular polarization and the alignment of the interwell excitons as their concentration increases are also completely unexplained. However, these results can be explained at least qualitatively as a consequence of the collective behavior of delocalized interwell excitons when a critical concentration and temperature are reached. It can be postulated these structures with narrow AlAs barriers exhibit large-scale fluctuations of the potential caused in particular by variations of the barrier width. These potential fluctuations are poorly screened, unlike the random potential, because of residual charge impurities which are also present in these structures. As a result of large-scale fluctuations of the potential relief in the quantum well plane photoexcited interwell excitons may accumulate in macroscopically extended regions with lateral confinement as far as several tenths of micron. In fact, variations of the effective lateral potential $U(r) = U(w(r))$ may be linked to variations of the quantum well width $w(r)$. Under quasiequilibrium conditions the lateral distribution of the exciton density will be determined by the equality $\mu(n(r)) + U(r) = \mu$, where μ is the chemical potential of the interwell excitons, which is related to their average density in the quantum wells, and $\mu(n)$ is the chemical potential of the homogeneous dielectric exciton phase in the lateral confinement region. Quite clearly $|\mu(r)| < |\mu|$ since $\mu(n) = -|E_{exc}| + |\delta U|$ (E_{exc} is the exciton binding energy) and the exciton density in the lateral confinement region may be substantially higher than the average density in the quantum well plane generated by the optical pumping. These are the regions where the main events associated with photoexcited interwell excitons take place. These events may evolve according to different scenarios but each is based on the assumption of collective interaction in a system of interwell excitons delocalized within macroscopically large lateral regions. In one scenario it could be postulated that on reaching critical densities and temperature the interwell excitons

condense to form a metallic *e–h* liquid with spatially separate electron–hole layers. If the narrow line in the spectrum is related to the metallic liquid phase, an upper estimate of its density can be obtained from the line width which, in this case, should be not less than the sum of the Fermi energies of the electrons and holes. Thus, the estimated density is 3×10^{10} cm^{-2} and the dimensionless parameter corresponding to this density is $r_s = 1/\sqrt{\pi n a_B^2} = 4$. Quite clearly the condensed phase in this approach is too "friable" to consider the Coulomb interaction in the interwell exciton completely screened. Hence at these concentrations excitons should retain their individuality so that the condensed phase cannot be metallic. This conclusion is also supported by studies of the photoluminescence spectra in a magnetic field in Faraday geometry. It was established (Section 5) that the narrow line, like a free exciton, splits into a Zeeman doublet with the intensity ratio of the σ^+ and σ^- components corresponding to the temperature and spin splitting, i.e., the individual properties of the exciton are completely conserved in the condensed phase.

We shall assume that strong narrowing of the interwell-exciton photoluminescence line and the sensitivity of this effect to density and temperature may be associated with the condensation of interwell excitons to form a collective dielectric exciton phase. It was shown in [14] that for specific values of the interwell-exciton dipole moment a fairly dense system of interwell excitons may condense to form a dielectric liquid despite the dipole–dipole repulsion between these excitons. It was also shown in [11] that when critical conditions are satisfied the collective phase of the interwell excitons is most likely to occur in regions with lateral confinement. According to our experiments using cw excitation such condensation occurs at $T < 5.5$ K and an average exciton concentration of 3×10^{10} cm^{-2}. Moreover, the narrowing of the photoluminescence line of the interwell excitons as their density increases at low temperature is accompanied by a systematic shift of this line toward lower energies (approximately by 1.5 meV). This behavior may be demonstrated by a fairly dense system of Bose quasiparticles having integer spin where $|\mu|/kT \longrightarrow 0$. However, the possibility of Bose–Einstein condensation of an exciton gas in systems of reduced dimensions is a very delicate and difficult problem [18]. Nevertheless, these observations serve as an independent argument in support of the assumption that in this particular case, we are dealing with a fairly dense dielectric collective phase in interwell excitons.

The condensed component of the excitons should be in phase within the coherence length. Spatial coherence should occur at least on scales of the de Broglie wavelength of an interwell exciton, λ_{ex} which at $T = 2$ K is $\lambda_{ex} = h/\sqrt{\pi m k T} = 1.5 \times 10^3$ Å and is more than an order of magnitude greater than the exciton Bohr radius ($a_B \sim 100$ Å). The exciton density under these conditions corresponds to the dimensionless parameter $r = n\lambda_{ex}^2 = 4$. Increased spatial coherence in the condensed phase is indicated by the observed threshold increase in the alignment of interwell excitons under conditions of resonant exciton by linearly polarized light. This alignment effect is directly related to the transverse relaxation and thus to the time of loss of phase coherence. A substantial increase in the alignment of interwell excitons which takes place as a threshold process may imply that the resultant collective exciton state has fairly long phase shift times (according to our estimates this time is around 1 ns at $T = 2$ K). Then the radiative decay of phase-correlated interwell excitons in the condensate should have significantly higher radiative probabilities compared with the photoluminescence of the uncondensed excitons. This conclusion also agrees with the experiment.

The assumption that interwell excitons condense to form a dielectric collective phase is nevertheless based on the unanswered important question concerning the spatial coherence scales of this collective state. This question may be answered using measurements of the photoluminescence intensity correlations under conditions when the postulated exciton condensation occurs.

ACKNOWLEDGMENTS

In conclusion, the authors thank S. Iordanskiĭ, A.S. Ioselevich, V.D. Kulakovskiĭ, R.E. Lozovik, V.G. Lysenko, and R.A. Suris for interesting discussions. This work was partly financed by the Russian Foundation for Basic Research (project no. 98-02-16656) and the Interdepartmental Program "Nanostructures."

REFERENCES

1. Yu. E. Lozovik and V. I. Yudson, Zh. Éksp. Teor. Fiz. **71**, 738 (1976) [Sov. Phys. JETP **44**, 389 (1976)].
2. T. Fukuzawa, E. E. Méndez, and J. M. Hong, Phys. Rev. Lett. **64**, 3066 (1990).
3. J. E. Golub, K. Kash, J. P. Harbison, and L. T. Flores, Phys. Rev. B **41**, 8564 (1990).
4. J. A. Kash, M. Zachau, E. E. Méndez, *et al.*, Phys. Rev. Lett. **66**, 2247 (1991).
5. L. V. Butov, A. Zrenner, G. A. Abstreiter, *et al.*, Phys. Rev. Lett. **73**, 304 (1994); L. V. Butov, in *Proceedings of the 23rd International Conference on Physics of Semiconductors, Berlin, 1996.*
6. V. B. Timofeev, A. I. Filin, A. V. Larionov, *et al.*, Europhys. Lett. **41**, 435 (1998).
7. V. B. Timofeev, A. V. Larionov, A. S. Ioselevich, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 580 (1998) [JETP Lett. **67**, 613 (1998)].
8. V. V. Krivolapchuk, E. S. Moskalenko, A. L. Zhmodikov, *et al.*, Solid State Commun. **111**, 49 (1999).

9. D. Yoshioka and A. H. MacDonald, J. Phys. Soc. Jpn. **59**, 4211 (1990).
10. X. M. Chen and J. J. Quinn, Phys. Rev. Lett. **67**, 895 (1991).
11. Xuejun Zhu, P. B. Littlewood, M. S. Hybersten, and T. Rice, Phys. Rev. Lett. **74**, 1633 (1995).
12. J. Fernández-Rossier and C. Tejedor, Phys. Rev. Lett. **78**, 4809 (1997).
13. Lerwen Liu, L. Swierkowski, and D. Nelson, Physica B (Amsterdam) **249–251**, 594 (1998).
14. Yu. E. Lozovik and O. L. Berman, Zh. Éksp. Teor. Fiz. **111**, 1879 (1997) [JETP **84**, 1027 (1997)].
15. V. B. Timofeev, A. V. Larionov, M. Grassi Alessi, *et al.*, Phys. Rev. B **60**, 8897 (1999).
16. *Optical Orientation*, *Modern Problem in Condensed Matter Sciences*, Ed. by F. Mayer and B. Zhakharchenya (Elsevier, Amsterdam, 1984), Vol. 8.
17. G. E. W. Bauer and T. Ando, Phys. Rev. B **37**, 3130 (1988); **38**, 6015 (1988).
18. J. M. Kosterlitz and D. J. Thouless, J. Phys. C **6**, 181 (1973).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 1105–1113.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1270–1279.

NONLINEAR PHYSICS

Stochastic Kadomtsev–Petviashvili Equation

V. A. Gorodtsov

Institute of Problems of Mechanics, Russian Academy of Sciences, Moscow, 117526 Russia
e-mail: teodor@ipmnet.ru

Received January 10, 2000

Abstract—The example of Kadomtsev–Petviashvili equations with a random time-dependent force (stochastic Kadomtsev–Petviashvili equations) is used to show that the theory of Brownian particle motion can be applied to the theory of the stochastic behavior of solitons of model hydrodynamic equations which are completely integrable in the absence of forces and interrelated by the generalized Galilean transformation. The Brownian motion of two-dimensional algebraic solitons of the Kadomtsev–Petviashvili equations with positive dispersion leads to their diffusion broadening similar to the broadening of one-dimensional solitons of other fully integrable hydrodynamic equations. However, for longer times the rate of decay of algebraic solitons is higher because of the degeneracy of the momentum integral for these solitons. The behavior of a periodic chain of algebraic solitons is established under the action of a random force. Tilted plane solitons of the Kadomtsev–Petviashvili equations with negative dispersion vary under the action of a random force similar to the solitons of the Korteweg–de Vries equation. Several of these solitons interact via "virtual solitons" and generate new solitons provided that resonance conditions are satisfied whose dimensions increase as a result of the influence of the random force. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The incompleteness of our knowledge of physical processes frequently compels us to make statistical analyses. However, the difficulties involved in making a detailed quantitative analysis are quite significant for nonlinear systems having an infinite number of degrees of freedom, typical of the approximation of a continuous medium. Recent success in solving nonlinear partial differential equations (integro-differential and differential-difference) for various media and dynamical processes has involved developing a representative class of completely integrable nonlinear equations and various methods of completely solving these (such as the method of the inverse scattering problem) and formulating the basic "soliton" concept. In some cases, stochastic equations can also be solved exactly if they are constructed using regular completely integrable equations with the addition of terms having random coefficient functions.

Analyses of stochastic evolution equations with a random force which depends only on a single time variable have also proved successful. The addition of this randomness to nonlinearity and dispersion is in fact equivalent to allowing for a diffusion type of dissipativeness which has already been demonstrated for spatially one-dimensional examples of the stochastic Korteweg–de Vries (KdV) equation [1, 2], the stochastic nonlinear Schrödinger equation [3], the stochastic Benjamin–Ono equation [4, 5], and the stochastic sine-Gordon equation [6]. In the present study, a similar analysis is made of the stochastic spatially two-dimensional Korteweg–de Vries equation, known as the Kadomtsev–Petviashvili equation. This is just as universal as the one-dimensional KdV equation. In cases of positive dispersion, the solutions of the Kadomtsev–Petviashvili equation in the form of plane solitons (KdV solitons) are unstable [7–11] and ultimately decay into a set of stable two-dimensionally localized solitons exhibiting power-law spatial decay [11]. The Brownian motion of these "algebraic solitons" and the stochastic spreading of their periodic sets under the influence of a random force will be discussed subsequently. In cases of negative dispersion, solutions of the Kadomtsev–Petviashvili equation in the form of tilted plane solitons whose interaction zone increases in size under the action of a random force become stable.

2. INHOMOGENEOUS KADOMTSEV–PETVIASHVILI EQUATIONS. PERTURBATION SPREADING UNDER THE ACTION OF RANDOM FORCES

When allowance is made for the action of an external time-dependent force, the two-dimensional generalizations of the Korteweg–de Vries equation are written, as in the original study [7], in the form of the system

$$\partial_t u + 2u\partial_x u + \partial_x^3 u = \partial_y \varphi + f(t), \quad \partial_x \varphi = \pm \partial_y u, \quad u = u(x, y, t), \quad \varphi = \varphi(x, y, t). \tag{2.1}$$

Using a simple change of variables which in fact involves going over to a coordinate system accelerating in the direction of the x-axis,

$$\tilde{x} = x - x_0(t), \quad \tilde{u}(\tilde{x}, y, t) = u(x, y, t) - u_0(t),$$

1063-7761/00/9006-1105$20.00 © 2000 MAIK "Nauka/Interperiodica"

$$\tilde{\varphi}(\tilde{x}, y, t) = \varphi(x, y, t),$$
$$\partial_t u_0(t) = f(t), \quad \partial_t x_0(t) = 2u_0(t),$$

the equations are reduced to homogeneous, i.e., ordinary Kadomtsev–Petviashvili equations with positive (KP1) or negative (KP2) dispersion for the corresponding sign in the second equation. As we well know [12, 13], these equations are completely integrable and differ primarily in respect of the stability of their solutions. The simplicity of this change of variables means that by using two integrations we can make this transformation explicitly for an arbitrary external force and, in particular, for each separate realization of the random process $f(t)$ in the case of a stochastic problem.

Thus, the problem of solving the stochastic Kadomtsev–Petviashvili equations with a random force which depends only on time can be divided into solving (for example, an initial) the problem for the deterministic homogeneous Kadomtsev–Petviashvili equations (the solution of the homogeneous equations will be denoted by v)

$$\partial_t v + 2v\partial_{\tilde{x}} v + \partial_{\tilde{x}}^3 v = \partial_y \varphi, \quad \partial_{\tilde{x}} \varphi = \pm \partial_y v, \quad (2.2)$$

a nonlinear change of variables of a purely algebraic nature which can also be conveniently written for the following analysis in terms of the shift operator:

$$\begin{aligned} u(x, y, t) &= u_0(t) + v(x - x_0(t), y, t) \\ &= u_0(t) + \exp(-x_0(t)\partial_x) v(x, y, t), \end{aligned} \quad (2.3)$$

and a stochastic problem (Langevin problem) for ordinary differential equations with a random force:

$$\begin{aligned} &\partial_t u_0(t) = f(t), \quad \partial_t x_0(t) = 2u_0(t), \\ &\langle f(t) \rangle = 0, \quad \langle f(t) f(t') \rangle = f_0^2 \delta(t - t'). \end{aligned} \quad (2.4)$$

For simplicity here and subsequently, we shall confine our analysis to Gaussian white noise with a zero average force. On account of the linearity of the relationships of the random displacements $x_0(t)$ and velocity $u_0(t)$ with the Gaussian random force function

$$u_0(t) = \int_0^t dt' f(t'), \quad x_0(t) = 2\int_0^t dt' f(t')(t - t'), \quad (2.5)$$

these are also Gaussian random functions with zero averages and the following simultaneous second moments:

$$\begin{aligned} &\langle x_0(t) \rangle = 0, \quad \langle u_0(t) \rangle = 0, \quad \langle u_0^2(t) \rangle = f_0^2 t, \\ &\langle u_0(t) x_0(t) \rangle = f_0^2 t^2, \quad \langle x_0^2(t) \rangle = \frac{4}{3} f_0^2 t^3. \end{aligned} \quad (2.6)$$

Higher moments can easily be expressed in terms of these. For example, we have

$$\langle x_0^n(t) \rangle = (2m - 1)!! \langle x_0^2(t) \rangle^m \delta_{n, 2m}.$$

By summing the results for these power expressions, we obtain a formula for averaging the exponential function:

$$\begin{aligned} \langle \exp\{a x_0\} \rangle &= \sum_{n=0}^{\infty} \frac{a^n}{n!} \langle x_0^n \rangle \\ &= \sum_{m=0}^{\infty} \frac{(2m-1)!!}{2m!} (a^2 \langle x_0^2 \rangle)^m = \sum_{m=0}^{\infty} \frac{a^{2m}}{2^m m!} \langle x_0^2 \rangle^m, \\ \langle \exp\{a x_0(t)\} \rangle &= \exp\left\{ \frac{1}{2} a^2 \langle x_0^2(t) \rangle \right\}, \end{aligned} \quad (2.7)$$

which can be used to find the characteristic function (the generating function of the single-point moments) of the random displacement process $x_0(t)$ simply by substituting $a = ik_0$ and using (2.6):

$$\begin{aligned} \varphi_t^{(x)}(k_0) &\equiv \langle \exp\{ik_0 x_0(t)\} \rangle \\ &= \exp\left\{ -\frac{k_0^2}{2} \langle x_0^2(t) \rangle \right\} = \exp\left\{ -\frac{2}{3} f_0^2 k_0^2 t^3 \right\}. \end{aligned} \quad (2.8)$$

The most comprehensive information on a random process is provided by its characteristic functional [14, 15]. For a delta-correlated Gaussian random force with zero mean value, this functional can easily be obtained explicitly [the calculations are exactly the same as those used to derive (2.7) and (2.8)]:

$$\begin{aligned} \Phi^{(f)}[k(t)] &\equiv \left\langle \exp\left\{ i \int_0^{\infty} dt k(t) f(t) \right\} \right\rangle \\ &= \exp\left\{ -\frac{f_0^2}{2} \int_0^{\infty} dt k^2(t) \right\}. \end{aligned} \quad (2.9)$$

As a result of the simple linear relationship between the force and the displacements (2.5), this can be used to express the characteristic displacement functional:

$$\begin{aligned} \Phi^{(x)}[k(t)] &\equiv \left\langle \exp\left\{ i \int_0^{\infty} dt k(t) x_0(t) \right\} \right\rangle \\ &= \Phi^{(f)}\left[2\int_t^{\infty} d\tau k(\tau)(\tau - t) \right] \\ &= \exp\left\{ -\frac{2}{3} f_0^2 \int_0^{\infty} dt_1 \int_0^{t_1} dt_2 k(t_1) k(t_2)(3t_1 t_2^2 - t_2^3) \right\}. \end{aligned} \quad (2.10)$$

Then by substituting $k(t) = k_0 \delta(\tau - t)$, we obtain a formula for the characteristic displacement functional (2.8).

In addition to the random displacements $x_0(t)$, the random velocities $u_0(t)$ are also involved in the transfor-

mation of the inhomogeneous equation into a homogeneous one using the substitution (2.3) and thus it is frequently more convenient to use joint distributions. Following the reasoning put forward above, for the joint characteristic function at a given time we find

$$\varphi_t(k_0, p_0) \equiv \langle \exp\{ik_0x_0(t) + ip_0u_0(t)\}\rangle$$
$$= \exp\left\{-f_0^2\left(\frac{2}{3}k_0^2t^3 + k_0p_0t^2 + \frac{1}{2}p_0^2t\right)\right\}.$$

For the joint characteristic functional, we have

$$\Phi[k(t), p(t)] \equiv \left\langle \exp\left\{i\int_0^\infty dt[k(t)x_0(t) + p(t)u_0(t)]\right\}\right\rangle$$
$$= \exp\left\{-f_0^2\int_0^\infty dt_1\int_0^{t_1} dt_2 \times\left[k(t_1)k(t_2)\frac{2}{3}(3t_1t_2^2 - t_2^3) + p(t_1)p(t_2)t_2\right]\right\}$$
$$\times \exp\left\{-f_0^2\int_0^\infty dt_1\int_0^{t_1} dt_2 \times [k(t_1)p(t_2)2t_1t_2^2 + (k(t_2)p(t_1) - k(t_1)p(t_2))t_2^2]\right\}.$$

The two-time correlation functions of the displacements and velocities

$$\begin{aligned}\langle u_0(t_1)u_0(t_2)\rangle &= f_0^2[t_1H(t_2 - t_1) + t_2H(t_1 - t_2)],\\ \langle x_0(t_1)u_0(t_2)\rangle &= f_0^2[t_1^2H(t_2 - t_1) + (2t_1t_2 - t_2^2)H(t_1 - t_2)],\\ \langle x_0(t_1)x_0(t_2)\rangle &= \frac{2}{3}f_0^2[(3t_2t_1^2 - t_1^3)H(t_2 - t_1)\\ &\quad + (3t_1t_2^2 - t_2^3)H(t_1 - t_2)]\end{aligned} \tag{2.11}$$

correspond to the second variational derivatives of this functional.

We can then derive an equation for the probability density of the joint distribution of the random quantities $x_0(t)$ and $u_0(t)$, having taken the time derivative of the random function

$$\pi \equiv \delta(x_0 - x_0(t))\delta(u_0 - u_0(t)),$$

using (2.4), and then averaging over random forces. This yields a coupling relationship between the two types of averaged functions $p(x_0, u_0, t) \equiv \langle\pi\rangle$ and $\langle f(t)\pi\rangle$:

$$\frac{\partial p}{\partial t} + 2u_0\frac{\partial p}{\partial x_0} + \frac{\partial}{\partial u_0}\langle f(t)\pi\rangle = 0.$$

Then one of these can be expressed in terms of the other using a formula valid for Gaussian random fields with zero averages and proven by analogy with (2.7):

$$\begin{aligned}\langle f(t)\exp w(t)\rangle &= \langle f(t)w(t)\rangle\exp\frac{\langle w^2(t)\rangle}{2}\\ &= \langle f(t)w(t)\rangle\langle\exp w(t)\rangle.\end{aligned} \tag{2.12}$$

Taking this into account and the relationships

$$\langle f(t)x_0(t)\rangle = 0, \quad \langle f(t)u_0(t)\rangle = f_0^2/2,$$

derived from (2.4), we finally have

$$\langle f(t)\pi(x_0, u_0, t)\rangle$$
$$= \left\langle f(t)\exp\left\{-x_0(t)\frac{\partial}{\partial x_0} - u_0(t)\frac{\partial}{\partial u_0}\right\}\right\rangle\delta(x_0)\delta(u_0)$$
$$= -\frac{f_0^2}{2}\frac{\partial}{\partial u_0}\langle\pi(x_0, u_0, t)\rangle.$$

We thus obtain a closed Fokker–Planck equation for the probability density distribution of the random displacements and velocities:

$$\frac{\partial p}{\partial t} + 2u_0\frac{\partial p}{\partial x_0} - \frac{f_0^2}{2}\frac{\partial^2 p}{\partial u_0^2} = 0,$$
$$p(x_0, u_0, t)\big|_{t=0} = \delta(x_0)\delta(u_0),$$

with a simple solution of the initial problem in the form of the two-dimensional Gaussian distribution

$$p(x_0, u_0, t) = \frac{\sqrt{3}}{2\pi f_0^2t^2}\exp\left\{-\frac{3x_0^2 - 6x_0u_0t + 4u_0^2t^2}{2f_0^2t^3}\right\}, \tag{2.13}$$

whose Fourier transformation with respect to the first two arguments is the characteristic function given earlier. By integrating over u_0 or x_0, we can then obtain the one-dimensional Gaussian distributions of the displacements or velocities:

$$\begin{aligned}p(x_0, t) &= \sqrt{\frac{3}{8\pi f_0^2t^3}}\exp\left(-\frac{3x_0^2}{8f_0^2t^3}\right),\\ p(u_0, t) &= \sqrt{\frac{1}{2\pi f_0^2t}}\exp\left(-\frac{u_0^2}{2f_0^2t}\right).\end{aligned} \tag{2.14}$$

Averaging the relationship between the solutions of the homogeneous and inhomogeneous equations (2.3) using the formula for averaging the exponential func-

tion (2.7) yields an important relation for the average flow rate of the medium:

$$\langle u(x, y, t)\rangle = \exp(\tau\partial_x^2)v(x, y, t),$$
$$\tau \equiv \frac{\langle x_0^2(t)\rangle}{2} = \frac{2}{3}f_0^2t^3, \tag{2.15}$$

in which the time variations associated with the regular and stochastic evolutions are separated. Differentiating this with respect to time, we obtain

$$\partial_t\langle u(x, y, t)\rangle$$
$$= D(t)\partial_x^2\langle u(x, y, t)\rangle + \exp(\tau\partial_x^2)\partial_t v(x, y, t), \tag{2.16}$$
$$D(t) \equiv \partial_t\langle x_0^2(t)\rangle/2 = 2f_0^2t^2,$$

which clearly demonstrates the diffusion role of the random forces. They give rise to diffusion spreading (as a result of the Brownian motion of the "center of gravity") of initially regular distributions whose rate increases with time as a result of the variability of the diffusion coefficient. In the particular cases of steady-state regular solutions of the homogeneous Kadomtsev–Petviashvili equations in the form of isolated or periodic waves $V(x', t')$, apart from transport, only diffusion spreading remains:

$$(\partial_t + \mathbf{c}\cdot\nabla)\langle u(x, y, t)\rangle = D\partial_x^2\langle u(x, y, t)\rangle,$$
$$\langle u(x', y', t)\rangle|_{t=0} = V(x', y'),$$
$$\langle u(x, y, t)\rangle = \exp(\tau\partial_x^2)V(x', y')$$
$$= \frac{1}{2\pi}\int d\tilde{k}\tilde{V}(k, y')\exp(-k^2\tau)\exp(ikx') \tag{2.17}$$
$$= \int dsV(s, y')g(x' - s, \tau),$$
$$x' \equiv x - c_xt, \quad y' \equiv y - c_yt, \quad g(x, \tau) = \frac{\exp(-x^2/4\tau)}{2\sqrt{\pi\tau}}.$$

Here, in addition to the exponential operator expression, other convenient representations are also given for the averaged solution in the form of a Fourier expansion and an integral convolution of the steady-state solution with the Green's function of the thermal conductivity operator $\partial_\tau - \partial_{x'}^2$. The self-similar time variation of the characteristic spatial scale in the Green's function leads to diffusion dispersal of the characteristic points of the average distribution (extremum, inflection, and zero points) according to the law $\propto t^{3/2}$.

We can then continue to make a similar derivation of relations for the moment and correlation characteristics of this stochastic flow field. On the basis of the expansion (2.3) and using relation (2.7), we obtain for the dispersion (second-order moment) of the velocity field u

$$\langle(u(x, y, t) - \langle u(x, y, t)\rangle)^2\rangle = \langle u_0^2(t)\rangle$$
$$- 2\langle x_0(t)u_0(t)\rangle\partial_x\langle u\rangle$$
$$+ \langle(v(x - x_0(t), y, t) - \langle v(x - x_0(t), y, t)\rangle)^2\rangle \tag{2.18}$$
$$= f_0^2t - 2f_0^2t^2\partial_x\langle u\rangle + \langle v^2(x - x_0(t), y, t)\rangle$$
$$- \langle v(x - x_0(t), y, t)\rangle^2.$$

Consequently, the dispersion of the flow field is made up of the dispersion of the background flow accelerated by a random force proportional to the correlation between the background velocity pulsations and the random displacements, and the dispersion of the velocity field v caused by the displacement dispersion. This last relationship is clearly illustrated by the formula for the simultaneous two-point correlation [this formula is obtained using (2.7):

$$\langle v(x_1 - x_0(t), y, t)v(x_2 - x_0(t), y, t)\rangle$$
$$= \exp[\tau(\partial_{x_1} + \partial_{x_2})^2]v(x_1, y, t)v(x_2, y, t)$$
$$= \sum_{n=0}^{\infty}\frac{(2\tau)^n}{n!}\frac{\partial^n\langle u(x_1, y, t)\rangle}{\partial x_1^n}\frac{\partial^n\langle u(x_2, y, t)\rangle}{\partial x_2^n},$$

which is even simpler in the single-point limit:

$$\langle v^2(x - x_0(t), y, t)\rangle = \exp[\tau\partial_x^2]v^2(x, y, t). \tag{2.19}$$

Here, by complete analogy with the relation for the average velocity (2.15), the stochastic and regular variations are factorized and thus differential corollaries of the type (2.16) and (2.17) are also valid. In the particular case of a regular solution of the Kadomtsev–Petviashvili equations in the form of a steady-state moving distribution $v(x, y, t) = V(x', y')$, we arrive at the simple problem of diffusion spreading of this quadratic characteristic of the field v:

$$(\partial_\tau - \partial_{x'}^2)\langle V^2(x' - x_0(t), y')\rangle = 0,$$
$$\langle V^2(x' - x_0(t), y')\rangle\big|_{\tau=0} = V^2(x', y'). \tag{2.20}$$

The behavior of the higher moment and correlation characteristics can be estimated without any particular difficulty. For example, again using (2.7) we transform the multipoint correlation

$$\langle v(x_1 - x_0(t), y, t)\ldots v(x_n - x_0(t), y, t)\rangle$$
$$= \langle\exp[-x_0(t)(\partial_{x_1} + \ldots + \partial_{x_n})]\rangle v(x_1, y, t)\ldots v(x_n, y, t)$$
$$= \langle\exp[\tau(\partial_{x_1} + \ldots + \partial_{x_n})^2]\rangle v(x_1, y, t)\ldots v(x_n, y, t).$$

Then, by going to the limit of the single-point moment which in fact corresponds to using the Leibnitz differentiation rule, we arrive at a simple generalization of

the previous result on the diffusion behavior of the moments:

$$\langle V^n(x - x_0(t), y, t)\rangle = \exp[\tau\partial_x^2] V^n(x, y, t). \quad (2.21)$$

In particular, for the higher moments, the relationship with the diffusion problem (2.20) becomes obvious.

3. ALGEBRAIC SOLITON UNDER THE ACTION OF A RANDOM FORCE

A stable algebraic soliton comprising a nonsingular two-parameter solution of the homogeneous KP1 equation [12, 16–18] may be represented as follows:

$$V(x', y') = s(X, Y) \equiv \frac{6}{(Y + iX)^2} + \text{c.c.},$$
$$\tilde{s}(k, Y) = 12\pi|k|\exp(-|k|Y),$$
$$X \equiv x' + py', \quad Y \equiv \sqrt{g^2 y'^2 + \frac{3}{q^2}}, \quad (3.1)$$
$$c_x = p^2 + q^2, \quad c_y = -2p.$$

Also given here is an expression for the Fourier transformation with respect to the first argument. Important characteristics of this soliton which distinguish it from a plane soliton (KdV soliton) are the power-law decay with respect to both spatial variables at infinity and, in addition to a positive maximum, the existence of two symmetrically positioned negative minima with integral compensation for the contributions of both signs (in particular, asymptotically negative values in the longitudinal direction correspond to positive ones in the transverse direction). In addition, if the algebraic soliton is invariant with respect to reflection $X \longrightarrow -X$, the maximum values are obtained at $X = 0$ and decay quadratically in the transverse direction, remaining positive, and specifically

$$\int dx' V(x', y') = 0, \quad s(X = 0, Y) = \frac{12}{Y^2}. \quad (3.2)$$

Under the action of a random force, the algebraic soliton begins to spread and after substituting the Fourier transformation from (3.1) the average distribution (2.17) is expressed in terms of the error function of the complex variable

$$\langle u(x, y, t)\rangle = \frac{3}{\tau}(1 - \sqrt{\pi} z e^{z^2}\operatorname{erfc} z) + \text{c.c.},$$
$$z \equiv \frac{Y + iX}{2\sqrt{\tau}}, \quad \operatorname{erfc} z \equiv \frac{2}{\sqrt{\pi}}\int_z^\infty ds e^{-s^2}. \quad (3.3)$$

The same result may be obtained by using the probability density function of the displacements (2.14) for the statistical weighting of the expression for the algebraic soliton (3.1) allowing for random longitudinal displacements of the type (2.3) caused by the action of a random force:

$$\langle u\rangle = \int dx_0 s(X - x_0, Y) p(x_0, t) = \int dk \tilde{s}(k, Y)\varphi_t^{(x)}(k).$$

Here, the transition to Fourier components using the characteristic function (2.8) directly yields the result (3.3).

Long times for fixed spatial coordinates x' and y' correspond to small z and using an expansion of the error function as a series [19], we find the law for degeneracy of the average velocity field of an algebraic soliton:

$$\langle u\rangle \approx \frac{6}{\tau}\left(1 - \frac{Y}{2}\sqrt{\frac{\pi}{\tau}} + \ldots\right), \quad \tau \gg X^2 + Y^2. \quad (3.4)$$

For short times (and fixed spatial coordinates) in accordance with the asymptotic behavior of the error function [19], we find the form of the small corrections to the initial distribution for an algebraic soliton:

$$\langle u\rangle \approx \left(\frac{6}{(Y + iX)^2} - \frac{6}{(Y + iX)^4} + \ldots\right) + \text{c.c.},$$
$$\tau \ll X^2 + Y^2. \quad (3.5)$$

The longitudinal symmetry of an algebraic soliton (invariance with respect to the substitution $X \longrightarrow -X$) and the zero longitudinal integral contribution are conserved with time under the action of a random force (for a KdV soliton the conserved integral contribution is nonzero) but the maximum values of the average velocity for a fixed transverse coordinate and the soliton peak in particular are obtained for $X = 0$ since the result for these reduces to the error function of the real variable:

$$\langle u\rangle|_{X=0} = \frac{6}{\tau}(1 - \sqrt{\pi}\xi e^{\xi^2}\operatorname{erfc}\xi), \quad \xi \equiv \frac{Y}{2\sqrt{\tau}}. \quad (3.6)$$

This function is monotonically decreasing ($\partial\langle u\rangle/\partial\xi < 0$) since the soliton peak occurs at $y' = 0$:

$$\xi = q\sqrt{3/(4\tau)}.$$

For long times, it decreases rapidly as t^{-3} (compare with the slower decrease $\propto t^{-3/2}$ for a soliton of the stochastic KdV equation [2]).

If the average velocity distribution is analyzed for fixed y' (thus, for $Y = \text{const}$), for long times Re z becomes small. Generally, neglecting this small real part, we can rewrite the basic formula for the average (3.3) in terms of the Dawson integral:

$$\langle u\rangle \approx \frac{6}{\tau}(1 - 2\eta F(\eta)) = \frac{6}{\tau}\frac{dF}{d\eta}, \quad Y^2 \ll \tau,$$
$$z \approx i\eta \equiv i\frac{X}{2\sqrt{\tau}}, \quad F(\eta) \equiv e^{-\eta^2}\int_0^\eta ds e^{s^2}. \quad (3.7)$$

Here, the zero of the average velocity corresponds to the only maximum point of the Dawson integral $\eta = \eta_m \approx 0.924$ for $\eta > 0$ [19]. For lower values, the average velocity is positive and for higher values it becomes negative. For long times, the two symmetrically posi-

tioned zeroes of the average velocity distribution $\eta = \pm\eta_m$ will move away from the centrally positioned soliton peak in accordance with

$$X = 2\eta_m\sqrt{\tau} \approx 1.85t^{3/2}. \quad (3.8)$$

The minimum points of the average velocity distribution corresponding to the point of inflection of the Dawson integral $\eta = \eta_n \approx 1.5$ behave similarly. For long times, these minimum points move away from the origin (from the maximum point) and its value decreases:

$$X = 2\eta_n\sqrt{\tau} \approx 3t^{3/2}, \quad \langle u\rangle \approx -2/\tau. \quad (3.9)$$

Thus, under the action of a random force, an algebraic soliton undergoes diffusion spreading with asymptotic dispersal of the characteristic points proportionately as $t^{3/2}$ and a decrease in the extreme values proportionately as t^{-3}. The main difference from a KdV soliton is the rate of decrease in the peak values. Since the final value of the integral momentum is conserved ("the soliton area"), the soliton amplitude varies in inverse proportion to its width ($\propto t^{3/2}$), i.e., as $t^{-3/2}$. For an algebraic KdV soliton, no such relationship exists because the conserved momentum integral goes to zero in accordance with (3.2) and the alternating-sign nature of the soliton is reflected in its faster degeneracy.

Among the moment characteristics of the flow field created by a random force about a moving soliton, we confine our analysis to the dispersion. For an algebraic soliton, the mean square of the velocity

$$\langle v^2\rangle = \langle s^2(X - x_0(t), Y)\rangle,$$

and the average velocity with its longitudinal gradient, which appear in the general expression (2.18), can be expressed by analogy with (3.3) in terms of the error function:

$$\partial_x\langle s(X - x_0(t), Y)\rangle$$

$$= -\frac{3}{2}\sqrt{\frac{\pi}{\tau^3}}\left((1 + 2z^2)e^{z^2}\operatorname{erfc} z - \frac{2z}{\sqrt{\pi}}\right) + \text{c.c.},$$

$$\langle s^2\rangle = \frac{9}{Y^2\tau} + \frac{3(z^2 + 2)}{\tau^2}$$

$$+ 3\sqrt{\pi}\left(\frac{3}{Y^3\sqrt{\tau}} - \frac{3z}{Y^2\tau} - \frac{z}{\tau^2}\left(z^2 + \frac{3}{2}\right)\right)e^{z^2}\operatorname{erfc} z + \text{c.c.}$$

For long times ($t \longrightarrow \infty$: $\tau \longrightarrow \infty$) and fixed coordinates X and Y, we will have $z \longrightarrow 0$, and asymptotic estimates yield the conclusion that the main contributions to the dispersion of the velocity field u are made by the dispersion of the background flow and the velocity dispersion $v = s(X - x_0(t), Y)$ associated with the random displacements of the soliton center of gravity:

$$\langle(u - \langle u\rangle)^2\rangle \approx \langle u_0^2(t)\rangle$$

$$+ \langle s^2(X - x_0(t), Y)\rangle \approx f_0^2 t + \frac{9\sqrt{6\pi}}{f_0 Y^3}t^{-3/2}. \quad (3.10)$$

The term from (2.18) proportional to the correlation of the background pulsations with the displacements is asymptotically smaller than the terms retained in (3.10). The increase in the background flow with time saturates if the dissipative properties of the medium are also taken into account, as is well known from the theory of Brownian motion [20].

Like the average velocity and the dispersion, the simultaneous two-point correlation functions can be expressed in terms of the error functions of the complex variable. For the longitudinal correlation function of the soliton component of the random velocity field, we have

$$\langle s(X_1 - x_0(t), Y)s(X_2 - x_0(t), Y)\rangle$$

$$= \exp[\tau(\partial_{X_1} + \partial_{X_2})^2]s(X_1, Y)s(X_2, Y)$$

$$= 18\sqrt{\frac{\pi}{\tau}}\partial_{X_1}\partial_{X_2}$$

$$\times\left(\frac{\varphi(z_2) - \varphi(\bar{z}_1)}{2Y - i(X_1 - X_2)} + i\frac{\varphi(z_2) - \varphi(z_1)}{X_1 - X_2}\right) + \text{c.c.},$$

$$\varphi(z) \equiv \exp(z^2)\operatorname{erfc} z, \quad 2z_n\sqrt{\tau} \equiv Y + iX_n.$$

For long times (and fixed other variables), this correlation decays with time as t^{-3}, i.e., by analogy with the average velocity. The spatial decrease in this longitudinal correlation is characterized by the asymptotic expansion

$$\langle ss\rangle \approx X_1^{-2}\frac{36}{\tau}(\sqrt{\pi}z_2\exp(z_2^2)\operatorname{erfc} z_2 - 1) + O(X_1^{-3}) + \text{c.c.},$$

$$X_1 \longrightarrow \infty.$$

4. PERIODIC SOLUTIONS CONSTRUCTED FROM ALGEBRAIC SOLITONS

For a periodic chain of Kadomtsev–Petviashvili solitons pulled in the longitudinal direction, the sum result may lead to a periodic solution of the Kadomtsev–Petviashvili equation expressed in the following general form (s denotes an isolated soliton and S denotes the sum periodic solution):

$$S(X^0, Y^0) = \sum_n s(X^0 + n\lambda, Y^0)$$

$$= S(X^0 + \lambda, Y^0) = \frac{1}{\lambda}\sum_n \tilde{s}(k_n, Y^0)e^{ik_nX^0},$$

$$\tilde{S}(k, Y^0) = \frac{2\pi}{\lambda}\sum_n \tilde{s}(k_n, Y^0)\delta(k - k_n), \quad (4.1)$$

$$k_n \equiv \frac{2\pi}{\lambda}n = k_1 n.$$

The spectrum of the periodic solution is discrete and the coefficients of the expansion of this solution as a Fourier series are proportional to the Fourier transform of the single-soliton solution.

In order to ensure that this superposition of solitons gives a solution of the KP1 equation for an algebraic soliton similar to (3.1),

$$s(X^0, Y^0) = \frac{6}{(Y^0 + iX^0)^2} + \text{c.c.},$$
$$\tilde{s}(k, Y^0) = 12\pi|k|\exp(-|k|Y^0), \tag{4.2}$$

the velocity of their combined motion should differ from that of a free soliton ($C \neq c$) and their transverse structure should be varied ($Y^0 \neq Y$ and the power-law asymptotic decay in the transverse direction is replaced by exponential):

$$X^0 \equiv x - C_x t + py' \neq X,$$
$$C_x - c_x = -k_1^2, \quad C_y = c_y,$$
$$k_1 Y^0 \equiv \operatorname{Arch}\{\alpha\coth(k_1 q y')\}, \tag{4.3}$$
$$\alpha = \sqrt{1 + 3k_1^2/q^2}.$$

The result of summing the periodic series of these solitons is expressed in terms of the elementary functions [21]:

$$S(X^0, Y^0) = \frac{\partial}{\partial X^0}\frac{6k_1\sin(k_1X^0)}{\alpha\coth(k_1qy') - \cos(k_1X^0)}$$
$$= 6k_1^2\frac{\cos(k_1X^0)\coth((k_1Y^0) - 1)}{(\coth(k_1Y^0) - \cos(k_1X^0))^2}. \tag{4.4}$$

Under the action of a random force, the operator representation and the Fourier expansion for the average velocity (2.17) for this particular case of a steady-state moving chain of algebraic solitons having the Fourier spectrum (4.1) and (4.2) are given by

$$\langle u\rangle = \exp(\tau\partial_X^2)S(X^0, Y^0)$$
$$= \frac{1}{\lambda}\sum_n \tilde{s}(k_n, Y^0)\exp(ik_nX^0)\exp(-\tau k_n^2) \tag{4.5}$$
$$= 12k_1^2\sum_{n=1}^{\infty} n\exp(-k_1Y^0n)\exp(-\tau k_1^2n^2)\cos(k_1X^0n).$$

The structure and behavior of this last series is similar to the structure and behavior of the Fourier series for theta Jacobi functions with an appreciable speed of convergence. For long times, the high-frequency Fourier components degenerate very rapidly and the first term of the series gives a fairly exact asymptotic estimate:

$$\langle u\rangle \approx 12k_1^2\exp(-k_1Y^0)\exp(-\tau k_1^2)\cos(k_1X^0). \tag{4.6}$$

In this approximation, the oscillations of the average field, periodic in the longitudinal direction, are harmonic, the amplitude decreases exponentially with time [more accurately, as $\exp(-2f_0^2k_1^2t^3)$], and they decay exponentially in the transverse direction

5. PLANE SOLITONS UNDER THE ACTION OF A RANDOM FORCE

Solutions in the form of plane solitons are stable relative to small perturbations for KP2 equations, i.e., for negative dispersion. Thus, only some modification of these (without rapid damage) can be predicted under the action of external noise.

Linearly extended soliton solutions of the regular KP2 equations (in this case, there are no algebraic solitons with a decaying dependence on coordinate in all directions) can extend in different directions:

$$v = \partial_x^2\ln(1 + e^\eta) = \frac{3}{2}k_x^2\operatorname{sech}^2\frac{\eta}{2},$$
$$\eta \equiv \mathbf{k}\cdot\mathbf{r} - \omega t + \eta^0, \quad \mathbf{k} = (k_x, k_y), \tag{5.1}$$
$$P(\mathbf{k}, \omega) \equiv k_x^4 + k_y^2 - \omega k_x = 0.$$

A similar three-parameter (k_x, k_y, η^0) solution only differs from the ordinary KdV soliton by the tilt (the additional parameter k_y). Thus, averaging in accordance with (2.14) using the Gaussian probability density function of the displacements over the x axis caused by a Gaussian random force does not give any significantly new result compared with that for the stochastic KdV equation. For the average velocity for a random force of the Gaussian white noise type, we obtain

$$\langle u(x, y, t)\rangle = \frac{3}{4\sqrt{\pi\tau}}$$
$$\times k_x^2\int dx_0\operatorname{sech}^2\frac{\eta - k_xx_0}{2}\exp\left(-\frac{x_0^2}{4\tau}\right). \tag{5.2}$$

This integral is slightly more complicated than that obtained earlier for an algebraic soliton but has a simple asymptotic estimate for long times:

$$\langle u(x, y, t)\rangle \approx \frac{3}{4\sqrt{\pi\tau}}k_x\exp\left(-\frac{\eta^2}{4k_x^2\tau}\right), \quad \tau \gg 1. \tag{5.3}$$

This reflects the fact that under the action of a random force, a plane soliton undergoes diffusion spreading according to the law $\propto t^{3/2}$ and its height decreases as $t^{-3/2}$. This correspondence of the asymptotic forms is obtained because the integral of the average velocity with respect to the longitudinal coordinate x conserves its nonzero value.

In the presence of several plane solitons, interaction will take place in their zones of intersection, the strength of this interaction depending on the angle of

intersection. It should be noted that the angles cannot be large because of the condition for derivation of the Kadomtsev–Petviashvili equations which assumes slow transverse variations (quasi-two-dimensionality) [7].

The two-soliton solution of the regular KP2 equations may be written in the form

$$v = 6\partial_x^2 \ln(1 + e^{\eta_1} + e^{\eta_2} + ae^{\eta_1+\eta_2}),$$
$$\eta_i \equiv \mathbf{k}_i \cdot \mathbf{r} - \omega_i t, \quad a = -\frac{P(\mathbf{k}_1 - \mathbf{k}_2, \omega_1 - \omega_2)}{P(\mathbf{k}_1 + \mathbf{k}_2, \omega_1 + \omega_2)}, \quad (5.4)$$
$$P(\mathbf{k}_i, \omega_i) \equiv k_{xi}^4 + k_{yi}^2 - \omega_i k_{xi}.$$

The character of this exact solution changes substantially as the parameters vary and in particular the phase shift parameter $\ln a$. The general picture becomes particularly clear for the asymptotically simplified representation of the solution [22–24]. If the angles between the solitons (directions of their wave vectors) are relatively large, to a first approximation the solution reduces to a simple sum of two plane solitons with a small modification in their zone of immediate intersection. As the angle decreases (and thus the phase shift increases), the interaction zone increases and the picture can be represented asymptotically as the scattering of free solitons with the phase shifts of these scattered free solitons parallel to that of the initial ones and a "virtual soliton" which links these in the interaction zone. In this case, the intermediate virtual soliton is not a true (free) soliton despite some similarity of its profile because the dispersion equation $P(\mathbf{k}, \omega) = 0$ is not satisfied for it.

In the limit of an infinite phase shift ($a \longrightarrow \infty$) which corresponds to a small critical angle, the scattered solitons are shifted to infinity and the infinite virtual soliton becomes free in the sense that the following resonance conditions [see (5.4)] are satisfied in this limit:

$$\mathbf{k} = \mathbf{k}_1 + \mathbf{k}_2, \quad \omega = \omega_1 + \omega_2, \quad P(\mathbf{k}, \omega) = 0. \quad (5.5)$$

Ultimately a soliton triad forms in which two initial plane solitons form a third plane soliton also pulled in the transverse direction. Formally, the Kadomtsev–Petviashvili equations can also have a similar resonant three-soliton solution with a longitudinally elongated third soliton (for $a \longrightarrow 0$ and opposite signs under resonance conditions). However, this contradicts the condition of quasi-two-dimensionality and thus this solution should not be taken into account. Finally for angles smaller than the critical angle, the solution is singular and its interpretation is not completely clear [22–25].

The result of averaging the field of the liquid flow velocities when a Gaussian white noise force acts on the two-soliton solution (5.4) may be written in the form of an integral similar to the single-soliton variant (5.2). The answer becomes complex. However, since the distributions undergo diffusion broadening (asymptotically $\propto t^{3/2}$) with this statistical Gaussian weighting, its influence on the asymptotic picture discussed above can easily be estimated. Since the angle between the intersecting plane solitons of the Kadomtsev–Petviashvili equations is small, their diffusion broadening leads to mutual overlap and merging near the intersection point, which increases in time and causes a rapid transverse shift of the overlap zone in the direction of the initial solitons with increasing velocity ($\propto t^{1/2}$). Ultimately the size of the virtual soliton will increase. The amplitudes of the wave perturbations will decrease accordingly ($\propto t^{-3/2}$). Similarly in the limiting resonant situation the increasing overlap of the two primary solitons will resemble an increase in the dimensions of the resonant soliton.

6. CONCLUSIONS

Inhomogeneous hydrodynamic equations in which external forces are taken into account possess generalized Galilean invariance. Specifically, on changing to a coordinate system moving at arbitrary acceleration, only the specific form of the external force changes in the equations. This means that for external forces which depend only on time the solutions of the problems for inhomogeneous and homogeneous equations can be related and in particular, the solutions of stochastic problems with a random force can be simplified. The two types of time variations (regular and stochastic) can then easily be separated, and in the case of completely integrable model homogeneous equations the solution of the stochastic nonlinear problem reduces to the solution of a set of linear problems. The statistical problem then involves the Brownian motion of randomly accelerated coordinate systems with linear stochastic ordinary differential equations. This was demonstrated above for the case of stochastic Kadomtsev–Petviashvili equations. Previously, these equations with positive dispersion had only been considered in a single publication [26] in which there were several errors which yielded incorrect asymptotic forms and numerical results and an incorrect conclusion on the strong oscillation of the average field.

In the hydrodynamic interpretation of the Kadomtsev–Petviashvili equations with positive dispersion, the action of a random force on algebraic solitons and periodic sets of these leads to the formation of a random background flow and diffusion spreading of soliton formations accompanied by Brownian motion of their centers of gravity. The broadening of the average field for Kadomtsev–Petviashvili solitons for long times takes place at the same rate ($\propto t^{3/2}$) as for KdV solitons whereas the degeneracy (asymptotic decrease in amplitude) of the Kadomtsev–Petviashvili solitons is faster ($\propto t^{-3}$ as opposed to $\propto t^{-3/2}$). This is because the conserved momentum integral for algebraic Kadomtsev–Petviashvili solitons goes to zero. The degeneracy of the average field for periodic chains of algebraic solitons is exponentially fast. The background pulsation flows and interference of the background with the

Brownian motion of the solitons will be reflected in changes in their moment (and correlation) characteristics.

The action of a random force on single tilted plane solitons of the Kadomtsev–Petviashvili equations with negative dispersion is exactly the same as the action of a random force on the solitons of the KdV equation. For the case of several solitons asymptotic broadening of the free soliton tails and a reduction in their amplitudes is accompanied by an increase in the interaction zone with increasing dimensions of virtual and resonant solitons.

Several problems with parametric stochasticity can be analyzed by analogy with the random force problems. For example, a solution of the stochastic Kadomtsev–Petviashvili equations with the random coefficient $\alpha(t)$

$$\partial_t u + \alpha(t)\partial_x u + 2u\partial_x u + \partial_x^3 u = \partial_y \varphi,$$
$$\partial_x \varphi = \pm \partial_y u$$

using the simple change of variables [see (2.3) and (2.4)]

$$u(x, y, t) = v(\tilde{x}, y, t), \quad \tilde{x} = x - x_0(t),$$
$$\partial_t x_0(t) = \alpha(t)$$

reduces to a solution of the regular Kadomtsev–Petviashvili equations with no variable coefficient and a stochastic Langevin problem with an ordinary stochastic differential equation for random displacements of the center of gravity of the distributions of initially regular flows.

ACKNOWLEDGMENTS

The author thanks A. I. Maĭmistov for critical comments which helped to improve the work.

The work was financed by the Russian Foundation for Basic Research (project no. 99-01–435).

REFERENCES

1. S. N. Gurbatov, E. N. Pelinovskiĭ, and A. I. Saichev, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **21**, 1485 (1978).
2. M. Wadati, J. Phys. Soc. Jpn. **52**, 2642 (1983).
3. I. M. Besieris, in *Nonlinear Electromagnetics*, Ed. by P. L. E. Uslengi (Academic, New York, 1980), p. 87.
4. V. A. Gorodtsov, Prikl. Mat. Mekh. **52**, 211 (1988).
5. Y. Matsuno, Phys. Rev. E **54**, 6313 (1996).
6. A. I. Maĭmistov and É. A. Manykin, Izv. Vyssh. Uchebn. Zaved., Fiz., No. 4, 91 (1987).
7. B. B. Kadomtsev and V. I. Petviashvili, Dokl. Akad. Nauk SSSR **192**, 753 (1970) [Sov. Phys. Dokl. **15**, 539 (1970)].
8. V. E. Zakharov, Pis'ma Zh. Éksp. Teor. Fiz. **22**, 364 (1975) [JETP Lett. **22**, 172 (1975)].
9. E. A. Kuznetsov and S. K. Turitsyn, Zh. Éksp. Teor. Fiz. **82**, 1457 (1982) [Sov. Phys. JETP **55**, 844 (1982)].
10. J. C. Alexander, R. L. Pego, and R. L. Sachs, Phys. Lett. A **226**, 187 (1997).
11. E. Infeld, A. Senatorski, and A. A. Skorupski, Phys. Rev. Lett. **72**, 1345 (1994).
12. S. Novikov, S. V. Manakov, L. P. Pitaevskiĭ, and V. E. Zakharov, *Theory of Solitons: the Inverse Scattering Method* (Nauka, Moscow, 1980; Consultants Bureau, New York, 1984).
13. J. Weiss, M. Tabor, and G. Carnevale, J. Math. Phys. **24**, 522 (1983).
14. A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics* (Nauka, Moscow, 1992, 1996; MIT Press, Cambridge, Mass., 1971,1975), Vols. 1 and 2.
15. V. I. Klyatskin, *Stochastic Equations and Waves in Randomly-Inhomogeneous Media* (Nauka, Moscow, 1980).
16. S. V. Manakov, V. E. Zakharov, L. A. Bordag, *et al.*, Phys. Lett. A **63**, 205 (1977).
17. M. J. Ablowitz and J. Satsuma, J. Math. Phys. **19**, 2180 (1978).
18. J. Satsuma and M. J. Ablowitz, J. Math. Phys. **20**, 1496 (1979).
19. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun (Dover, New York, 1971; Nauka, Moscow, 1979).
20. S. Chandrasekhar, *Stochastic Problems in Physics and Astronomy* (AIP, New York, 1943; Inostrannaya Literatura, Moscow, 1947).
21. A. A. Zaĭtsev, Dokl. Akad. Nauk SSSR **272**, 583 (1983) [Sov. Phys. Dokl. **28**, 720 (1983)].
22. J. W. Miles, J. Fluid Mech. **79**, 157 (1977); **79**, 171 (1977).
23. N. C. Freeman, in *Advances in Applied Mechanics*, Ed. by C.-S. Yi (Academic, New York, 1980), p. 1.
24. K. Okhuma and M. Wadati, J. Phys. Soc. Jpn. **52**, 749 (1983).
25. A. C. Newell and L. G. Redekopp, Phys. Rev. Lett. **38**, 377 (1977).
26. V. Y. Belashov, Phys. Lett. A **197**, 282 (1995).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 1114–1121.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1280–1288.

MISCELLANEOUS

Electrodiffusion Instability in Spherical Cells

V. G. Kamenskiĭ

Landau Institute for Theoretical Physics, Russian Academy of Sciences,
Chernogolovka, Moscow oblast, 142432 Russia
e-mail: kamensky@landau.ac.ru

Received February 14, 2000

Abstract—An electrodiffusion model is used to analyze the behavior of an isolated spherical cell. It is shown that under the influence of fluctuations of the membrane protein density instability may develop, leading to the evolution of ion currents in the cell. Instability thresholds and characteristic harmonic frequencies are determined. General expressions are obtained for the instability thresholds for a system of two spherical cells and an analysis is made of the case of fairly distant cells which allows fundamental qualitative conclusions to be drawn on the nature of the ensuing instabilities. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The physical principles determining the behavior of biological objects such as membranes, cells, and nerve fibers have been studied intensively over many years. It has been established that cell growth and division processes, information transfer along nerve fibers, exchange phenomena in living organisms, and many other properties of biological systems can be explained in terms of physical models which more or less accurately describe this range of phenomena. These processes are determined by the elastic properties of the membranes forming the biological objects and by their electrical properties. Taking into account all important factors simultaneously is naturally extremely difficult, so that frequently fairly approximate models which only take into account some of these factors are considered. However, bearing in mind the wide diversity of biological objects and also the broad spread of characteristic parameters, it is to be hoped that even very rough models can give a correct qualitative description of observed phenomena.

Ion currents play an important role in the behavior of biological objects and in many respects determine processes such as cell growth and division. The existence of these currents has been observed experimentally [1–4]. Various models have been proposed to explain the appearance of ion current in cells, allowing in various ways for the cell geometry, the elastic and electrical properties of the membranes forming the cell surface, chemical reactions, and other factors which influence the processes being studied [5–8]. A common feature of all models is that the ion currents are the result of dynamical instability established when various external control parameters are varied.

Experiments [9] have shown that in some cases appreciable ion currents occur at the early stages of instability evolution (before the initial cell shape is deformed) which allows us to consider models in which the cell can be represented as a fixed geometric surface. One of the classical objects studied is the cell of the Fucus brown algae whose initial shape is very close to spherical [10]. The physical model to describe this cell is a sphere of radius $R \sim 10^{-3}$ cm whose surface is formed by a two-layer lipid membrane of thickness $d \sim 10^{-7}$ cm. Electrolyte is present inside and outside the cell. The membrane itself is impermeable to electrolyte ions but proteins are built into the membrane plane and under the action of external factors (such as the membrane potential, ion concentration, chemical processes, and so on) these can actively (pumps) or passively (channels) transport electrolyte ions through the membrane. The proteins can migrate in the membrane plane both as a result of diffusion and under the action of electric forces since they have charged ends on the outside and inside of the membrane. In the initial steady state, the proteins are distributed uniformly over the cell surface and do not transport ion currents. However, when fluctuations of the protein density occur, a situation may arise where proteins located in the fluctuation region begin to transport ion currents which will then promote further fluctuation of the protein density and therefore increase the currents. As a result of this process, the system becomes unstable.

2. ISOLATED SPHERICAL CELL

The following equations are used to describe the electrodiffusion phenomena in the system.

(1) Nernst–Planck equation for an ion flux of species j and concentration C_j:

$$\mathbf{J}_j = -D_j\left(\nabla C_j + C_j\frac{ez_j}{kT}\nabla\Phi\right), \qquad (1)$$

where D_j is the diffusion coefficient, ez_j is the ion charge, and Φ is the electric potential.

(2) Equation of continuity:

$$-\mathrm{div}\mathbf{J}_j = \frac{\partial C_j}{\partial t} = D_j\left[\Delta C_j + \frac{ez_j}{kT}\mathrm{div}(C_j\nabla\Phi)\right]. \quad (2)$$

(3) Poisson's equation:

$$\Delta\Phi = -\frac{\rho}{\epsilon}, \quad (3)$$

where $\rho = \sum_j N_a ez_j C_j$ is the space charge, N_a is Avogadro's number, and ϵ is the dielectric constant of the electrolyte.

(4) Equation of motion for the proteins:

$$\frac{\partial N}{\partial t} = D_p\left\langle \Delta_S N + \frac{e}{kT}\mathrm{div}_S[N\nabla_S(Z_i\Phi_i + Z_e\Phi_e)]\right\rangle, \quad (4)$$

where N *i*s the surface density of the proteins assuming a continuous distribution, $eZ_{i,e}$ are their charges inside and outside the membrane, $\Phi_{i,e}$ are the corresponding potentials near the membrane, D_p is the protein diffusion coefficient, and subscript S indicates the execution of a mathematical operation in the membrane plane.

These equations, together with the boundary conditions for the potentials and the ion concentrations and also the conditions for transport of currents through the membrane, in principle, completely describe the model. For simplicity, we shall subsequently consider the case of a two-component system of electrolyte ions. We shall also assume that the diffusion coefficients for both species of ions are the same and do not differ on different sides of the membrane. (This approximation naturally only corresponds to a qualitative description of the phenomena taking place in the cell.)

The initial steady state is obtained subject to the conditions N = const, $\mathbf{J}_j = 0$. The condition for the absence of ion currents gives

$$C_j^{i,e} = C_{j0}^{i,e}\exp\left[-\frac{ez_j}{kT}(\Phi_{i,e} - \Phi_{i,e}^0)\right], \quad (5)$$

where $C_{j0}^{i,e}$ and $\Phi_{i,e}^0$ are the concentration and potential at large distances from the membrane (at infinity outside the cell and at the center of the cell). When $ez_j(\Phi_{i,e} - \Phi_{i,e}^0)/kT \ll 1$ (which is confirmed experimentally), the concentration C_j may be represented in the form

$$C_j^{i,e} = C_{j0}^{i,e}\left[1 - \frac{ez_j}{kT}(\Phi_{i,e} - \Phi_{i,e}^0)\right],$$

and the charge density

$$\rho_{i,e} = \rho_{i,e}^0 - \epsilon\chi_{i,e}^2(\Phi_{i,e} - \Phi_{i,e}^0), \quad (6)$$

where $\chi^{-1} = \lambda_D$ is the Debye length,

$$\chi^2 = \frac{N_a e^2}{\epsilon kT}\sum_j z_j^2 C_{j0}.$$

The natural boundary conditions at infinity are $\Phi_e^0 = 0$ and the electroneutrality condition $\rho_e^0 = 0$. As far as the conditions inside the cell are concerned, unlike [5] where the case of complete electroneutrality $\rho \equiv 0$ was considered, a more general approach is presented in [11] where only finite values of Φ and ρ inside the cell are taken as boundary conditions. We shall subsequently adopt this model. Substituting into (3) expressions (6) for ρ taking into account these boundary conditions gives

$$\Delta\Phi_e = \chi_e^2\Phi_e, \quad (7)$$

$$\Delta\Phi_i = -\frac{\rho_i^0}{\epsilon} + \chi_i^2(\Phi_i - \Phi_i^0). \quad (8)$$

The solutions of (6), (7), and (8) have the form

$$\Phi_e = a_e\frac{\exp(-\chi_e r)}{\chi_e r},$$

$$\Phi_i = \Phi_i^0 - \frac{\rho_i^0}{\epsilon\chi_i^2}\left(\frac{\sinh(\chi_i r)}{\chi_i r} - 1\right),$$

$$\rho_e = -a_e\epsilon\chi_e^2\frac{\exp(\chi_e r)}{\chi_e r}, \quad \rho_i = \rho_i^0\frac{\sinh(\chi_i r)}{\chi_i r},$$

and the unknown constants may be obtained from the matching conditions at the membrane boundaries which are determined by the surface charge distribution of the membrane. However, in any case it follows from these solutions that under the condition $\chi R \gg 1$ which corresponds to most experimental situations, the main changes in the potential and charge density take place in a narrow (of the order of a few Debye lengths) region near the membrane boundaries. We shall subsequently assume $\chi_i = \chi_e$ which negligibly influences the final results.

In order to study the possible occurrence of instability with respect to fluctuations of any physical variables of the system in the linear approximation, we express these in the following form:

$$C_j = C_j^{\mathrm{st}} + c_j, \quad \Phi = \Phi^{\mathrm{st}} + \phi, \quad N = N^{\mathrm{st}} + n_p,$$

$$\rho = \rho^{\mathrm{st}} + \delta\rho, \quad \delta\rho = \sum_j N_a ez_j c_j.$$

The subscript st denotes the initial steady state.

Substitution of these expressions into equations (1)–(4) and linearization of resulting equations lead to the following system of equations:

$$\frac{\partial\delta\rho}{\partial t} = D(\Delta\delta\rho - \chi^2\delta\rho), \quad (9)$$

$$\frac{\partial n_p}{\partial t} = D_p\Delta_S\left[n_p + \frac{eN^{\mathrm{st}}}{kT}(Z_i\phi_i + Z_e\phi_e)\right], \quad (10)$$

$$\Delta\phi = -\frac{\delta\rho}{\epsilon}. \quad (11)$$

(We use the Poisson's equation since we are subsequently only interested in the initial stage of instability development, i.e., the quasi-steady-state case.) In addition, here and subsequently we shall confine our analysis to fluctuations having the characteristic wave vectors $k \ll \chi$. The boundary conditions for $\delta\rho$ and ϕ far from the membrane are the same as before and the boundary conditions for the potential at the membrane, as shown in [11], have the form

$$\phi_i - \phi_e = -\alpha\nabla_{\mathbf{n}}\phi_i, \quad \nabla_{\mathbf{n}}\phi_i = \nabla_{\mathbf{n}}\phi_e, \quad (12)$$

where $\nabla_{\mathbf{n}}$ is the derivative with respect to the normal to the surface, $\alpha = d\epsilon/\epsilon_m$, and ϵ_m is the dielectric constant of the membrane. All the quantities in (12) are taken at $r = R$, i.e., the membrane is assumed to be infinitely thin. The conditions for continuity of the ion currents through the membrane should also be satisfied at the membrane boundaries:

$$j_{\mathbf{n},i} = j_{\mathbf{n},e} = I, \quad (13)$$

where

$$\mathbf{j} = -\sum_j z_j D_j\left(\nabla c_j + C_{j0}\frac{ez_j}{kT}\nabla\phi\right) = -D\nabla(\delta\rho + \epsilon\chi^2\phi)$$

is the ion current in the electrolyte, and $I = I_0 n_p/N^{\mathrm{st}}$ is the ion current via channels and pumps (to be specific we shall only consider the dependence of this current on the protein density in the membrane).

An analysis of equations (9)–(11) shows that the main results obtained in [11] can be derived by a simpler method. Since we are only interested in the region near the instability threshold, i.e., $\omega \sim 0$, and the characteristic value $D\chi^2 \sim 10^{12}$–10^{13} s^{-1}, it follows from equation (9) that

$$\Delta\delta\rho \approx \chi^2\delta\rho.$$

Substituting this expression into (11) gives

$$\phi = \phi^0 - \frac{\delta\rho}{\epsilon\chi^2}, \quad (14)$$

where ϕ^0 is the solution of the equation $\Delta\phi^0 = 0$.

From the conditions at the boundary (12) and (13), it follows that

$$\nabla_{\mathbf{n}}\delta\rho_i = \nabla_{\mathbf{n}}\delta\rho_e, \quad (15)$$

and thus $\delta\rho_i = -\delta\rho_e$, and from (12)–(15)

$$\nabla_{\mathbf{n}}\phi_i^0 = \nabla_{\mathbf{n}}\phi_e^0, \quad \phi_i^0 - \phi_e^0 \approx -\alpha\nabla_{\mathbf{n}}\phi_i^0 + \frac{\alpha\delta\rho_i}{\epsilon\chi}. \quad (16)$$

This last expression was derived assuming that for typical membranes $\alpha\chi \gg 1$. Finally, condition (13) gives

$$-\epsilon\chi^2 D\nabla_{\mathbf{n}}\phi_i^0 = \frac{I_0 n_p}{N^{\mathrm{st}}}. \quad (17)$$

Substituting (14) and the second equality (16) into (10) taking into account that

$$Z_i\phi_i^0 + Z_e\phi_e^0 \gg \frac{(Z_i - Z_e)(\phi_i^0 - \phi_e^0 + \alpha\nabla_{\mathbf{n}}\phi_i^0)}{\alpha\chi},$$

since $\alpha\nabla_{\mathbf{n}}\phi_i^0 \sim \alpha\phi_i^0/R \ll \alpha\chi\phi_i^0$, ultimately yields

$$\frac{\partial n_p}{\partial t} = D_p\Delta_S\left[n_S + \frac{eN^{\mathrm{st}}}{kT}(Z_i\phi_i^0 + Z_e\phi_e^0)\right]. \quad (18)$$

Consequently, instead of solving the system of equations (9)–(11) with the boundary conditions (12) and (13) at the membrane surface (12), (13), it is sufficient to analyze the equation of motion of the membrane protein in the field generated by a potential satisfying the Laplace equation and the boundary conditions (16) and (17). Physically, this means that when fluctuations of the membrane protein density occur, the change in the charge density only takes place in a narrow $\sim\lambda_{\mathrm{D}}$ region near the membrane and in the remaining part of the space $\delta\rho = 0$. The change in potential is associated with a change in the boundary conditions at the membrane. The form of the boundary conditions (16) shows that the membrane and its $\sim\lambda_{\mathrm{D}}$ vicinity are effectively an electric double layer with a varying dipole density. These fluctuations generate the potential ϕ^0. It should be noted that in accordance with the definition (13), the condition $\delta\rho = 0$ far from the membrane by no means implies that ion currents are absent.

The solution of the Laplace equation for an isolated spherical cell is well known. Confining ourselves for simplicity to the azimuthally symmetric case and using the boundary condition (12), we have

$$\phi_i^0 = \sum_l b_l r^l P_l(\cos\theta),$$

$$\phi_e^0 = -\sum_l b_l\frac{l}{l+1}r^{-l-1}R^{2l+1}P_l(\cos\theta).$$

Substituting these expressions and (17) into equation (18), we finally obtain

$$\omega_l = \frac{l(l+1)D_p}{R^2}\left[\frac{eRI_0}{\epsilon\chi^2 DkT}\left(\frac{Z_i}{l} - \frac{Z_e}{l+1}\right) - 1\right]. \quad (19)$$

An analysis of this expression shows that for fixed $Z_{i,e}$ the first harmonic is the most unstable ($l = 0$ would correspond to a change in the total number of proteins in the membrane). The main control parameters responsible for the occurrence of instability are the cell size R and the conducting properties of the pumps and the

channels which are determined by the value of I_0. The dependence of I_0 on the state of the system and the external parameters is fairly complex and requires special study so that it is very difficult to check these results experimentally. However, the direct proportionality of the instability threshold to the cell size (naturally assuming that the other parameters are constant) may be checked qualitatively by means of simple observations. The available experimental data suggest that large cells are most liable to the formation of ion currents.

3. SYSTEM OF TWO CELLS

Naturally, in order to check the applicability of this model (in view of some difficulties in the experimental investigations) it would be useful to have other results obtained using it which would allow qualitative conclusions to be drawn. For this purpose, it is interesting to study a system of identical cells in order to identify their mutual influence on the possible occurrence of the instability considered above.

We shall consider two spherical cells spaced at some finite distance and having the same characteristic parameters. Quite clearly, the initial state of this system will be determined by conditions of its formation. As the cells slowly approach each other from infinity, obviously neither the potential distribution nor the surface protein density of each cell will possess the spherical symmetry assumed as the basis of the preceding analysis. However, if we envisage a situation where one cell is positioned fairly rapidly in the vicinity of another (which can be achieved experimentally), or some screening element exists in a system of two cells in the initial state, the initial steady state of each cell may be taken to be the same as in the previous analysis. The subsequent state of the system in the presence of fluctuations may be determined by a mechanism which we shall now analyze.

In order to solve this problem, we shall use the bispherical coordinates:

$$x = \frac{a\sin\theta\cos\varphi}{\cosh\eta - \cos\theta}, \quad y = \frac{a\sin\theta\sin\varphi}{\cosh\eta - \cos\theta},$$
$$z = \frac{a\sinh\eta}{\cosh\eta - \cos\theta},$$

where a is a constant parameter, $-\infty < \eta < \infty$, $0 \le \theta \le \pi$, $0 \le \varphi \le 2\pi$, and the surfaces $\eta = \pm\eta_0 = \text{const}$ correspond to two spheres of radii $R = a/\sinh\eta_0$ having centers at the points ($x = 0$, $y = 0$, $z = \pm R\cosh\eta_0$).

The solutions of the Laplace equation in these coordinates have the form

$$\phi^0_{n,m} = \sqrt{\cosh\eta - \cos\theta}\exp[\pm(n + 1/2)\eta] \times P^m_n(\cos\theta)\exp(\pm im\varphi).$$

Confining ourselves for simplicity to the case $m = 0$, i.e., solutions which are symmetric relative to rotation about the z axis and taking into account the boundary condition $\phi^0_e = 0$ for $r \longrightarrow \infty$ ($\cos\theta,\ \cosh\eta \longrightarrow 1$) and the fact that ϕ^0_i does not diverge inside the cells (i.e., $\eta \longrightarrow \pm\infty$ for any θ), we write the expressions for the potential ϕ^0 in three regions:

$$\phi^0_{i1} = \sqrt{\cosh\eta - \cos\theta}\sum_{n=0}^{\infty} A_n \exp\left[\left(n + \frac{1}{2}\right)\eta\right] \times P_n(\cos\theta)\exp(\omega t),$$

$$\phi^0_e = \sqrt{\cosh\eta - \cos\theta}\sum_{n=0}^{\infty}\left\{C_n\exp\left[\left(n + \frac{1}{2}\right)\eta\right] + D_n\exp\left[-\left(n + \frac{1}{2}\right)\eta\right]\right\}P_n(\cos\theta)\exp(\omega t), \tag{20}$$

$$\phi^0_{i2} = \sqrt{\cosh\eta - \cos\theta}\sum_{n=0}^{\infty} B_n \exp\left[-\left(n + \frac{1}{2}\right)\eta\right] \times P_n(\cos\theta)\exp(\omega t).$$

The surface of cell 1 corresponds to the coordinate $\eta = -\eta_0$ and the surface of cell 2 corresponds to the coordinate $\eta = \eta_0$. Using the expressions obtained for the potentials, we write the equation of motion for the membrane proteins (18) and the boundary conditions for the potentials at the membrane. Condition (16) at the surface of cell 1 $\nabla_{\mathbf{n}}\phi^0_{i1} = \nabla_{\mathbf{n}}\phi^0_{e1}$ gives the expression

$$\sum_{n=0}^{\infty} A_n \exp\left[-\eta_0\left(n + \frac{1}{2}\right)\right]P_n(\cos\theta) \times [(2n + 1)\cosh\eta_0 - \sinh\eta_0 - (2n + 1)\cos\theta]$$
$$= \sum_{n=1}^{\infty}\left\{C_n\exp\left[-\eta_0\left(n + \frac{1}{2}\right)\right]\right. \times [(2n + 1)\cosh\eta_0 - \sinh\eta_0 - (2n + 1)\cos\theta]$$
$$\left. - D_n\exp\left[\eta_0\left(n + \frac{1}{2}\right)\right][(2n + 1)\cosh\eta_0 + \sinh\eta_0 - (2n + 1)\cos\theta]\right\}P_n(\cos\theta). \tag{21}$$

Equation (18) for cell 1 has the form

$$\sum_{n=0}^{\infty}\left\{2\gamma a\left[n(n + 1)(\cosh\eta_0 - \cos\theta)^2 - \frac{1}{4}(4\cosh\eta_0\cos\theta - \cos^2\theta - 3)\right](Z_i\phi^0_{i1} + Z_e\phi^0_{e1}) + \phi^0_{i1}W(\theta, \eta_0, n)\right\} = 0. \tag{22}$$

Here, we have

$$\gamma = \frac{eI_0}{\epsilon\chi^2 DkT},$$

$$W(\theta, \eta_0, n) = \frac{1}{4}(4\cosh\eta_0\cos\theta - \cos^2\theta - 3)$$
$$\times[(2n+1)\cosh\eta_0 - \sinh\eta_0 - (2n+1)\cos\theta]$$
$$+ (n+1)(\cosh\eta_0 - \cos\theta)^2[(2n+1)(n+2)\cos\theta$$
$$- n(2n+1)\cosh\eta_0 + n\sinh\eta_0]$$
$$- 2n(2n+1)(\cosh\eta_0 - \cos\theta)^2 P_{n-1}(\cos\theta)/P_n(\cos\theta).$$

We immediately note that because of the technical complexity of the calculations, we merely confine ourselves to finding the instability appearance threshold, i.e., the case $\omega = 0$. Equations for cell 2 are obtained from formulas (21) and (22) by replacing A with B and $C \rightleftarrows D$. Expressions (21) and (22) contain products of the powers $\cos\theta$ and the Legendre polynomials and thus are not true expansions in terms of orthogonal polynomials so that such an expansion must be made. Multiplying the left- and right-hand sides of (21) by $P_m(\cos\theta)$ and integrating with respect to $\cos\theta$, we obtain

$$\begin{aligned}
&[m(A_m - A_{m-1})\exp\eta_0 + (m+1)(A_m - A_{m+1})\\
&\times\exp(-\eta_0)]\exp\left[-\left(m+\frac{1}{2}\right)\eta_0\right]\\
&= [m(C_m - C_{m-1})\exp\eta_0\\
&+ (m+1)(C_m - C_{m+1})\exp(-\eta_0)]\\
&\times\exp\left[-\left(m+\frac{1}{2}\right)\eta_0\right] - [m(D_m - D_{m-1})\exp(-\eta_0)\\
&+ (m+1)(D_m - D_{m+1})\exp\eta_0]\exp\left[\left(m+\frac{1}{2}\right)\eta_0\right].
\end{aligned} \tag{23}$$

The fact that this expression contains the coefficients A, C, and D with different indices implies that different harmonics may be excited simultaneously in the system although their amplitudes and growth rates will be regulated by exponential factors which depend on η_0.

Expanding equation (22) in terms of the Legendre polynomials gives

$$\begin{aligned}
&A_{m-3}T_{-3} + A_{m-2}T_{-2} + A_{m-1}T_{-1} + A_mT_0 + A_{m+1}T_1\\
&+ A_{m+2}T_2 + A_{m+3}T_3 + \gamma Z_i a[A_{m+2}t_2\exp(-2\eta_0)\\
&+ A_{m+1}t_1\exp(-\eta_0) + A_mt_0 + A_{m-1}t_{-1}\exp\eta_0\\
&+ A_{m-2}t_{-2}\exp(2\eta_0)] + \gamma Z_e a\exp\left[\left(m+\frac{1}{2}\right)\eta_0\right]\\
&\times\left\{t_2\left(C_{m+2}\exp\left[-\left(m+\frac{5}{2}\right)\eta_0\right]\right.\right.\\
&\left.\left.+ D_{m+2}\exp\left[\left(m+\frac{5}{2}\right)\eta_0\right]\right) + t_1\left(C_{m+1}\exp\left[-\left(m+\frac{3}{2}\right)\eta_0\right]\right.\right.\\
&\left.\left.+ D_{m+1}\exp\left[\left(m+\frac{3}{2}\right)\eta_0\right]\right)\right.\\
&+ t_0\left(C_m\exp\left[-\left(m+\frac{1}{2}\right)\eta_0\right]\right.\\
&\left.+ D_m\exp\left[\left(m+\frac{1}{2}\right)\eta_0\right]\right) + t_{-1}\left(C_{m-1}\exp\left[-\left(m-\frac{1}{2}\right)\eta_0\right]\right.\\
&\left.+ D_{m-1}\exp\left[\left(m-\frac{1}{2}\right)\eta_0\right]\right)\\
&+ t_{-2}\left(C_{m-2}\exp\left[-\left(m-\frac{3}{2}\right)\eta_0\right]\right.\\
&\left.\left.+ D_{m-2}\exp\left[\left(m-\frac{3}{2}\right)\eta_0\right]\right)\right\}.
\end{aligned} \tag{24}$$

Expressions for the coefficients T_k and t_k are given in the Appendix.

We shall first consider the symmetric case relative to the axis $\eta = 0$ when $A = B$ and $C = D$. Equation (24) has the form

$$\begin{aligned}
&A_{m-3}T_{-3} + A_{m-2}T_{-2} + A_{m-1}T_{-1} + A_mT_0\\
&+ A_{m+1}T_1 + A_{m+2}T_2 + A_{m+3}T_3\\
&+ \gamma Z_i a[A_{m+2}t_2\exp(-2\eta_0) + A_{m+1}t_1\exp(-\eta_0)\\
&+ A_mt_0 + A_{m-1}t_{-1}\exp\eta_0 + A_{m-2}t_{-2}\exp(2\eta_0)]\\
&+ 2\gamma Z_e a\exp\left[\left(m+\frac{1}{2}\right)\eta_0\right]\left\{t_2C_{m+2}\cosh\left[\left(m+\frac{5}{2}\right)\eta_0\right]\right.\\
&+ t_1C_{m+1}\cosh\left[\left(m+\frac{3}{2}\right)\eta_0\right] + t_0C_m\cosh\left[\left(m+\frac{1}{2}\right)\eta_0\right]\\
&+ t_{-1}C_{m-1}\cosh\left[\left(m-\frac{1}{2}\right)\eta_0\right]\\
&\left.+ t_{-2}C_{m-2}\cosh\left[\left(m-\frac{3}{2}\right)\eta_0\right]\right\}.
\end{aligned} \tag{25}$$

After simple calculations, formula (23) gives

$$\begin{aligned}
&2(m+1)(C_{m+1} - C_m)\sinh\left[\left(m+\frac{3}{2}\right)\eta_0\right]\\
&= (m+1)(A_m - A_{m+1})\exp\left[-\left(m+\frac{3}{2}\right)\eta_0\right]\\
&+ \frac{\sinh\eta_0}{\sinh\left[\left(m+\frac{1}{2}\right)\eta_0\right]}[(A_1 - A_0) + 2(A_2 - A_1)\\
&+ 3(A_3 - A_2) + \ldots + m(A_m - A_{m-1})].
\end{aligned} \tag{26}$$

Substitution of this expression into (25) results in the elimination of the variables C with the result that (25) is converted into a system of equations for A_m with known coefficients determined by the coefficients T and t, exponential functions, and hyperbolic functions of η_0. The condition that the determinant of this system of equations goes to zero gives a set of threshold values for the onset of instability for different harmonics. The determinant is an "almost triangular" matrix with three diagonals above the principal diagonal. For any values of η_0, this matrix can, in principle, be reduced to a triangular form. Then, the excitation thresholds of any harmonics will be determined by its principal diagonal. The corresponding analytic expressions are too cumbersome so we shall not give them here. In order to make a qualitative analysis of the results, we utilize the fact that for distances between the cells of the order of twice their diameters the value $\eta_0 \sim 2$ is achieved so that to within the second order the hyperbolic functions can be replaced by corresponding exponential functions which appreciably simplifies the expressions given above. In this approximation the line in the system of equations (25) corresponding to the index m has the following form:

$$\sum_{k=0}^{m-4}\frac{2\exp(2\eta_0)Z_e}{2m-1}A_k+\left[T_{-3}+\frac{2\exp(2\eta_0)Z_e}{2m-1}\right]A_{m-3}+\sum_{k=-2}^{2}\left[T_k+\gamma a t_k\exp(-k\eta_0)\left(Z_i-\frac{m+k}{m+k+1}Z_e\right)\right]\times A_{m+k}+T_3A_{m+3}. \tag{27}$$

Here, we have retained the previous notation of T_k and t_k for the time being. Formula (27) can be used to write the matrix of interest to us in a fairly simple form and to analyze it. For this purpose it is convenient to introduce the notation

$$M_k(m) = T_k(m)+\gamma a t_k(m)e^{-k\eta_0}\left(Z_i-\frac{m+k}{m+k+1}Z_e\right)\quad(k=0,\pm1,\pm2)$$

for the elements of the principal diagonal and the two nearest above and below. It is deduced from the form of the coefficients T_k and t_k that M_{-2}, M_{-1}, $M_0 \sim \exp(3\eta_0)$ whereas $M_1 \sim \exp(\eta_0)$ and $M_2 \sim \exp(-\eta_0)$.

The procedure for reducing this matrix to triangular form suggests that the elements of the principal diagonal have the form

$$M_0(m)-M_{-1}(m)\frac{M_1(m-1)}{M_0(m-1)}, \tag{28}$$

and, consequently, the correction to M_0 will be of the order $M_0\exp(-2\eta_0)$. Hence, in the principal order, the excitation thresholds of the corresponding harmonics are determined by the diagonal elements

$$T_0+\gamma a t_0\left(Z_i-\frac{m}{m+1}Z_e\right)=0, \tag{29}$$

which for $\eta_0 \longrightarrow \infty$ and the corresponding representation of the coefficients T_0 and t_0 naturally yields (19).

It can be seen from the preceding analysis that additional information determining the influence of the second cell on the onset of instability for $\eta_0 > 1$ can only be obtained allowing for the next correction to M_0 in terms of $\exp(-2\eta_0)$. Naturally, the corrections in the expansion of the coefficients T_k and t_k and also the corresponding terms of the expansion of the hyperbolic functions in (25) and (26) must be taken into account accurately. We shall return to their calculation later but we now consider another solution for the asymmetric case. This solution is obtained as follows. We write equations (23) for both cells, subtract one from the other, and introduce the variables $U = A - B$ and $K = C - D$. In these variables, we have

$$\begin{aligned}2(m+1)(K_m-K_{m+1})\cosh\left[\left(m+\frac{3}{2}\right)\eta_0\right]&=(m+1)(U_m-U_{m+1})\exp\left[-\left(m+\frac{3}{2}\right)\eta_0\right]\\&+\frac{\sinh\eta_0}{\cosh\left[\left(m+\frac{1}{2}\right)\eta_0\right]}[(U_1-U_0)+2(U_2-U_1)\\&+3(U_3-U_2)+\ldots+m(U_m-U_{m-1})].\end{aligned}$$

In this case, equation (24) gives

$$\begin{aligned}&U_{m-3}T_{-3}+U_{m-2}T_{-2}+U_{m-1}T_{-1}+U_mT_0+U_{m+1}T_1\\&+U_{m+2}T_2+U_{m+3}T_3+\gamma Z_i a[U_{m+2}t_2\exp(-2\eta_0)\\&+U_{m+1}t_1\exp(-\eta_0)+U_mt_0+U_{m-1}t_{-1}\exp(\eta_0)\\&+U_{m-2}t_2\exp(2\eta_0)]-2\gamma Z_e a\exp\left[\left(m+\frac{1}{2}\right)\eta_0\right]\\&\times\left\{K_{m+2}t_2\sinh\left[\left(m+\frac{5}{2}\right)\eta_0\right]\right.\\&+K_{m+1}t_1\sinh\left[\left(m+\frac{3}{2}\right)\eta_0\right]+K_mt_0\sinh\left[\left(m+\frac{1}{2}\right)\eta_0\right]\\&+K_{m-1}t_{-1}\sinh\left[\left(m-\frac{1}{2}\right)\eta_0\right]\\&\left.+K_{m-2}t_{-2}\sinh\left[\left(m-\frac{3}{2}\right)\eta_0\right]\right\}.\end{aligned} \tag{30}$$

The same procedure as in the symmetric case yields expression (29), i.e., an equality in the principal order of the harmonic excitation thresholds for both cases. This is a natural result which merely implies that when sufficiently far apart, the cells are almost independent. We shall now calculate the next correction in terms of the parameter $\exp(-2\eta_0)$. We express the control parameter γ in the form

$$\gamma = \gamma_0(1+\delta), \tag{31}$$

where

$$\gamma_0 = \frac{m(m+1)}{R[(m+1)Z_i - mZ_e]}$$

is its value for $\eta_0 \longrightarrow \infty$, i.e., corresponding to expression (19). Substituting (31) into (25) and (26) where the coefficients T_k and t_k and the hyperbolic functions are taken to within terms of the expansion of the order $\exp(-2\eta_0)$ and confining ourselves to the case $m < \exp(2\eta_0)$, we obtain the matrix elements of M_k with the same accuracy. Substituting these elements into expression (28) and equating to zero, we obtain as a result

$$\delta = \frac{\exp(-2\eta_0)}{m^2(m+1)}\Bigg\langle \gamma_0 Z_e R\Bigg[\frac{1}{2} + 2m + \frac{10}{3(m+1)} + \frac{1}{12(2m-1)} + \frac{3}{(2m-1)(2m+3)}\Bigg]$$

$$+\Bigg[\frac{m^4}{2} + 6m^3 + 10m^2 + 11m + \frac{19}{4} + \frac{3}{2(2m-1)} + \frac{1}{4(2m+3)}\Bigg] \pm \gamma_0 Z_e R m^2 \exp[(-2m+3)\eta_0]\Bigg\rangle.$$

The plus and minus signs before the last term refer to the symmetric and asymmetric cases, respectively. It can be seen from this expression that the terms in brackets are always positive for $\gamma_0 Z_e > 0$ (which corresponds to the experimental situation) and are the same for both cases. This part of the correction reflects the fact that the occurrence of protein density fluctuations in one cell leads to a change in potential in all space and consequently changes the protein-density of the second cell. Naturally, this process requires a higher energy compared to the case of an isolated cell, i.e., increases the excitation threshold. The last term of the correction introduces a difference between the symmetric and asymmetric cases. Under these assumptions on the sign of $\gamma_0 Z_e$ the asymmetric case is preferable. This conclusion qualitatively explains the experimental observation [12] that in groups of cells (albeit those with strongly developed ion currents), the most stable and long-lived configurations are those corresponding to the symmetric case. The reduction in the difference between the two states with increasing harmonic number has an obvious explanation since for higher harmonics the protein density distribution over the cell surface effectively becomes increasingly uniform.

We have therefore shown that despite being fairly arbitrary, the electrodiffusion model can, under certain conditions, predict the possible appearance of instability against the development of ion currents in spherical cells and can also provide a qualitative description of the behavior of a system of two such cells.

ACKNOWLEDGMENTS

The author thanks E. Dubois-Violette and M. Leonetti for proposing the problem and useful discussions.

The work was financed by the Russian Foundation for Basic Research (project no. 00-02-17785).

APPENDIX

$$T_{-3} = \frac{m(m-1)(m-2)(2m-3)}{2(2m-1)}\exp(3\eta_0),$$

$$T_3 = \frac{(m+1)(m+2)(m+3)(2m+5)}{2(2m+3)}\exp(-3\eta_0),$$

$$T_{-2} = -\frac{m(m-1)}{2m-1}\exp(2\eta_0) \times\Bigg[(6m^2-11m+8)\cosh\eta_0 + \frac{(2m-3)}{2}\exp(-\eta_0)\Bigg],$$

$$T_2 = -\frac{(m+1)(m+2)}{2m+3}\exp(-2\eta_0) \times\Bigg[(6m^2+21m+20)\cosh\eta_0 + \frac{(2m+5)}{2}e^{-\eta_0}\Bigg],$$

$$T_{-1} = m\exp(\eta_0)\Bigg[\frac{7}{2} + 3m(m-1) - \frac{1}{2m-3} + \frac{3(m^2-2)(2m+1)^2}{2(4m^2-9)} + \frac{6m^3-7m^2+5m-2}{2m-1} \times\cosh(2\eta_0) + \frac{2m^2-2m+1}{2m-1}\exp(-2\eta_0)\Bigg],$$

$$T_1 = (m+1)\exp(-\eta_0) \times\Bigg[3m^2+7m+5 + \frac{6m^3+11m^2+14m-10}{2(2m-1)} + \frac{6m^3+21m^2+25m+10}{2m+3}\cosh(2\eta_0) + \frac{2m^2+6m+5}{2m+3}\exp(-2\eta_0)\Bigg],$$

$$T_0 = \frac{1}{2}[\sinh\eta_0 - (2m+1)\cosh\eta_0] \times [4m(m+1)\cosh^2\eta_0 + 3] + \frac{2m^2+2m-1}{2(2m-1)(2m+3)}[(2m+1)\sinh\eta_0 - (12m^2+28m+21)\cosh\eta_0],$$

$$t_0 = 2\left[2 + m(m+1)(1+2\cosh^2\eta_0) + \frac{1}{(2m-1)(2m+3)}\right],$$

$$t_{-1} = -\frac{4m(2m^2-2m+1)}{2m-1}\cosh\eta_0,$$

$$t_1 = -\frac{4(m+1)(2m^2+6m+5)}{2m+3}\cosh\eta_0,$$

$$t_{-2} = \frac{m(m-1)(2m-3)}{2m-1},$$

$$t_2 = \frac{(m+1)(m+2)(2m+5)}{2m+3}.$$

REFERENCES

1. N. A. R. Gow, Adv. Microb. Physiol. **30**, 89 (1989).
2. B. Fontaine and J. P. Changeux, J. Cell Biol. **108**, 1025 (1989).
3. C. M. Colbert and D. Johnston, J. Neurosci. **16**, 6676 (1996).
4. F. M. Harold and J. H. Caldwell, *Tip Grows in Plant and Fungal Cells* (Academic, New York, 1990).
5. R. Larter and P. Ortoleva, J. Theor. Biol. **96**, 175 (1982).
6. L. F. Jaffe, K. R. Robinson, and R. Nuccitelli, Ann. N. Y. Acad. Sci. **9**, 372 (1974).
7. K. Toko, H. Chosa, and K. Yamafuji, J. Theor. Biol. **114**, 127 (1985).
8. P. Pelce, Phys. Rev. Lett. **71**, 1107 (1993).
9. M-m Poo and K. R. Robinson, Nature **265**, 602 (1977).
10. D. L. Kropf, Microbiol. Rev. **56**, 316 (1992).
11. M. Léonetti and E. Dubois-Violette, Phys. Rev. E **56**, 4521 (1997); Phys. Rev. Lett. **81**, 1977 (1998).
12. L. F. Jaffe, Adv. Morphog. **7**, 295 (1968).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 919–926.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1059–1066.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Compton Mechanism of Neutrino and Axion Production on an Effectively Two-Dimensional Magnetized Fermi Gas

V. V. Skobelev
Moscow State Industrial University, Moscow, 109280 Russia
Fax: (095)275-22-56
Received January 24, 2000

Abstract—The Compton channels for the production of axions $\gamma e \longrightarrow ea$ and neutrinos $\gamma e \longrightarrow e\nu\bar{\nu}$ in a magnetic field are examined on the basis of a two-dimensionally covariant formalism developed in this paper. Expressions are obtained for the cross sections of the processes as well as the power of the radiation per unit volume in a degenerate and nondegenerate electron gas. It is shown that the axion luminosity of white dwarfs on account of the Compton generation mechanism is at least four orders of magnitude less than the photon luminosity, and the axion luminosity for magnetic neutron stars approaches the neutrino luminosity in magnitude.

1. INTRODUCTION

An axion is a pseudo-Goldstone boson, a type of particle that appears as a result of spontaneous breaking of global symmetries (for an axion this is $U(1)_{PQ}$ symmetry, introduced by Peccei and Quinn [1]). The coupling of these bosons with known particles is suppressed by the large energy scale f for breaking of the corresponding symmetry and has the form

$$\mathcal{L} = \frac{1}{f}\frac{\partial\phi}{\partial x^\mu}J^\mu, \tag{1}$$

where J^μ is the particle current in the standard model. Specifically, the Lagrangian of the axion-electronic interaction is [2]

$$\mathcal{L}_{ae} = \frac{c_e}{2f}\frac{\partial a}{\partial x^\mu}(\bar{\Psi}\gamma^\mu\gamma^5\Psi) \tag{2a}$$

(c_e is a model-dependent parameter of order 1), which in the case considered in the present paper is equivalent to a Lagrangian with pseudoscalar coupling

$$\mathcal{L}_{ae} = -ic_e\frac{m}{f}a(\bar{\Psi}\gamma^5\Psi) \tag{2b}$$

(m is the electron mass).

The possible existence of an axion is extremely important for a natural explanation of the exact *CP* invariance of the strong interactions [1, 3] and for astrophysics, since the small axion mass m_a, which appears as a result of mixing with π_0, could be a substantial part of the cold, dark matter [4]. This gives the lower limit of the mass $m_a \geq 10^{-5}$ eV; the upper limit $m_a \leq 10^{-2}$ eV is determined from the condition that the existing ideas about the evolution of collapsed astrophysical objects remain unchanged [2, 5, 6]. Specifically, a natural requirement is that the axion luminosity should not exceed the neutrino luminosity at the early stages of the evolution of neutron stars and the surface photon luminosity of white dwarfs. The approximate "window" of mass values

$$10^{-5}\text{ eV} \lesssim m_a \lesssim 10^{-2}\text{ eV} \tag{3}$$

determined in this manner also fixes the range of values of the energy scale f by virtue of the relation between them [2, 3]

$$m_a \simeq 0.6\times10^{-3}\text{ eV}\times(10^{10}\text{ GeV}/f). \tag{4}$$

Since the values of the axion parameters are based on astrophysical estimates, any accompanying factor, such as strong magnetic fields, which can radically change the characteristics of processes involving axions and neutrinos and open up new reaction channels, must be taken into account. Of the latter reactions, we note the neutrino and axion synchrotron radiation [7–9], which could increase the upper limit of the mass m_a [9], and a new decay channel for the axion in a superstrong magnetic field $a \longrightarrow 3\gamma$ [10], whose probability is greater than the probability of the two-photon channel $a \longrightarrow 2\gamma$, which is considered to be the main channel.

The Compton mechanism $\gamma e \longrightarrow ea(\nu\bar{\nu})$ is one of the main axion and neutrino production mechanisms. The process $\gamma e \longrightarrow ea$ and its contribution to the axion luminosity of stars have been studied in [11] on the basis of the Primakov effect (where, according to the scheme (1), the axion is coupled only with heavy fermions with the consequent effective $(a\gamma\gamma)$-coupling by means of the triangle anomaly) and as a result of direct *aee*-coupling (2a). However, the results of these works pertain to special configurations of the momenta and the magnetic induction, while the field should have a determining effect if the electrons occupy the ground Landau level.

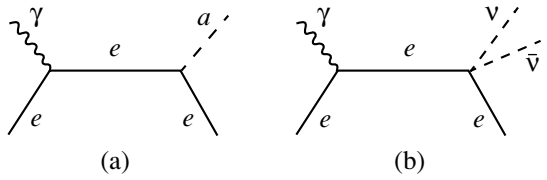


Fig. 1.

In this paper the Compton mechanism of axion production in an external magnetic field is examined using direct *aee*-coupling (diagram in Fig. 1a) and neutrino production in the contact approximation according to the "weak" vertex (diagram in Fig. 1b). A two-dimensionally covariant method of calculating Feymann diagrams with electron lines is used. The foundations for this method are laid in [12] (see also a further elaboration of this method in [10] for diagrams with electron loops).

The exposition below is organized as follows. The two-dimensionally covariant technique for diagrams without excitation of the electron–positron vacuum is presented in Section 2. An application of this technique to the calculation of the matrix elements and the cross sections for the processes $\gamma e \longrightarrow ea$ and $\gamma e \longrightarrow e\nu\bar{\nu}$ is presented in Section 3. The intensity of the axion and neutrino radiations from a unit volume in a degenerate and nondegenerate electron gas is calculated in Section 4. The astrophysical aspects of the results in application to the luminosity of magnetic neutron stars and white dwarfs are discussed in Section 5.

2. WAVEFUNCTION AND GREEN'S FUNCTION FOR THE DIRAC EQUATION IN A CONSTANT AND UNIFORM MAGNETIC FIELD

The solution of the Dirac equation

$$(i\hat{\partial} - e\hat{A} - m)\Psi = 0 \tag{5}$$

for an electron in a constant and uniform magnetic field with induction **B** in the special gauge

$$A_\alpha = Bx_1 g_{\alpha 2} \tag{6}$$

has been known for a long time and is a generalization of the corresponding solution obtained by Landau for the Schrödinger equation (see, for example, [13]):

$$\Psi = \frac{(\gamma\pi)^{1/4}}{(2p_0L_2L_3)^{1/2}}\exp\left[-\frac{\xi_x^2}{2} + i(p_2x_2 + p_3x_3)\right]u_n, \tag{7}$$

where the spinor u_n is expressed in terms of the Hermite polynominals with the argument

$$\xi_x = x_1\sqrt{\gamma} + p_2/\sqrt{\gamma}, \quad \gamma = |eB|. \tag{8}$$

Here L_2 and L_3 are normalization lengths, the quasimomentum p_2 gives the position of the center of the wave packet along the 1-axis, and the possible values of the energy are determined by the relation

$$p_0 = (m^2 + 2\gamma n + p_3^2)^{1/2}, \tag{9}$$

where p_3 is the momentum along the 3-axis, and the quantum number $n = 0, 1, 2, \ldots$.

In the ground Landau level $n = 0$ the electron spin is oriented in a direction opposite to the field, and the spinor u_0 degenerates into a two-component spinor satisfying the equations

$$\begin{aligned}(\hat{p}_\parallel - m)u_0 &= 0, \quad \hat{p}_\parallel = p_0\gamma^0 + p_3\gamma^3,\\ \bar{u}_0u_0 &= 2m, \quad \frac{1 - i\gamma_1\gamma_2}{2}u_0 = u_0,\\ u_0\bar{u}_0 &= \frac{1 - i\gamma_1\gamma_2}{2}(\hat{p}_\parallel + m).\end{aligned} \tag{10}$$

In this connection, it is convenient to introduce two-dimensional matrices $\tilde{\gamma}_0$ and $\tilde{\gamma}_3$ in the subspace (0, 3), which can be chosen, for example, in the form

$$\tilde{\gamma}^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \tilde{\gamma}^3 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{11}$$

and the two-dimensional spinor $v(p)$, $p = (p_0, p_3)$, which in the representation (11) has the form

$$v(p) = \frac{1}{\sqrt{p_0 + m}}\begin{pmatrix} p_0 + m \\ -p_3 \end{pmatrix}. \tag{12}$$

Irrespective of the representation, $v(p)$ satisfies the equation

$$(\check{p} - m)v = 0, \quad \check{p} = p_0\tilde{\gamma}^0 + p_3\gamma^3 \tag{13}$$

with the invariant, in (0, 3), normalization condition

$$\bar{v}v = 2m, \quad \bar{v} = v^+\tilde{\gamma}^0 \tag{14a}$$

and density matrix

$$v\bar{v} = \check{p} + m. \tag{14b}$$

We introduce the properties of the $\tilde{\gamma}$ matrices that will be helpful below:

$$\tilde{\gamma}_\alpha\tilde{\gamma}_\beta + \tilde{\gamma}_\beta\tilde{\gamma}_\alpha = 2\tilde{g}_{\alpha\beta}, \tag{15a}$$

$$\tilde{\gamma}^\alpha(\tilde{\gamma}_{\alpha_1}\ldots\tilde{\gamma}_{\alpha_{2n+1}})\tilde{\gamma}_\alpha = 0, \tag{15b}$$

$$\tilde{\gamma}_\alpha\tilde{\gamma}_\beta = \tilde{g}_{\alpha\beta} + \tilde{\gamma}^5\varepsilon_{\alpha\beta}. \tag{15c}$$

Here $\tilde{g}_{\alpha\beta} = (1, -1)$ is the metric tensor in the subspace (0, 3), $\varepsilon_{\alpha\beta}$ is an absolutely antisymmetric tensor ($\varepsilon_{30} = -\varepsilon_{03} = 1$, $\varepsilon_{00} = \varepsilon_{33} = 0$), $\tilde{\gamma}^5 = \tilde{\gamma}^0\tilde{\gamma}^3$ is the analog of the

matrix γ^5 in four-dimensional space, and $(\tilde{\gamma}^5)^2 = 1$. Using the relations

$$\varepsilon^{\alpha\beta}\varepsilon^{\rho\sigma} = \tilde{g}^{\alpha\sigma}\tilde{g}^{\beta\rho} - \tilde{g}^{\alpha\rho}\tilde{g}^{\beta\sigma} \tag{16}$$

and the reduction formula (15c), the traces of any number of $\tilde{\gamma}$ matrices can be calculated in an elementary way, for example,

$$\frac{1}{2}\mathrm{Tr}(\tilde{\gamma}^\alpha\tilde{\gamma}^\beta\tilde{\gamma}^\rho\tilde{\gamma}^\sigma) = \tilde{g}^{\alpha\beta}\tilde{g}^{\rho\sigma} + \tilde{g}^{\alpha\sigma}\tilde{g}^{\beta\rho} - \tilde{g}^{\alpha\rho}\tilde{g}^{\beta\sigma}, \tag{17a}$$

$$\frac{1}{2}\mathrm{Tr}(\tilde{\gamma}^5\tilde{\gamma}^\alpha\tilde{\gamma}^\beta\tilde{\gamma}^\rho\tilde{\gamma}^\sigma) = \tilde{g}^{\alpha\beta}\varepsilon^{\rho\sigma} + \varepsilon^{\alpha\beta}\tilde{g}^{\rho\sigma} \tag{17b}$$

and so on.

The solution of the singular Dirac equation in a constant and uniform magnetic field

$$(i\hat{\partial}_x - e\hat{A} - m)G(x, y) = \delta(x - y) \tag{18}$$

under certain conditions likewise becomes effectively two-dimensional in the subspace (0, 3). Let us take, for example, the solution in the form obtained by squaring equation (18) [14]:

$$\begin{aligned} G(x, y) &= \frac{(2\gamma)^{1/2}}{(2\pi)^{3/2}}(i\hat{\partial}_x - e\hat{A} + m) \\ &\times \int_{-\infty}^{\infty} dp_2 \exp(ip_2(x_2 - y_2)) \\ &\times \sum_{n=0}^{\infty} \frac{1}{n!} D_n(\sqrt{2}\xi_x) D_n(\sqrt{2}\xi_y) \\ &\times \left[\frac{1 - i\gamma_1\gamma_2}{2} G_{2n}(x - y) + \frac{1 + i\gamma_1\gamma_2}{2} G_{2(n+1)}(x - y)\right], \end{aligned} \tag{19}$$

where the value of $\xi_{x,y}$ is determined by the formula (8), D_n are the parabolic cylinder functions related with the Hermite polynomials [15], and $G_\nu(x - y)$ is the Green's function of the Klein–Gordon equation in the subspace (0, 3)

$$\left(-\frac{\partial^2}{\partial x_0^2} + \frac{\partial^2}{\partial x_3^2} - m_\nu^2\right) G_\nu(x - y) = \delta^{(0,3)}(x - y), \tag{20}$$

$$m_\nu^2 = m^2 - \nu\gamma,$$

and has the form

$$\begin{aligned} G_\nu(z) &= \frac{1}{(2\pi)^2}\int \frac{d^2p}{p_\parallel^2 - m_\nu^2} e^{-i(pz)_{0,3}}, \\ p_\parallel^2 &= p_0^2 - p_3^2. \end{aligned} \tag{21}$$

The causal Green's function can be obtained from equation (21) in accordance with the rule for handling the poles $p_0 = \pm\sqrt{p_3^2 + m_\nu^2}$ [16] and can be expressed in terms of zero-order cylinder functions

$$\begin{aligned} G_\nu^{(c)} &= \frac{1}{4}\theta(z^2)[J_0(m_\nu\sqrt{z^2}) - iN_0(m_\nu\sqrt{z^2})] \\ &+ \frac{i}{2\pi}\theta(-z^2)K_0(m_\nu\sqrt{-z^2}), \end{aligned} \tag{22}$$

where $\theta(x)$ is a step function.

For diagrams without excitation of the vacuum, the value of $p_\parallel^2$ in equation (21) is fixed and can be expressed in terms of the momenta of the external lines. Then, provided that

$$(p_\parallel^2 - m^2) \ll \gamma, \tag{23}$$

the ground Landau level $n = 0$ makes the main contribution to the sum over n in the expression (19). In diagrams with electron loops the condition (23) holds when the momentum integrals of the loop on the electron mass converge and with the additional restriction

$$B \gg B_0, \quad B_0 = \frac{m^2}{|e|} = 4.41 \times 10^{13}\ \mathrm{G}. \tag{24}$$

The leading contribution of the $n = 0$ state can be represented in the form

$$G(x, y) = \exp\left[-\frac{i\gamma}{2}(x_1 + y_1)(x_2 - y_2)\right] G(x - y), \tag{25a}$$

$$G(z) = \frac{1}{(2\pi)^4}\int d^4p \exp(-i(pz))G(p), \tag{25b}$$

$$\begin{aligned} G(p) &= (1 - i\gamma_1\gamma_2)\exp\left(-\frac{p_\perp^2}{\gamma}\right)\frac{\hat{p}_\parallel + m}{p_\parallel^2 - m^2}, \\ p_\perp^2 &= p_1^2 + p_2^2. \end{aligned} \tag{25c}$$

Using equations (10), this actually decreases the dimension of the space of the mathematical apparatus to (0, 3), since vertex factors of the form

$$(1 - i\gamma_1\gamma_2)(\gamma_\alpha, \gamma_\alpha\gamma^5)(1 - i\gamma_1\gamma_2)$$

are different from zero only for $\alpha = 0, 3$, and $\gamma^0\gamma^3$ plays the role of the matrix γ^5. After calculating the integrals over the coordinates and the perpendicular momentum $p_\perp$, the projection operator $(1 - i\gamma_1\gamma_2)/2$ in the matrix element can be dropped, switching to two-dimensional matrices and convolutions in the subspace (0, 3). Introducing the two-dimensional Green's function

$$\tilde{G}(p) = \frac{\not{p} + m}{p^2 - m^2}, \quad p^2 = p_0^2 - p_3^2 \tag{26}$$

and using the formulas (13) and (14), the matrix element acquires an explicitly two-dimensionally covariant form.

The expression (26) must be modified [17] taking into account the influence of the medium and temperature effects. In studying the processes $\gamma e \longrightarrow ea$ and $\gamma e \longrightarrow e\nu\bar{\nu}$ below, we shall be interested in the case of a degenerate effectively-dimensional electron gas

$$T \ll \mu - m, \tag{27a}$$

$$\mu^2 - m^2 < 2\gamma, \tag{27b}$$

where μ is the chemical potential, which is essentially equal to the Fermi energy $E_F = \sqrt{p_F^2 + m^2}$, and the Fermi momentum p_F is given by the expression [18]

$$p_F = \frac{2\pi^2 n_e}{\gamma}, \tag{28}$$

where n_e is the density. Then the generalized form of $\tilde{G}$ can be obtained from equation (26) using the substitution (see, for example, [19])

$$\frac{1}{p^2 - m^2} \longrightarrow \frac{1}{p^2 - m^2} + 2\pi i\delta(p^2 - m^2)\theta(p_0)\theta(E_F - p_0).$$

However, in the cases considered in the present paper, electrons in the intermediate state are off the mass shell, and the last term makes no contribution in this expression. Thus, the medium changes only the phase volume of the initial and final states (see Section 4).

3. GENERAL EXPRESSIONS FOR THE MATRIX ELEMENTS AND CROSS SECTIONS OF THE PROCESSES $\gamma e \longrightarrow ea$ AND $\gamma e \longrightarrow e\nu\bar{\nu}$

We shall determine the matrix element M of the process $\gamma e \longrightarrow ea$ (diagram in Fig. 1a) in terms of the S matrix element as follows:

$$\langle f|S|i\rangle = i(2\pi)^3\delta^{(0,2,3)}(p + \kappa - p' - k) \times \frac{M}{(2p_0 2\kappa_0 2p_0' 2k_0)^{1/2} L_2 L_3 V}, \tag{29}$$

where p and p' are the momenta of the initial and final electrons and κ and k are the momenta of the photon and axion.

Using the "axion" Lagrangian (2a) and (2b) and the Lagrangian of the electrodynamic interaction

$$\mathcal{L}_e = e(\bar{\Psi}\gamma^\alpha\Psi)A_\alpha, \tag{30}$$

we obtain for M, using the scheme presented in Section 2, a two-dimensionally covariant expression in the subspace (0, 3)

$$M = \frac{i\sqrt{4\pi}emc_e}{f}\bar{v}(p')[\tilde{\gamma}^5\tilde{G}(p + \kappa)\tilde{e} + \tilde{e}\tilde{G}(p' - \kappa)\tilde{\gamma}^5] \times v(p)\exp\left[-i\frac{p_2}{\gamma}(\kappa_1 - k_1)\right], \tag{31}$$

where $\tilde{G}(p)$ is the two-dimensional electron propagator (26), the properties of the two-dimensional spinor $v(p)$ are given by equations (13), (14a), and (14b), and e is the photon polarization vector. The phase factor in the expression (31) has a symbolic significance; its role will be clarified in Section 4.

The formula (31) holds under the condition (23), which in the case at hand assumes the form

$$(p\kappa),\ \kappa^2 \ll \gamma. \tag{32}$$

The matrix element is different from zero only for photon polarization states with electric vector in the momentum–field plane, given in the two-dimensionally covariant notation by the expression

$$e_\alpha = (\varepsilon\kappa)_\alpha/\sqrt{\kappa^2}. \tag{33}$$

The state with orthogonal polarization is sterile.

Making some transformations, the squared matrix element calculated using the properties of $\tilde{\gamma}$ matrices (Section 2) and equations (14b), (16), and (33) can be represented in the form

$$|M|^2 = \frac{16\pi\alpha mc_e^2}{f^2}k^2\left(\frac{p}{\Delta} + \frac{p'}{\Delta'}\right)^2, \tag{34}$$

$$\Delta = 2(p\kappa) + \kappa^2, \quad \Delta' = -2(p'\kappa) + \kappa^2. \tag{34a}$$

The cross section for the process $\gamma e \longrightarrow ea$ can be determined in the standard manner in terms of the S matrix element and is

$$\sigma_a = \frac{1}{8(p\kappa)(2\pi)^2} \times \int\frac{dp_3'}{2p_0'}\int dp_2'\int\frac{d^3k}{2k_0}\delta^{(0,2,3)}(p + \kappa - p' - k)|M|^2, \tag{35}$$

where the integration over dp_2' is trivial and the $\delta^{(2)}$ function vanishes. In the expression (35) we formally averaged over the photon polarization states, which, since the orthogonal polarization is sterile, reduces to division by 2. Integrating over d^3k, we obtain, using $|M|^2$ from equation (34), the distribution over p_3'

$$\sigma_a = \frac{\alpha m^2 c_e^2}{2f^2(p\kappa)}\int_{(\Gamma)}\frac{dp_3'}{2p_0'}k^2\left(\frac{p}{\Delta} + \frac{p'}{\Delta'}\right)^2 \tag{36}$$

with the two-dimensionally covariant region of integration

$$\Gamma = \{k^2 = (p + \kappa - p')^2 \geq m_a^2\}, \tag{37}$$

where m_a is the axion mass.

We note that the matrix element M (with the exception of the formal phase factor) and the cross section do not contain a field dependence; the role of the latter

reduces to "two-dimensionalization of the motion" of an electron, which is a general property of diagrams without excitation of the vacuum.

For $\kappa^2 = \kappa_\perp^2 \geq 4m^2$ the integrand in equation (36) possesses in the region of integration a singularity related with the possibility of e^+e^- pair production by a photon and axion in a magnetic field [20]; in principle, this can be eliminated by taking into account the finite width relative to this "decay."

An analytic calculation can be performed for nonrelativistic electrons and for $\kappa^2 \ll m^2$. In this approximation the cross section has the form

$$\sigma_a = \frac{\alpha c_e^2 \kappa^2}{15 f^2 m^2}, \tag{38}$$

where and below the small axion mass is neglected. Similar results can also be obtained for the photoproduction of a neutrino pair $\gamma e \longrightarrow e\nu\bar{\nu}$ (diagram in Fig. 1b) on an electron in the contact approximation of the Weinberg–Salam model with the effective Lagrangian interaction

$$\mathcal{L}W = -\frac{G}{\sqrt{2}} \times (\bar{\Psi}_e \gamma^\mu (C_V + C_A\gamma^5)\Psi_e)(\bar{\Psi}_\nu \gamma_\mu (1+\gamma^5)\Psi_\nu), \tag{39}$$

where

$$C_V^e = \frac{1}{2} + \sin^2\theta_W, \quad C_A^{(e)} = \frac{1}{2},$$

(θ_W is the Weinberg angle) for electronic neutrinos and $C^{(\mu,\tau)} = C^{(e)} - 1$ for μ and τ neutrinos.

The matrix element M is determined in terms of the S matrix element similarly to the formula (29) with the substitutions

$$k \longrightarrow k + k', \quad 2k_0 \longrightarrow 2k_0 2k_0', \quad V \longrightarrow V^{3/2},$$

where k and k' are the neutrino and antineutrino momenta. Once again we obtain

$$\begin{aligned} M = {} & (2\pi)^{1/2} eG[\bar{u}_\nu(k)\gamma^\alpha(1+\gamma^5)u_\nu(-k')] \\ & \times \exp\left[-\frac{ip_2}{\gamma}(\kappa_1 - k_1 - k_1')\right] \\ & \times \bar{v}(p')[\tilde{\gamma}_\alpha(C_V + C_A\tilde{\gamma}^5)\tilde{G}(p+\kappa)\tilde{e} \\ & + \tilde{e}\tilde{G}(p'-\kappa)\tilde{\gamma}_\alpha(C_V + C_A\tilde{\gamma}^5)]v(p), \end{aligned} \tag{40}$$

where the matrices in the neutrino bracket are, of course, four-dimensional, and the phase factor plays the same role as in equation (31).

The cross section averaged over the photon polarization states can be represented as a distribution over p_3' and has the form ($m_\nu = 0$)

$$\sigma_\nu = \frac{1}{8(2\pi)^5 (p\kappa)} \int\limits_{(Q^2 \geq 0)} \frac{dp_3'}{2p_0'} I, \tag{41}$$

$$Q = p + \kappa - p',$$

$$I = \int \frac{d^3k}{2k_0} \int \frac{d^3k'}{2k_0'} |M|^2 \delta^{(0,3)}(Q - k - k'). \tag{42}$$

The result of the integration over the neutrino momenta is

$$\begin{aligned} I = {} & \frac{8}{3}(2\pi)^3 \alpha G^2 m^2 Q^2 \left\{ (C_V^2 - C_A^2) Q^2 \left(\frac{p}{\Delta} + \frac{p'}{\Delta'}\right)^2 \right. \\ & + 2C_A^2 \left[2(m^2 - (pp'))\left(\frac{p}{\Delta} + \frac{p'}{\Delta'}\right)^2 - \frac{2\kappa^2 (pp')}{\Delta\Delta'} \right. \\ & \left. \left. - 2m^2\left(\frac{p'\kappa}{\Delta^2} - \frac{p\kappa}{\Delta'^2}\right) + (2(pp') + m^2)\left(\frac{1}{\Delta} + \frac{1}{\Delta'}\right)\right]\right\}. \end{aligned} \tag{42a}$$

This same quantity was calculated in [21] but without using the two-dimensionally covariant separation with respect to polarization states (33), and consequently it has a very complicated form.

Just as in the preceding case, an exact calculation of the integral in the expression (41) is impossible, with similar remarks. In the nonrelativistic approximation, however, we have

$$\sigma_\nu = \frac{\alpha G^2 \kappa^2 (p\kappa)^2}{105\pi^2 m^4}(C_V^2 + C_A^2). \tag{43}$$

4. RADIATION INTENSITY WITH AXION AND NEUTRINO PHOTOPRODUCTION ON A NONDEGENERATE AND A DEGENERATE ELECTRON GAS

The key question in astrophysical applications is the neutrino and axion luminosity of stars. To this end we shall find the intensity of the radiation from a unit volume on account of the axion and neutrino production mechanisms considered above.

For a nondegenerate electron gas, in the nonrelativistic approximation the radiation intensity can be found from the expression

$$S = n_e \int \frac{2d^3\kappa}{(2\pi)^3} \kappa_0 \sigma f_B(\kappa_0), \tag{44}$$

where n_e is the electron density,

$$f_B = (\exp(\kappa_0/T) - 1)^{-1}$$

is the Bose–Einstein distribution function ($T \ll m$). Using equations (38) and (43) we obtain for the axion and neutrino radiation intensities

$$S_a = \frac{(2\pi)^4}{2835}\frac{\alpha c_e^2 n_e}{f^2 m^2}T^6, \tag{45a}$$

$$S_\nu = \frac{(2\pi)^4}{4725}\frac{\alpha G^2 n_e}{m^2}T^8(C_V^2 + C_A^2). \tag{45b}$$

To determine the radiation intensity for a degenerate electron Fermi gas, we note first that when integrating over the states of the initial electron, in accordance with the form of the phase factor in the expression (31), there appears an integral of the form

$$J = \frac{1}{L_1}\int_{-\infty}^{\infty}\frac{dp_2}{2\pi}\exp\left[\frac{ip_2}{\gamma}(\kappa_1 - \kappa_1)\right], \tag{46}$$

where L_1 is the effective normalization length along the 1 axis. Formal integration gives

$$J = \frac{1}{L_1}\delta\left(\frac{\kappa_1 - \kappa_1}{\gamma}\right) = \frac{\gamma}{L_1}\delta(0) = \frac{\gamma}{2\pi}. \tag{46a}$$

Then the intensity of the axion radiation from a unit volume is independent of the normalization lengths and is

$$S_a = \frac{2\gamma}{(2\pi)^5}\int d^3\kappa\, f_B(\kappa_0)\int_{-\infty}^{\infty} dp_3 f_F(p_0) \times \int_{(\Gamma)} dp_3'\frac{d\sigma_a}{dp_3'}(p_0 + \kappa_0 - p_0')(1 - f_F(p_0')). \tag{47}$$

Here

$$f_F(p_0) = \left(\exp\frac{(p_0 - \mu)}{T} + 1\right)^{-1}$$

is the Fermi velocity distribution function of the initial electron and $f_F(p_0')$ is the same for the final electron; the distribution of the cross section over p_3' is given by the expression (36).

For what follows, we note that when

$$\left(\frac{p_0}{|p_3|}\right)\left(\frac{\kappa_0}{m}\right)\left(\frac{p_0}{m}\right) \ll 1 \tag{48}$$

the integral over dp_3' in equation (47) can be rewritten in the form

$$J_a = \int_{(\Gamma)} dp_3'\frac{d\sigma_a}{dp_3'}(p_0 + \kappa_0 - p_0')(1 - f_F(p_0')) \simeq \frac{\alpha c_e^2 m^4 \kappa^2}{8 f^2 (p\kappa)^5}\int_0^{x_+} dx[\kappa^2 + 2x(p\varepsilon\kappa) - x^2 m^2]^2 \times (\kappa_0 - p_3 x)\left[1 - \left(\exp\left(\frac{p_0 + p_3 x - \mu}{T}\right) + 1\right)^{-1}\right], \tag{49}$$

where the substitution

$$p_3' = p_3 + x p_0 \tag{50}$$

is made when integrating over dp_3' with the upper limit of integration for x

$$x_+ = \frac{1}{m^2}[(p\kappa) + (p\varepsilon\kappa)] \tag{50a}$$

($\varepsilon_{\alpha\beta}$ is the antisymmetric tensor in (0, 3), introduced in Section 2). The derivation of J_a also took account of the fact that the substitutions $p_3 \longrightarrow -p_3$ and (or) $\kappa_3 \longrightarrow -\kappa_3$ do not change anything because of the subsequent integration over $d^3\kappa$ and dp_3 in equation (47).

Next, we switch to the nonrelativistic limit, i.e., we assume

$$p_F \ll m, \quad T \ll p_F^2/2m \tag{51}$$

together with the condition for applicability of our two-dimensionally covariant method

$$p_F^2 < 2\gamma. \tag{52}$$

Since in this case $\mu \simeq E_F$, we find that the conditions (27) hold automatically, and the restriction (48) is equivalent to the inequality

$$p_F \gg T, \tag{53}$$

which also follows from equation (51).

Thus, the applicability of the formulas obtained below for the intensity of radiation in a degenerate electron gas is limited by the inequalities (51) and (52).

It is easy to see that $|p_3 x|_{\text{eff}} \ll T$ and the terms $p_3 x$ can be dropped in the parentheses in equation (49) and in the argument of the exponential function. Then, using equation (16), the integral over dx in equation (49) is elementary, and in the calculation of the integral over dp_3 the limit $|p_3| \longrightarrow p_F$ can be used everywhere with the exception of the argument of the exponential func-

tions in $f_F(p_0) \simeq f_F(p_0')$, which in the region of overlapping of the Fermi steps is

$$\frac{p_0 - E_F}{T} \simeq \frac{p_3^2 - p_F^2}{2mT} \simeq \frac{p_F}{mT}(|p_3| - p_F).$$

In consequence, the integrals over $d^3\kappa$ and dp_3 in equation (47) factorize and likewise can be easily calculated. The final result for the intensity of the axion radiation in a degenerate nonrelativistic electron gas has the form

$$S_a = \frac{2(2\pi)^2}{2835}\frac{\alpha c_e^2 T^7 \gamma}{f^2 m p_F}, \tag{54}$$

where the value of the Fermi momentum is presented in equation (28).

The intensity of the neutrino radiation S_ν is determined by a formula of the form (47) with the cross section taken from equations (41) and (42a), and the integral similar to the expression (49) in the approximation (48) has the form

$$J_\nu \simeq \frac{\alpha G^2 m^4 (C_V^2 + C_A^2)\kappa^2}{12(2\pi)^2 (p\kappa)^5} \times \int_0^{x_+} dx[\kappa^2 + 2x(p\varepsilon\kappa) - x^2 m^2]^3 \times (\kappa_0 - p_3 x)\left[1 - \left(\exp\frac{(p_0 + p_3 x - \mu)}{T} + 1\right)^{-1}\right]. \tag{55}$$

The radiation intensity with the inequalities (51) and (52) is

$$S_\nu = \frac{2(2\pi)^2}{4725}\frac{\alpha G^2 (C_V^2 + C_A^2) T^9 \gamma}{m p_F}. \tag{56}$$

5. DISCUSSION

For astrophysical applications it is convenient to represent the restrictions (51) and (52) (using equation (28)) for our results for the axion and neutrino radiation intensities in a degenerate, nonrelativistic, effectively two-dimensional, electron gas in the following form:

$$(n_e \times 10^{-30}\ \text{cm}^3) \ll \frac{B}{B_0}, \tag{57a}$$

$$\frac{T}{m} \ll 0.6(n_e \times 10^{-30}\ \text{cm}^3)\left(\frac{B}{B_0}\right)^2, \tag{57b}$$

$$0.64(n_e \times 10^{-30}\ \text{cm}^3)^2 < \left(\frac{B}{B_0}\right)^3. \tag{57c}$$

The surface photon luminosity of white dwarfs for the dominant cooling mechanism per unit mass of the star is [6]

$$\frac{\mathcal{L}_\gamma}{M} \simeq 3.3 \times 10^{-3}\left(\frac{T}{10^7\ \text{K}}\right)^{7/2}\ \text{erg/(g s)}. \tag{58}$$

The average density of white dwarfs is $\rho \simeq 10^6$ g/cm^3, and the corresponding axion luminosity S_a/ρ can be rewritten, using equation (54), in the form

$$\frac{S_a}{\rho} \simeq 9.2 \times 10^{-12} c_e^2 \left(\frac{T}{10^7\ \text{K}}\right)^7 \left(\frac{B}{10^{13}\ \text{G}}\right)^2 \times \frac{(10^{10}\ \text{GeV}/f)^2}{n_e \times 10^{-30}\ \text{cm}^3}\ \text{erg/(g s)}. \tag{59}$$

For the average temperature $T \simeq 10^7$ K and electron density $n_e \simeq 10^{29}$ cm^{-3} we assume that the induction in the interior regions is $B \simeq 10^{13}$ G. Then our restrictions (57) hold approximately, and since there is no synchrotron mechanism $e \longrightarrow ea$ (all electrons are in the ground Landau level) the Compton mechanism of axion emission in principle can be the dominant mechanism. From the condition that the existing notions concerning the character of the evolution of white dwarfs remain unchanged, the inequality

$$\mathcal{L}_\gamma/M \gtrsim S_a/\rho,$$

should be satisfied, whence we obtain in order of magnitude

$$f \gtrsim 0.5 c_e \times 10^7\ \text{GeV}. \tag{60}$$

This is two orders of magnitude less than the lower limit on f, determined from equations (3) and (4), and is equal in order of magnitude to the results obtained in [9], where the synchrotron mechanism of axion emission in neutron stars was studied (equation (26) of [9] contains an obvious error: GeV should appear instead of eV). If the lower limit $f \gtrsim 0.6 \times 10^9$ GeV obtained from equations (3) and (4), is correct, then our result indeed means that the axion luminosity of white dwarfs under the conditions considered is at least four orders of magnitude weaker than the photon luminosity.

For the shells of magnetic neutron stars, the possible ranges of the parameters are as follows: $n_e \sim 10^{30\text{–}37}$ cm^{-3}, $T \sim 10^{8\text{–}10}$ K, and $B \sim 10^{12\text{–}14}$ G. The limitations (57) introduced above are approximately satisfied, if $n_e \simeq 10^{30}$ cm^{-3}, $T \simeq 10^8$ K, and $B \simeq 3B_0$. Since the luminosity of neutron stars at the early stages of their evolution is essentially neutrino luminosity, the quantities S_a (54) and S_ν (56) should be equal:

$$\frac{S_a}{S_\nu} = \frac{5}{3}\frac{c_e^2}{G^2 T^2 (C_V^2 + C_A^2) f^2}. \tag{61}$$

It is interesting that this relation is the same for the nondegenerate case determined by equations (45a) and

(45b). We find from the condition $S_a \lesssim S_\nu$ for the indicated values of the parameters

$$f \gtrsim \left(\frac{c_e^2}{C_V^2 + C_A^2}\right)^{1/2} \times 1.3 \times 10^{10} \text{ GeV}. \qquad (62)$$

This falls in the range of the possible values of f (3) and (4), and therefore the axion luminosity of magnetic neutron stars with a degenerate, effectively two-dimensional, electron gas can approach the neutrino luminosity of these stars.

REFERENCES

1. R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977); Phys. Rev. D **16**, 1791 (1977).
2. G. G. Raffelt, Phys. Rep. **198**, 1 (1990).
3. S. Weinberg, Phys. Rev. Lett. **40**, 223 (1978); F. Wilczek, Phys. Rev. Lett. **40**, 279 (1978).
4. J. Preskill, M. Wise, and F. Wilczek, Phys. Lett. B **120**, 127 (1983); L. Abbott and P. Sikivie, Phys. Lett. B **120**, 133 (1983); M. Dine and W. Fischler, Phys. Lett. B **120**, 137 (1983); M. S. Turner, Phys. Rev. D **33**, 889 (1986).
5. M. S. Turner, Phys. Rep. **197**, 67 (1990).
6. G. Raffelt, *Stars as Laboratories for Fundamental Physics* (Univ. of Chicago Press, Chicago, 1996).
7. A. V. Borisov and V. Yu. Grishina, Zh. Éksp. Teor. Fiz. **106**, 1553 (1994) [JETP **79**, 837 (1994)].
8. M. Kachelriess, S. Wilke, and G. Wunner, Phys. Rev. D **56**, 1313 (1997).
9. V. V. Skobelev, Zh. Éksp. Teor. Fiz. **112**, 25 (1997) [JETP **85**, 13 (1997)].
10. V. V. Skobelev, Zh. Éksp. Teor. Fiz. **116**, 26 (1999) [JETP **89**, 13 (1999)]; Zh. Éksp. Teor. Fiz. **116**, 2271 (1999) [JETP **89**, 1200 (1999)].
11. A. V. Borisov and K. V. Zhukovskiĭ, Yad. Fiz. **58**, 1298 (1995) [Phys. At. Nucl. **58**, 1218 (1995)]; A. V. Borisov and V. Yu. Grishina, Vestn. Mosk. Univ., Ser. 3: Fiz., Astron., No. 4, 24 (1996); A. V. Borisov and V. Yu. Grishina, Zh. Éksp. Teor. Fiz. **110**, 1575 (1996) [JETP **83**, 868 (1996)].
12. V. V. Skobelev, Zh. Éksp. Teor. Fiz. **71**, 1263 (1976) [Sov. Phys. JETP **44**, 660 (1976)]; Zh. Éksp. Teor. Fiz. **72**, 1298 (1977) [Sov. Phys. JETP **45**, 682 (1977)]; Doctoral Dissertation in Mathematical Physics (Mosk. Gos. Univ., Moscow, 1982).
13. N. P. Klepikov, Zh. Éksp. Teor. Fiz. **26**, 19 (1954).
14. Yu. M. Loskutov and V. V. Skobelev, Vestn. Mosk. Univ., Ser. 3: Fiz., Astron., No. 6, 111 (1977).
15. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Nauka, Moscow, 1971; Academic, New York, 1980).
16. V. B. Berestetskii, E. M. Lifshitz and L. P. Pitaevskiĭ, in *Course of Theretical Physics*, Vol. 4: *Quantum Electrodynamics*, 2nd ed. (Pergamon, Oxford, 1982; Nauka, Moscow, 1980).
17. Yu. M. Loskutov, A. S. Vshivtsev, and V. V. Skobelev, Teor. Mat. Fiz. **84**, 372 (1990).
18. Yu. M. Loskutov and V. V. Skobelev, Yad. Fiz. **43**, 1495 (1986) [Sov. J. Nucl. Phys. **43**, 964 (1986)].
19. A. V. Borisov and P. E. Sizin, Zh. Éksp. Teor. Fiz. **115**, 3 (1999) [JETP **88**, 1 (1999)].
20. Yu. M. Loskutov and V. V. Skobelev, Vestn. Mosk. Univ., Ser. 3: Fiz., Astron., No. 4, 387 (1976); V. V. Skobelev, Yad. Fiz. **61**, 2236 (1998) [Phys. At. Nucl. **61**, 2123 (1998)].
21. Yu. M. Loskutov and V. V. Skobelev, Vestn. Mosk. Univ., Ser. 3: Fiz., Astron., No. 6, 4 (1980).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 927–938.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1067–1079.

NUCLEI, PARTICLES, AND THEIR INTERACTION

The Process $e^+e^- \longrightarrow \omega\pi^0$ near the ϕ Resonance

V. M. Aulchenko, M. N. Achasov, K. I. Beloborodov, A. V. Berdyugin, A. G. Bogdanchikov, A. V. Bozhenok, A. D. Bukin, D. A. Bukin, S. V. Burdin, A. V. Vasiljev, I. A. Gaponenko, V. B. Golubev, T. V. Dimova, A. A. Drozdetsky, V. P. Druzhinin*, M. S. Dubrovin, P. M. Ivanov, V. N. Ivanchenko, I. A. Koop, A. A. Korol, S. V. Koshuba, G. A. Kukartsev, A. P. Lysenko, A. V. Otboev, E. V. Pakhtusova, A. A. Salnikov, S. I. Serednyakov, V. A. Sidorov, Z. K. Silagadze, A. N. Skrinsky, V. V. Shary, and Yu. M. Shatunov

Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences, and Novosibirsk State University, Novosibirsk, 630090 Russia
**e-mail: druzhinin@inp.nsk.su*

Received February 3, 2000

Abstract—The reaction $e^+e^- \longrightarrow \omega\pi^0$ near a ϕ resonance was studied with a spherical neutral detector at VEPP-2M e^+e^- collider. Both main modes of decay of a ω meson were investigated: $\omega \longrightarrow \pi^+\pi^-\pi^0$ and $\omega \longrightarrow \pi^0\gamma$. The probability of decay $\phi \longrightarrow \omega\pi^0$ was obtained from the magnitude of the interference wave in the cross section for the reaction $e^+e^- \longrightarrow \omega\pi^0 \longrightarrow \pi^+\pi^-\pi^0\pi^0$: $B(\phi \longrightarrow \omega\pi^0) = (5.5^{+1.6}_{-1.4} \pm 0.3) \times 10^{-5}$. The ratio of the partial widths of the ω meson was obtained from the ratio of the cross sections for the two modes: $\Gamma(\omega \longrightarrow \pi^0\gamma)/\Gamma(\omega \longrightarrow \pi^+\pi^-\pi^0) = 0.0994 \pm 0.0036 \pm 0.0038$.

1. INTRODUCTION

Substantial progress has been made in the last few years in the study of rare decays of the ϕ meson in experiments at VEPP-2M collider with a spherical neutral detector (SND) and CMD-2 detector. Among the rare decays, the class of OZI suppressed and *G*-parity breaking decays $\phi \longrightarrow X$, where $X = \omega\pi^0$ [1], $\pi^+\pi^-\pi^+\pi^-$ [2], and $\pi^+\pi^-$ [2, 3], stands out. Due to the presence of a relatively large amplitude for the corresponding nonresonance processes $e^+e^- \longrightarrow X$, these decays are manifested in the form of an interference pattern in the energy dependence of the cross section near the ϕ resonance. The Born cross section, taking account of interference, can be represented in the form

$$\sigma(E) = \sigma_0(E)\left|1 - Z\frac{m_\phi\Gamma_\phi}{D_\phi}\right|^2, \qquad (1)$$

where $\sigma_0(E)$ is the cross section for the nonresonance process, Z is a complex interference parameter that is equal to the ratio of the decay amplitude to the amplitude of the nonresonance process, m_ϕ and Γ_ϕ are the mass and width and $D_\phi = m_\phi^2 - E^2 - iE\Gamma_\phi(E)$ is the inverse propagator of the ϕ meson, and E is the energy of e^+e^- in the center-of-mass system. In this parameterization all energy dependences except the resonance dependence are neglected in the second term. The relative probability of the decay $\phi \longrightarrow X$ is proportional to the squared modulus of the interference parameter and the magnitude of the nonresonance cross section at the maximum of the ϕ resonance:

$$B(\phi \longrightarrow X) = \frac{\sigma_0(m_\phi)|Z|^2}{\sigma_\phi}. \qquad (2)$$

Here

$$\sigma_\phi = 12\pi B(\phi \longrightarrow e^+e^-)/m_\phi^2 = 4220 \pm 110 \text{ nb}$$

is the cross section for the production of the ϕ meson in e^+e^- collisions [15]. The imaginary and real parts of the interference parameter can be determined by analyzing the experimentally measured cross section, i.e., one can measure not only the decay probability but also the relative phase of the resonance and nonresonance amplitudes.

The simplest and most natural mechanism for *G*-parity breaking in decays of the ϕ meson is the electromagnetic transition $\phi - \gamma - \rho, \rho'$. This mechanism contributes only to the real part of the interference parameter and is identical for all processes under discussion:

$$\text{Re}(Z)_\gamma = 3B(\phi \longrightarrow e^+e^-)/\alpha = 0.123.$$

Other decay mechanisms contribute to the real and imaginary parts of the parameter Z. The experimental separation of the real and imaginary parts of the amplitude makes it possible to clarify the question of the contribution of various mechanisms to the decay.

The present work is devoted to a study of the process $e^+e^- \longrightarrow \omega\pi^0$ near the ϕ resonance. The only previous measurement of the decay $\phi \longrightarrow \omega\pi^0$ was made in an

experiment at the VEPP-2M e^+e^- collider with the SND detector while studying the reaction

$$e^+e^- \longrightarrow \omega\pi^0 \longrightarrow \pi^+\pi^-\pi^0\pi^0$$

(see [1]). The real and imaginary parts of the parameter Z were measured and the following value was obtained for the decay probability:

$$(4.8^{+1.9}_{-1.7} \pm 0.8) \times 10^{-5}.$$

The neutral channel

$$e^+e^- \longrightarrow \omega\pi^0 \longrightarrow \pi^0\pi^0\gamma$$

has also been studied using an SND detector [4]. In this channel, together with decay $\phi \longrightarrow \omega\pi^0$, the decay $\phi \longrightarrow \rho^0\pi^0$, $\rho^0 \longrightarrow \pi^0\gamma$ contributes to interference. In [4] the interference in the reaction $e^+e^- \longrightarrow \omega\pi^0 \longrightarrow \pi^0\pi^0\gamma$ was observed at the level of three standard deviations and its value was consistent with the sum of the two contributions mentioned. Both analyses [1, 4] were based on the data accumulated in 1996 by the SND detector. In the present paper we present the results on an analysis of the 1998 data to check the preceding measurements and to improve the accuracy.

2. DETECTOR, EXPERIMENT

The SND detector [5] is a universal nonmagnetic detector, whose main part is a three-layer calorimeter based on 1630 NaI(Tl) crystals. The energy resolution of the calorimeter for photons is

$$\sigma_E/E = 4.2\%/\sqrt[4]{E[\text{GeV}]}$$

and the angular resolution is about 1.5°. The solid angle of the calorimeter is 90% of 4π. The angles of the charged particles are measured by a system of two drift chambers. The measurement accuracy is about 0.5° for the azimuthal angle and about 2° for the polar angle. The solid angle of the system of drift chambers is 95% of 4π.

The data sample accumulated in 1998 at the VEPP-2M e^+e^- collider was investigated in this analysis. The experiment was performed by the method of scanning of the energy interval E = 984–1060 MeV. The data were collected with 16 values of the beam energy. Two scans were made with a total integrated luminosity of about 8 pb^{-1}, corresponding to about 1.2×10^7 produced ϕ mesons. The luminosity was measured by using the processes $e^+e^- \longrightarrow e^+e^-$ and $e^+e^- \longrightarrow \gamma\gamma$. The systematic error of the normalization to the luminosity was 2%.

3. THE REACTION $e^+e^- \longrightarrow \omega\pi^0 \longrightarrow \pi^0\pi^0\gamma$

3.1. Selection of Events

The preliminary selection of events of the reaction

$$e^+e^- \longrightarrow \omega\pi^0 \longrightarrow \pi^0\pi^0\gamma \qquad (3)$$

was made according to the following criteria:

(1) the number of photons is greater than or equal to five and there are no tracks in the drift chamber;

(2) the energy deposition in the calorimeter is greater than $0.7E$;

(3) the total momentum in an event, measured using the calorimeter, is less than $0.15E$.

Besides the events corresponding to the reaction under investigation, events of the following processes also fall into the same class:

$$e^+e^- \longrightarrow \phi \longrightarrow K_SK_L \longrightarrow \textit{neutrals}, \qquad (4)$$

$$e^+e^- \longrightarrow \phi \longrightarrow \eta\gamma,\ \eta \longrightarrow 3\pi^0, \qquad (5)$$

$$e^+e^- \longrightarrow \phi \longrightarrow \pi^0\pi^0\gamma. \qquad (6)$$

Even though the probability is low ~10^{-4} [6, 7], the decay (6) is one of the main background processes, since its final state is the same as for the process (3). To simulate process (6), we used the model of the transition through the intermediate state $f_0\gamma$, where f_0 is a scalar particle, and the recoil photon spectrum was taken from [6]. Since the recoil photon in the process (6) is predominantly soft, the interference between the processes (6) and (3) is weak.

A high beam background in the detector results in the appearance of extra photons in some events. The fraction of such events in the 1998 experiment was about 8%. An additional background from the following processes arose because of superpositions:

$$e^+e^- \longrightarrow 2\gamma,\ 3\gamma. \qquad (7)$$

The angular and energy distributions of the spurious photons were studied using special events with random triggering from an external generator. These events were admixed to the main flux of events during the experiment. Analysis of these events showed that the energy of the most spurious photons lies near the detecting threshold 20 MeV and is concentrated near the edges of the calorimeter at small angles with respect to the axis of the beam. Information about the detector channels triggered in events with triggering from the generator was used to simulate the process under study and the background processes in order to imitate the superpositions which occur in a real experiment.

The procedure of kinematic reconstruction was applied to the events which satisfied the preliminary selection conditions, i.e., the particle parameters (angles and energies) for which the laws of conservation of energy and momentum as well as the condition for the presence of intermediate particles in an event are satisfied, were found by the maximum likelihood method. The kinematic reconstruction was performed for two hypotheses:

(1) an event refers to the process $e^+e^- \longrightarrow 3\gamma$;

(2) an event refers to the process $e^+e^- \longrightarrow \pi^0\pi^0\gamma$.

In events in which the number of photons was greater than the requirement for the hypothesis considered, the extra photons were assumed to be spurious and were

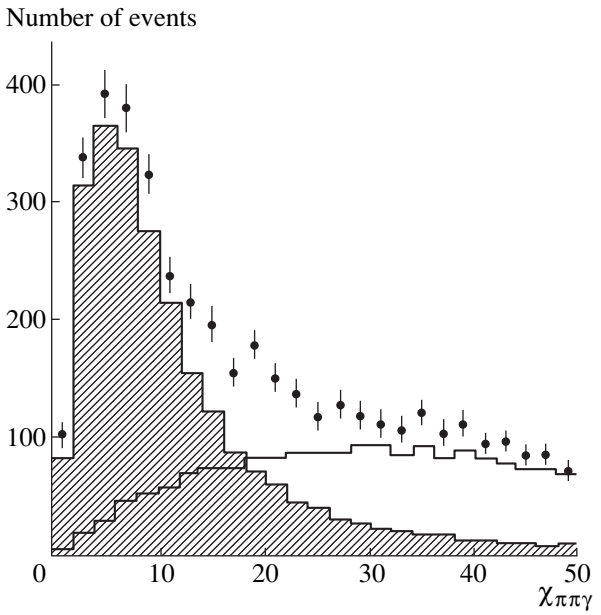


Fig. 1. Distributions over the parameter $\chi_{\pi\pi\gamma}$ for the experimental events (points with error bars) and events from the simulation of processes (3) (hatched region) and (5) (line).

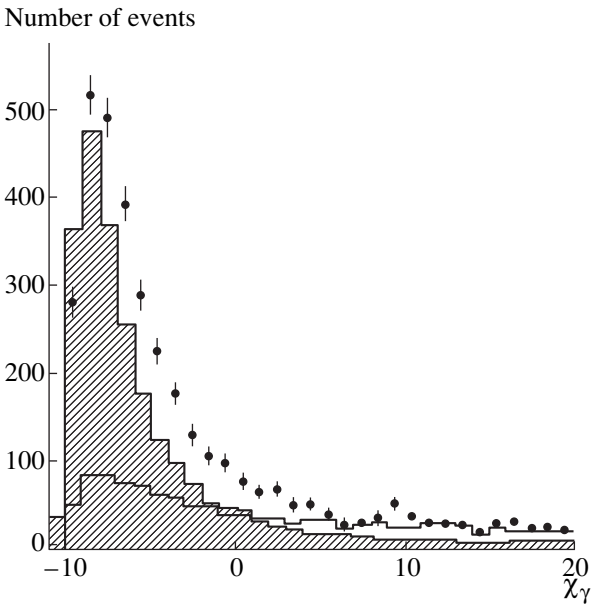


Fig. 2. Distributions over the parameter χ_γ for the experimental events (points with error bars) and events from the simulation of processes (3) (hatched region) and (5) (line).

discarded. All possible combinations with the correct number of photons were checked and the best combination, which minimized the value of the function χ^2, was chosen. As a result of the kinematic reconstruction, the values $\chi_{3\gamma}$ and $\chi_{\pi\pi\gamma}$ of the χ^2 functions for both hypotheses and the recoil mass $M_{\pi\gamma}$ of π^0 closest to the mass of the ω meson for the process $e^+e^- \longrightarrow \pi^0\pi^0\gamma$ were calculated. The condition

$$\chi_{3\gamma} > 25$$

was used to suppress the background from the processes (7).

Subsequently, the events where the number of photons was equal to and greater than five were analyzed separately. The distributions over the parameter $\chi_{\pi\pi\gamma}$ for the experimental events and events simulating the processes (3) and (5) with five photons are presented in Fig. 1. The parameter $\chi_{\pi\pi\gamma}$ was required to satisfy the restriction

$$\chi_{\pi\pi\gamma} < 40.$$

The criteria described above select events satisfying the laws of conservation of energy and momentum. Consequently, the selected events of the process (5), in which there are seven photons in the final state, become part of the class of five-photon events mostly because close photons merge. This is also true of the process (4). Events of this process, where the K_L meson decayed inside the calorimeter into three π^0 mesons (the K_S meson decays into two π^0 mesons inside the vacuum chamber), are selected; the ten photons from π^0 decays with low probability can give a configuration similar to a five-photon event. Additional suppression of the background from the processes (5) and (4) was achieved by analyzing the transverse distributions of the energy deposition in showers from the detected photons. A logarithmic likelihood function for the transverse distribution of the energy deposition was constructed for each photon in an event [8]. The parameter χ_γ was the maximum value of the logarithmic likelihood function among five photons. The distributions over this parameter for the experimental events and the simulated events, which passed through the preceding selections, are presented in Fig. 2. This parameter was required to satisfy the stringent restriction

$$\chi_\gamma < -4.$$

The agreement between experiment and the simulation near the value $\chi_\gamma = -4$ is not very good, and consequently the class of events with $\chi_\gamma > -4$ was also analyzed but with more stringent selection conditions with respect to other parameters.

The distributions over the parameter $M_{\pi\gamma}$ for the experimental and simulated events are presented in Fig. 3. It is evident that the agreement between the simulated and experimental distributions is good both on the pedestal and in the region of the peak. In calculating the distribution of background events, the distributions of all background processes described above were summed. At this stage of selection the background is determined by the processes (5) and (6) in approximately the same proportion. Events satisfying the restriction

$$|M_\omega - 782| < 50$$

were used for subsequent analysis.

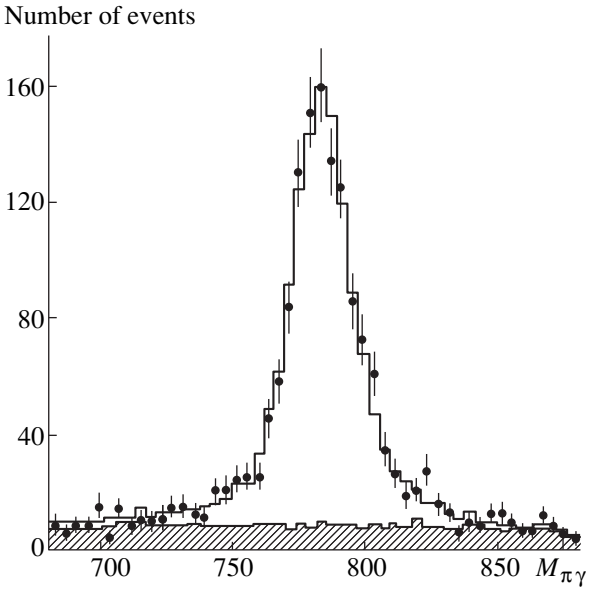


Fig. 3. Distribution over the parameter $M_{\pi\gamma}$ for the experimental events (points with error bars) and events from the simulation of the background processes (hatched histogram). The line is the sum of process (3) and the background.

Events with more than five photons were required to satisfy the following conditions:

$$\chi_{\pi\pi\gamma} < 20, \quad |M_\omega - 782| < 25.$$

The same restrictions were used for the five-photon events with $\chi_\gamma > -4$.

The detection efficiency of the events of process (3) with the conditions

$$\chi_{\pi\pi\gamma} < 40, \quad |M_\omega - 782| < 50 \qquad (8)$$

and with the number of photons greater than or equal to five was calculated using simulation and was $\varepsilon = 38\%$. To estimate the systematic error in this value, events with seven or more photons were analyzed with the selection conditions close to (8). The restrictions were imposed on the likelihood function and the recoil mass of the photon, which were obtained in the kinematic reconstruction using the hypothesis $e^+e^- \longrightarrow 3\pi^0\gamma$. The measured value of the decay probability

$$B(\phi \longrightarrow \eta\gamma) = (1.26 \pm 0.01)\%$$

was found to be equal to the tabulated value $1.26 \pm 0.06\%$. Consequently, we believe that the systematic error in determining the detection efficiency for the process (3) with the conditions (8) does not exceed the total error of the measured and tabulated values of $B(\phi \longrightarrow \eta\gamma)$, which is equal to 5%.

3.2. Analysis of the Data

The events selected according to the criteria described above were divided into six classes:

1. $N_\gamma = 5$, $\chi_\gamma < -4$, $\chi_{\pi\pi\gamma} < 20$, $|M_{\pi\gamma} - 782| < 25$,
2. $N_\gamma = 5$, $\chi_\gamma < -4$, $\chi_{\pi\pi\gamma} < 20$, $|M_{\pi\gamma} - 782| > 25$,
3. $N_\gamma = 5$, $\chi_\gamma < -4$, $\chi_{\pi\pi\gamma} > 20$, $|M_{\pi\gamma} - 782| < 25$,
4. $N_\gamma = 5$, $\chi_\gamma < -4$, $\chi_{\pi\pi\gamma} > 20$, $|M_{\pi\gamma} - 782| > 25$,
5. $N_\gamma = 5$, $\chi_\gamma > -4$, $\chi_{\pi\pi\gamma} < 20$, $|M_{\pi\gamma} - 782| < 25$,
6. $N_\gamma = 5$, $\chi_{\pi\pi\gamma} < 20$, $|M_{\pi\gamma} - 782| < 25$.

Class number 1 contains about half the events selected according to the condition (8). This class is used as the main class for determining the parameters of the process (3) under investigation. The number of events of the process (3) is 10 times less in the class 2; this makes it possible to determine from the experimental data the magnitude of the resonance background. The background level in class 1 can be estimated using the ratio of the levels of the resonance background in the classes 1 and 2, as calculated from the simulation. The value of this ratio was found to be close to 1:

$$k_{12} = 0.97 \pm 0.07$$

The error presented is determined by the statistics of the simulation of the background processes. The systematic error was estimated by varying over wide limits the restrictions on the parameters $\chi_{\pi\pi\gamma}$ and χ_γ and did not exceed 3%. Since quite stringent restrictions on the parameters $\chi_{\pi\pi\gamma}$ and χ_γ are used to select events in classes 1 and 2, the probability that events for the process under study fall into these classes, as determined by the simulation, can contain a large systematic error. The classes 3–6 were used to eliminate the influence of this factor.

The cross section for detecting the experimental events in each class was approximated by the following formula:

$$\sigma_{vis}(E) = \alpha_i \varepsilon \sigma_{\omega\pi}(E)(1 + \delta_{\omega\pi}) + \sigma_{\phi i}(E),$$

where $\sigma_{\omega\pi}(E)$ is the Born cross section of the process $e^+e^- \longrightarrow \omega\pi^0 \longrightarrow \pi^0\pi^0\gamma$, $\delta_{\omega\pi}$ is the radiative correction to this cross section [9], and $\sigma_{\phi i}(E)$ is the cross section for the resonance background in the ith class (the shape of the excitation curve for the ϕ meson was described by the energy dependence of the cross section of the main background process $e^+e^- \longrightarrow \phi \longrightarrow \eta\gamma$ taking into account radiation corrections), ε is the detection efficiency for events of the process (3) with the conditions (8), and α_i is the probability that events of the process (3) which were selected according to the conditions (8) fall into the ith class. The coefficients α_i are related with one another by the following relation:

$$(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)\left(1 + \frac{\alpha_5 + \alpha_6}{\alpha_1}\right) = 1. \qquad (9)$$

The second cofactor in this relation compensates the loss of events due to additional requirements as compared with the selection (8):

$$N_\gamma = 5, \quad \chi_\gamma < -4.$$

Table 1. The results of the fit to the cross section for process (3)

	$\sigma_0(m_\phi)$, nb	A	Re Z	Im Z	χ^2/ND
Model 1	0.744 ± 0.021	-0.114 ± 0.07	0.01 ± 0.05	-0.19 ± 0.07	82.2/74
Model 2	0.745 ± 0.023	-0.150 ± 0.07	0.01 ± 0.05	-0.19 ± 0.07	83.5/74

The calculation of the correction by this method presumes that the distributions over the parameters N_γ and χ_γ do not depend on the restrictions on $\chi_{\pi\pi\gamma}$ and $M_{\pi\gamma}$. In reality, this dependence is present, and taking it into account by means of simulation changes the efficiency ε by 2%. This value was also added to the systematic error in the efficiency.

The formula (1) was used to describe the cross section $\sigma_{\omega\pi}(E)$. In our preceding work [1], linear and quadratic functions were used to parameterize the nonresonant cross section $\sigma_0(E)$. The dependence on the choice of function describing the nonresonant cross section gave the largest contribution to the systematic error of the parameters of the decay $\phi \longrightarrow \omega\pi$. Consequently, in the present work we focused on the selection of a model for $\sigma_0(E)$. It is known that the process $e^+e^- \longrightarrow \omega\pi^0$ in the energy range 1–2 GeV is described by a sum of the contributions of $\rho(770)$ and ρ' mesons with relative phase 180° [10], so that its cross section can be represented in the form

$$\sigma_0(E) = \frac{4\pi\alpha^2}{E^3}\left(\frac{g_{\rho\omega\pi}}{f_\rho}\right)^2\left|\frac{m_\rho^2}{D_\rho} + A\frac{m_{\rho'}^2}{D_{\rho'}}\right|^2 P_f(E), \qquad (10)$$

where $\alpha = 1/137$, $g_{\rho\omega\pi}$ is the constant for the transition $\rho \longrightarrow \omega\pi$, f_p is the coupling constant between the ρ meson and photon and can be calculated from the width of the decay $\rho \longrightarrow e^+e^-$:

$$\Gamma_{\rho ee} = 4\pi m_\rho \alpha^2/3 f_\rho^2;$$

m_ρ, $m_{\rho'}$, D_ρ, and $D_{\rho'}$ are, respectively, the masses and inverse propagators for the ρ and ρ' mesons:

$$D_\rho = m_\rho^2 - E^2 - iE\Gamma_\rho(E);$$

A is a negative real number, equal to the ratio of the coupling constants of ρ and ρ' mesons,

$$A = \frac{g_{\rho'\omega\pi}}{g_{\rho\omega\pi}}\frac{f_\rho}{f_\rho'}.$$

The factor $P_f(E)$ describes the energy dependence of the phase volume of the final state. In the approximation of an infinitely narrow ω meson

$$P_f(E) = (1/3)p_\omega^3 B(\omega \longrightarrow \pi^0\gamma),$$

where p_ω is the momentum of the ω meson, $B(\omega \longrightarrow \pi^0\gamma)$ is the relative probability of the decay $\omega \longrightarrow \pi^0\gamma$. However, in the energy range which we studied, and especially in the range near the threshold of the reaction $e^+e^- \longrightarrow \omega\pi^0$, more accurate formulas that take into account the finite width of the ω meson must be used. It should be noted that the energy behavior of the factor $P_f(E)$ is different for the two decay channels studied for the decay of the ω meson because of the strong dependence of the width of the decay $\omega \longrightarrow \rho\pi$ on the mass of ω. The energy dependence of the factor $P_f(E)$ for the process $e^+e^- \longrightarrow \omega\pi^0 \longrightarrow \pi^0\pi^0\gamma$ was calculated using the formulas from [4].

The main parameters of a ρ' meson, such as, the mass, width, and relative decay probabilities into various final states, are known only very poorly [15]. This is due to not only the lack of accurate experimental data but also the model uncertainty in their description. For example, the values obtained in the recent work [11] for the mass and width of ρ' are substantially different from the tabulated values, and the difference increases substantially when the mass dependence of the width of the resonances is taken into account. We used two variants for the parameters of a ρ' meson:

(1) $m_{\rho'} = 1465$ MeV, $\Gamma_{\rho'} = 310$ MeV [15]; $\Gamma_\rho(E) = \Gamma_\rho(m_\rho)(m_\rho/E)(p_\pi(E)/p_\pi(m_\rho))^3$, where $p_\pi(E)$ is the momentum of the pion in the decay $\rho \longrightarrow 2\pi$. The energy dependence of $\Gamma_{\rho'}$ was neglected.

(2) $m_{\rho'} = 1530$ MeV, $\Gamma_{\rho'} = 430$ MeV. The energy dependence of Γ_ρ and $\Gamma_{\rho'}$ was neglected.

The second model describes well the spectrum of invariant masses of the state $\omega\pi$ in the decay $\tau \longrightarrow \omega\pi\nu_\tau$ in the range from 0.9 to 1.7 GeV [11]. Using two models makes it possible to estimate the systematic error associated with the choice of parameters in the formula (10).

A fit to the detection cross section was obtained in all six classes simultaneously. In each class the cross section was measured at 16 energy points. In addition, our preliminary data on the measurement of the cross section $e^+e^- \longrightarrow \omega\pi^0 \longrightarrow \pi^0\pi^0\gamma$ at 24 points in the range 920–1400 MeV [12] were included in the approximation. The free parameters of the approximation were $\sigma_0(m_\phi)$, A, ReZ, ImZ, α_i, and $\sigma_{\phi i}(m_\phi)$ with the exception of α_3 and $\sigma_{\phi 1}(m_\phi)$. The value of $\sigma_{\phi 1}(m_\phi)$ was found from the relation

$$\sigma_{\phi 1}(m_\phi) = (0.97 \pm 0.08)\sigma_{\phi 2}(m_\phi),$$

where α_3 was obtained from equation (9). The detection efficiency ε and the coefficients α_i were essentially energy-independent in the range studied. A total of 14 parameters were determined in the fit. The values of the four main parameters for both models are presented in Table 1. Of all parameters of the fitting function, only A depends on the model. The values obtained for the

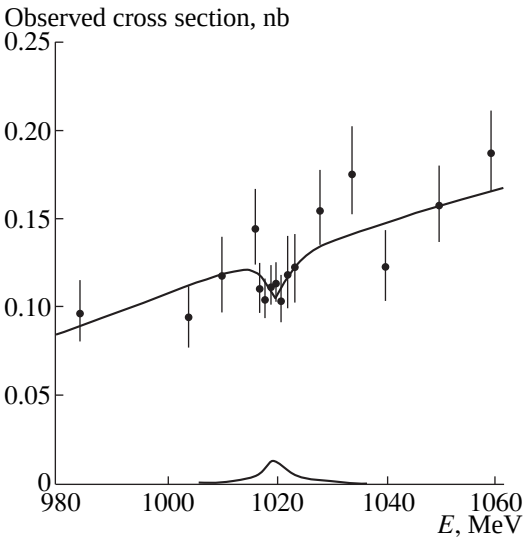


Fig. 4. Energy dependence of the cross section for detecting class-1 events and the fitted curve. The curve at the bottom was obtained by fitting the cross section for the resonance background.

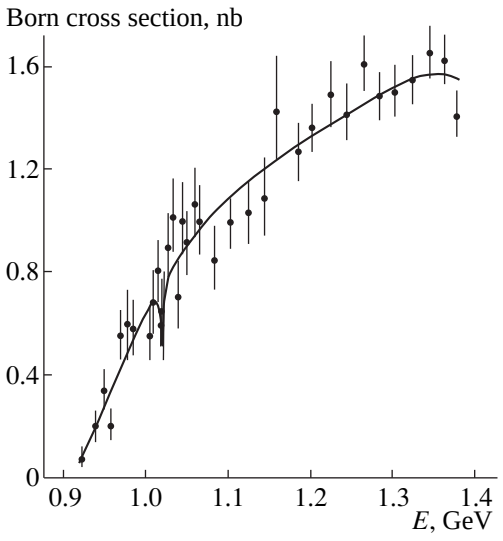


Fig. 5. Energy dependence of the Born cross section for the reaction $e^+e^- \longrightarrow \omega\pi^0 \longrightarrow \pi^0\pi^0\gamma$ and the fitted curve.

resonance cross section $\sigma_{\phi i}(m_\phi)$ agree statistically with the simulation calculation. The agreement with the simulation is worse for the coefficients α_i. The difference is primarily due to the above-discussed inadequate accuracy in the simulation of the distribution with respect to the parameter χ_γ. The experimental data and fitting curves for the narrow energy range near the ϕ meson and the complete interval from 920 to 1400 MeV are presented in Figs. 4 and 5. The curves show the result of the fit in the model 2. It is evident that in both cases the fitting function describes the experimental data well. The parts referring to class 1 and to the energy range outside of the ϕ meson resonance can be extracted from the total value of χ^2: 9.6/12 and 20/22, respectively. Here the calculation of the number of degrees of freedom took into account the fact that the class-1 data determined primarily the values of four parameters: ReZ, ImZ, $\sigma_0(m_\phi)$, and α_1; the region outside of the resonance determines two parameters: $\sigma_0(m_\phi)$ and A.

We present as the result of the analysis the values of the three basic parameters describing the cross section of the process $e^+e^- \longrightarrow \omega\pi^0 \longrightarrow \pi^0\pi^0\gamma$ near the ϕ meson resonance:

$$\sigma_0(m_\phi) = 0.74 \pm 0.02 \pm 0.04 \text{ nb},$$
$$\text{Re}Z = 0.01 \pm 0.05, \qquad (11)$$
$$\text{Im}Z = -0.19 \pm 0.07.$$

The first of the errors presented for $\sigma_0(m_\phi)$ is a statistical error; the second error is a systematic error, which is determined by the error in the detection efficiency. In the values obtained for the parameters Re(Z) and Im(Z), the statistical error is larger than the systematic error. It is evident from Fig. 4 that, despite the stringent conditions for selection in class 1, the resonance background level is approximately half the amplitude of the interference pattern. The accuracy of its subtraction is determined by the statistical error in the magnitude of the resonance background in the second class $\sigma_{\phi 2}(m_\phi) = 0.013 \pm 0.04$ nb. The main source of systematic error in the interference parameters—the uncertainty in the coefficient k_{12}—was taken into account in the fitting and was included in the errors presented above for the parameters ReZ and ImZ.

4. THE REACTION $e^+e^- \longrightarrow \omega\pi^0 \longrightarrow \pi^+\pi^-\pi^0\pi^0$

4.1. Selection of Events

Events with two charged particles and four or five photons were selected to study the reaction

$$e^+e^- \longrightarrow \omega\pi^0 \longrightarrow \pi^+\pi^-\pi^0\pi^0. \qquad (12)$$

The additional photons appear in the event because of, in the first place, the nuclear interaction of the pions in the calorimeter and, in the second place, the superposition of the beam background on the event. For the process under study, the probability of finding an additional spurious photon in the event was about 20%. To suppress beam background events, it was necessary that the reconstructed point of emergence of the charged particles be displaced from the encounter location by not more than 0.3 cm in a plane perpendicular to the axis of the beams and by not more than 7.5 cm along the axis of the beams.

The main source of the background for the process (12) were decays of a ϕ meson:

$$e^+e^- \longrightarrow \phi \longrightarrow K^+K^-, \qquad (13)$$

$$e^+e^- \longrightarrow \phi \longrightarrow K_SK_L, \quad K_S \longrightarrow \pi^+\pi^-, \tag{14}$$

$$e^+e^- \longrightarrow \phi \longrightarrow \pi^+\pi^-\pi^0, \tag{15}$$

$$e^+e^- \longrightarrow \phi \longrightarrow \eta\gamma, \quad \eta \longrightarrow \pi^+\pi^-\pi^0. \tag{16}$$

To suppress the background from the reaction (13), in which the charged K mesons are quite slow ($\beta \approx 0.25$), it was necessary that for both charged particles the average ionization losses in the drift chamber not exceed $4(dE/dx)_{\min}$, where $(dE/dx)_{\min}$ represents the average ionization losses of a minimally ionizing particle. A substantial suppression of the background from the process (14) was obtained by limiting the spatial angle between the charged particles: $\Delta\psi < 140°$. In the process (14) the minimum angle between the charged pions from the decay of a K_S meson is about 150°.

The kinematic reconstruction procedure was applied to the event satisfying the conditions described above. Three hypotheses were considered:

(1) an event refers to the process $e^+e^- \longrightarrow \pi^+\pi^-\pi^0$;

(2) an event refers to the process $e^+e^- \longrightarrow \pi^+\pi^-\pi^0\gamma$;

(3) an event refers to the process $e^+e^- \longrightarrow \pi^+\pi^-\pi^0\pi^0$.

In checking the hypothesis $e^+e^- \longrightarrow \pi^+\pi^-\pi^0\pi^0$ photons with energy less than 50 MeV, lying outside the polar angle range $30° < \vartheta < 150°$, were dropped. This requirement decreased the probability of an incorrect reconstruction and decreased the background from the processes (15) and (16), which arises because of the superpositions of the beam background on the events corresponding to these processes. The values of the functions χ^2 for all three hypotheses were calculated as the result of the kinematic reconstruction: $\chi_{3\pi}$, $\chi_{3\pi\gamma}$, and $\chi_{4\pi}$, and the parameters M_{rec} and $M_{3\pi}$—the recoil mass of the photon in the reaction $e^+e^- \longrightarrow \pi^+\pi^-\pi^0\gamma$ and the recoil mass of the π^0 meson closest to the mass of ω for the reaction $e^+e^- \longrightarrow \pi^+\pi^-\pi^0\pi^0$. These parameters were required to satisfy the following restrictions:

$$\begin{aligned} &\chi_{4\pi} < 40, \quad |M_{3\pi} - 782| < 100, \\ &\chi_{3\pi} > 25, \quad M_{rec} > 620. \end{aligned} \tag{17}$$

The first two conditions select correctly reconstructed events of the process $e^+e^- \longrightarrow \omega\pi^0$, the third condition is directed against the background from the process (15), and the fourth condition suppresses events due to the process (16). The distributions over the parameter $\chi_{3\pi}$ for the experimental events and the simulation events for the processes (15) and (12) are presented in Fig. 6. A peak with low values of $\chi_{3\pi}$, corresponding to the contribution of the process (15), is clearly seen in the experimental distribution. The total computed number of events for the process (15) agrees to within about 10% with experiment, but the forms of the distributions are different, especially at the edge, for $\chi_{3\pi} \sim 20$. Consequently, we do not assume that simulation can give the correct estimate of the background with $\chi_{3\pi} > 25$. The distributions over the parameter $\chi_{4\pi}$ for the experimental events and the modeling events for the process (12) are presented in Fig. 7. The distributions were obtained using the energy points at the edges of the range studied, where the resonance background is negligible. It is evident that the agreement between the experimental data and the simulation is quite good.

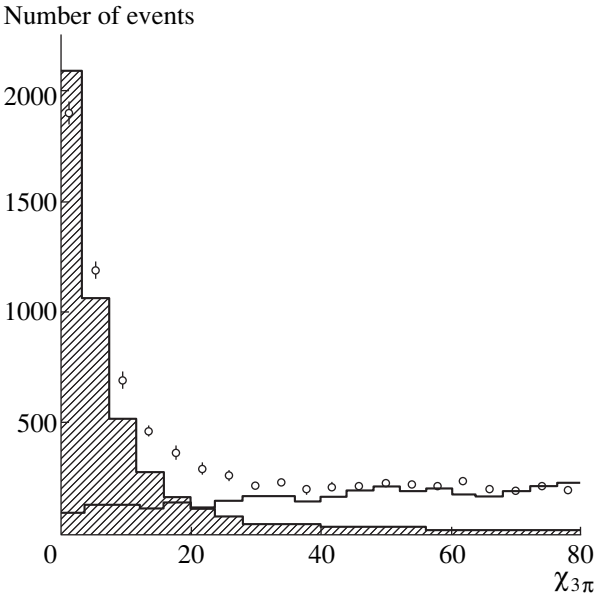


Fig. 6. Distribution over the parameter $\chi_{3\pi}$ for the experimental events (points with error bars) and the events from the simulation of process (15) (hatched histogram) and (12) (line).

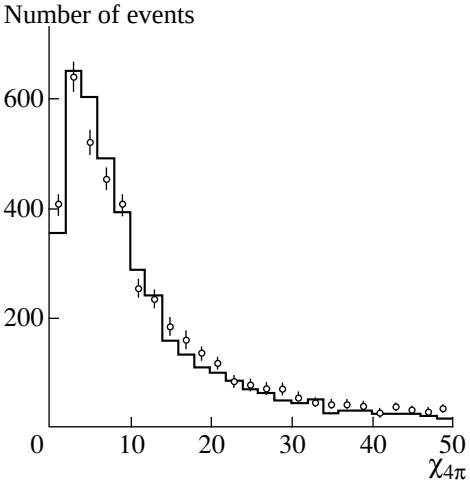


Fig. 7. Distribution over the parameter $\chi_{4\pi}$ for the experimental events (points with error bars) and the events from the simulation of process (12) (line).

The spectra of the recoil masses of the π^0 meson for the experimental and modeled events are presented in Fig. 8. The hatched histogram shows the contribution of background processes. The contribution of the following two nonresonance processes was taken into

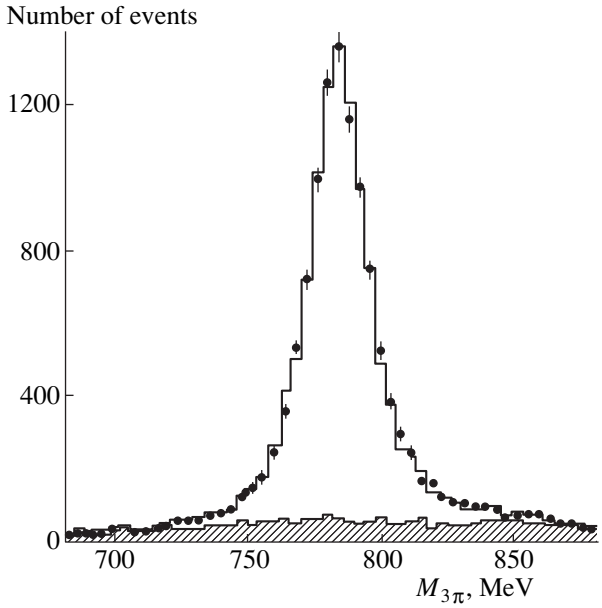


Fig. 8. Spectra of the recoil masses of a π^0 meson for experimental and simulation events. The points with error bars are experimental points, the hatched histogram represents the simulation of the background processes, and the line is the sum of the events of process (12) and the background.

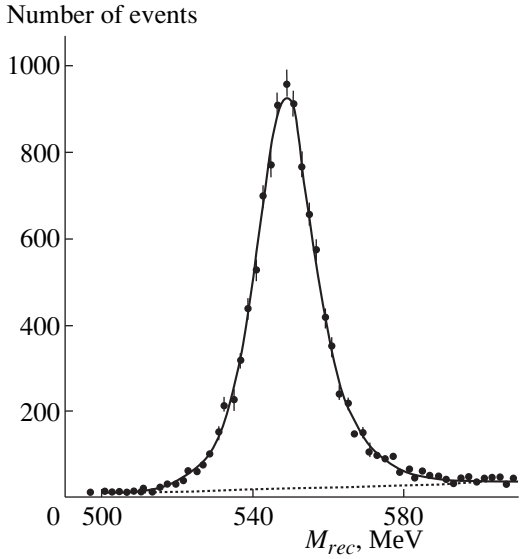


Fig. 9. Spectrum of the photon recoil masses M_{rec} for the experimental events of the reaction $e^+e^- \longrightarrow \eta\gamma \longrightarrow \pi^+\pi^-\pi^0\gamma$. The solid line is the result of a fit of the spectrum and the dotted line is the contribution of background processes.

account in the calculation of the background in addition to the above-described decays of the ϕ meson:

$$e^+e^- \longrightarrow \omega\gamma \longrightarrow \pi^+\pi^-\pi^0\gamma, \quad (18)$$

$$e^+e^- \longrightarrow \pi^+\pi^-\pi^0\pi^0. \quad (19)$$

A model of the intermediate state $\alpha_1(1260)\pi$ was used to simulate the reaction (19). As shown in [13], this state dominates in the reaction $e^+e^- \longrightarrow 4\pi$ in the range 1.0–1.4 GeV. In the $a_1\pi$ model the cross section of the reaction (19) is uniquely related with the cross section of the reaction $e^+e^- \longrightarrow \pi^+\pi^-\pi^+\pi^-$, which was measured in the energy range of interest to us [2]. There is virtually no interference between the processes (19) and (12). According to our estimate, made assuming a $a_1\pi$ model for the reaction (19), the interference contribution does not exceed 0.2% of the cross section of the reaction (12).

Approximately 85% of the 1.2×10^4 discarded events refer to the process (12); the resonance background makes a contribution of 6.5%; and, the processes (19) and (18) contribute 6.5 and 1.7%, respectively. The resonance background level calculated according to the simulation was found to be 1.8 ± 0.2 times less than the value determined from the experimental data with a fit of the cross section. Consequently, the resonance background in the distributions presented in Fig. 8 is included with a factor of 1.8.

Just as in the analysis of the neutral channel, the decay $\phi \longrightarrow \eta\gamma$ was studied in order to estimate the systematic error in the efficiency. Events with two charged particles and three or four photons with the same preliminary conditions as the events for the process (12) were selected. A kinematic reconstruction in the hypotheses $e^+e^- \longrightarrow 3\pi\gamma$ and $e^+e^- \longrightarrow 3\pi$ was performed for them. The background from the decay $\phi \longrightarrow 3\pi$ was suppressed using the restriction $\chi_{3\pi} > 25$. The parameter $\chi_{3\pi\gamma}$ was required to satisfy the restriction $\chi_{3\pi\gamma} < 35$. The spectrum of photon recoil masses M_{rec} for the event selected in this manner is presented in Fig. 9. There are about 10^4 events in the interval $520 < M_{rec} < 580$, and 95% of these events are events due to the process $\phi \longrightarrow \eta\gamma$. The background level was determined by fitting the recoil mass spectrum of the photon. The probability of the decay $\phi \longrightarrow \eta\gamma$, obtained taking into account the detection efficiency calculated according to the simulation, was $(1.25 \pm 0.02)\%$. The error presented is determined mainly by the inaccuracy in the subtraction of the background; the statistical error is about 1%. The tabulated value of the probability is $(1.26 \pm 0.06)\%$. Since the conditions for selecting the events for the decay $\phi \longrightarrow \eta\gamma$ were close to those used to select the events of the reaction (12) under study, we introduced into the detection efficiency obtained by simulation of the process (12) the correction 0.99 ± 0.05. Taking this correction into account, the detection efficiency for the process (12) was $\varepsilon = 0.21 \pm 0.01$.

4.2. Data Analysis

The analysis of the charged mode $e^+e^- \longrightarrow \omega\pi^0 \longrightarrow \pi^+\pi^-\pi^0\pi^0$ is similar to the above- described analysis of the neutral mode $e^+e^- \longrightarrow \omega\pi^0 \longrightarrow \pi^0\pi^0\gamma$. The selected events were divided into four classes:

$$1.\ \chi_{4\pi} < 20, \quad |M_{3\pi} - 782| < 35,$$

Table 2. The results of the fit to the cross section for process (12)

	$\sigma_0(m_\phi)$, nb	A	Re Z	Im Z	χ^2/ND
Model 1	7.34 ± 0.14	-0.08 ± 0.07	0.112 ± 0.019	-0.129 ± 0.025	38.1/54
	7.32 ± 0.14	-0.114 ± 0.007	0.110 ± 0.019	-0.127 ± 0.025	38.4/55
Model 2	7.34 ± 0.14	-0.11 ± 0.08	0.113 ± 0.019	-0.129 ± 0.025	38.1/54
	7.32 ± 0.14	-0.150 ± 0.007	0.110 ± 0.019	-0.127 ± 0.025	38.4/55

2. $\chi_{4\pi} < 20$, $|M_{3\pi} - 782| > 35$,
3. $\chi_{4\pi} > 20$, $|M_{3\pi} - 782| < 35$,
4. $\chi_{4\pi} > 20$, $|M_{3\pi} - 782| > 35$.

The class 1 contains about 80% of the events due to the reaction (12), which were selected according to the conditions (17). This is the main class for determining the interference parameters. The difference in the ratios of the levels of the nonresonant cross sections in classes 1 and 2 from the ratio expected for the process (12) makes it possible to estimate the contribution of the background process (19). Fitting the cross section in the classes 2–4 permits determining the total cross sections for the background resonance processes: $\sigma_{\phi 2}$, $\sigma_{\phi 3}$, and $\sigma_{\phi 4}$. The cross section for the resonance background in the first class was estimated using the formula

$$\sigma_{\phi 1} = \sigma_{\phi 2}(\sigma_{\phi 3}/\sigma_{\phi 4}).$$

This method employs the natural assumption that the distributions over the parameters $\chi_{4\pi}$ and $M_{3\pi}$ are independent for the background from the decays (13)–(16) and is model-independent. The validity of the formula used was checked on the simulation events and specially selected experimental events for background processes with a statistical accuracy of about 15%.

The cross section for detecting experimental events in each class was approximated by the formula

$$\sigma_{vis}(E) = \alpha_i \varepsilon \sigma_{\omega\pi}(E)(1 + \delta_{\omega\pi}) + R_{4\pi}\beta_i\varepsilon_{4\pi}\sigma_{4\pi}(E)(1 + \delta_{4\pi}) + \sigma_{\omega\gamma i}(E) + \sigma_{\phi i}(E),$$

where $\sigma_{\omega\pi}(E)$ and $\sigma_{4\pi}(E)$ are the Born cross sections of the processes (12) and (19), $\sigma_{\phi i}(E)$ is the cross section for the resonance background in class i, $\sigma_{\omega\gamma i}(E)$ is the visible cross section of the process (18) in class i, calculated by simulation, ε and $\varepsilon_{4\pi}$ are the detection efficiencies for the processes (12) and (19) with the conditions (17), and $\delta_{\omega\pi}$ and $\delta_{4\pi}$ are radiation corrections. The coefficients α_i and β_i are the probabilities that events due to the processes (12) and (19) fall into ith class $\left(\sum \alpha_i = 1, \sum \beta_i = 1\right)$. The cross section $\sigma_{\omega\pi}(E)$ was approximated by the formulas (1) and (10). The energy dependence of the factor $P_f(E)$ for the process $e^+e^- \longrightarrow \omega\pi^0 \longrightarrow \pi^+\pi^-\pi^0\pi^0$ was calculated using the formulas from [14]. The Born cross section of the process (19) was also represented in the form (1). It was obtained by multiplying the cross section of the process $e^+e^- \longrightarrow \pi^+\pi^-\pi^+\pi^-$ by the factor 0.4, calculated assuming $a_1\pi$ dominance in the reaction $e^+e^- \longrightarrow 4\pi$ [13]. The parameters for the nonresonance cross section and the values of ReZ and ImZ for the process $e^+e^- \longrightarrow \pi^+\pi^-\pi^+\pi^-$ were taken from [2]. The parameter $R_{4\pi}$ shows the difference between the values of the cross section for the process (19) obtained by fitting the experimental data and the computed cross section.

The fit was made in all four classes simultaneously. The free parameters were $\sigma_0(m_\phi)$, A, ReZ, ImZ, $R_{4\pi}$, α_1, α_3, and $\sigma_{\phi i}(m_\phi)$ with the exception of $\sigma_{\phi 1}$. The parameters α_2, β_1, β_2, and β_3 were calculated from the relations

$$\alpha_2 = 0.06\alpha_1, \quad \beta_2 = 1.13\beta_1,$$

$$\beta_3/\beta_1 = \beta_4/\beta_2 = (\alpha_3 + \alpha_4)/(\alpha_1 + \alpha_2).$$

The first two relations were obtained from the simulation. The third relation is based on the natural assumptions that the distribution over the parameter $\chi_{4\pi}$ is identical for the processes (19) and (12) and that for the process (19) the distribution over the parameter $\chi_{4\pi}$ does not depend on $M_{3\pi}$. The validity of these assumptions was checked on the simulated events with a statistical accuracy of better than 5%. The coefficients α_i and β_i are essentially energy-independent; their energy dependences were calculated by simulation and fits were made using linear functions.

A fit to the cross section was made in the same two models which were used to describe the data in the neutral channel of the reaction $e^+e^- \longrightarrow \omega\pi$. The results are presented in Table 2. The second and fourth rows in the table correspond to the cases where the values of the parameter A were fixed at the values obtained in the analysis of the neutral mode. All parameters except A are essentially model-independent. The values obtained for α_1 and α_3 in the fit agree well with the calculation based on the simulation. The cross sections for the resonance background in the classes 2–4, as already mentioned, were approximately two times greater than the estimates obtained with simulation. The background level from the process (19) agreed with the calculations:

$$R_{4\pi} = 1.0 \pm 0.2.$$

The visible cross section of the experimental events in class 1 and the fitting curve obtained in the second model with $A = -0.150 \pm 0.007$ are presented in Fig. 10. The value of χ^2/ND for this curve is 8.7/11. For $E = m_\phi$

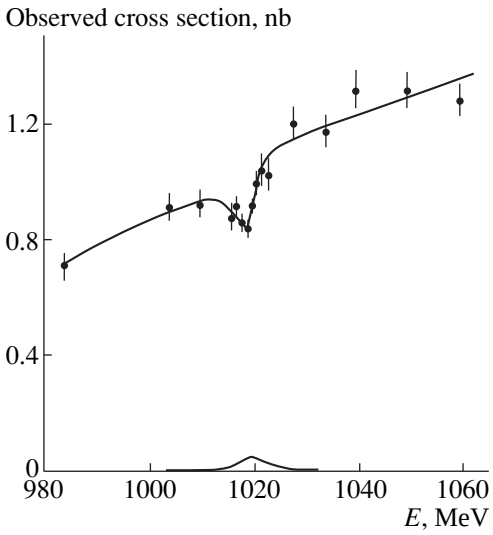


Fig. 10. Energy dependence of the cross section for detecting class-1 events and the fitted curve. The curve at the bottom was obtained by fitting the cross section for the resonance background.

the resonance background level in the first class is 5% of the cross section of the process (12) under study; the background level from the process (19) corresponds to 2.3%, and the background level from the process (18) is about 1%. We present at the result the values of the three basic parameters:

$$\sigma_0(m_\phi) = 7.32 \pm 0.14 \pm 0.38 \text{ nb},$$
$$\mathrm{Re}\,Z = 0.110 \pm 0.019 \pm 0.003, \qquad (20)$$
$$\mathrm{Im}\,Z = -0.127 \pm 0.025 \pm 0.005.$$

The first of the errors presented is statistical, while the second is systematic. The main source of the systematic error in the nonresonance cross section $\sigma_0(m_\phi)$ is the error in determining the detection efficiency 5%. The next, in magnitude, factor is the uncertainty in the cross section for detecting the background process $e^+e^- \longrightarrow \omega\gamma$. When this quantity is varied by a factor of 2, the nonresonance cross section changes by 1%. The model dependence, the uncertainty in the calculation of the cross section for $e^+e^- \longrightarrow \omega\gamma$, and the possible inaccuracy in the estimate of the resonance background in class 1 are included in the systematic errors for Re*Z* and Im*Z*.

5. DISCUSSION

The main results of the analysis performed in this work and the result of previous measurements [1, 4] are presented in Table 3. A comparison shows that all parameters for the two experiments agree well with one another. The approach used with normalization to the process $\phi \longrightarrow \eta\gamma$ made it possible to decrease substantially the systematic error in the measurements of the nonresonance cross section. The accuracy in measuring Re*Z* and Im*Z* for the charged mode was improved. Even though the luminosity integral in the 1998 experiment is larger, the interference parameters measured in the neutral channel are less accurate than the result of the previous experiment. This is because in the present analysis the resonance background could not be suppressed to the level obtained in the analysis of the 1996 data.

Since a substantial fraction of the systematic error in the nonresonance cross sections in the charged and neutral modes is a general error and is related with the accuracy of the tabulated value of $B(\phi \longrightarrow \eta\gamma)$, it is interesting to present the measured ratio of these cross sections:

$$\frac{\sigma_0^{\pi^0\pi^0\gamma}(m_\phi)}{\sigma_0^{\pi^+\pi^-\pi^0\pi^0}(m_\phi)} = 0.1017 \pm 0.0037 \pm 0.0039. \qquad (21)$$

The computed value of this quantity 0.098 ± 0.006 is 1.023 times greater than the ratio of the widths

$$\Gamma(\omega \longrightarrow \pi^0\gamma)/\Gamma(\omega \longrightarrow \pi^+\pi^-\pi^0) = 0.096 \pm 0.006$$

Table 3. The values of the main parameters for the process $e^+e^- \longrightarrow \omega\pi^0$. The results of the present analysis, the results of our previous measurements, and the average values over two measurements are presented

$e^+e^- \longrightarrow \omega\pi^0 \longrightarrow \pi^+\pi^-\pi^0\pi^0$			
	$\sigma_0(m_\phi)$, nb	Re *Z*	Im *Z*
1998	$7.32 \pm 0.14 \pm 0.38$	$0.110 \pm 0.019 \pm 0.003$	$-0.129 \pm 0.025 \pm 0.005$
1996 [1]	$7.28 \pm 0.18 \pm 0.80$	$0.104 \pm 0.028 \pm 0.006$	$-0.118 \pm 0.030 \pm 0.009$
1996 + 1998		0.108 ± 0.016	-0.125 ± 0.02
$e^+e^- \longrightarrow \omega\pi^0 \longrightarrow \pi^0\pi^0\gamma$			
	$\sigma_0(m_\phi)$, nb	Re *Z*	Im *Z*
1998	$0.74 \pm 0.02 \pm 0.04$	0.01 ± 0.05	-0.19 ± 0.07
1996 [4]	0.64 ± 0.08	0.04 ± 0.05	-0.19 ± 0.06
1996 + 1998		0.025 ± 0.035	-0.19 ± 0.05

(see [15]). This difference is due to the difference in the behavior of the phase volumes for the decays $\omega \longrightarrow \pi^0\gamma$ and $\omega \longrightarrow \pi^+\pi^-\pi^0$ as a function of the mass of the ω meson. Taking into account the factor of 1.023, the ratio of the widths can be obtained from the ratio of the cross sections:

$$\frac{\Gamma(\omega \longrightarrow \pi^0\gamma)}{\Gamma(\omega \longrightarrow \pi^+\pi^-\pi^0)} = 0.0994 \pm 0.0036 \pm 0.0038, \quad (22)$$

which is the most accurate current measurement of this quantity. The total nonresonant cross section for $e^+e^- \longrightarrow \omega\pi^0$ with $E = m_\phi$ is

$$\sigma_0(m_\phi)_{\omega\pi} = 8.25 \pm 0.14 \pm 0.43 \text{ nb}. \quad (23)$$

This value was obtained as a sum of the measured cross sections in the charged and neutral modes taking into account the contribution of other decay modes of the ω meson: $(2.3 \pm 0.3)\%$. The result for the interference parameter in the charged channel can be presented in the different form $Z = |Z|e^{i\psi}$:

$$|Z| = 0.168 \pm 0.022 \pm 0.004,$$
$$\psi = -49^\circ \pm 7^\circ \pm 1^\circ.$$

The decay probability can be calculated from the value of $|Z|$ and the value of the nonresonant cross section (23) using equation (2):

$$B(\phi \longrightarrow \omega\pi) = (5.5^{+1.6}_{-1.4} \pm 0.3) \times 10^{-5}. \quad (24)$$

The result of the previous measurement [1] is

$$B(\phi \longrightarrow \omega\pi) = (4.8^{+1.9}_{-1.7} \pm 0.8) \times 10^{-5}.$$

Since the decay probabilities in the two experiments are the same, we can present an average value obtained over all data collected with the SND detector near the ϕ resonance:

$$B(\phi \longrightarrow \omega\pi) = (5.2^{+1.3}_{-1.1}) \times 10^{-5}. \quad (25)$$

The average values of the interference parameters over the two experiments are presented in Table 3.

As already mentioned in the introduction, the decay $\phi \longrightarrow \rho\pi^0 \longrightarrow \pi^0\pi^0\gamma$ contributes to the interference in the neutral channel. A calculation of this contribution in the vector dominance model was made in [4]:

$$\mathrm{Re}\,Z = -0.079, \quad \mathrm{Im}\,Z = -0.053.$$

The sum of these quantities with the interference parameters measured in the charged mode (20) is

$$\mathrm{Re}\,Z = 0.032 \pm 0.016, \quad \mathrm{Im}\,Z = -0.180 \pm 0.020$$

and agrees well with the measurement in the neutral mode.

The main diagrams for the decay $\phi \longrightarrow \omega\pi$, which were studied in the theoretical works [16–18], are shown in Fig. 11. The diagram in Fig. 11a describes the ϕ–ρ transition. Aside from the single-photon mechanism, the successive ϕ–ω–ρ transition was taken into account for this diagram. Substantial discrepancies

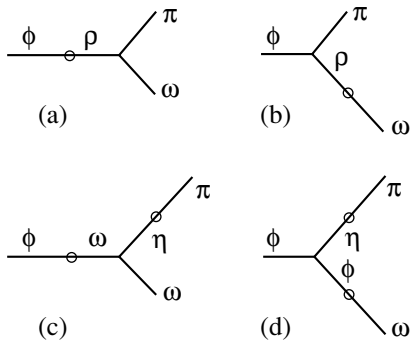


Fig. 11. Diagrams describing the contributions to the amplitude of the decay $\phi \longrightarrow \omega\pi$.

exist between the theoretical works under discussion here. The value $g_{\rho\omega\pi} = 14.3$ GeV^{-1}, calculated from the decay $\omega \longrightarrow 3\pi$, was used in [16, 18] to calculate the nonresonant cross section for the process $\rho \longrightarrow \omega\pi$. The nonresonance cross section ~5 nb obtained in the process is much smaller than the measured value (23). The value $g_{\rho\omega\pi} = 16.5$ GeV^{-1} was used in [17]. Our value of $g_{\rho\omega\pi}$, obtained in the model 2, which describes the cross section for $e^+e^- \longrightarrow \omega\pi$ well in a wide range of energies, is 15.5 GeV^{-1}, and the contribution of ρ' to the cross section in the region of the ϕ meson is about 25%. The contribution of the ρ' meson was neglected in all of the works [16–18]. Large discrepancies also exist in the calculation of the contributions of the diagrams in Figs. 11b–11d. They are mainly due to the different treatment of the ϕ–ω mixing and the different manner in which the mass dependence of the widths of the resonances was taken into account. The probabilities obtained in [16, 17] for the decay $\phi \longrightarrow \omega\pi$ are $(8–9) \times 10^{-5}$ and exceed the measured value. In [18] a combined analysis of the decays $\phi \longrightarrow \omega\pi$ and $\phi \longrightarrow \pi^+\pi^-$ is given and predictions are made for the parameters ReZ and ImZ. None of the proposed sets of parameters agree with our measurement. The use of a large value for the nonresonance cross section can substantially increase the contribution of the diagram in Fig. 11a in [16, 18] and change the predictions. Another conclusion can be drawn on the basis of the theoretical works. According to [16–18], the contributions of the diagrams in Figs. 11b–11d and of the ϕ–ω–ρ transition should increase the value of ReZ = 0.123, obtained taking into account only the single- photon mechanism. To explain the small experimental value Re$Z = 0.108 \pm 0.016$, in our view, either the presence of an additional contribution to the ϕ–ρ transition, comparable to the single-photon contribution, or the existence of the direct decay $\phi \longrightarrow \omega\pi$ is required.

6. CONCLUSIONS

In conclusion, we shall list the basic results obtained in this work. In the energy range near the ϕ resonance the cross section for the process $e^+e^- \longrightarrow \omega\pi$ was measured for two decay modes of the ω meson: $\omega \longrightarrow \pi^0\gamma$

and $\omega \longrightarrow \pi^+\pi^-\pi^0$. The parameters of the interference pattern associated with the decay $\phi \longrightarrow \omega\pi$ were measured for both cross sections. The probability of this decay was obtained from the magnitude of the interference pattern in the cross section for the process $e^+e^- \longrightarrow \omega\pi \longrightarrow \pi^+\pi^-\pi^0\pi^0$:

$$B(\phi \longrightarrow \omega\pi) = (5.5^{+1.6}_{-1.4} \pm 0.3) \times 10^{-5}.$$

This value agrees with our first measurement [1] but it is more accurate. The small real part of the measured interference parameter indicates either the presence of an additional contribution to the ϕ–ρ transition, comparable in magnitude to the single-photon contribution, or the existence of a direct decay $\phi \longrightarrow \omega\pi$. The ratio of the partial widths of the two main decay modes of the ω meson was obtained from the ratio of the cross sections in the charged and neutral modes:

$$\Gamma(\omega \longrightarrow \pi^0\gamma)/\Gamma(\omega \longrightarrow \pi^+\pi^-\pi^0)$$
$$= 0.0994 \pm 0.0036 \pm 0.0038.$$

At the present time this is the most accurate measurement of this quantity.

ACKNOWLEDGMENTS

This work was supported by the STP Foundation "Integratsiya" (grant no. 274) and the Russian Foundation for Basic Research (grant nos. 99-02-16815 and 99-02-16813).

REFERENCES

1. M. N. Achasov, S. E. Baru, A. V. Berdyugin, *et al.*, Phys. Lett. B **449**, 122 (1999).
2. R. R. Akhmetshin, G. A. Aksenov, E. V. Anashkin, *et al.*, Preprint No. 99-11 (Budker Institute of Nuclear Physics, Novosibirsk, 1999).
3. M. N. Achasov, V. M. Aulchenko, K. I. Beloborodov, *et al.*, hep-ex/9910070.
4. M. N. Achasov, V. M. Aulchenko, A. V. Berdyugin, *et al.*, hep-ex/9907026.
5. M. N. Achasov, V. M. Aulchenko, S. E. Baru, *et al.*, hep-ex/9909015.
6. M. N. Achasov, V. M. Aulchenko, A. V. Berdyugin, *et al.*, Phys. Lett. B **440**, 442 (1998).
7. R. R. Akhmetshin, E. V. Anashkin, M. Arpagaus, *et al.*, Phys. Lett. B **462**, 380 (1999).
8. A. V. Bozhenok, V. N. Ivanchenko, and Z. K. Silagadze, Nucl. Instrum. Methods Phys. Res. A **379**, 507 (1996).
9. É. A. Kuraev and V. S. Fadin, Yad. Fiz. **41**, 733 (1985) [Sov. J. Nucl. Phys. **41**, 466 (1985)].
10. A. B. Clegg and A. Donnachie, Z. Phys. C **62**, 455 (1994).
11. K. W. Edwards, R. Janicek, P. M. Patel, *et al.*, hep-ex/9908024.
12. M. N. Achasov, V. M. Aulchenko, K. I. Beloborodov, *et al.*, hep-ex/9910057.
13. R. R. Akhmetshin, E. V. Anashkin, M. Arpagaus, *et al.*, Preprint No. 98-83 (Budker Institute of Nuclear Physics, Novosibirsk, 1998); hep-ex/9904024.
14. E. L. Bratkovskaya, E. A. Kuraev, Z. K. Silagadze, and O. V. Teryaev, Phys. Lett. B **338**, 471 (1994).
15. Review of Particles Physics, Eur. Phys. J. C **3**, 1 (1998).
16. V. A. Karnakov, Yad. Fiz. **42**, 1001 (1985) [Sov. J. Nucl. Phys. **42**, 634 (1985)].
17. H. Genz and S. Tatur, Phys. Rev. D **50**, 3263 (1994).
18. N. N. Achasov and A. A. Kozhevnikov, Int. J. Mod. Phys. A **7**, 4825 (1992).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 939–949.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1080–1093.

SOLIDS
Structure

Softening of Ionic Crystals as a Result of a Change in the Spin States of Structural Defects under Paramagnetic Resonance Conditions

Yu. I. Golovin*, R. B. Morgunov, V. E. Ivanov, and A. A. Dmitrievskiĭ

Tambov State University, Tambov, 392622 Russia
**e-mail: golovin@tsu.tmb.ru*

Received May 21, 1999

Abstract—The purpose of this work was to investigate the plastic properties of Ca- or Eu-doped NaCl and KCl single crystals in crossed constant and microwave magnetic waves under paramagnetic resonance conditions. It was found that when the photon energy of the microwave field equals the Zeeman splitting of the electronic spin sublevels, resonance softening of the crystals, manifested as an increase in the free path of individual dislocations and the macroplastic flow velocity as well as a decrease of the microhardness of the crystals, is observed. It was established that metastable Ca- and Eu-impurity complexes, which are also sensitive to the constant magnetic field in the absence of the microwave field, as well as complexes formed by dislocations and point defects are responsible for resonance softening. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The traditional approach to describing the macroplastic properties of real crystals is limited, as a rule, to the analysis of mesoscopic-scale processes which occur in the structural-defect subsystem (dislocations, twins, crowdions, and so on). Sometimes the behavior of individual atomic clusters (inflections and steps on dislocations, point defects, and others) is analyzed. At the same time, a variety of data indicate that only an analysis of defects at the electronic level can give an adequate description of many observed effects in plasticity. For example, deformation luminescence [1], exoemission [2], changes in the luminescence spectra of point centers during plastic deformation [3], photoconductivity [4], and EPR [5] and NMR [6] signals as well as the wide spectrum of experimental data on the effect of light on the plasticity of wide-gap crystals [7] and the magnetic field and current on the mechanical properties of metals [8] attest to the diverse manifestations of the electronic subsystem of a crystal in the formation of their plastic properties.

The recent observation of a number of paradoxical, from the standpoint of equilibrium thermodynamics, effects, which include the influence of a magnetic field with induction ~1 T on the plasticity of dielectric ionic crystals [9–32], polymers [33–36], semiconductors [37], molecular crystals [38], and diamagnetic metals [39–43] at temperatures close to room temperature, makes it necessary to seek new approaches for developing the physics of plasticity at the electronic level of analysis. With respect to the magnetoplastic effects listed above, it is now believed that these are spin-dependent processes, i.e., they require not only an analysis of the coordinate part of the wave function of the elections (transitions between bands and localized states, Coulomb interaction of defects, and so on) but also an investigation of the influence of the spin degrees of freedom on the formation of plasticity [10–15, 44–46]. This, initially hypothetical, point of view has been confirmed experimentally very recently [47]. This has led to a new class of problems which require analysis at an even deeper level—the spin level.

However, the not entirely clear relation between the various hierarchical levels of the analysis of plasticity (macro-, meso-, dislocation, atomic, electron-spin) raises a number of fundamental questions. How can a weak, from the energy standpoint, magnetic field compete with thermal fluctuations at high temperatures? How do spin transitions in structural defects influence the atomic state of the defects and the mobility of dislocations? What are magnetically sensitive defects in crystals and how can they be produced? Why are the standard experimental methods too insensitive to observe these defects while the plastic properties are sensitive to their presence in a crystal?

Similar problems arose several decades ago in the chemistry of magnetically sensitive radical reactions, where they were solved using new methods of spectroscopy, one of which is RYDMR (reaction yield detected magnetic resonance)—paramagnetic resonance detected not by the absorption of an electromagnetic wave but by the "yield" of a chemical reaction, i.e., according to the change in the rate and equilibrium constants [48], the change in the fluorescence intensity [49], and the change in the photoconductivity [50] and other macroproperties of crystals. This method makes it possible to detect and investigate at room temperature less than 10^3 pairs of intermediate short-lived paramagnetic particles, present

in a sample simultaneously and entering together into different spin-dependent reactions [51, 52].

In [53] it was shown, theoretically, for the first time that resonance softening of crystals is possible if the frequency ν of the ac field satisfies the paramagnetic resonance condition $h\nu = g\mu B_0$ in a radical pair formed by an impurity center on a dislocation and on a local stop (h is Planck's constant, g is the spectroscopic splitting factor, μ is the Bohr magneton, and B_0 is the constant magnetic induction). The first experimental observation of resonance softening of ionic crystals is described in [47]. Our objective in the present work was to investigate in greater detail the mobility of individual edge dislocations, macroplastic flow, and the microhardness of ionic crystals under conditions where a constant magnetic field and a high-frequency magnetic field crossed with it act simultaneously, i.e., to endeavor to transfer the ideology and approaches of the RYDMR spectroscopy to plasticity physics, where the change in any plastic characteristic of a crystal can serve as a response to resonance, i.e., the yield.

2. EXPERIMENTAL PROCEDURE

Three standard procedures were used to investigate the plasticity of crystals in external magnetic fields: (1) measurement of the free paths L of individual edge dislocations by the method of double selective etching, (2) detection of the change in the macrodeformation diagram of crystals in a constant magnetic field at the moment when the microwave field is switched on, and (3) measurement of the Vickers microhardness H of crystals after the crystals are exposed to a magnetic field (0.2 N load on the indenter; 20 s load duration). The experiments were performed on $3 \times 3 \times 5$ mm Ca- of Eu-doped NaCl and Ca-doped KCl single crystals annealed at 700 K and cooled to 293 K in 10 h. The mass fraction of the impurity was ~0.1% for investigation of macroplasticity and microhardness and ~0.01% for determining the free path of individual dislocations. The samples were positioned at a node of the magnetic field of a H_{102} standing wave in a cavity, connected with a ~0.1 W klystron, which operated at the frequency ν = 9.5 GHz. The cavity was placed between the poles of an electromagnet, which produced a constant magnetic field with induction B_0 ranging from 0 to 0.8 T in the region of the crystal. In all experiments the vector $\mathbf{B}_0$ was directed along the [001] crystallographic direction.

3. RESULTS

The effect of combined constant and microwave magnetic fields on the mobility of individual edge dislocations was investigated in the first series of experiments. After fresh dislocations were introduced and the first etching, which revealed the initial positions of the dislocations, the NaCl : Ca crystals were placed inside a cavity where they were exposed to crossed magnetic fields for 15 min. Next, the crystal was removed from the cavity and a second etching revealed the new positions of the dislocations. The double etching procedure in the absence of a magnetic field itself led to an identical, in all experiments, displacement of the dislocations under the action of internal stresses by an average distance $L_0 = 12 \pm 1$ μm. The exposure of crystals simultaneously to a constant magnetic field and a microwave field (in the configuration $\mathbf{B}_1 \perp \mathbf{B}_0$, where $\mathbf{B}_1$ is the induction of the magnetic microwave field) increased L with $B_0 = B_{res1} = 0.32 \pm 0.03$ T, $B_0 = B_{res2} = 0.18 \pm 0.02$ T, and $B_0 = B_{res3} = 0.12 \pm 0.02$ T (Fig. 1). These values correspond to magnetic fields $B_0 = h\nu/\mu g$, (ν = 9.5 GHz)in which resonance transitions between the electronic sublevels split in a constant magnetic field occur. The corresponding effective g factors are $g_1 = 2.1 \pm 0.2$, $g_2 = 3.8 \pm 0.3$, and $g_3 = 5.7 \pm 0.7$. Measurements near the peaks were performed especially carefully: the smallest step in B_0 and a larger statistically sample (up to 500 measurements of the free path at each point) were used. When the constant and microwave fields were applied in the configuration $\mathbf{B}_1 \parallel \mathbf{B}_0$, the peaks vanished (Fig. 1). In the absence of a microwave field, the constant magnetic field with $B_0 > 0.35$ T also gave rise to dislocation displacements which were greater than the background free paths L_0, caused by etching in the absence of a magnetic field, but the field dependence $L(B_0)$ was monotonic in this case, i.e., there were no peaks (Fig. 1).

It was established that in crossed magnetic fields, under resonance conditions, the individual dislocations are, on the average over the crystal, equally likely to move in the crystallographic directions $[\bar{1}10]$, $[1\bar{1}0]$, $[\bar{1}\,\bar{1}0]$, and [110] over identical average distances on all faces of a sample. Therefore, in our experiments the role of the external fields reduced to depinning of dislocations from stops, while dislocation motion occurred under the action of random internal mechanical stresses. Investigation of NaCl : Ca crystals in a standard EPR spectrometer did not show any peaks of resonance absorption of the electromagnetic wave at the values of B_{res} mentioned above.

In [17] it was established experimentally that ionic crystals contain metastable complexes of point defects, whose decomposition can be stimulated by a constant magnetic field in the absence of a microwave field. Similar data on the possibility of initiating the decomposition of impurity complexes by an external magnetic field were obtained in [54–57] by independent methods. Resonance plastification could also be caused by a change in the probability of detachment of dislocations (D) from paramagnetic point defects (P) [10–16, 44–46]. Resonance can arise in pairs of spin carriers, which are present in one impurity complex $P\ldots P$, in the first case and in the pairs $P\ldots D$, which are formed by paramagnetic centers localized in a point stop and a dislocation nucleus, in the second case. It is obvious that these possibilities can be distinguished if before a crystal con-

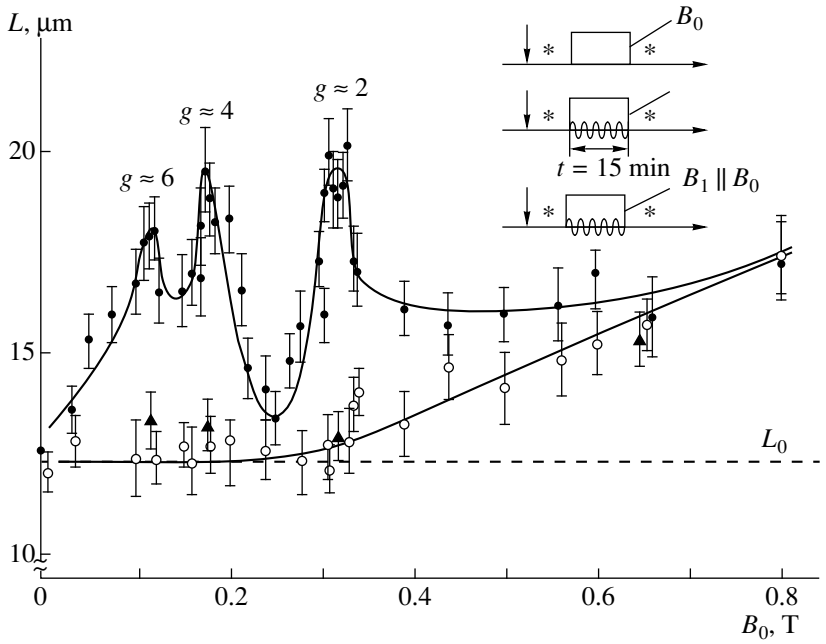


Fig. 1. Mean free path L of edge dislocations in NaCl : Ca crystals versus the induction of a constant magnetic field B_0 applied for 15 min: (○) no microwave field (ν = 9.5 GHz), (●) simultaneous application of microwave and constant magnetic fields in the configuration $\mathbf{B}_1 \perp \mathbf{B}_0$, where $\mathbf{B}_1$ is the induction of the microwave magnetic field, (▲) simultaneous application of microwave and constant magnetic fields in the configuration $\mathbf{B}_1 \parallel \mathbf{B}_0$. The dashed line shows the dislocation free path L_0 due to the action of an etchant in the absence of external magnetic fields. Inset: Sequence of procedures: (arrow) introduction of dislocations, (asterisk) etching, (rectangle) exposure of crystals in a magnetic field.

taining dislocations is placed inside a cavity the magnetically sensitive complexes of point defects are "removed" from the crystal by transforming them into complexes which are insensitive to a magnetic field. Then the appearance of resonance will be possible only in the pairs *P*…*D*.

To produce such an experimental situation, the crystals were exposed to several pulses of a magnetic field with amplitude B_0 = 7 T and duration 10^{-2} s (in the absence of a microwave field) before dislocations are introduced. As shown in [17], these pulses irreversibly remove the sensitivity of point-defect complexes to subsequent magnetic-field pulses. Next, fresh edge dislocations were introduced into the crystals and the crystals were placed inside a cavity. It was found that the peaks at $B_0 = B_{res2} = 0.18 \pm 0.02$ T and at $B_0 = B_{res3} = 0.12 \pm 0.02$ T vanish completely, and the peak at $B_0 = B_{res1} = 0.32 \pm 0.03$ T decreases by a factor of 2 (Fig. 2a). Therefore, resonance in the pairs *P*…*P* leads to the appearance of three softening peaks at B_{res1}, B_{res2}, and B_{res3}, and resonance in the pairs *P*…*D*, apparently, gives one peak at B_{res1}. Another method for "removing" metastable point defects which are sensitive to a magnetic field is to anneal the crystals at a high temperature or prolonged "aging" of the crystals at T = 293 K [17]. Investigation of the displacement of dislocations in crystals, which were "aged" for two years, under conditions of paramagnetic resonance at frequency ~152 MHz revealed a single peak at B = 5 mT, which corresponds to an effective g factor $g \approx 2$ (Fig. 2b). Just as in previous experiments, the increase in the free path caused by resonance magnetic fields depended on the angle φ between the vectors $\mathbf{B}_1$ and $\mathbf{B}_0$. At $\varphi = 0$ the peak vanished (see inset in Fig. 2b). We note that in these experiments the magnetoplastic effect appeared at record low magnetic fields (~10^{-3} T).

Thus, a large part of the observed resonance softening of crystals is due to paramagnetic resonance in previously observed point-defect complexes which are sensitive to a constant magnetic field, i.e., in the pairs *P*…*P*. Another part, to which the peak at $g \approx 2$ attests, which cannot be removed by a quasistationary magnetic field pulse, is probably due to resonance in *D*…*P* pairs.

The combined effect of a constant magnetic field and a microwave field with frequency ν = 9.5 GHz, which are applied in the configuration $\mathbf{B}_1 \perp \mathbf{B}_0$, on the rate of macroplastic flow of NaCl : Ca crystals was investigated in a second series of experiments. The macroplastic deformation of the crystals was conducted in a "soft" machine with quartz rods. This machine gave a mechanical compression stress increasing linearly with time: $\sigma = qt$, where q = const [18]. The length of the sample was measured continuously with an induction sensor to within ±0.1 μm; this made it possible to construct a load diagram on an *X*–*Y* plotter, i.e., the dependence of the relative strain ε on σ or on the running time t, elapsed from the onset of load application (Fig. 3a).

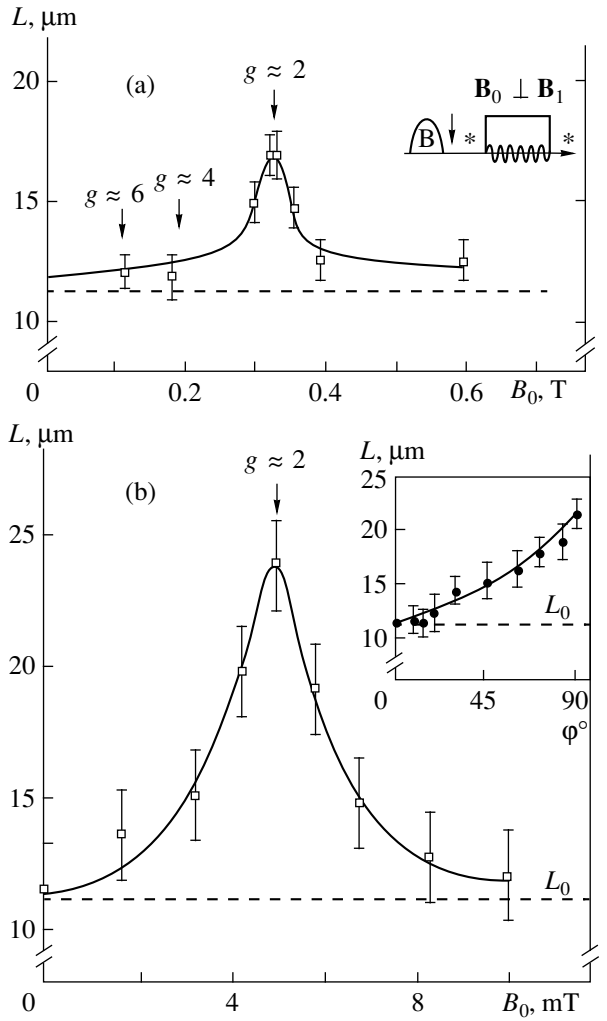


Fig. 2. Mean free path L of edge dislocations versus the induction B_0 of a constant magnetic field applied for 15 min together with a microwave field in the configuration $\mathbf{B}_1 \perp \mathbf{B}_0$: (a) for freshly quenched from 700 K NaCl : Ca crystals subjected to a quasistationary magnetic field (three pulses with amplitude $B_0 = 7$ T and duration 10 ms) before the samples were placed in a cavity and exposed in crossed magnetic fields ($\nu = 9.5$ GHz); (b) for NaCl : Ca crystals ($\nu = 0.152$ GHz) "aged," after quenching, for two years. Inset in Fig. 2b: Dependence of the dislocation free path L on the angle φ between the vectors $\mathbf{B}_1$ and $\mathbf{B}_0$ under resonance conditions.

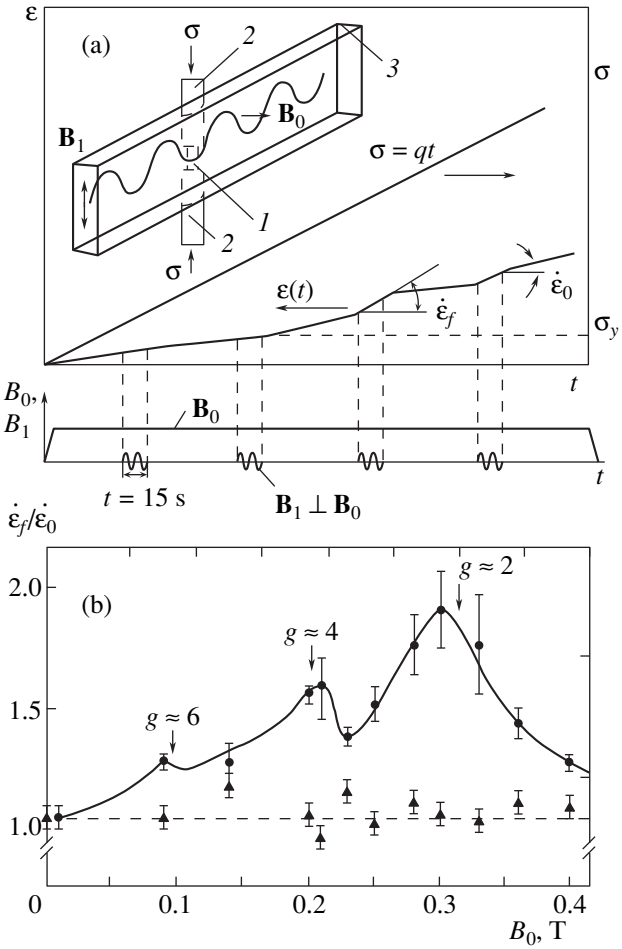


Fig. 3. (a) Typical dependences of the relative strain ε and mechanical stresses σ on the deformation time t are shown schematically at the moments when the microwave field $\mathbf{B}_1$ ($\nu = 9.5$ GHz) is switched on and off. A constant magnetic field with induction $\mathbf{B}_0$ was applied throughout the entire deformation process (σ_y is the yield point). Inset in Fig. 3a: Diagram of the experimental technique: (*1*) sample, (*2*) quartz rods, (*3*) cavity. (b) Dependence of the softening of NaCl : Ca crystals (ratio of the plastic flow rate after a microwave field is switched on $(d\varepsilon/dt)_f$ to the plastic flow rate before the microwave field is switched on $(d\varepsilon/dt)_0$) on the induction B_0 of the constant magnetic field in which deformation is conducted: (▲) before passage of the yield point with relative strains $\varepsilon < 0.03\%$, (●) after passage of the yield point in the range of relative strains $0.1\% < \varepsilon < 0.5\%$ ($\nu = 9.5$ GHz).

The application of a microwave field in the absence of a constant magnetic field did not result in any changes in the strain rate $d\varepsilon/dt$. This made it possible to assume that any changes which a microwave field could introduce in the deformation diagrams under conditions of a continually acting constant magnetic field can be interpreted as being due to the combined effect of the constant and microwave magnetic fields. Subsequently, each sample was deformed for 10–15 min in the presence of a continually acting constant magnetic field up to a relative strain $\varepsilon = 0.5\%$. The microwave field was applied for 20 s several times during the deformation process. This ruled out possible softening of the crystals under the action of only the constant magnetic field [17], and it also made it possible to avoid artifacts associated with mechanical oscillations of the setup at the moment the electromagnetic is switched on. To decrease the variance in the data, the combined action of the fields on the macroplasticity was investigated in a comparatively narrow range of strains ($0.1\% < \varepsilon < 0.5\%$) at the easy-slip stage.

It was established that switching on a microwave field far below the yield point σ_Y did not change the $\varepsilon(t)$ diagram, while near the yield point a magnetic field

decreased σ_Y by 15–20%. Adding a microwave field to a constant magnetic field for $\sigma > \sigma_Y$ increased $d\varepsilon/dt$ (Fig. 3a). The ratio of the rate $(d\varepsilon/dt)_f$ of plastic flow of crystals after a microwave field is switched on to the rate $(d\varepsilon/dt)_0$ of plastic flow before the microwave field is switched on, i.e., $\gamma = (d\varepsilon/dt)_f/(d\varepsilon/dt)_0$, was chosen as the quantitative characteristic of softening. The maximum softening was attained with $B_0 = B_{res1}$ and $B_0 = B_{res2}$ (Fig. 3b). Switching off the magnetic field, as a rule, restored the rate of plastic flow (Fig. 3b). Thus, the combined effect of microwave and constant magnetic fields on macroplastic flow is also of a resonance character, and the positions of the softening maxima are close to the positions observed in the investigation of the free paths of individual dislocations. This means that even with macroplastic deformation the softening of crystals occurs as a result of resonance in *P*…*P* and *P*…*D* pairs.

In the third series of experiments the microhardness *H* of Eu- and Ca-doped NaCl crystals was measured after the crystals were exposed in crossed magnetic fields. Since fresh dislocations appeared under indentation only after exposure of a crystal in a magnetic field, the magnetic field could not have affected the processes in *D*…*P* pairs. Therefore the results obtained in this series of experiments refer to processes in *P*…*P* pairs.

It was observed that several resonance softening peaks can be observed in NaCl crystals with both types of impurities (Fig. 4). There were eight peaks in Eu-doped crystals and three peaks in Ca-doped crystals (Fig. 4), and in the latter case the number and position of the peaks are the same as those revealed by chemical etching (Fig. 1) or by a change in the macrodeformation diagram (Fig. 3), and they also agree with the spectrum obtained for KCl : Ca crystals. Thus, measurements of various characteristics of plasticity of crystals permit recording a spectrum of electronic transitions in a subsystem of structural defects that depends on the type of the main impurity. Therefore Ca and Eu ions are present in *P*…*P* pairs, and the evolution of the complexes investigated depends on the spin state of these ions. This shows that the complexes can possess a different structure, depending on the type of the main impurity in the crystal; this advances the investigation in the direction of identifying point defects sensitive to a magnetic field.

The change in microhardness *H* under resonance conditions with $B_0 = 0.32$ T as compared with the change in H_0 for crystals exposed in crossed magnetic field depended on the exposure duration *t* (Fig. 5). At $T = 293$ K the deviation of *H* from the control value H_0 reached a maximum value in $t \approx 10^3$ s. For $t > 10^3$ s the dependence $H(t)$ saturated. Therefore, the accumulation of a definite number of events initiated by a magnetic field is required in order to change the plasticity of crystals under paramagnetic resonance conditions. This explains partially why our attempt to record paramagnetic resonance spectra in NaCl : Ca crystals using

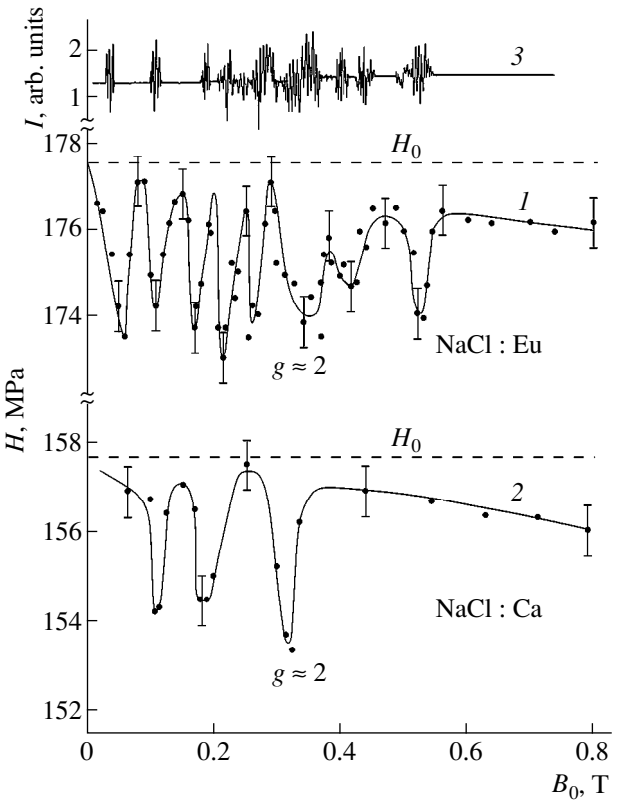


Fig. 4. Microhardness *H*, measured after 15-min exposure of crystals in crossed constant and microwave magnetic fields ($\nu = 9.5$ GHz) versus the induction B_0 of a constant magnetic field: (*1*) in NaCl : Eu, (*2*) in NaCl : Ca. The dashed line shows the values of the microhardness H_0 in crystals which were not subjected to magnetic fields, (*3*) dependence of the signal intensity *I* of a standard EPR spectrometer on B_0, obtained on freshly quenched NaCl : Eu (0.1%) crystals [49].

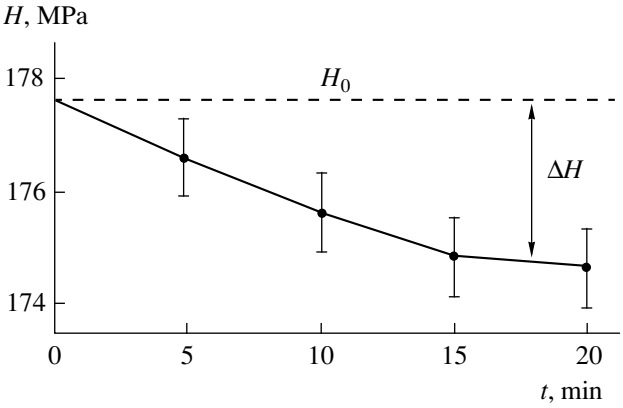


Fig. 5. Microhardness of NaCl : Ca crystals versus their exposure time *t* in crossed constant and microwave magnetic fields under resonance conditions with $B_0 = 0.32$ T and $\nu = 9.5$ GHz. The dashed line shows the microhardness H_0 in crystals not subjected to magnetic fields.

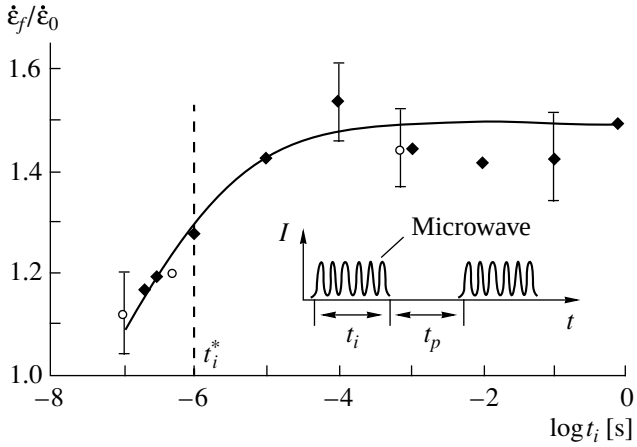


Fig. 6. Softening of NaCl : Eu crystals (ratio of the macroplastic flow rate $(d\varepsilon/dt)_f$ after exposure to a microwave field to the plastic flow rate $(d\varepsilon/dt)_0$ before the exposure to a microwave field) versus the duration t_i of square microwave pulses which modulate the microwave field: ◆—with pause duration between pulses $t_p = 10^{-3}$ s, ○—$t_p = 10^{-7}$ s.

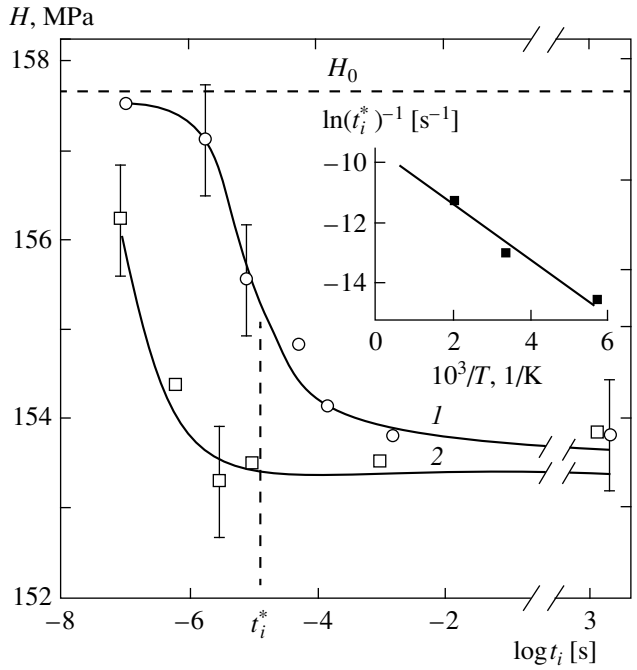


Fig. 7. Microhardness of NaCl : Ca crystals at $T = 293$ K after exposure in a constant magnetic field and a perpendicular microwave field, modulated by square pulses, versus the pulse duration t_i for different temperatures T of exposure in a magnetic field: (*1*) $T = 170$ K, (*2*) $T = 490$ K. The frequency of the microwave field is $\nu = 9.5$ GHz, the constant magnetic induction $B_0 = 0.32$ T, the total duration of application of the microwave field $\sum t_i = 2 \times 10^3$ s is the same in all experiments. The dashed line shows the microhardness H_0 of crystals which have not been exposed to magnetic fields. Inset: Threshold duration t_i^* of microwave packets versus the reciprocal of the temperature in semilogarithmic coordinates.

the standard method (absorption of an electromagnetic wave) was unsuccessful: to detect a spectrum, 10^{12}–10^{13} paramagnetic particles must appear in the sample simultaneously in the microwave spectrometer, while in our experiments the appearance of these particles could be distributed over the time interval ~10^3 s.

A fourth series of experiments was performed to check this conjecture and to measure the expectation time of elementary events occurring in magnetically sensitive complexes. The effect of amplitude modulation of the microfield by square pulses on the rate of macroplastic flow of NaCl : Eu crystals and the microhardness of NaCl : Ca crystals under paramagnetic resonance conditions with $B_0 = 0.32$ T, i.e., at a softening maximum corresponding to $g \approx 2$ (see Figs. 3, 4b), was investigated. Thus, in this series of experiments the resonance softening of the crystals in a constant magnetic field was stimulated by microwave packets with different duration t_i, comparable to the duration of possible electronic and atomic processes in complexes of point defects (see inset in Fig. 6).

It was found that at $T = 293$ K in a wide range, 10^{-5} s $< t_i < 10^3$ s, under otherwise equal conditions, the same softening of the crystals is attained under resonance conditions. For $t_i < 10^{-5}$ s the effect of microwave and constant magnetic fields on the rate of microplastic flow of NaCl : Eu crystals is weaker than in the case of continual microwave generation. At $t_i < 10^{-7}$ s the softening of crystals in crossed magnetic fields is no longer observed, i.e., the ratio $\gamma = (d\varepsilon/dt)_f/(d\varepsilon/dt)_0$ of the rate of plastic flow after the action of a microwave field $(d\varepsilon/dt)_f$ to the rate of strain $(d\varepsilon/dt)_0$ before the action of the field becomes close to 1 (Fig. 6). It was specially checked that varying the pause duration t_p between pulses from 10^{-3} to 10^{-7} s with constant t_i does not change the dependence $\gamma(t_i)$ to within the experimental error (Fig. 6). This means that the decrease in the softening effect of a magnetic field $t_i < 10^{-5}$ s is due to the fact that the duration of the microwave packets becomes inadequate for appearance of an elementary transformation of structural defects.

Similar results were obtained in the measurement of the H microhardness of NaCl : Ca crystals exposed in crossed amplitude-modulated microwave and constant magnetic fields with induction $B_0 = 0.32$ T, i.e., under resonance conditions. The total exposure duration of crystals in the microwave field in these experiments was maintained constant $\sum t_i = 2 \times 10^3$ s; this corresponded to saturation of softening with respect to the residence time in crossed magnetic fields at $T = 170$ K and higher temperatures (see Fig. 5). When t_i decreased to 10^{-4} s, the softening ΔH decreased and vanished at $t_i = 10^{-6}$ s, if the exposure in crossed magnetic fields occurred at $T = 170$ K (Fig. 7). Increasing the temperature at which the crystals are exposed in crossed magnetic fields from 170 to 490 K decreased by a factor of

~100 the critical value t_i^* in which the dependences $H(t_i)$ assumed a value at the center of the descending section (Fig. 7). Increasing the temperature greatly accelerates the change in the microhardness in a constant magnetic field in the absence of a microwave field (Fig. 8). The results of this series of experiments attest to the fact that as the temperature increases, events which are favorable for changes to occur in the complexes of points defects under the action of a magnetic field occur more often.

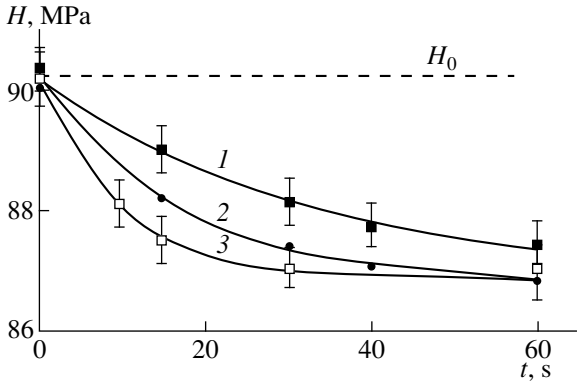


Fig. 8. Microhardness of KCl : Ca crystals versus their exposure time t in a constant magnetic field with induction $B_0 = 2.3$ T in the absence of a microwave field at various temperatures: (*1*) $T = 233$ K, (*2*) $T = 388$ K, (*3*) $T = 523$ K. The dashed line shows the microhardness H_0 in crystals not exposed to magnetic fields.

4. DISCUSSION

It follows from what we have said above that the results obtained are similar to the observations made in EPR spectroscopy, with the exception that the response was not the absorption of an electromagnetic wave but rather a change in the characteristics of the plasticity of a crystal. The correspondence between the values of B_{res} found experimentally and the standard values of the effective g factors of electrons localized on defects in crystals as well as the vanishing of the softening of crystals for a parallel orientation of the vectors $\mathbf{B}_1$ and $\mathbf{B}_0$ unequivocally attest to the fact that electron spin resonance is responsible for the plastification of the crystals. We shall discuss the change occurring in the properties of metastable complexes of point defects $P\ldots P$ in crossed magnetic fields, and we shall not discuss the reasons for the possible appearance of a resonance in $D\ldots P$ pairs.

The sensitivity of the detected spectra to the type of main impurity convincingly shows that impurity ions are constituents of the magnetically sensitive defect complexes which were investigated. The three resonance maxima in NaCl : Ca crystals (Figs. 1, 3, 4) could correspond to "allowed" and "forbidden" transitions between $2s + 1$ states in a complex of particles with total spin $s = 3/2$. In this case the maximum at $g \approx 2$ corresponds to allowed and the maxima at $g \approx 4$ and $g \approx 6$ corresponds to forbidden transitions. The complexes investigated can be formed by Ca^+ ions and (or) neutral atoms Ca^0, since the total spin of the Ca^{++} ion is zero. Since the complexes are sensitive to an electric field [20], Ca^+ ions are the correct choice. In Eu-doped crystals the spectrum of electronic transitions is even richer, as usually happens for rare-earth elements. Comparing the positions of the lines in a standard EPR spectrometer, measured in strongly doped quenched NaCl : Eu crystals [58], with the position of the maxima obtained in our experiments on measuring the microhardness (Fig. 4) reveals that the spectra are qualitatively the same.

We shall now discuss the possible reasons for the appearance of a threshold modulation frequency of the microwave field (Figs. 6, 7). In the most general case the sequence of events can be represented in terms of the following chain of processes: (1) events preceding the spin stage (excitation of a complex by fluctuations or mechanical stresses, encounter of paramagnetic particles, and so on); (2) a spin stage, which could consist of a change in the multiplet structure of the nonequilibrium pair, which in itself is insufficient for atomic rearrangement of complex; (3) electronic rearrangement of a complex, being a consequence of a change in the multiplet structure in a pair of particles; and, (4) a change in the atomic configuration of a complex (dissociation, reorientation, reconnection, and change in the length of a covalent bond, and so on). A constant magnetic field or the combined effect of a constant magnetic field and a microwave magnetic field can change something only at the second stage of this chain.

Generally speaking, any of the four stages enumerated above can be the limiting stage. However, it is necessary to take account of the fact that the duration τ_{S-T} of the intercombination transition cannot depend on the temperature, in contrast to the times t_i^* obtained in our experiments (see Fig. 7). Therefore, the presence of a threshold in the dependence $H(t_i)$ cannot be attributed to the delay of intercombination transitions, and the limiting stage in the evolution of complexes seems to precede the spent stage or appears after this stage is completed.

Assuming an Arrhenius temperature dependence for t_i^* (see inset in Fig. 7), the activation energy can be estimated as $E_{fl} = 0.15 \pm 0.05$ eV, which characterizes the height of the potential barrier that must be overcome in order to excite a complex into a magnetically sensitive state. Thermal activation analysis of the relaxation kinetics of complexes, followed according to the change in the microhardness (Fig. 8) or free path of individual dislocations [23] in only a constant magnetic field in the absence of a microwave field, yields the same value of E_{fl}. The value of E_{fl} is of the same order of magnitude as the experimentally determined value [59]

and the theoretically computed values [60] of the binding energy of complexes consisting of 2–4 impurity-vacancy dipoles. This also agrees with the conclusions drawn above that the ions of the main impurity are constituents of the complexes investigated.

To explain the role of magnetic fields in the change in the spin states of defects, the basic assumptions of thermodynamics and the theory of spin-dependent magnetically sensitive chemical reactions must be taken into account [50, 51, 61]:

(1) a magnetic field can change the multiplet structure of a pair of defects only if the energy difference between its triplet and singlet states—$E_T - E_S$—is comparable to the energy $U_e \approx \mu Bg$ which the field transfers to the magnetic moment of the electron, i.e., for a weakened covalent bond;

(2) in statistics, a constant magnetic field can influence the multiplet structure only of spin-correlated pairs of paramagnetic defects, i.e., occupying predominantly the singlet S or triplet T states, since for an equally probably outcome of the filling of S and T levels the frequency of S–T and T–S transitions will be the same and the resulting distribution of pairs over levels will not change in the magnetic field (to within the Zeeman splitting, which can explain only a part $\sim \mu g B_0/kT \sim 10^{-3}$ of the total effect observed);

(3) in order for a weak constant magnetic field, imparting energy less than kT by two–three orders of magnitude to the spins, to be able to change the macroproperties of crystals by tens and hundreds of percent, the lifetime τ_{ex} of the magnetically sensitive state must satisfy the inequalities $\tau_{S-T} < \tau_{ex} < \tau_r$, where τ_{S-T} is the duration of the intercombination transition and τ_r is the spin relaxation time. In our experiments it follows from the half-width of the peaks that $\tau_{ex} \sim h/\mu g \Delta B \sim 10^{-8}$ s, which can easily be less than τ_r at room temperature. On the other hand, for $B_0 \sim 1$ T and $\Delta g \sim 10^{-2}$ in the absence of a magnetic field $\tau_{S-T} \sim h/\mu \Delta g B_0 \sim 10^{-8}$ s, i.e., the transition can easily occur. Thus, only the satisfaction of the inequalities mentioned above gives isolation of the spin system from thermal fluctuations and opens up the possibility for a weak magnetic field to interfere in the spin evolution of the subsystem of defects which are far from thermodynamic equilibrium.

Two, at first glance contradictory, requirements follow from the conditions enumerated above. On the one hand, it is necessary to explain the nature of the magnetic sensitivity of $P\ldots P$ complexes, which is observed over a period of several months and years, which is incomparably longer than any reasonable duration of a spin-lattice relaxation. On the other hand, the magnetically sensitive states must be short-lived and must be produced in a spin-polarized state.

In our view, these requirements can be reconciled if it is assumed that the complexes $P\ldots P$ are present in a crystal in a dynamical equilibrium between (at least) two electronic states. In one state the spins form a stable covalent bond whose energy is much greater than kT and therefore "preserves" the spin polarization. In this state, neither a "weak" magnetic field nor thermal fluctuations can change the multiplet structure of the complexes. In the other, magnetically sensitive, state the complexes $P\ldots P$ must have a weakened (almost broken) bond, for which the energy difference of the singlet S and triplet T states is comparable to $U_e = \mu g B$, so that even a "weak" magnetic field can lead to S–T transitions. The second state must have a short enough lifetime so that during the life of the state spin-lattice relaxation does not change the total spin of a pair in the absence of a magnetic field. In other words, magnetically sensitive states of $P\ldots P$ complexes seem to be "scintillating." We note that other point defects which behave similarly are also now known. They include, for example, metastable V_k centers, in which the covalent bond between the parts is spontaneously broken by thermal fluctuations and is restored once again, so that the defect is a pair of "flickering" spins [62].

In our experiments the absence of a source generating magnetically sensitive short-lived states $P\ldots P$ during exposure of crystals in a magnetic field attests to the existence of internal factors which are responsible for the transitions of complexes between "preserved" and magnetically sensitive states. The increase in the rate of magnetically stimulated relaxation of complexes with increasing temperature (Figs. 7, 8) shows that these transitions are generated by thermal fluctuations. If the thermal fluctuations did not play any role in the observed processes, the only result of increasing the temperature would be a decrease in the spin-lattice relaxation time, which would only degrade the conditions for a magnetic field to influence a complex and would decrease the magnetoplastic effect. At the same time, it is obvious from Fig. 8 that an increase in temperature not only does not decrease the efficiency of magnetically stimulated relaxation of complexes, but it even increases the rate of such relaxation.

The results obtained makes it possible to choose the state (S or T) that is predominantly populated at the moment the short-lived states of complexes are created. A microwave field can initiate transitions only between states with nonzero total spin: for example, in a pair of two particles with spin 1/2 a microwave field mixes the states T_+ and T_- (with projections of the total spin in the direction $\mathbf{B}_0$ equal to +1 and –1, respectively) and the state T_0 (with zero projection of the spin on $\mathbf{B}_0$) [50, 51]. Therefore, the role of the microwave field is to "pump" electrons from the T_0 state and indirectly initiate S–T_0 transitions in the case when S states are predominantly filled and their transitions are retarded in the case when the T states are overfilled. A constant magnetic field can only increase the rate of S–T_0 transitions, assuming a Δg mixing mechanism, which best corresponds to our situation and takes into account the difference of the g factors of the paramagnetic parts of the complex, Δg. According to the experimental data, a constant magnetic field increases the free path of dislocations. The

same effect is observed under the combined action of constant and microwave magnetic fields. This shows that the *S* state is predominantly filled at the moment the magnetically sensitive states are created.

The experimental data and theoretical considerations enumerated above taken together make it possible to represent schematically a possible mechanism by which magnetic fields influence the evolution of metastable complexes (Fig. 9). Thermal fluctuations with frequency ν_1 excite a complex, for example, by stretching the covalent bond (or by changing other configuration coordinates *r*, for example, the valence angles) from the initial state *S* into the excited state *S**. In the absence of a magnetic field, the elastic forces exerted by the crystal lattice return the complex into the initial *S* state by virtue of the fact that a change in the total spin is forbidden, so that the complex is in a dynamical equilibrium established between the *S* and *S** states.

A magnetic field partially lifts the forbiddenness and a complex, whose multiplet structure has changed, with frequency $\nu_2 = \mu_e B\Delta g/h$ (for the Δg mechanism of mixing of the states which is most likely under our conditions [61, 63]) is transferred into a new electronic state *T**. Next, the elastic forces exerted by the crystal lattice cause a return motion of the nuclei (the parts *P*…*P*). The equilibrium state between them, R_T, is larger than in the singlet state, R_S, since the negative value of the exchange integral *j* causes parts of the complex to repel one another. Thus, a relatively long-lived *T* state, in which the total binding energy of the parts of a complex is ΔE = 0.1–1 eV less than in the *S* state, is formed with frequency ν_3. The complexes which are "opened" in this manner are less stable than the initial complexes, and the random motions of nuclei can cause them to decay with frequency ν_4, which is accompanied by the system leaving a local energy minimum and further relaxing or returning with frequency ν_5 into the initial *S* state.

The products of decay of complexes with free energy less than that of the initial complexes differ from the latter also by the dilatation properties and therefore the efficiency of the interaction with dislocations. As a rule, the decay of complexes of point defects results in the formation of weaker stops for dislocations [64]; this agrees with the experimental data with respect to the softening effect of the exposure of ionic crystals in magnetic fields before fresh dislocations are introduced into the crystals.

The ideas developed above could be universal, since for a definite choice of the frequencies ν_1, ν_2, ν_3, ν_4, and ν_5 they can be used to describe magnetically stimulated relaxation of complexes of various kinds in diverse materials [17–30, 37, 38, 54–57]. Of course, the real number of stages of the above-described spin-dependent relaxation process can be much larger than the number considered, and further detailing of their sequence is certainly of interest. Some post-spin stages are investigated in detail in [17, 20, 21].

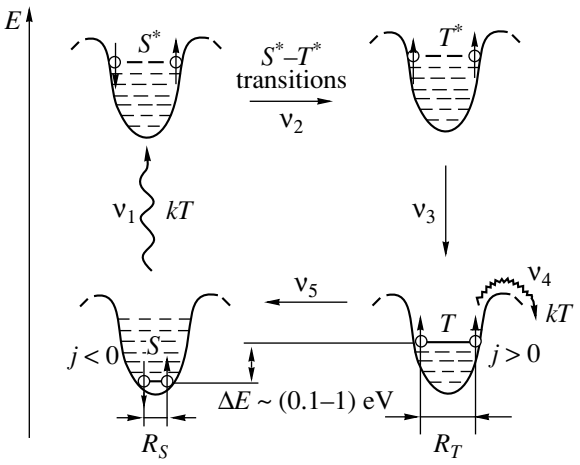


Fig. 9. Schematic diagram of the sequence of processes occurring in complex point defects in a magnetic field on the scale of the energies *E* of the complex. (*S*) *0*long-lived metastable singlet state of the complex, (*S**) intermediate singlet state of a complex excited by thermal fluctuations, (*T**) intermediate excited triplet state, into which a transition is allowed only in the presence of a magnetic field, (*T*) long-lived metastable triplet state. The local minimum characterizes the profile of the elastic interaction between the parts of a complex in a metastable state. The solid and wavy lines connecting parts of a complex represent a covalent bond, respectively, in equilibrium and excited states; (*kT*) thermally stimulated process, (*j*) exchange integral, (ΔE) difference of exchange energies in *S* and *T* states of a complex, (ν_i) frequencies of transitions between states.

We note that considerations similar to those developed above could be applicable for discussing the effect of a magnetic field on the multiplet structure of *D*…*P* type pairs. In so doing, it is necessary to take account of the fact that besides thermal fluctuations internal mechanical stresses, stretching the covalent bond between a dislocation and a paramagnetic center in the bulk, operate on them. This factor turns out to be ~*d*/*a* times stronger for *D*…*P* pairs than for *P*…*P* pairs (*d* is the distance between the pinning points of dislocations and *a* is the lattice parameter). For lightly doped crystals, in which magnetoplastic effects are observed, $d/a \sim 10^2$–10^3, which can render the contribution to the thermal activation in the depinning of a dislocation from a stop in a magnetic field negligible and the magnitude of the magnetoplastic effect essentially temperature-independent, as shown experimentally in [10].

5. CONCLUSIONS

In summary, softening of crystals under the conditions of electron paramagnetic resonance was observed. It was established experimentally that in ionic crystals spin-dependent magnetically sensitive reactions in the subsystem consisting of paramagnetic structural defects make an appreciable contribution to their plastic properties, a weak constant magnetic field can be used to regulate the kinetics of these reactions effectively, and the com-

bined use of a constant magnetic field together with a microwave field can be used to do this even more effectively. Record-breaking short stages of plastic flow of ionic crystals ($\sim 10^{-9}$–10^{-6} s), which play a decisive role in controlling the mechanical properties via a change in the spin states of structural defects, were identified.

It was shown that the plasticity parameters of crystals can be used as indicators for the spin resonance in nonthermalized short-lived complexes of paramagnetic defects. The results obtained can serve as a basis for a new high-frequency method for investigating paramagnetic structural defects in crystals that makes it possible to establish directly the relationship between their plastic properties and the electronic state of defects. The proposed method of investigation can be extended to a wide range of materials in which spin-dependent reactions between defects accompany the process of plastic deformation.

We note that the facilitation of depinning of dislocations from stops under the action of a magnetic field can lead to, besides magnetoelastic effects, "magnetoelastic" effects which are not associated with lattice magnetostriction. Indeed, it is known from investigations of the internal friction under a sign-alternating load that an increase of the dislocation damping decrement is always associated with a corresponding decrease in the modulus of elasticity: the "defect of the modulus." A change in Young's modulus in ionic crystals in a constant magnetic field with induction 0.3 T was observed in [31, 32]. Evidently, the possibility of effects of this kind must be taken into account in the operation of precision physical apparatus (vibrational systems, pendulum suspensions, gravity antennas, and so on), which weak magnetic fields (for example, the earth's magnetic field or technogenic magnetic fields) can influence, even if the deformations achieved are far from the limit of macroscopic elasticity.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 97-02-16074) and the program "Universities of Russia—Fundamental Research" (project no. 381).

REFERENCES

1. S. Z. Shmurak, Izv. Akad. Nauk SSSR, Ser. Fiz. **40**, 1886 (1976).
2. V. A. Zakrevskiĭ and A. V. Shuldiner, Philos. Mag. **71**, 127 (1995).
3. M. Czapelski and M. Suszynska, Phys. Status Solidi B **132**, 409 (1985).
4. L. G. Kirichenko and V. F. Petrenko, Fiz. Tverd. Tela (Leningrad) **22**, 1590 (1980) [Sov. Phys. Solid State **22**, 929 (1980)].
5. M. V. Vlasova and N. G. Kakazeĭ, *Electron Paramagnetic Resonance in Failed Solids* (Naukova Dumka, Kiev, 1979).
6. W. Raupach, Phys. Status Solidi A **58**, 435 (1980).
7. Yu. A. Osip'yan and V. F. Petrenko, Dokl. Akad. Nauk SSSR **226**, 803 (1976).
8. M. I. Kaganov, V. Ya. Kravchenko, and V. D. Natsik, Usp. Fiz. Nauk **111**, 655 (1973) [Sov. Phys. Usp. **16**, 878 (1973)].
9. V. I. Al'shits, E. V. Darinskaya, T. M. Perekalina, and A. A. Urusovskaya, Fiz. Tverd. Tela (Leningrad) **29**, 467 (1987) [Sov. Phys. Solid State **29**, 265 (1987)].
10. V. I. Al'shits, E. V. Darinskaya, and E. A. Petrzhik, Fiz. Tverd. Tela (Leningrad) **33**, 3001 (1991) [Sov. Phys. Solid State **33**, 1694 (1991)].
11. V. I. Al'shits, E. V. Darinskaya, and O. L. Kazakova, Pis'ma Zh. Éksp. Teor. Fiz. **62**, 352 (1995) [JETP Lett. **62**, 375 (1995)].
12. A. A. Urusovskaya, V. I. Al'shits, A. E. Smirnov, and N. N. Bekkauer, Pis'ma Zh. Éksp. Teor. Fiz. **65**, 471 (1997) [JETP Lett. **65**, 497 (1997)].
13. V. I. Al'shits, N. N. Bekkauer, A. E. Smirnov, and A. A. Urusovskaya, Zh. Éksp. Teor. Fiz. **115**, 951 (1999) [JETP **88**, 523 (1999)].
14. V. I. Al'shits, E. V. Darinskaya, and O. L. Kazakova, Fiz. Tverd. Tela (St. Petersburg) **40**, 81 (1998) [Phys. Solid State **40**, 70 (1998)].
15. V. I. Al'shits, E. V. Darinskaya, E. Yu. Mikhina, and E. A. Petrzhik, Fiz. Tverd. Tela (St. Petersburg) **38**, 2426 (1996) [Phys. Solid State **38**, 1333 (1996)].
16. V. I. Al'shits, E. V. Darinskaya, and O. L. Kazakova, Zh. Éksp. Teor. Fiz. **111**, 615 (1997) [JETP **84**, 338 (1997)].
17. Yu. I. Golovin and R. B. Morgunov, Zh. Éksp. Teor. Fiz. **115**, 605 (1999) [JETP **88**, 332 (1999)].
18. Yu. I. Golovin and R. B. Morgunov, Pis'ma Zh. Éksp. Teor. Fiz. **61**, 583 (1995) [JETP Lett. **61**, 596 (1995)].
19. Yu. I. Golovin, R. B. Morgunov, D. V. Lopatin, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **40**, 2065 (1998) [Phys. Solid State **40**, 1870 (1998)].
20. Yu. I. Golovin, R. B.Morgunov, A. V. Tyutyunnik, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **40**, 2184 (1998) [Phys. Solid State **40**, 1981 (1998)].
21. Yu. I. Golovin, R. B. Morgunov, S. E. Zhulikov, *et al.*, Fiz. Tverd. Tela (Leningrad) **39**, 634 (1997) [Phys. Solid State **39**, 554 (1997)].
22. Yu. I. Golovin and R. B. Morgunov, Pis'ma Zh. Éksp. Teor. Fiz. **58**, 189 (1993) [JETP Lett. **58**, 191 (1993)].
23. Yu. I. Golovin and R. B. Morgunov, Fiz. Tverd. Tela (St. Petersburg) **37**, 1239 (1995) [Phys. Solid State **37**, 674 (1995)].
24. Yu. I. Golovin, R. B. Morgunov, and S. E. Zhulikov, Fiz. Tverd. Tela (St. Petersburg) **39**, 495 (1997) [Phys. Solid State **39**, 430 (1997)].
25. Yu. I. Golovin, R. B. Morgunov, and V. E. Ivanov, Fiz. Tverd. Tela (St. Petersburg) **39**, 630 (1997) [Phys. Solid State **39**, 550 (1997)].
26. Yu. I. Golovin and R. B. Morgunov, Izv. Akad. Nauk, Ser. Fiz. **61**, 850 (1997).
27. Yu. I. Golovin, R. B. Morgunov, and S. E. Zhulikov, Izv. Akad. Nauk, Ser. Fiz. **61**, 965 (1997).
28. Yu. I. Golovin and R. B. Morgunov, Dokl. Akad. Nauk **354**, 632 (1997).

29. Yu. I. Golovin, R. B. Morgunov, M. V. Badylevich, and S. Z. Shmurak, Fiz. Tverd. Tela (St. Petersburg) **39**, 1389 (1997) [Phys. Solid State **39**, 1232 (1997)].
30. Yu. I. Golovin and R. B. Morgunov, Chem. Rev. **23**, 23 (1998).
31. É. P. Belozerova, A. A. Svetashev, and V. L. Krasnikov, Izv. Akad. Nauk, Ser. Fiz. **61**, 291 (1997).
32. N. A. Tyapunina, V. L. Krasnikov, and E. P. Belozerova, Fiz. Tverd. Tela (St. Petersburg) **41**, 1035 (1999) [Phys. Solid State **41**, 942 (1999)].
33. Yu. I. Golovin, R. B. Morgunov, and S. Yu. Liksutin, Vysokomol. Soed., Ser. B **40**, 373 (1998).
34. V. A. Zhorin, L. L. Mukhina, and I. V. Razumovskaya, Vysokomol. Soed., Ser. B **40**, 1035 (1998).
35. V. A. Zhorin, L. L. Mukhina, and I. V. Razumovskaya, Vysokomol. Soed., Ser. B **40**, 1213 (1998).
36. N. N. Peschanskaya, V. Yu. Surovova, and P. N. Yakushev, Fiz. Tverd. Tela (St. Petersburg) **34**, 2111 (1992) [Sov. Phys. Solid State **34**, 1127 (1992)].
37. Yu. I. Golovin, R. B. Morgunov, A. A. Baskakov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 114 (1999) [JETP Lett. **69**, 127 (1999)].
38. Yu. A. Osip'yan, R. K. Nikolaev, S. Z. Shmurak, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 110 (1999) [JETP Lett. **69**, 123 (1999)].
39. V. I. Al'shits, E. V. Darinskaya, I. V. Gektina, and F. F. Lavrent'ev, Kristallografiya **35**, 1014 (1990) [Sov. Phys. Crystallogr. **35**, 597 (1990)].
40. V. I. Al'shits, R. Voska, E. V. Darinskaya, and E. A. Petrzhik, Fiz. Tverd. Tela (St. Petersburg) **35**, 70 (1993) [Phys. Solid State **35**, 37 (1993)].
41. V. I. Al'shits, E. V. Darinskaya, and E. A. Petrzhik, Fiz. Tverd. Tela (St. Petersburg) **34**, 155 (1992) [Sov. Phys. Solid State **34**, 81 (1992)].
42. O. I. Datsko and V. I. Alekseenko, Fiz. Tverd. Tela (Leningrad) **39**, 1234 (1997) [Phys. Solid State **39**, 1094 (1997)].
43. O. I. Datsko, V. I. Alekseenko, and A. D. Shakhova, Fiz. Tverd. Tela (St. Petersburg) **38**, 1799 (1996) [Phys. Solid State **38**, 992 (1996)].
44. M. I. Molotskiĭ and V. N. Fleurov, Phys. Rev. Lett. **78**, 2779 (1997).
45. M. I. Molotskiĭ and V. N. Fleurov, Phys. Rev. B **52**, 15829 (1995).
46. M. I. Molotskiĭ, R. E. Kris, and V. N. Fleurov, Phys. Rev. B **51**, 12531 (1995).
47. Yu. I. Golovin, R. B. Morgunov, V. E. Ivanov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 400 (1998) [JETP Lett. **68**, 426 (1998)].
48. E. L. Frankevich, V. I. Lesin, and A. I. Pristupa, Zh. Éksp. Teor. Fiz. **75**, 415 (1978) [Sov. Phys. JETP **48**, 208 (1978)].
49. E. L. Frankevich, A. I. Pristupa, and V. I. Lesin, Chem. Phys. Lett. **47**, 304 (1977).
50. E. L. Frankevich, A. I. Pristupa, M. M. Tribel', and I. A. Sokolik, Dokl. Akad. Nauk SSSR **236**, 1173 (1977).
51. Ya. B. Zel'dovich, A. L. Buchachenko, and E. L. Frankevich, Usp. Fiz. Nauk **155**, 3 (1988) [Sov. Phys. Usp. **31**, 385 (1988)].
52. A. L. Buchachenko, Usp. Fiz. Nauk **62**, 1139 (1993).
53. M. I. Molotskiĭ and V. N. Fleurov, Philos. Mag. Lett. **73**, 11 (1996).
54. G. I.Distler, V. M. Kanevskiĭ, V. V. Moskvin, *et al.*, Dokl. Akad. Nauk SSSR **268**, 591 (1983) [Sov. Phys. Dokl. **28**, 43 (1983)].
55. M. N. Levin, Yu. O. Lichmanov, and V. M. Maslovskiĭ, Pis'ma Zh. Tekh. Fiz. **20** (4), 27 (1994) [Tech. Phys. Lett. **20**, 145 (1994)].
56. V. M. Maslovskiĭ, Yu. A. Klimov, N. S. Samsonov, and E. V. Simanovich, Fiz. Tekh. Poluprovodn. (St. Petersburg) **28**, 772 (1994) [Semiconductors **28**, 453 (1994)].
57. M. N. Levin and B. A. Zon, Zh. Éksp. Teor. Fiz. **111**, 1373 (1997) [JETP **84**, 760 (1997)].
58. H. Opyrchal, K. D. Nierzewski, and H. Drulis, Phys. Status Solidi B **118**, K125 (1983).
59. C. Zaldo and F. Agullo-López, Philos. Mag. A **53**, 51 (1986).
60. N. M. Bannon, J. Corish, and P. W. M. Jacobs, Philos. Mag. A **51**, 797 (1985).
61. K. M. Salikhov, R. Z. Sagdeev, and A. L. Buchachenko, *Spin Polarization and Magnetic Effects in Radical Reactions*, Ed. by Yu. N. Molin (Nauka, Novosibirsk, 1978; Elsevier, Amsterdam, 1984).
62. Yu. R. Zakis, L. N. Kantorovich, and E. A. Kotomin, *Model of Progress in Wide-Gap Solids with Defects* (Zinatne, Riga, 1991).
63. A. P. Tyutnev, A. V. Vannikov, G. S. Mingaleev, and V. S. Saenko, *Electric Phemonena in Polymer Radiation* (Énergoatomizdat, Moscow, 1985).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 950–963.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1094–1109.

SOLIDS
Structure

Molecular Transport in Subnanometer Channels

V. D. Borman*, **, V. V. Teplyakov*, V. N. Tronin*, I. V. Tronin*, and V. I. Troyan***

**Moscow State Engineering-Physics Institute, Moscow, 115409 Russia*
***e-mail: borman@bvd.mephi.ru*
****Topchiev Institute of Petrochemical Synthesis, Russian Academy of Sciences, Leninskiĭ pr. 29, Moscow, 117912 Russia*
Received December 27, 1999

Abstract—The density functional method is used to show that the description of molecular transport in subnanometer channels reduces to the description of diffusion in a one-dimensional system where strong density fluctuations with a finite cluster lifetime. A new diffusion mechanism is proposed; it makes it possible to explain the transition from activated diffusion of single particles in a channel at a low filling factor to fast barrier-free diffusion, which consists of the propagation of density disturbances at high filling factor. It is shown that as the filling factor increases, the attraction between the molecules (the effective attraction of molecules—hard spheres) causes the energy barrier for diffusion along the channel axis to vanish. Another consequence of the "effective" attraction between the molecules is the formation of molecular clusters in the channel, which possess a finite lifetime because of the one-dimensionality of the system. The size and lifetime of the clusters increase with the filling factor of the channel. The diffusion of particles in clusters is a barrier-free process of propagation of density disturbances. The dependences obtained for the diffusion coefficient on the pressure, temperature, and filling factor make it possible to describe, even in the hard-sphere model, all experimental data known to the authors. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Molecular transport in nanometer channels in porous bodies is currently drawing a great deal of interest from the standpoint of fundamental science [1–5] and because of the many applications of membrane and nanotechnologies in various fields ranging from nuclear power to ecology [6, 7]. Polycrystalline ceramic membranes consisting of complex oxides (zeolites), possessing subnanometer channels ranging in diameter from 0.3 to 1.4 nm, have been synthesized recently [6]. On account of the high selectivity of molecular transport in zeolite membranes, as compared with the well-known polymeric membranes [7–9], new technologies for separation, reprocessing, and utilization of materials are being intensively developed on the basis of zeolite membranes [7–9].

When the channel diameters in membranes decrease to the nanometer scale, molecular transport is determined by Knudsen flow in the central zone of the channel, free of interaction of molecules with walls, and particle diffusion in the surface force field [5]. In subnanometer channels the interaction potentials between molecules and the opposite walls overlap and molecular transport occurs under conditions of a constant interaction of molecules with a solid. Consequently, the diffusion coefficient in the limit of small filling numbers of a channel is determined by the relaxation of particles on phonons and surface defects [3]. For molecules, with the exception of light particles (H, He), the channel walls are impenetrable, and consequently molecular transport is possible only along the channel axes. In this sense it is different from diffusion in solids and may be assumed to be one-dimensional. In contrast to surface diffusion in channels with diameter $d > 1$ nm, a fundamentally new property of molecular transport should appear in subnanometer channels. As the filling factor of a channel or the diameter (σ) of molecules increases, so that $d < 2\sigma$, each molecule can block the motion of two molecules. Consequently, it can be expected that as the diameter of the molecules or the external gas pressure increases, the molecular flux in membranes with subnanometer channels should decrease. However, experiments have shown that the diffusion flux of a number of molecules (CH_4, C_2H_6, and others) in a ZSM-5 zeolite membrane with channel diameters 0.54–0.57 nm does not decrease but rather increases as the external gas pressure increases [9]. It has also been established that the diffusion coefficient for these gases increases by more than an order of magnitude as the filling factor of the channels with molecules increases. It has been found that for a number of gases the temperature dependence of the flux possesses a maximum and a minimum. A microscopic theory explaining these laws is not available in the literature.

In the present paper transport of a single-component molecular gas in subnanometer-diameter channels is investigated theoretically. The analysis of the experimental data, performed in Section 4, showed that for most gases investigated the sorption and transport can be treated as a process in a one-dimensional (1D) system. In statistical physics the theoretical models of 1D systems are classical examples of exactly solvable models [10]. It is well known [11, 12] that in 1D systems a phase transition (condensation) does not occur

as density increases. This means that, specifically, no critical nuclei are present in the system and stable nuclei of a new phase with macroscopically long lifetimes do not appear in the system. At the same time, the system of 1D channels in zeolite membranes consists of channels of finite length, where for a sufficiently high filling factor clusters with sizes comparable to the channel length can form. Thus, the description of molecular transport in zeolite membranes reduces to describing transport in a 1D system, where there are strong density fluctuations with a finite lifetime of clusters. The density functional method [4] makes it possible to calculate the spectrum of density fluctuations and the diffusion coefficient for particles with arbitrary density and arbitrary laws of interaction with one another and with the channel walls. Since the flux measured in the experiments performed in [7] is determined by the diffusion coefficient and the filling factor of the channels in the membrane, the sorption isotherm of a one-component molecular gas is calculated separately in Section 2. For the one-dimensional channels considered in the present paper, this problem is solved exactly for an arbitrary intermolecular interaction potential. A new feature here is an analysis of the dependence of the lifetime and size of the 1D clusters on the degree of filling of the channel with particles. A description of molecular transport for an arbitrary filling factor (θ) of a channel is given in Section 3. It was found that as the filling factor of a channel approaches 1, so that blocking of the relative motion of the particles becomes substantial, the diffusion coefficient increases. A new diffusion mechanism is proposed. It makes it possible to explain the transition from activated diffusion of single particles in a channel with a low filling factor to collective fast barrier-free diffusion, which consists of the propagation of density disturbances for large values of θ. As shown below, in the hard-sphere approximation, as θ increases, the well-known [12] effective attraction of the molecules—hard spheres—results in vanishing of the energy barrier for diffusion along the channel axis. Another consequence of the effective attraction of molecules is the formation of molecular clusters in the channel, which have a finite lifetime because of the one-dimensionality of the system. The size and lifetime of the clusters increase as θ increases. The diffusion of particles in clusters is described as a barrier-free process of propagation of density disturbances. The pressure, temperature, and filling factor dependences of the flux of molecules make it possible, even in the high-sphere model, to describe the experimental data currently known to the authors (Section 4).

2. SORPTION ISOTHERM IN A 1*D* CHANNEL

Let us consider a surface in contact with an ideal single-component gas with temperature T and pressure P. Let us assume that particles located on the surface do not interact with one another. We shall also assume that the energy of the gas molecules on the surface is ε_0. Let there be N adsorption centers on the surface. Then the average number $\bar{N}_1$ of particles adsorbed on the surface as a function of the gas pressure and temperature has the form (Langmuir isotherm)

$$\frac{\bar{N}_1}{N} = \frac{P}{P + P_0(T)}, \qquad P_0(T) \equiv T\left(\frac{\hbar^2}{2\pi m T}\right)^{-3/2} \exp(-\beta\varepsilon_0), \tag{2.1}$$

where m is the mass of a gas molecule, $\hbar$ is Planck's constant, and $\beta = T^{-1}$ is the reciprocal of the temperature.

The method used to derive the Langmuir isotherms (2.1) admits wide extensions. Specifically, this method can be used to solve the problem of the filling factor of a cylindrical channel with diameter d taking into account the interparticle interaction in the channel. Let the channel diameter be comparable to the maximum diameter of a gas molecule. Let us consider the equilibrium of the gas with the surface on which k of the channels described above emerge. Let ε_1 be the binding energy of a particle at the entrance to the channel. If the energy ε_1 is negative, then it is energetically favorable for the gas molecule to enter the channel. Let q be the total number of particles in a channel of length L, n the total number of particles in the channel and on the surface, and N_0 the number of settling locations in the channel. Then the partition function for the grand canonical ensemble, taking into account the interaction of the gas particles in the channel, is

$$\Theta = \sum \frac{(N-k)!\exp[\beta\varepsilon_0(N_1-n)]}{(N_1-n)!(N-k-(N_1-n))!} \times \frac{k!\exp[\beta\varepsilon_1(n-q)]}{(n-q)!(k-(n-q))!} \times \frac{N_0!\exp(\beta\varepsilon_2 q)}{q!(N_0-q)!}\exp(\beta\mu N_1) Z_{\text{int}}(q), \qquad N_1 + q = n. \tag{2.2}$$

Here ε_2 is the binding energy of the particles in the channel and $Z_{\text{int}}(q)$ is the partition function, corresponding to taking account of the interaction of the gas particles in the channel. Since Z_{int} depends on the number q of particles in the channel, it is impossible to calculate the partition function (2.2) in the grand canonical ensemble in the general case. However, in the problem of filling of a channel the states for which the number of particles in the channel $q \gg 1$ should make the main contribution to the partition function (2.2). Then the quantity $Z_{\text{int}}(q)$ can be replaced by the partition function of interacting particles in the channel, calculated with the average number $\bar{q}$ of particles in the channel,

$$Z_{\text{int}}(q) \approx Z_{\text{int}}(\bar{q}). \tag{2.3}$$

The approximation (2.3) corresponds to the thermodynamic limit ($q \gg 1$) for gas molecules located in the channel. Using the relation (2.3), the partition function (2.2) can be easily calculated. Using the condition for equilibrium between the gas and the surface, we obtain for the average number $\bar{q}$ of particles in the channel

$$\frac{\bar{q}}{N_0} = \frac{p}{p + \tilde{p}(T, \bar{q})}, \quad \tilde{p}(\bar{q}, T) = \frac{\exp[-\beta\varepsilon(\bar{q}, T)]}{\alpha(T)},$$
$$(\alpha(T))^{-1} = T\left(\frac{\hbar^2}{2m\pi T}\right)^{-3/2}, \quad \varepsilon = \varepsilon_1 - F_{\text{int}}(\bar{q}, T), \tag{2.4}$$
$$F_{\text{int}}(\bar{q}, T) \equiv -T\ln Z_{\text{int}}(q, \bar{T}).$$

Here $F_{\text{int}}(\bar{q}, T)$ is the free energy of interaction per gas particle in the channel. Instead of the number N_0 of settling locations in the channel, it is convenient to introduce the average distance between settling locations in the channel ($\eta = L/N_0$) and to replace $\bar{q}$ by the filling factor $\theta \equiv \bar{q}\sigma/L$ of particles in the channel. Then, we obtain from equation (2.4) an equation determining the pressure and temperature dependences of the filling factor θ of the channel:

$$\frac{\eta\theta}{\sigma} = \frac{p}{p + \tilde{p}(T, \theta)}. \tag{2.5}$$

The relations (2.4) show that the problem of obtaining the equation for the adsorption $\theta(p, T)$, determining the filling factor of a channel at various pressures and temperatures of the gas above the surface, reduces to calculating the quantity $F_{\text{int}}(\theta, T) \equiv -T\ln Z_{\text{int}}(\theta, T)$, determining the "correction" to the pressure as a result of the interaction of the particles in the channel.

Thus, to obtain the adsorption isotherms in the system under study it is necessary to calculate the total free energy of the particles in a subnanometer channel taking into account their interaction with the channel wall and with one another. For this, we shall consider N particles in a channel whose size is comparable to the average particle diameter. In the general case the total potential energy of such a system can be written as

$$E(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \sum_{i,j=1}^{N} V(\mathbf{r}_i - \mathbf{r}_j) + \sum_{i=1}^{N} U(\mathbf{r}_i). \tag{2.6}$$

Here $V(\mathbf{r}_i - \mathbf{r}_j)$ is the potential energy of the pair interaction of the particles located at points with coordinates $\mathbf{r}_i$ and $\mathbf{r}_j$; $U(\mathbf{r}_i)$ is the interaction energy between a particle located at the point $\mathbf{r}_i$ and the channel walls. We shall transform the expression for the pair interaction potential of the particles using the fact that the channel diameter is comparable to the particle diameter. Then the interaction potential depends only on the particle coordinates along the channel r. We write the expression for $V(\mathbf{r}_i - \mathbf{r}_j)$ in the form

$$V(\mathbf{r}_i - \mathbf{r}_j) \equiv V(x_i - x_j, y_i - y_j, r_i - r_j) \approx V(r_i - r_j, 0, 0) \equiv \Phi(r_i - r_j). \tag{2.7}$$

This relation holds when the channel diameter is comparable to the particle diameter and the pair interaction potential $V(\mathbf{r}_i - \mathbf{r}_j)$ of the particles is not a long-range potential. It is convenient to represent the interaction energy between a particle and the channel walls in the form

$$U(\mathbf{r}_i) = U(\mathbf{r}_i', r_i). \tag{2.8}$$

Here $\mathbf{r}_i'$ is a dimensionless two-dimensional radius vector in a plane perpendicular to the channel. It is convenient to write the partition function of such a system in the form

$$Z_N \propto \int d\mathbf{r}_1' \ldots d\mathbf{r}_N' dz_1 \ldots dz_N \exp(-\beta U_N) \times \exp\left[-\beta \sum_{i=1}^{N} U(\mathbf{r}_i', r_i)\right]. \tag{2.9}$$

Here

$$U_N = \sum_{(1 \le i, j \le N)} \Phi(r_i - r_j).$$

The integration over the particle coordinates in a plane perpendicular to the channel can be performed exactly

$$\int d\mathbf{r}_1' \ldots d\mathbf{r}_N' \exp\left[-\beta \sum_{i=1}^{N} U(\mathbf{r}_i', r_i)\right] = \int d\mathbf{r}_1' \exp[-\beta U(\mathbf{r}_1', r_1)] \int d\mathbf{r}_2' \exp[-\beta U(\mathbf{r}_2', r_2)] \times \ldots \times \int d\mathbf{r}_N' \exp[-\beta U(\mathbf{r}_N', r_N)] \equiv \exp\left[-\beta \sum_{i=1}^{N} \Psi(r_i)\right]. \tag{2.10}$$

Here $\Psi(r_i)$ is the effective potential in which a particle moves in the channel as a result of the interaction of the particle with the walls:

$$\exp[-\beta\Psi(r)] \equiv \int \exp[-\beta U(\mathbf{r}', r)] d\mathbf{r}'. \tag{2.11}$$

Let us assume that the energy of a gas molecule interacting with the channel surface is constant and equal to $\varepsilon_1 = \Psi(r_{\min}) = \Psi(0)$. Then

$$Z_N' \propto \exp(-\beta N \varepsilon_2) \int dr_1 \ldots dr_N \exp(-\beta U_N). \tag{2.12}$$

The energy ε_1 physically corresponds to the binding energy of a particle in the surface potential $\Psi(r)$. Thus, under the assumptions made above, the gas in the channel can be assumed to be one-dimensional (the channel diameter is comparable to the maximum diameter of a

gas molecule). It is well known [10, 13] that the partition function of a one-dimensional gas with an arbitrary interaction potential can be calculated exactly under certain assumptions, which are formulated below. Indeed, let us consider an equilibrium system of N particles in a channel, which possess only one degree of freedom per particle and are located in the segment $[0, L]$ of the r axis. The total partition function of the system has the form

$$\tilde{Z}_N = \left(\frac{mT}{2\pi\hbar^2}\right)^{N/2}\frac{Q_N}{N!}\exp(-\beta N\varepsilon_1), \tag{2.13}$$

where

$$Q_N = \int_0^L\ldots\int_0^L \exp(-\beta U_N)dr_1\ldots dr_N, \tag{2.14}$$

$$U_N = \sum_{(1\le i\le N-1)} \Phi(r_{i+1}-r_i). \tag{2.15}$$

Here $\Phi(r)$ is the pair interaction potential of the particles in the channel. To calculate F_{int} the cofactor responsible for the partition function of an ideal gas must be eliminated from the partition function (2.13). In our problem the cofactor corresponding to a gas of particles in a channel without a pair interaction potential between the particles is already included in the expression (2.2). Thus, since the partition function of a one-dimensional gas without an interaction has the form

$$Z_{\text{id}} = \left(\frac{mT}{2\pi\hbar^2}\right)^{N/2}\frac{L^N}{N!}\exp(-\beta N\varepsilon_2), \tag{2.16}$$

we obtain from equation (2.2)

$$F_{\text{int}} = -T\frac{\ln\tilde{Z}_{\tilde{N}}(\theta, T)}{Z_{\text{id}}} = -T\ln\left(\frac{Q_N}{L^N}\right). \tag{2.17}$$

Thus, the problem has been reduced to calculating the configuration integral

$$\begin{gathered} Q_N = \int\ldots\int \exp[-\beta\{\Phi(\xi_1) \\ + \Phi(\xi_2) + \ldots + \Phi(\xi_N)\}]d\xi_1\ldots d\xi_N, \\ \xi_i \equiv r_{i+1} - r_i. \end{gathered} \tag{2.18}$$

In studying the configuration space of a system of particles in a one-dimensional channel, it should be kept in mind that the quantities ξ_i are not independent. They are related as

$$\sum_{j=1}^{N}\xi_j = L, \tag{2.19}$$

where L is the total length of the channel. Physically, this relation corresponds to the impossibility of particles penetrating one another and the condition of "blocking" of particles with diameter σ in a channel whose diameter $d < 2\sigma$. An explicit expression for the configuration integral can be obtained if the explicit form of the pair interaction potential of the particles is known. We shall consider a simple but nontrivial case: a system of hard spheres with diameter σ. The expression for the interatomic interaction energy in this case is

$$\Phi_i(\xi) = \begin{cases}\infty & \text{for } \xi < \sigma, \\ 0 & \text{for } \xi \ge \sigma.\end{cases} \tag{2.20}$$

We obtain for F_{int} in the hard-sphere approximation

$$F_{\text{int}}(\theta, T) = -T\ln(1-\theta). \tag{2.21}$$

Thus, in the hard-sphere approximation the adsorption isotherm for a gas in a subnanometer channel has the form

$$\begin{gathered} \frac{\eta\theta}{\sigma} = \frac{(1-\theta)p}{(1-\theta)p + \tilde{p}_0(T)}, \\ \tilde{p}_0(T) \equiv T\left(\frac{\hbar^2}{2\pi mT}\right)^{-3/2}\exp(-\beta\varepsilon_1). \end{gathered} \tag{2.22}$$

The relation (2.22) makes it possible to construct the dependence of the filling factor of a channel on the gas pressure for various molecules. It should be kept in mind that the sign of the energy ε_1 determines the possibility or impossibility of a particle entering a channel: for $\varepsilon_1 > 0$ it is energetically favorable for a gas particle to enter the channel for any external pressure greater than p_0. For $\varepsilon_1 < 0$ the particles must overcome a potential barrier to enter the channel.

The configuration integral can be calculated exactly for an arbitrary interaction between the particles in the channels [13]. This makes it possible to obtain an equation of state of a 1D gas for an arbitrary interaction. Thus, for an interparticle interaction potential of the form

$$\Phi(z) = \begin{cases}+\infty & \text{for } r \le \sigma, \\ -\varepsilon & \text{for } \sigma < r \le \sigma + R, \\ 0 & \text{for } r > R + \sigma,\end{cases} \tag{2.23}$$

where R is the effective radius of attraction, the equation of state of a one-dimensional gas is

$$p_{1D}\sigma\left(\frac{1}{\theta} - 1\right) = T - p_{1D}R\left[\frac{\exp(\beta p_{1D}R)}{1-\exp(-\beta\varepsilon)} - 1\right]^{-1}. \tag{2.24}$$

In the limits $\varepsilon \longrightarrow 0$ and $R \longrightarrow 0$, this equation becomes the equation of state of a system of hard spheres. The isotherms calculated numerically starting from equation (2.24) are presented in Fig. 1. It is also evident in this figure that at high temperatures the gas behaves almost as an ideal gas ($p_{1D} \propto \theta$). At low temperatures the isotherms seem to consist of two parts. For high density $\theta \approx$ const, which is typical for a condensed phase, whereas at pressure less than a characteristic value a gas-like phase for which $p_{1D} \propto \theta$ appears.

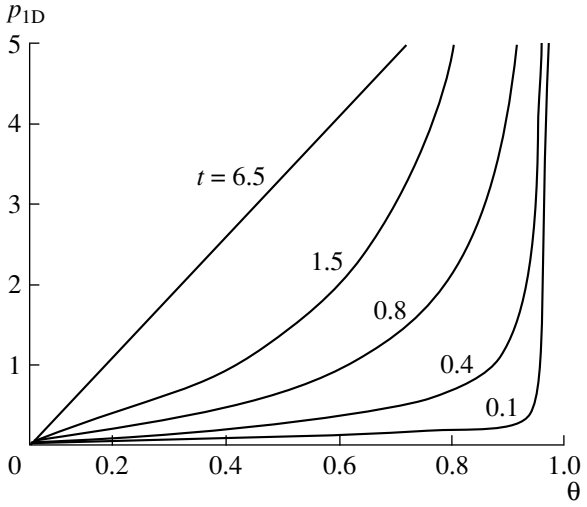


Fig. 1. Reduced "pressure" (p_{1D}) (2.24) of a one-dimensional gas in a channel versus the filling factor (θ) of the channel at various temperatures $t = T/T_0$.

It can be shown that at a transition from one regime to another the free energy of the system has no singularities. For this reason, there is no exact analogy with phase transitions. This assertion agrees completely with the Landau–van Hove theorem [10], according to which any one-dimensional model of a gas with a finite interaction radius does not undergo phase transitions. On the other hand, it should be expected that for sufficiently low temperatures, as the filling factor θ increases, the gas in the channel tends to form clusters whose size increases with the filling factor. The number and size of the clusters grow with the total number of particles in the system in a manner so as to ensure the existence of such clusters in the thermodynamic limit, where $N \longrightarrow \infty$, $L \longrightarrow \infty$, and N/L = const. This agrees completely with the formation of nuclei of a new phase in three-dimensional systems with first-order phase transitions [11, 12]. On the other hand, in our one-dimensional system, in contrast to three-dimensional systems, a continuous behavior of thermodynamic quantities determined as derivatives of the free energy of the system with respect to the temperature, pressure, and filling factor, should be expected in the entire range of variation of the thermodynamic parameters. For example, the specific heat calculated using the relation (2.24) will possess a maximum at the "critical" point. This fact has been observed experimentally [1], which attests to the applicability of the one-dimensional model employed above.

3. TRANSPORT IN A DENSE 1D SYSTEM

The gas flux through a membrane is the main experimentally measurable parameter (Section 4). Particle transport is usually described using a relation between the outgoing particle flux and the parameters of the problem, such as, for example, the difference of the gas pressures on different sides of the membrane. This description is based on Fick's relation between the gas flux density **J**, the gradient of the gas concentration, and the diffusion coefficient D, using the conservation law for the number of particles:

$$\mathbf{J} = -\tilde{D}\nabla n, \tag{3.1}$$

$$\partial n/\partial t + \mathrm{div}\mathbf{J} = 0. \tag{3.2}$$

In the relations (3.1) and (3.2) **J** is the flux, $\tilde{D}$ is the diffusion coefficient, and ∇n is the gradient of the particle concentration. The relation (3.1) is actually a definition of the diffusion coefficient $\tilde{D}$. Thus, the problem of determining the particle flux through a channel reduces to calculating the diffusion coefficient. A more general method of describing transport is to calculate the characteristic relaxation time of density fluctuations arising or specially created in the system under study [14]. Formally, the problem of determining this time reduces to calculating the characteristic frequency or spectrum $\omega(\mathbf{k})$ for the system under study. Thus, when the particle density is low and there is no interparticle interaction, an explicit expression for the spectrum $\omega(\mathbf{k})$ can be easily obtained from the relation (3.1). Indeed, let n_0 be the average particle concentration in the system. From equation (3.1) we obtain an equation determining the dynamics of the density fluctuations $\delta n(\mathbf{r}, t) = n - n_0$:

$$\partial(\delta n)/\partial t = \tilde{D}\Delta(\delta n).$$

Switching to a Fourier representation of the density fluctuations

$$\delta n(\mathbf{r}, t) = \frac{1}{(2\pi)^4}\int d\mathbf{k}d\omega \exp(-i\mathbf{k}\cdot\mathbf{r} - i\omega t)\delta n(\mathbf{k}, \omega),$$

we obtain an expression determining the reciprocal of the relaxation time of the density fluctuations δn in the case at hand:

$$\omega_0(\mathbf{k}) = -i\tilde{D}k^2.$$

It is obvious that a spectrum of this type is characteristic for systems where relaxation to equilibrium occurs by diffusion. Let us consider a system of interacting particles with arbitrary density. Assume that we have been able to calculate the characteristic relaxation time ω^{-1} of density fluctuations in the system. If the spectrum of relaxation times in such a system has the form

$$\omega(\mathbf{k}) = -iD(\mathbf{k}, n_0)k^2, \tag{3.3}$$

then it is natural to consider the quantity $D(\mathbf{k}, n_0)$ to be the diffusion coefficient in the system, provided that in the limit of low densities n_0 the quantity $D(\mathbf{k}, n_0)$

becomes the diffusion coefficient for a low-density system of particles:

$$\lim_{n_0 \to 0} D(\mathbf{k}, n_0) = \tilde{D}.$$

This definition of the diffusion coefficient means that the system possesses a diffusion characteristic mode. Proceeding from the definition (3.3), to calculate the diffusion coefficient $D(\mathbf{k}, n_0)$ it is necessary to know the characteristic modes of a weakly nonequilibrium system. If these characteristic modes have the form (3.3), then $D(\mathbf{k}, n_0)$ will be the diffusion coefficient of the system. It is obvious that the diffusion coefficient in the general case is nonlocal and is a functional of the density of the diffusing particles. The expression for the flux of diffusing particles will have a form generalizing equation (3.1):

$$\mathbf{J} = -\int d\mathbf{r}' D(\mathbf{r}, \mathbf{r}', n(\mathbf{r}', t))\nabla n(\mathbf{r}', t). \quad (3.4)$$

The equation, obtained using this relation, for the particle number density in the system is in general a nonlinear integrodifferential equation. When studying the diffusion of a single-component gas in subnanometer channels, it should be kept in mind that the pair interaction potential between the particles is short-range and no phase transitions occur in the system of particles in the channel (see preceding section). In this case it can be assumed that the diffusion coefficient $D(\mathbf{r}, \mathbf{r}', n(\mathbf{r}, t))$ is local:

$$D(\mathbf{r}, \mathbf{r}', n(\mathbf{r}', t)) = D(n(\mathbf{r}, t))\delta(\mathbf{r} - \mathbf{r}'). \quad (3.5)$$

The corresponding expression for the flux of diffusing particles and the equation for their density become

$$\begin{aligned} \mathbf{J} &= -D(n)\nabla n, \\ \partial n/\partial t &= \nabla(D(n)\nabla n). \end{aligned} \quad (3.6)$$

For low densities the relation (3.6) becomes equation (3.1), which is the first term in the expansion of the flux in odd powers of the density gradient.

Thus, for systems where the relaxation spectrum of the fluctuations is of the form (3.3), the diffusion coefficient can be determined starting from the explicit form of this spectrum. The expression obtained for the diffusion coefficient in this case is a generalization of the expressions obtained for the diffusion coefficient in various models, specifically, in the Maxwell–Stefan model [9].

It is convenient to calculate the relaxation spectrum of the fluctuations using the density functional formalism [4]. We shall present the basic assumptions of the density functional formalism following [4]. Let us assume that a state that can be characterized by the moments of the multiparticle distribution function has been established in the system. Let us assume that the state of the system can be described by giving the single-particle distribution function $F_1(\mathbf{x}, t)$, where $\mathbf{x} \equiv (\mathbf{r}, \mathbf{p})$ and $\mathbf{r}$ and $\mathbf{p}$ are, respectively, the coordinates and momentum of a particle. In [4] such systems are called gas-like. Let the gas in the channel be such a system. Then it can be assumed that the distribution function is a product of the longitudinal ($f_1(x, t)$, where $x = (r, p)$, r, and p are the coordinate and momentum of the particle along the channel) and transverse single-particle distribution functions. The distribution function in the transverse direction relaxes in phonon times $t \approx \tau_{ph}$ [3] and therefore for $t \gg \tau_{ph}$ it is an equilibrium system. Under these conditions ($t \gg \tau_{ph}$) the state of the system of particles in the channel can be described by giving the longitudinal single-particle distribution function $f_1(x, t)$. Here the variable x is a one-dimensional coordinate along the channel and the momentum in the direction of this coordinate. Let us consider the relaxation of weakly nonequilibrium states of such a system. Weakly nonequilibrium states are states for which local equilibrium has been established with respect to the momentum. Then the free energy $F(t)$, which is a functional of the distribution function $f_1(x, t)$ of the system [4], can be introduced:

$$F(g, t) = F_0[f_1] + \frac{1}{2}\int dx_1 dx_2 \int_0^g f_2(\lambda, x_1, x_2) V d\lambda. \quad (3.7)$$

Here f_2 is a binary distribution function. The time dependence of the functional $F(g, t)$ is determined completely by the time dependence of the distribution function $f_1(x, t)$. According to [4], for example, states for which the distribution function $f_1(x, t)$ is a functional of its first moments are considered to be weakly equilibrium states. Then, the description of relaxation to equilibrium based on the functional (3.7) is valid only at the final stage of the evolution of the system. We now introduce the average value of the free energy of a nonequilibrium system:

$$\tilde{\Delta}_g[f_1] = \lim_{\tau \to \infty} \tau^{-1} \int_0^\tau F(g, t) dt. \quad (3.8)$$

The limit in the relation (3.8) corresponds to a description of the relaxation kinetics of the system over times much longer than the characteristic local equilibration time. The functional (3.8) is an extremal for an equilibrium ($f_1 = f^{EQ}$) state of the system with a variable number of particles:

$$\left.\frac{\delta \tilde{\Delta}_g}{\delta f}\right|_{f = f^{EQ}} = 0. \quad (3.9)$$

If the number of particles is fixed, then it is convenient to introduce instead of equation (3.8) the functional

$$\Delta_g[f_1] = \lim_{\tau \to \infty} \tau^{-1} \left[\int_0^\tau F(g, t) dt - \mu \int_0^\tau f_1 dx dt \right], \quad (3.10)$$

where μ is a Lagrange multiplier, whose significance is similar to that of the chemical potential of an equilib-

rium system. The functional (3.10) is extremal for an equilibrium system with a fixed number of particles $\delta\Delta_g/\delta f_1 = 0$.

The functionals (3.7)–(3.10) make it possible to determine the generalized susceptibility of the system under study. Let us place the system into a weak, variable, external field $eV_{\text{ext}}(x, t)$ (e is the effective charge), and let the characteristic times of variation of the field be such that the condition of weak nonequilibrium of the system is satisfied. We shall define the response function χ of the system as

$$\delta f(x, t) = \int \chi(x, x', t, t') eV_{\text{ext}}(x', t') dx' dt'. \quad (3.11)$$

If $f_0 = f_0(t)$, which corresponds to a locally equilibrium state of the system, the response function χ depends on the temporal arguments t and t' and not only on their difference. It is convenient to represent this dependence in the form

$$\chi(x, x', t, t') = \chi(x, x', t - t', t),$$

where the dependence of the response function on the second argument is determined only by the time dependence of the single- particle distribution function $f_1(x, t)$, whose functional form is χ. In the presence of an external field $eV_{\text{ext}}(x, t)$ the functional $F^e(g, t)$ which is the free energy of a nonequilibrium state in the limit $e \longrightarrow 0$, is determined by the equation

$$F^e(g, t) = F(g, t) + \int eV_{\text{ext}} f_1 dx. \quad (3.12)$$

Then the functional Δ_g^e has the form

$$\Delta_g^e = \lim_{\tau \to \infty} \tau^{-1} \left[\int_0^\tau F^e(g, t) dt - \mu \int_0^\tau f_1 dx dt \right]. \quad (3.13)$$

Varying Δ_g^e with respect to the distribution function f_1 and using the condition (3.9) for the functional to be an extremal with the equilibrium value $f_1 = f_0$, we obtain an equation for the response function χ_g of a nonequilibrium system described by the functional Δ_g^e:

$$\chi(g, y, y') = \chi_0(y, y') + \int dy dy' \chi_0(y, y_1) R(g, y_1, y_2) \chi(g, y_2, y'), \quad (3.14)$$

$$R(g) \equiv \frac{\delta^2}{2\delta f_1^2} \int_0^\infty d\tau \int_0^g d\lambda V f_2(\lambda, x_1, x_2) dx_1 dx_2, \quad (3.15)$$

$$\chi(g) \equiv -(\delta^2 \Delta_g / \delta f_1^2)^{-1}, \quad y \equiv (x, t),$$

and χ_0 is the response function of a system of noninteracting particles. The pair distribution function can be related with the response function via the fluctuation–dissipation theorem [12]:

$$\langle \delta f(x, t) \delta f(x', t) \rangle = -T \int_0^\infty \chi(x, x' \tau, t) d\tau,$$

$$\langle \delta f(x, t) \delta f(x', t) \rangle = f_2(x, x', \lambda) - f_1(x, t) f_1(x, t) + \delta(x - x') f_1(x, t). \quad (3.16)$$

The fluctuation–dissipation theorem in its thermodynamic variant (3.16) is valid only for states near thermal equilibrium, where the locally equilibrium momentum distribution of the particles has been established. The relaxation of the system in this case proceeds via the spatial variation of a macroscopic characteristics of the system. In the relation (3.16) the pair distribution function f_2 and the response function χ depend on the time t via the single-particle distribution function $f_1(x, t)$, of which they are functionals. Using equations (3.15) and (3.16), we find finally

$$R(g) = gV(\mathbf{r} - \mathbf{r}') \delta(t - t') - \frac{T}{2} \frac{\delta^2}{\delta f_1^2} \left[\int_0^g d\lambda \int_0^\infty d\tau V \chi(\lambda, x_1, x_2) dx_1 dx_2 \right]. \quad (3.17)$$

We now switch from the response function χ to the density response function $\beta(r, r', t, t')$. The response function $\beta(r, r', t, t')$, determining the change $\delta n(r, t)$ in the density of the system when a weak external field $eV_{\text{ext}}(\mathbf{r}, t)$ is switched on, can be determined from the relation

$$\delta n(r, t) = \int \beta(r, r', t, t') eV_{\text{ext}}(r, t') dr' dt'. \quad (3.18)$$

The response function $\beta(r, r', t, t')$ is related with the response function, as

$$\chi(\lambda, x, x' t, t') = \beta(\lambda, r, r', t, t') \varphi(p, t) \delta(p - p'). \quad (3.19)$$

Here $\varphi(p, t)$ is the locally equilibrium momentum distribution function normalized to 1:

$$\int \varphi(p, t) dp = 1. \quad (3.20)$$

We note that in this case, when the momentum distribution along the channel has been established, the response function $\beta(r, r', t, t')$ can be obtained by directly varying the average value of the free energy (3.7) which is a functional of the density $n(r, t)$ [4]:

$$\beta(g, r, r', t, t') = -\left(\frac{\delta^2 \Delta_g^F}{\delta n(r, t) \delta n(r', t')} \right)^{-1},$$

$$\Delta_g^F[f_1] = \lim_{\tau \to \infty} \tau^{-1} \int_0^\tau F(g, t) dt, \quad (3.21)$$

$$F(g, t) = F_0[n] + \frac{1}{2} \int dr_1 dr_2 \int_0^g d\lambda V g_2[\lambda, n],$$

where g_2 is the pair correlation function. Substituting the expression (3.19) into equations (3.14) and (3.17) and integrating over the momenta we obtain for a one-dimensional channel

$$\begin{aligned}\beta(g, r, r', t, t') &= \beta_0(r, r', t, t') \\ &+ \int dr_1 dr_2 dt_1 dt_2 \beta_0(r, r_1, t, t_1) \\ &\times R(g, r_1, r_2, t_1, t_2)\beta(g, r_2, r', t_2, t'), \qquad (3.22) \\ R(g, r, r', t, t') &= gV(r-r')\delta(t-t') \\ &- \frac{T}{2}\frac{\delta^2}{\delta n^2}\int_0^g d\lambda V\beta(\lambda)dr_1 dr_2 dt_1 dt_2.\end{aligned}$$

Here $\beta_0(r, r', t, t')$ is the response function of a system of noninteracting particles, which will be calculated below. To obtain a closed equation we shall use a local approximation [4], which in the present case, for an arbitrary functional $\Phi[x_1, n(x)]$ has the form

$$\frac{\delta^2\Phi[x_1, n]}{\delta n(x)\delta n(x')} \approx \delta(x_1 - x)\delta(x_1 - x')\frac{d^2\Phi}{dn^2}. \qquad (3.23)$$

Using equation (3.23) we find closed equations for the response function:

$$\begin{aligned}\beta(g, r, r', t, t') &= \beta_0(r, r', t, t') \\ &+ \int dr_1 dr_2 dt_1 dt_2 \beta_0(r, r_1, t, t_1) \\ &\times R(g, r_1, r_2, t_1, t_2)\beta(g, r_2, r', t_2, t') \qquad (3.24) \\ R(g, r, r', t, t') &= gV(r-r')\delta(t-t') - \frac{T}{2}\frac{d^2}{dn^2}\int_0^g d\lambda V\beta(\lambda).\end{aligned}$$

Assuming the characteristics of the ground state of the system, such as the temperature and average particle number density, to be time-independent, it is convenient to switch from equations (3.24) to the equations for the Fourier components of the response function:

$$\begin{aligned}\beta(g, k, \omega) &= \beta_0(k, \omega) + \beta_0(k, \omega)R(g, k, \omega)\beta(g, k, \omega), \\ R(g, k, \omega) &= gV(k) \qquad (3.25) \\ &- \frac{T}{2}\frac{d^2}{dn^2}\int_0^g \frac{d\lambda dp}{2\pi}V(k-p)\beta(\lambda, k, \omega).\end{aligned}$$

The equations (3.25) make it possible to determine the response function $\beta(g)$ of a weakly nonequilibrium system of particles in a channel with an arbitrary interaction by "dressing" the noninteracting particles using the pair interaction potential $V(r - r')$ of the molecules [4]. In accordance with [4], the relaxation spectrum $\omega(k)$ of the system is determined by the relation

$$\beta^{-1}(g = 1, k, \omega) = 0. \qquad (3.26)$$

We shall use the density functional formalism to calculate the response function and the diffusion coefficient of a dense gas in a subnanometer channel. We shall calculate first the response function $\beta_0(k, \omega)$ and the relaxation spectrum of such a gas when there is no interaction between the particles. We write the diffusion equation for the filling factor θ of the channel with molecules in the presence of a weak perturbing external field $eV_{\text{ext}}(r, t)$ in the form

$$\frac{\partial\theta}{\partial t} = D_0\nabla\left(\nabla\theta + \frac{\theta}{T}\nabla eV_{\text{ext}}\right), \quad \nabla \equiv \frac{\partial}{\partial r}. \qquad (3.27)$$

Here D_0 is the diffusion coefficient for noninteracting particles. Let θ_0 be the equilibrium filling factor of the channel. We shall seek the solution of equation (3.27) in the form

$$\theta = \theta_0 + \delta\theta(r, t). \qquad (3.28)$$

Then, we obtain from equation (3.27), assuming the external field eV_{ext} to be weak, an equation for $\delta\theta(r, t)$:

$$\frac{\partial\delta\theta}{\partial t} = D_0\Delta\delta\theta + \frac{D_0\theta_0\Delta eV_{\text{ext}}}{T}. \qquad (3.29)$$

In deriving equation (3.29), only first-order infinitesimals in $\delta\theta$ and eV_{ext} were retained in all terms. We shall seek the solution of equation (3.29) in the form of a Fourier integral

$$\begin{aligned}\delta\theta(r, t) &= \frac{1}{(2\pi)^2}\int e^{i\mathbf{k}\cdot\mathbf{r}}e^{i\omega t}\delta\theta(k, \omega)dkd\omega, \\ eV_{\text{ext}}(r, t) &= \frac{1}{(2\pi)^2}\int e^{i\mathbf{k}\cdot\mathbf{r}}e^{i\omega t}eV_{\text{ext}}(k, \omega)dkd\omega.\end{aligned} \qquad (3.30)$$

Substituting the expressions (3.30) into equation (3.29) and solving the resulting linear equation for $\delta\theta(\mathbf{k}, \omega)$, we obtain

$$\delta\theta(k, \omega) = \beta_0(k, \omega)eV_{\text{ext}}(k, \omega). \qquad (3.31)$$

Here

$$\beta_0(k, \omega) = -\frac{\theta_0 k^2 D_0}{T(i\omega + k^2 D_0)}, \qquad (3.32)$$

is the response function of a gas of noninteracting diffusing particles. Indeed, as $\omega \longrightarrow 0$, the function $\beta_0(k, \omega)$ reduces to the well-known response function of an ideal equilibrium gas with density θ_0 at temperature T [14, 4]:

$$\beta_0(k, 0) = -\frac{\theta_0}{T}. \qquad (3.33)$$

The relaxation spectrum of such a gas is determined from the relation (3.26) and has the form

$$\omega_0(k) = -iD_0k^2. \qquad (3.34)$$

In accordance with the definition (3.2) the quantity D_0 is the diffusion coefficient. The relations obtained make

it possible to calculate the relaxation spectrum $\omega(k)$ by solving equations (3.25). Thus, we find for the response function of a one-dimensional system

$$\beta(k, \omega) = \frac{\beta_0(k, \omega)}{1 - \beta_0(k, \omega)R(k, \omega)}. \quad (3.35)$$

Using the relations (3.32) and (3.35), we obtain an equation determining the relaxation spectrum of the system under study:

$$1 + \frac{\theta_0}{T}\frac{\omega_0}{\omega + \omega_0}R(k, \omega) = 0. \quad (3.36)$$

Equation (3.36) can be solved in a general form in the quasistatic case $\omega \longrightarrow 0$. Indeed, in this case we find from equation (3.36)

$$\omega(k) = -iD(\theta_0, k)k^2, \quad (3.37)$$

$$D(\theta_0, k) = D_0\left[1 + \frac{\theta_0}{T}R(k, 0)\right]. \quad (3.38)$$

It is convenient to rewrite the relation (3.38), introducing the pair distribution function $n_2(r - r') = \theta_0^2[1 + \nu(r - r')]$. We employ a relation, following from equation (3.16), between the response function in the static limit $\omega = 0$ and the pair distribution function:

$$T\beta(k, 0) = \theta_0^2\delta(k) - n_2(k) - \theta_0. \quad (3.39)$$

Then we obtain from the first of equations (3.25) and the relation (3.32)

$$R(k, 0) = -\frac{T\nu(k)}{1 + \theta_0\nu(k)}. \quad (3.40)$$

Substituting the expression (3.40) into equation (3.38) gives a relation determining the relaxation spectrum of a dense gas in a one-dimensional channel:

$$\omega(k) = -\frac{iD_0k^2}{1 + \nu(k)}. \quad (3.41)$$

We shall now calculate the diffusion coefficient of a gas in a one-dimensional channel taking into account the interaction between hard-sphere particles. In this case, the correlation function $n_2(r) = \theta_0^2[1 + \nu(r)]$ has been calculated exactly [13] with an arbitrary filling factor of the channel:

$$n_2(r) = \begin{cases} 0, & r < \sigma, \\ \dfrac{1}{\theta_0}\displaystyle\sum_m \frac{(r\sigma^1 - m)^{m-1}}{(m-1)!(\theta_0^{-1} - 1)^m}\exp\left(-\frac{r\sigma^{-1} - m}{\theta_0^{-1} - 1}\right), & r > \sigma. \end{cases} \quad (3.42)$$

Using the relations (3.41) and (3.42), we find

$$\omega(k) = -iD(\theta_0)k^2, \quad D(\theta_0) = D_0(1 + \theta_0^2/(1 - \theta_0)^2). \quad (3.43)$$

The relations (3.43) can be obtained by a shorter method. From the definition (3.21) follows a relation between the response function and the free energy of the system [4]:

$$\frac{1}{\beta(\mathbf{k}, \omega)} = -\frac{\delta^2\Delta^F}{\delta\theta(\mathbf{k}, \omega)\delta\theta(\mathbf{k}, \omega)}. \quad (3.44)$$

Here the functional Δ^F is determined by the relation (3.21). The free energy of the gas in a channel can be calculated in the hard-sphere approximation using the methods described in Section 2, and it has the form

$$F(t) = F_0(t) - \frac{TN}{L}\int \ln[1 - \theta(x, t)]dx. \quad (3.45)$$

Here F_0 is the free energy of a gas of interacting particles, N is the total number of particles in the channel, and L is the length of the channel. Using the relations (3.44) and (3.45) we obtain for the response function of a gas of hard spheres in a channel

$$\frac{1}{\beta(k, \omega)} = -\frac{T}{\theta_0}\frac{i\omega + D_0k^2}{D_0k^2} - \frac{T\theta_0}{(1 - \theta_0)^2}. \quad (3.46)$$

The derivation of equation (3.46) used the equality

$$\frac{\delta^2\Delta^F}{\delta\theta(k, \omega)\delta\theta(k', \omega')} = \int \frac{\delta^2\Delta^F}{\delta\theta(x, t)\delta\theta(x', t')} \times \exp(ikx - i\omega t)\exp(ikx' - i\omega t')dxdx'dtdt'.$$

The relations (3.43) can be obtained by equating the expression (3.46) to zero and calculating the spectrum $\omega(k)$. It follows from equation (3.43) that the general equation describing particle transport in a dense 1D system, in accordance with equation (3.6), has the form

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial x}\left[D_0\left(1 + \frac{\theta^2}{(1 - \theta)^2}\right)\frac{\partial\theta}{\partial x}\right]. \quad (3.47)$$

This equation can be obtained from equation (3.43) if the characteristic local equilibration times in a dense system of particles in a channel are short compared with the characteristic propagation times of disturbances along such a system. For the cases considered below this assumption obviously holds. We shall consider solutions of equation (3.47) in the case where complete equilibrium has been established in the channel and the filling factor of the channel $\theta = \theta_0(p, T)$, the function $\theta_0(p, T)$ being determined by the relation (2.23). Transport in a channel in this case will be determined by the dynamics of the motion of disturbances of the filling factor of the channel. The solution of equation (3.47) in this case should be sought in the form $\theta = \theta_0(p, T) + \delta\theta(x, t)$. It follows from equation (3.47) that the relaxation spec-

trum of disturbances in this case has the form (3.43), where $\theta_0 = \theta_0(p, T)$. Thus, taking account of the interaction of hard-sphere particles in a channel does not change the character of the relaxation of its weakly nonequilibrium state: the relaxation spectrum (3.43) remains a diffusion spectrum with diffusion coefficient $D(\theta_0)$. However, it is important to underscore that transport in the channel in this case is a collective effect and proceeds via transport of disturbances of the equilibrium density $\theta_0(p, t)$. The diffusion coefficient $D(\theta_0)$ in this case is the diffusion coefficient of disturbances of the equilibrium density. It is convenient to rewrite the relation (3.43) for the diffusion coefficient in a different form. Using the Arrhenius character of the diffusion coefficient for noninteracting particles,

$$D_0 = \tilde{D}_0 \exp(-E/T) \tag{3.48}$$

($\tilde{D}_0$ is proportional to the product of the squared lattice constant of the wall material of the channel and the relaxation frequency of particles in a channel on defects and lattice phonons and is calculated in [3], $E \approx \Psi(r_{\max})$ (2.12) is the activation energy of diffusing of noninteracting particles in the channel), we shall rewrite equation (3.43) in the form

$$D(\theta_0) = \tilde{D}_0 \exp\left(-\frac{\tilde{E}(\theta_0)}{T}\right),$$
$$\tilde{E}(\theta_0) \equiv E - T\ln\left[1 + \frac{\theta_0^2}{(1-\theta_0)^2}\right]. \tag{3.49}$$

This way of writing the diffusion coefficient for a gas of interacting particles in a channel makes it possible to give a physical interpretation for the change in the diffusion coefficient accompanying a change in the filling factor θ_0 of the channel. The relation (3.49) shows that the interaction of particles in a channel decreases the activation energy E of the motion of particles, even in the hard-sphere model. For an interaction between hard-sphere particles such that there is no direct attraction between the particles, the diffusion activation energy decreases with increasing filling factor θ_0 of the channel as a result of the effective interaction (see below). Physically, this corresponds to a change, because of the presence of another particle in a neighboring potential well, in the parameters of the potential in which a gas particle moves. Since the gas particles are assumed to be indistinguishable, diffusion with the filling factor of the channel $\theta_0 \sim 1$ (when the effective diffusion activation energy $\tilde{E}(\theta_0)$ becomes comparable to the temperature of the system) can be interpreted as a transfer of "excitation" of the density along a chain of close gas particles. It is obvious that the motion of such an "excitation" will occur with substantial velocities. This leads to a large increase in the diffusion coefficient when $\theta_0 \sim 1$. For $E = 0$ the possibility of such a diffusion mechanism is played out in [2]. We note that for all gases investigated the barrier E is different from 0. This is indicated by the presence of a maximum in the temperature dependence of the flux in a zeolite membrane [9]. It can be assumed that the increase in the diffusion coefficient in a channel observed with degrees of filling $\theta_0 \geq 0.5$ is related with the formation of clusters in the channel whose sizes increase with the filling factor. The transport of gas in the channel containing such clusters is determined by the motion of "excitation" in a finite-size cluster. The formation of clusters in a gas consisting of particles with a hard-sphere pair interaction potential is related with the well-known [12] appearance of an effective attraction between such particles. This effect is manifested in the appearance of a maximum in the correlation function $n_2(r)$, describing the probability of finding the "first" particle at a distance r from the "second" particle.

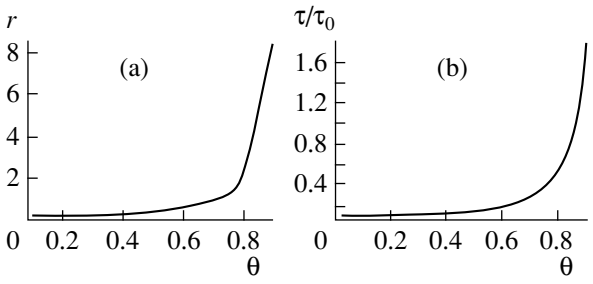


Fig. 2. (a) Size ($r = R/\sigma$) and (b) lifetime (τ/τ_0, $\tau_0 = \sigma^2/D_0$) of clusters versus the filling factor of the channel.

The relaxation spectra of the system for an arbitrary wave vector k can be found from equations (3.41) and (3.42). It is found that the spectrum $\omega(k)$ contains a real part, which corresponds to transport of an "excitation" in the system. The imaginary part of the spectrum obtained has a minimum at a definite value of the wave vector $k = k_{\min}$, which depends on the filling factor of the channel θ_0. It is natural to interpret the value $R = 2\pi/k_{\min}(\theta_0)$ as the characteristic size of a cluster with a given filling factor and the value $\tau = \mathrm{Im}\,\omega^{-1}(k_{\min}(\theta_0))$ as its characteristic lifetime. As the filling factor of the channel increases, the value of the effective attraction between the particles increases [12]; this could lead to the appearance of clusters consisting of two or more particles. However, it is found that as a result of the one-dimensionality of the channel, the lifetime of the clusters that are formed is finite, which corresponds to the absence, as noted above, of a phase transition in such systems. The dependences of the lifetime and size of the clusters formed on the filling factor θ_0 are presented in Fig. 2. It is evident in the figure that increasing the filling factor of the channel increases the lifetime and the size of the clusters formed. In finite-size channels, for a definite value of the filling factor $\theta_0 = \theta_c$ a cluster equal in size to the length of the channel is formed. It is obvious that the diffusion coefficient in

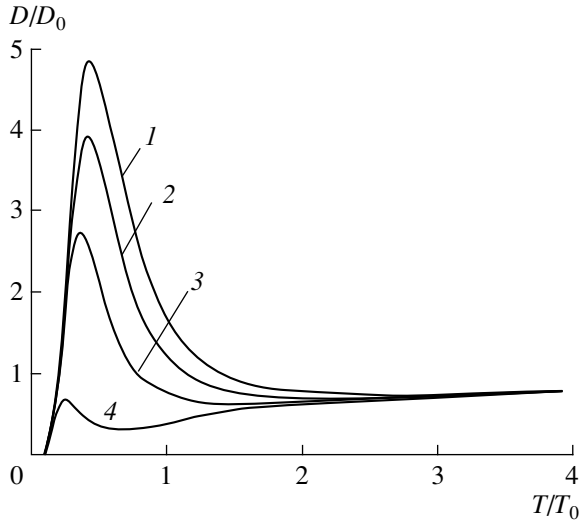


Fig. 3. Theoretical curves of the relative diffusion coefficient D/D_0 versus the relative temperature T/T_0 for various values of the relative pressure p/p_0: (*1*) 3, (*2*) 2, (*3*) 1, (*4*) 0.1.

such a channel increases without bound as $\theta_0 \longrightarrow \theta_c$, even though the lifetime of such a cluster is finite. We note that the density functional method can be used to take into account the interaction of particles in a channel for a potential that is different from a hard-spheres interaction, similarly to the way this was done above. Analysis of the experimental data showed that for high filling factor of the channel $\theta_0 \sim 1$ the dependence of the diffusion coefficient $D(\theta_0)$, calculated from the relation (3.43), leads to a discrepancy between theory and experiment (Section 4). This discrepancy could be due, in our opinion, to the above-described influence of a finite channel length, the asphericity of the gas molecules, and the possible breakdown of one-dimensionality of the problem. These effects lead to the following dependence of the diffusion coefficient on the filling factor of the channel:

$$D(\theta_0) = D_0\left(1 + \frac{\theta_0^2}{(1-\zeta\theta_0)^2}\right). \qquad (3.50)$$

Here ζ is a coefficient that takes into account the finite size of the channel, the possible deviation of the channel from one-dimensionality, and the asphericity of the molecules. The relations (3.43), (3.50), and (2.23) make it possible to describe transport in a subnanometer channel for different values of parameters such as the gas pressure and the pressure difference outside the membrane, the temperature of the membrane, and the type of gas molecules. As an example, we present the dependence, calculated according to equations (3.43), (3.50), and (2.23), of the diffusion coefficient in a subnanometer channel on the relative temperature T/T_0 for various relative pressures p/p_0 (Fig. 3). It is evident in the figure that the character of the temperature dependence of the relative diffusion coefficient in subnanometer channels is substantially different at different pressures p. For example, if the pressure is high ($p/p_0 > 1$), the dependence of D/D_0 on T/T_0 has a pronounced maximum at temperatures $T \sim 0.5T_0$ (curve *1*). As pressure decreases, the magnitude of this maximum decreases and the maximum itself shifts into the region of lower temperatures (curves *2*, *3*). A further decrease of pressure to values $p/p_0 < 1$ results in the appearance of a minimum in the dependence of D/D_0 on T/T_0 in the temperature range $T \sim 0.5T_0$. The maximum occurs at $T \sim 0.3T_0$. Since for $p/p_0 < 0.1$ the flux depends on p and the average distance between particles in the channel is much greater than the diameter of the particles, the increase in the diffusion coefficient for $T > 0.5T_0$ is due to the temperature dependence of the diffusion coefficient D_0 of individual molecules in the channel. At high temperatures all curves saturate at a value corresponding to $D/D_0 = 1$. In the next section, the theoretical laws obtained above will be compared with the experimental data.

4. ANALYSIS OF THE EXPERIMENTAL DATA. COMPARISON OF THEORY WITH EXPERIMENT

Quite extensive experimental investigations of the sorption and transport properties of a series of organic molecules (CH_4, C_2H_6, *n*-C_4H_8, *i*-C_4H_8, and others) and the inert gases Ar, Ne, and Kr in ZSM-5 zeolite membranes (MFI, silicate-1) have now been performed. The results are presented in [8, 15, 16]. According to [17], ZSM-5 zeolite membranes have a complicated chemical and crystalline structure. The chemical structure of zeolite membranes is given by the formula

$$Na_n[Si_{96-n}Al_nO_{192}] \cdot 16H_2O, \quad n \le 8.$$

For small values of n zeolite membranes of this type are called Silicate-1 membranes. The crystalline structure of ZSM-5 membranes has been studied quite well. It consists of a 3D structure, consisting of sinusoidal channels with a circular cross section (0.54 ± 0.02 nm) parallel to the *a*-axis [100], which intersect straight channels with an elliptical cross section (0.57–0.58) × (0.51–0.52) nm^2, parallel to the *b*-axis [010] [15, 18]. Cavities ~0.9 nm in size form at the intersection of the channels. The sorption capacity of ZSM-5 zeolite is determined by the number of sorbed (entering the channels) molecules per cell of a crystal. According to [15, 19], the cell parameters are: a = 2.007 nm, b = 1.992 nm, c = 1.342 nm. A single cell is a structure consisting of four segments of 0.46 nm linear channels, four segments of 0.66 nm sinusoidal channels, and four 0.54 nm intersections. Depending on the structure of a molecule, sorption of one or two molecules per intersection is possible. In [15] the sorption capacity of ZSM-5 was measured for a number of molecules. It was concluded that the molecules CH_4, C_3H_8, and

n-C_4H_{10} are sorbed with one molecule per intersection, while nitrogen, *n*-hexane, and *p*-xylene are sorbed with two molecules per intersection. Thus, for gases of the type CH_4, n-C_3H_8, and C_4H_{10} channels in the crystal structure of ZSM-5 can be treated as one-dimensional. For nitrogen-type molecules two molecules can be arranged at an intersection and therefore molecules can change places at an intersection (absence of blocking). Nonetheless, since there are significantly more nitrogen molecules in a channel (24 molecules per unit cell [15]), the one-dimensional model can also be used for this gas.

Analysis of the experimental data presented [9, 19–22] shows that all gases investigated can be conditionally divided into two groups. The first group of gases, containing, specifically, Ar, Kr, Ne, and CH_4, is characterized by a linear pressure dependence of the filling factor θ, i.e., Henry's law holds for them [9]. For these gases, the pressure dependence of the flux is nearly linear, and the diffusion coefficient is essentially independent of the filling factor [19]. The temperature dependence of the flux is characterized by the presence of a minimum at temperatures $T \sim 400$ K (for example, for Ar and Kr), while a maximum is not observed in the diffusion coefficient [21, 22]. Here neon, whose diffusion coefficient increases with temperature [22], is an exception. The second group of gases, containing, specifically, i-C_4H_{10}, C_2H_4, C_2H_6, and C_4H_8, is characterized by a dependence of the filling factor θ on the pressure p in the form of a curve that saturates [9]. For these gases the pressure dependence of the flux is likewise characterized by a curve that emerges at a value that is pressure-independent [19]. For this group of gases the diffusion coefficient depends strongly on the filling factor [20]. The temperature dependence of the flux is characterized by the existence of a maximum and minimum at high ($T > 500$ K) temperature [21].

The model, constructed in this paper, of the sorption and transport properties makes it possible to describe these dependences of transport in one-dimensional channels on the basis of general assumptions. Figure 4 shows the dependence of the diffusion coefficient D on the filling factor θ (solid line), calculated using equation (3.50), and the experimental data obtained in [21] for various gases. We note that the dependence (3.50) presented in Fig. 4 takes into account the asphericity of the molecules, equal to $\zeta = 0.86$. For inert gases and methane (first group of gases) the diffusion coefficient measured in the range of filling factor $\theta_0 < 0.5$ depends weakly on θ_0. A large increase in the diffusion coefficient at $\theta_0 \sim 1$ is observed for gases of the type C_2H_6 and n-C_4H_8 (second group of gases). This behavior of the diffusion coefficient $D(\theta)$ can be satisfactorily described on the basis of the model proposed, which takes account of the pair interaction of particles with a very simple form of the interaction potential for hard spheres. It is obvious that this coefficient is different for different molecules. The physical mechanism leading to an increase of the diffusion coefficient consists in a decrease of the activation barriers for diffusion, which is due to an interaction of the particles for high filling factor. Indeed, it follows from the relation (3.49) for the effective diffusion activation energy $\tilde{E}(\theta_0)$ that in this case, when the filling factor of a channel is low, $\theta_0 \ll 1$, gas diffusion in the channel can be studied in the single-particle approximation neglecting the pair interaction of the gas particles with one another. As the filling factor increases, the diffusion activation energy decreases as a result of the effective attraction of the gas particles to one another. This can be interpreted as the formation of widely separated clusters. Diffusion in each cluster can be treated as a transfer of "excitation" along a chain of close gas particles. A further increase of the filling factor increases the cluster lifetime and decreases the average intercluster distance and, in consequence, increases the diffusion coefficient. For $\theta_0 \sim 1$ the diffusion process can be regarded as a motion of this "excitation" along the entire channel, when the arrival of a particle at the channel entrance results in the end-most particle leaving the channel.

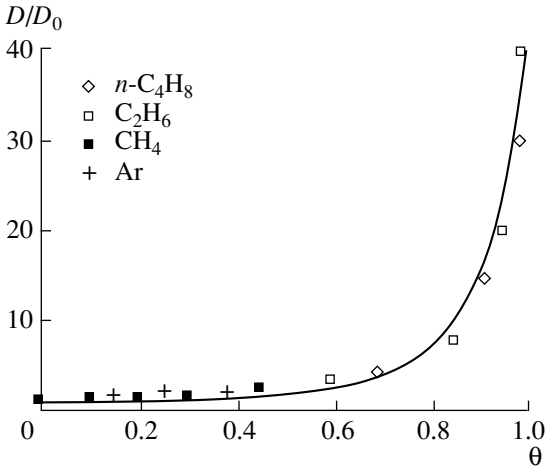


Fig. 4. Relative diffusion coefficient versus the filling factor of the channel. Solid curve—theoretical dependence (3.50) with $\zeta = 0.86$; dots—experimental data from [21].

Figure 5a shows the temperature dependences of the gas flux through a membrane for ethane and argon, calculated using the formulas (2.5) and (3.50) and a relation following from equation (3.6): $J \propto D(\theta_0(T, p))\theta_0(T, p)$. These dependences were compared with the experimental values obtained in [8]. It is evident that the theoretical and experimental results agree satisfactorily with one another. The temperature dependence can also be understood on the basis of the model developed in the present work. Indeed, it follows from the relation (2.5) that at low temperatures ($T \longrightarrow 0$) the filling factor of the channel $\theta_0 \sim 1$. The effective diffusion activation energy $\tilde{E}(\theta_0)$ is E_0 at zero temperature (3.27). The diffusion coefficient of the gas particles in the channel (3.50) approaches zero as a result of the "freezing out"

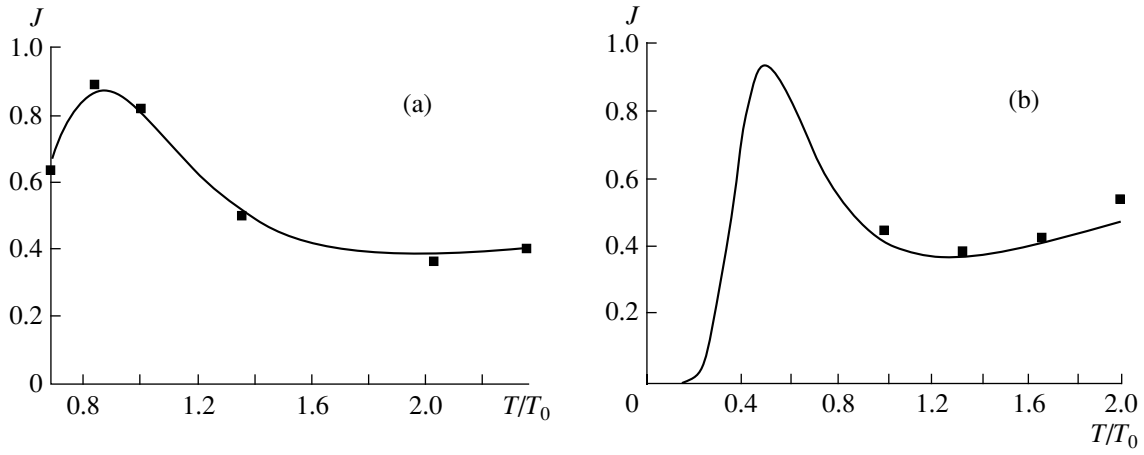


Fig. 5. Relative flux versus the relative temperature T/T_0, T_0 = 296 K. Solid lines—theoretical curves calculated using equations (2.5), (3.6), and (3.50); dots—experimental data (a) for ethane [8] and (b) argon [22].

of the thermal motion of the particles in the potential field of the channel surface. Increasing the temperature decreases $\tilde{E}(\theta_0)$ without appreciably changing the filling factor of the channel. The value of $\tilde{E}(\theta_0)$ reaches a minimum at a certain temperature T_0. It is obvious that the diffusion coefficient reaches its maximum value precisely at this temperature. It follows from equations (2.5) and (3.49) that a further increase of temperature decreases the filling factor of the channel θ_0, and this is accompanied by an increase in the effective diffusion activation energy right up to the value E_0 which is attained for $\theta_0(P, T) \approx \theta_0 \ll 1$ at some temperature $T = T_1(p)$. The temperature T_1 corresponds to a minimum in the function $D(T)$. A further increase of the diffusion coefficient with increasing temperature corresponds to the variation of the coefficient according to the Arrhenius law $D \propto \exp(-E_0/T)$. Figure 5b shows the temperature dependence of the flux for argon, calculated using equation (3.50) and the data from [21]. We note that the theory developed in the present paper predicts that the presence of a minimum at $T \sim 400$ K should be accompanied by the appearance of a maximum at $T \sim 200$ K. In this connection, for further elaboration of the theory it is of interest to make an experimental search for a maximum in the temperature dependence of the flux for argon near $T \sim 200$ K.

Thus, two types of behavior of particles in a channel, which have different diffusion mechanisms, can be distinguished in the pressure and temperature ranges investigated. For the first group this is the diffusion of single particles in a channel, and for the second group it is diffusion as a result of a collective interaction of particles in a completely filled channel. Such behavior of single-component gases is a direct consequence of the real one-dimensionality of subnanometer channels, where molecules cannot change places with one another. The proposed model is based on two basic assumptions: (1) The pair interaction between gas particles plays the decisive role in the description of the state and transport phenomena in a one-component gas in subnanometer channels and (2) as a result of the fact that the channel diameter is comparable to the diameter of the gas molecules, such channels can be assumed to be one-dimensional.

The one-dimensionality of the system studied is actually due to the fact that the gas molecules do not have classical transverse degrees of freedom. This difference from the conventional systems plays a fundamental role in the analysis of all phenomena described above.

We note that equation (3.47) derived above possesses, besides the solution described in the text, strongly nonuniform nonstationary soliton-like solutions. It should be expected that an experimental consequence of the existence of such solutions will be the appearance of substantial fluctuations of the gas flux through a membrane.

The proposed model satisfactorily describes the equilibrium properties of a gas in a channel (specifically, the filling factor) and the weakly nonequilibrium properties of the gas, such as the diffusion coefficient and flux. It should be noted that a satisfactory description of the equilibrium properties, for example, the adsorption isotherm, has been obtained previously on the basis of a number of phenomenological models [8]. A qualitative analysis of the maximum (but not minimum) in the temperature dependence of the relative diffusion coefficient is presented in [9]. A model describing the dependence of the diffusion coefficient on the filling factor on the basis of a one-particle approximation by introducing into the diffusion rate jumps of individual particles a phenomenological correction factor $\chi = (1 - \theta_0)^{-1}$, has been proposed in [2]. However, even though each of the phenomena indicated above can be described on the basis of individual models, the approximations used in so doing

either are not physically substantiated or completely contradict one another.

ACKNOWLEDGMENTS

We thank INTAS (grant no. 99-01744) for partial financial support of this work.

REFERENCES

1. V. N. Bogomolov, Phys. Rev. B **51**, 17040 (1995).
2. S. Yu. Krylov, A. V. Prosyanov, and J. J. M. Beenakker, J. Chem. Phys. **107**, 6970 (1997).
3. J. J. M. Beenakker, V. D. Borman, and S. Yu. Krylov, Phys. Rev. Lett. **103**, 4622 (1994).
4. Yu. N. Devyatko and V. N. Tronin, Zh. Éksp. Teor. Fiz. **98**, 1570 (1990) [Sov. Phys. JETP **71**, 880 (1990)].
5. V. D. Borman, S. Yu. Krylov, and A. V. Prosyanov, Zh. Éksp. Teor. Fiz. **97**, 1795 (1990) [Sov. Phys. JETP **70**, 1013 (1990)].
6. K. K. Sirkar, Chem. Eng. Commun. **157**, 145 (1997).
7. S.-T. Hwang and K. Kammermeyer, *Membranes in Separations* (Wiley, New York, 1975; Khimiya, Moscow, 1981).
8. J. M. van de Graaf, Thesis (University of Delft, Delft, 1999).
9. F. Kapteijn, W. J. W. Bakker, G. Zheng, *et al.*, Chem. Eng. J. **57**, 145 (1995).
10. J. M. Ziman, *Models of Disorder: The Theoretical Physics of Homogeneously Disordered Systems* (Cambridge Univ. Press, Cambridge, 1979; Mir, Moscow, 1982).
11. L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 3rd ed. (Nauka, Moscow, 1976; Pergamon Press, Oxford, 1980), Part 1.
12. R. Balescu, *Equilibrium and Nonequilibrium Statistical Mechanics* (Wiley, New York, 1975; Mir, Moscow, 1978).
13. I. Z. Fisher, *Statistical Theory of Liquids* (Fizmatgiz, Moscow, 1961).
14. P. Resibois and M. De Leener, *Classical Kinetical Theory of Fluids* (Wiley, New York, 1977; Mir, Moscow, 1980).
15. R. E. Richards and L. V. C. Rees, Langmuir **3**, 335 (1987).
16. W. J. W. Bakker, G. Zheng, F. Kapteijn, *et al.*, in *Precision Process Technology*, Ed. by M. P. C. Weijnen and A. A. H. Drinkenburh (Kluwer, Amsterdam, 1993), p. 425.
17. *Introduction to Zeolite Science and Practice*, Ed. by H. van Bekkam, E. M. Flanigen, and J. C. Jancen (Elsevier, Amsterdam, 1991), Vol. 58.
18. G. T. Kokotailo, S. L. Lawton, D. H. Olson, and W. M. Meier, Nature **272**, 437 (1978).
19. D. H. Olson, W. D. Haag, and R. M. Lago, J. Catal. **61**, 390 (1980).
20. R. Krishna, Gas Sep. Purif. **7**, 91 (1993).
21. W. J. W. Bakker, G. Zheng, F. Kapteijn, *et al.*, Microporous Mater. **3**, 227 (1994).
22. W. J. W. Bakker, F. Kapteijn, K. J. Jansen, *et al.*, Process Technol. **12**, 7 (1993).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 964–973.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1110–1121.

SOLIDS
Structure

Universality of the Ratio of the Critical Amplitudes of the Magnetic Susceptibility in a Two-Dimensional Ising Model with Nonmagnetic Impurities

O. A. Vasilyev and L. N. Shchur*

Landau Institute for Theoretical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia
**e-mail: lev@itp.ac.ru*

Received January 18, 2000

Abstract—The behavior of the magnetic susceptibility of a two-dimensional Ising model with nonmagnetic impurities is investigated numerically. A new method for determining the critical amplitudes and critical temperature is developed. The results of a numerical investigation of the ratio of the critical amplitudes of the magnetic susceptibility are presented. It is shown that the ratio of the critical amplitudes is universal right up to impurity concentrations $q \le 0.25$ (the percolation point of a square lattice is $q_c = 0.407254$). The behavior of the effective critical exponent $\gamma(q)$ of the magnetic susceptibility is discussed. Apparently, a transition from Ising-type universal behavior to percolation behavior should occur in a quite narrow concentration range near the percolation point of the lattice. © *2000 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

The effect of impurities on the critical behavior of magnetic systems is one of the most important subjects in the theory of phase transitions [1–3]. The Harris criterion [4] answered the fundamental question of the change occurring in the critical behavior when a small quantity of stationary ("frozen-in") impurities is introduced. According to this criterion, if $d\nu > 2$, where d is the dimension of the system and ν is the critical exponent of the correlation length, then impurities do not change the critical exponents.[1] The Harris criterion is inapplicable to the two-dimensional Ising model because for this model $d\nu = 2$, i.e., it is a boundary case.

The problem of the critical behavior of the specific heat in the two-dimensional Ising model in which a low concentration q of impurity bonds is introduced was solved by the Dotsenko brothers [5]. They discovered that as long as the correlation length ξ does not exceed the characteristic distance between impurities $l_i \simeq \exp(\text{const}/q)$, the behavior of the specific heat is identical to the case of a pure Ising model, i.e., the specific heat $C(\tau)$ diverges logarithmically: $C(\tau) \simeq \ln(1/|\tau|)$, where $\tau = (T - T_c)/T_c$ is the reduced temperature. However, for values of the temperature closer to the critical temperature, so that the correlation length ξ is much greater than the impurity length l_i, the critical behavior of the specific heat changes to a double logarithmic behavior: $C(\tau) \simeq \ln\ln(1/|\tau|)$. The correlation length acquires the logarithmic factor $\xi \simeq \sqrt{\ln|\tau|}/|\tau|$. The critical behavior of the magnetic susceptibility $\chi(\tau) \simeq \xi(\tau)^{-7/4} \simeq |\tau|^{-7/4}(\ln|\tau|)^{7/8}$ was also subsequently clarified independently by Shalaev, Shankar, and Lüdwig (see the review [6]).

These results were confirmed numerically in the works of many authors for the model of random weakened bonds (see the review [2]). The theoretical results presented above (the Dotsenko– Dotsenko–Shlaev–Shankar–Lüdwig theory [5, 6]) have been definitely checked for the model of random ferromagnetic bonds and for the model of impurity nonmagnetic sites and low impurity concentrations.

The results obtained by different authors on the basis of numerical simulation for finite concentrations of nonmagnetic impurities occupying sites in a square lattice are not so unequivocal. The result presented above that the critical exponents of the magnetization and susceptibility are independent of the degree of disorder is confirmed in [7, 8]. At the same time, other authors assert that the susceptibility exponent depends on the degree of disorder and a "weak universality"—independence of the ratio γ/ν of the susceptibility and correlation length exponents from impurities—is observed [9–11].

It is well known that the universality class is characterized not only by the critical exponents by also by numerous combinations of critical amplitudes (see the review [12]). The difference between the values of the critical exponents for various universality classes is often small, while the ratios of the amplitudes can differ substantially.

In the present work we posed the problem of determining numerically the ratio of the critical amplitudes of the magnetic susceptibility for the Ising model with impurities in order to check the question of whether or

[1] Systems that undergo first-order phase transitions in the absence of impurities are not considered in this paper.

not the universality class changes from Ising-type (with impurity concentration $q = 0$) to percolation-type (with $q = q_c$).

We have developed a numerical method for determining simultaneously the effective critical exponent γ of the susceptibility, the critical amplitudes Γ and Γ' of the susceptibility, and the "critical temperature" of a sample of finite size. We used two methods to determine the quantities of interest. In the first method the values obtained for the amplitudes were averaged over realizations of the impurity distribution in the samples. In the second method we first averaged the magnetic susceptibility over the impurity distribution and then determined the critical amplitudes. Both methods lead to the same results.

Analysis of the numerical data shows that the ratio of the critical amplitudes of the susceptibility remains unchanged in a quite wide range of impurity concentrations investigated $0 \le q \le 0.25$, while the amplitude values themselves undergo a several-fold change. At the same time, the effective critical exponent γ for the susceptibility varies continuously, reaching a maximum at the concentration of nonmagnetic sites $q \approx 0.2$. This result can apparently be explained by the fact that in the indicated impurity concentration range the system possesses an Ising-type universality. The change in the susceptibility exponent can be attributed to the corrections, which are unknown to us, to the scaling behavior. Nonetheless, a numerical investigation of the critical exponents and amplitudes close to the percolation point of the lattice is required in order to obtain a more reliable answer.

This paper is organized as follows. The model and the parameters for which the numerical simulation was performed are described in Section 2. A new method for determining the critical amplitudes, which makes it possible to investigate the ratio of the critical amplitudes for strongly diluted systems, is presented in Section 3. This method also makes it possible to determine the critical temperature. The computational results are presented in Section 4, and these results are discussed in Section 5.

2. MODEL AND NUMERICAL METHOD

Spins σ_i, assuming the values $\sigma_i = +1$ and -1, and nonmagnetic impurities ($\sigma_i \equiv 0$) occupy the sites i of an $L \times L$ square lattice with periodic boundary conditions. The nonmagnetic impurities are stationary (quenched disorder). In what follows we shall call a lattice with a fixed distribution of nonmagnetic impurities the sample. The bond energy between two sites is 0 if a nonmagnetic impurity occupies at least one site and J if magnetic spins occupy both sites. The Hamiltonian of such a system can be written in the form

$$H = -J\sum_{i,j}\sigma_i\sigma_j \qquad (1)$$

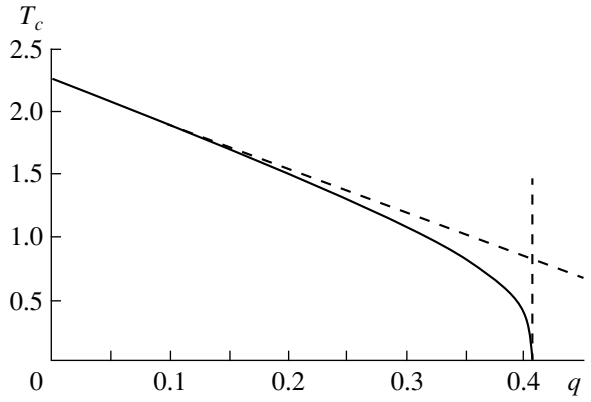


Fig. 1. Phase diagram of the two-dimensional Ising model with nonmagnetic impurities: T_c is the critical temperature and q is the concentration of nonmagnetic impurities.

(in what follows we assume $J = 1$ and we work with the dimensionless temperature). The concentration of magnetic spins is determined by summing the absolute value of the spin at all lattice sites:

$$p = \frac{1}{L^2}\sum_{i=1}^{L^2}|\sigma_i|. \qquad (2)$$

Then the value $p = 1$ corresponds to a pure Ising model and $p = 0$ corresponds to an empty, nonmagnetic lattice. Evidently, $p = 1 - q$.

As the concentration of magnetic sites decreases, there exists a critical impurity concentration q_c down to which a cluster of magnetic sites touching the sample boundaries is always present in the system [13]. In this case, a phase transition can occur in the system at a finite temperature into a phase with nonzero total magnetic moment of the system. The phase diagram is shown schematically in Fig. 1. The vertical asymptote passes through the percolation point $q_c = 0.407254$ [14]. The straight line is the asymptote of the phase curve for low impurity concentrations, $T_c(q) = T_c(0)(1 - 1.565q)$ [1].

For a concentration of magnetic sites below the critical value, $p < p_c$, the lattice separates into finite clusters with characteristic size of the order of the percolation correlation length $\xi_p \propto (p - p_c)^{-\nu_p}$, where ν_p is the critical exponent of the percolation length (in the two-dimensional case $\nu_p = 4/3$) [13]. Since the magnetic moments of individual clusters are uncorrelated, the total magnetic moment will be 0 and a phase transition at $p < p_c$ is impossible [1].

When the sample is produced, unbound geometric clusters of magnetic sites arise in the lattice. Far from the critical point, for $p > p_c$, a cluster of magnetic sites extending from edge to edge and n_s clusters containing a relatively small number of sites s are always present in a sufficiently large lattice $L \gg \xi_p \propto (p - p_c)^{-\nu_p}$:

$$n_s \sim (p - p_c)^2 \exp[-s(p - p_c)]. \qquad (3)$$

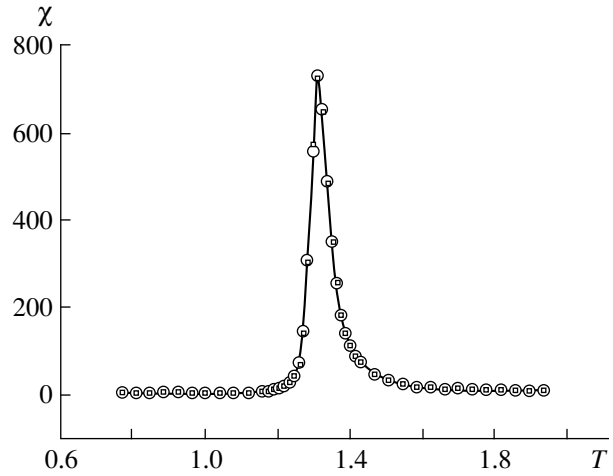


Fig. 2. Susceptibility χ of the maximum cluster (squares and solid line) and of the entire sample (circles) as a function of temperature T. The impurity concentration $q = 25\%$.

In the limit of an infinite lattice, only correlated magnetic moments of an infinite, passing cluster will make a singular contribution to the magnetic susceptibility and the specific heat. Consequently, when calculating the critical characteristics it is reasonable to neglect the contribution from sites which have no bonds with the passing cluster. In this manner, the "noise" from magnetic moments of finite-size clusters can be decreased.

The practical details for implementing the algorithm are as follows.

When the sample is produced, frozen-in nonmagnetic impurities occupy sites in an $L \times L$ lattice. The coordinates of the impurities are a sequence of pseudorandom numbers, which were obtained using a shift-register type generator [15] with 9689 and 471 cycle lengths [16]. Generators with such a large length decrease the systematic error arising as a result of the use of pseudorandom number generators in Monte Carlo cluster algorithms [17, 18]. Using the expression obtained in [17] to estimate the maximum error δE for the energy E

$$\delta E < L^{-0.84} p^{-0.52},$$

we obtain for $L \geq 10$ that the expected systematic error in determining the energy is $\delta E < 0.00005$, which is less than the statistical error obtained in our calculations. In other words, the systematic error introduced by the pseudorandom number generator in our case is indistinguishable against the background due to the statistical error.

Monte Carlo calculations show that small clusters make a very small contribution to the thermodynamic quantities and do not exhibit singular behavior at the critical point. For this reason, to investigate the singular behavior of the susceptibility all spin clusters except an "infinite" cluster were replaced by nonmagnetic impurities. In what follows, the thermodynamic characteristics of only a cluster of maximum size were simulated. In the first place, this decreased the computational time and, in the second place, it made it possible to filter immediately the "noise," introduced into the susceptibility by many small clusters. Technically, this was done as follows. A fixed number of nonmagnetic impurities, equal to the integer part of $L^2(1 - p)$, were distributed on the lattice. The coordinates of the impurities were chosen using the pseudorandom number generator. Next, using the Hoshen–Kopelman algorithm [19], the sample obtained was divided into clusters of uncoupled spins. Then, all clusters, with the exception of the largest cluster, were replaced by nonmagnetic impurities. Figure 2 shows the susceptibility of one sample with $p = 0.75$ and the susceptibility of the largest cluster for the same sample. It is evident that removing the very small clusters has no effect on the critical behavior.

A cluster algorithm for the Ising model was proposed in [20]. A more efficient variant of the algorithm is proposed in [21]. The spin variables $\sigma_{i,j}$, assuming the values $\sigma_{i,j} = \pm 1$, are distributed at the sites of an $L \times L$ square lattice. Let any initial configuration of the spins $\Xi = \{\sigma_{i,j}\}$, $i, j = 1, \ldots, L$, be given. We introduce a set of variables on the bonds $\Omega = \{\omega_{i,j}\}$, $i, j = 1, \ldots, L$. Each variable can assume the values $\omega_{i,j} = 0$ (broken bond) and $\omega_{i,j} = 1$ (closed bond). The collection of spins, bound with one another by closed bonds ($\omega_{i,j} = 1$), is called a spin cluster.

We shall now consider the process of producing a configuration of variables on bonds Ω in a spin configuration Ξ. If two spins are antiparallel, then the bond between them is always broken. If the spins are parallel, then the bond between them is closed with probability $1 - \exp(-2\beta)$ and broken with probability $\exp(-2\beta)$. This process of correlated percolation along bonds [20] divides the lattice into so-called "physical" clusters (in contrast to the procedures where bonds between parallel spins are always closed, in which case "geometric" clusters are generated). If we now assign to the spins of each physical cluster the values ± 1 with probability 1/2, then we obtain a new spin configuration Ξ'. This procedure is the Swendsen–Wang algorithm.

The Wolf method differs by the fact that a spin is chosen arbitrarily in the lattice, a physical cluster to which the spin belongs is constructed, and then the entire constructed cluster is flipped. A substantiation of cluster algorithms is proposed in [22].

Here we present the Wolf variant of a Monte Carlo algorithm in the form which we employed.

(1) Two random numbers fix the coordinates (i, j) of a lattice site. If a nonmagnetic impurity occupies this site, then new random numbers are generated and they will continue to be generated until the coordinates of the "present" spin, which we shall call the central spin, are generated. The central spin is flipped—its value is reversed.

(2) The nearest "neighbors" of the central spin are examined. If a magnetic spin, which is parallel to the

unflipped central spin, occupies a neighboring site, then this spin also flips with probability $1 - \exp(-2\beta)$, where $\beta = J/kT$, and its coordinates are stored in a stack.

(3) After all neighboring sites have been checked (and, if necessary, neighboring spins have been flipped), the spin whose coordinates were stored in the stack last becomes the "central" spin and the operation 2 is repeated).

(4) The procedure of flipping the spins is terminated when the stack is empty. This process is called flipping of a "cluster," and all flipped spins are said to belong to the "cluster" (Wolf). We note that when a cluster is flipped, using the procedure described above, the same spin can be checked several times, but it can be flipped only once.

This algorithm implements a Markov process, and spin configurations are generated with the corresponding probability. The energy and magnetization of a given spin configuration n are calculating using the formulas

$$e_n = -\sum_{i,j}\sigma_i\sigma_j, \quad m_n = \sum_i \sigma_i. \tag{4}$$

To decrease the correlations, the energy e and the magnetization m of the spin configuration were calculated every three flips of a Wolf cluster—we call this one Monte Carlo step. At the outset of the calculations all spins are parallel (so-called "cold start"—this corresponds to the state of the system at $T = 0$). To obtain a spin configuration characteristic for a given temperature, a certain number of clusters must be flipped. This process is called thermalization. In our calculations thermalization constituted 10^4 Monte Carlo clusters. Then, the thermodynamic characteristics for a given sample were calculated by averaging over $N = 10^5$ configurations using the formulas

$$E = \langle e\rangle = \frac{1}{N}\sum_{n=1}^{N} e_n,$$

$$M = \langle |m|\rangle = \frac{1}{N}\sum_{n=1}^{N} |m_n|,$$

$$C = \beta^2\left(\frac{1}{N}\sum_{n=1}^{N} e_n^2 - \langle e\rangle^2\right), \tag{5}$$

$$\chi = \begin{cases} \beta\left(\dfrac{1}{N}\displaystyle\sum_{n=1}^{N} m_n^2 - \langle |m|\rangle^2\right), & \beta > \beta_c, \\ \beta\dfrac{1}{N}\displaystyle\sum_{n=1}^{N} m_n^2, & \beta < \beta_c, \end{cases}$$

where n is the number of a given spin configuration and $\langle\ldots\rangle$ is the thermodynamic average over the spin configurations.

3. DATA PROCESSING METHOD

3.1. Analysis of Existing Methods

We recall that our problem is to calculate the ratio $R = \Gamma/\Gamma'$ of the critical amplitudes Γ and Γ' of the susceptibility. To determine the critical amplitudes we must know the critical temperature, and using this temperature we must interpolate the numerical data in order to determine the critical amplitudes and critical exponents.

We shall now analyze the possibility of such an analysis for the example of an impurity-free two-dimensional Ising model, for which an analytic solution is known.

For the Ising model, smoothing [23] of the singularities of the susceptibility in a certain range $[-\tau_r(L), \tau_r(L)]$ of the reduced temperature $\tau = (T - T_c(q))/T$ occurs on a finite-size lattice. Here and below we shall use a definition of the reduced temperature where the denominator contains not T_c but rather T, which is generally accepted for analysis of the behavior of the susceptibility [24]. The "smoothing" radius $\tau_r(L)$ of the singularities of the susceptibility is inversely proportional to the size of the system: $\tau_r(L) \simeq 1/L$. This is because for $|\tau| \sim \tau_r(L)$ the correlation length $\xi(\tau) \simeq \tau^{-\nu}$ ($\nu = 1$) reaches the lattice size L. Correspondingly, the values of the thermodynamic quantities within the smoothing radius, $|\tau| < \tau_r(L)$, are bounded and differ from their values for an infinite sample. The smoothing effect has been demonstrated numerically in [25].

For the Ising model without impurities ($T_c(0) = 2/\ln(\sqrt{2} + 1)) = 2.269185314\ldots$), the dependence of the susceptibility $\chi(\tau)$ near the critical point for $|\tau| \ll 1$ is known [24]:

$$T_\chi(\tau) = \begin{cases} \Gamma(\tau)^{-\gamma}(1 + e_\chi\tau) + \ldots, & \tau > 0, \\ \Gamma'(-\tau)^{-\gamma'}(1 + e_\chi\tau) + \ldots, & \tau < 0, \end{cases} \tag{6}$$

where $\gamma = \gamma' = 7/4$, $\Gamma = 0.9625817322\ldots$, $\Gamma' = 0.0255369719\ldots$, $e_\chi = 0.07790315\ldots$. It is easy to see that, despite the smallness of the coefficient e_χ, the correction $e_\chi\tau$ to the scaling dependence $\chi_s(\tau) \simeq |\tau|^{-\gamma}$ can become substantial for comparatively small values of the reduced temperature τ and reaches 1% even for $\tau_a \simeq 0.13$.

Thus, the critical behavior of the magnetic susceptibility, $\tau_r(L) < |\tau| < \tau_a$, can be observed in the temperature ranges $\chi(\tau) \simeq \chi_s(\tau)$. The existence of the critical region can be proved by constructing plots of the ratio of the susceptibility to its singular part $\chi_s \simeq \tau^{-\gamma}$ [26]. An example is shown in Fig. 3 for a system with linear size $L = 256$. It is evident from the figure that in the critical region $0.01 \le \tau \le 0.1$ there exists a "plateau"—a section where the ratio of the numerical data obtained for the susceptibility χ to $\chi_s \simeq \tau^{-\gamma}$ is quite close to the theoretical value of the critical amplitudes (see expression (6)).

The plot of the susceptibility with concentration p of magnetic sites on a lattice with a finite linear size L

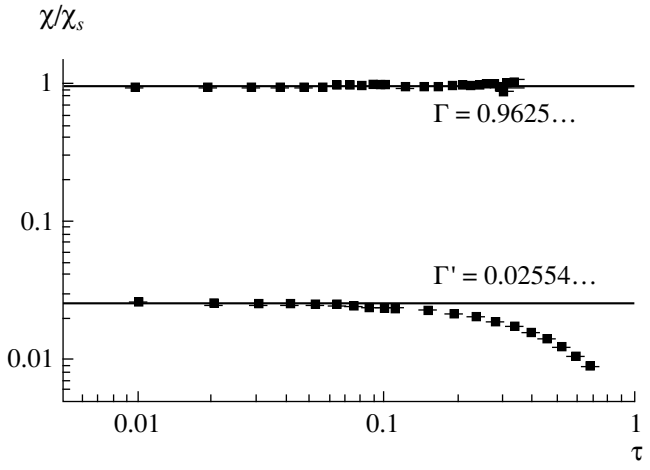


Fig. 3. Ratio of the computed susceptibility $\chi(|\tau|)$ to the singular part of the susceptibility $\chi_s(|\tau|)$in the critical region of the pure Ising model. The exact value of the critical temperature $T_c = 2.26918\ldots$ was used to calculate τ. The solid lines show the theoretical values of the critical amplitudes for an infinite lattice.

possesses a maximum at a certain value of the temperature $T_{\max}(p; L)$; see [7]. As the impurity concentration q increases, the critical temperature $T_c(q)$ decreases monotonically to 0 [1] when the critical impurity concentration $q_c = 1 - p_c$ is reached, as shown in the phase diagram in Fig. 1.

Thus, to determine the critical amplitudes we must first calculate the values of the susceptibility $\chi_{\text{comp}}(T, p)$ for a fixed impurity concentration $q = 1 - p$ for several values of the temperature T in the low- and high-temperature critical regions. Next, we must know the critical temperature $T_c(p)$ for each fixed value of the concentration p of the magnetic sites. It is necessary to know $T_c(p)$ in order to calculate the reduced temperature $\tau = (T - T_c(p))/T$ and then fit the numerical data for the susceptibility $\chi_{\text{comp}}(T, p)$. The fit in turn is necessary in order to determine the critical amplitudes of the susceptibility, $\Gamma(p; L)$ and $\Gamma'(p; L)$, and the critical exponents $\gamma(p; L)$ and $\gamma'(p; L)$ assuming

$$T\chi_{\text{comp}}(\tau, p) \simeq \begin{cases} \Gamma(p; L)(\tau)^{-\gamma(p; L)}, & \tau > 0, \\ \Gamma'(p; L)(-\tau)^{-\gamma(p; L)}, & \tau < 0. \end{cases} \quad (7)$$

The Monte Carlo calculation yields a set of numerical values of the susceptibility $\chi(T, p; L, i)$, where i is the number of the sample, $i = 1, N_L$, and N_L is the number of samples in the simulation of a system of size L. The realization of disorder, i.e., the difference in the distribution of the impurities over the sites of an $L \times L$ square lattice, distinguishes the samples i.

The values of the critical temperature $T_c(p)$ for the Ising model in the presence of impurities are not known analytically. We need a method to determine $T_c(p)$ from the experimental numerical data. The following methods are most commonly used:

The first method is Binder's cumulant method [3]

$$U_L(T, p) = 1 - \frac{\langle m^4(T, p; L)\rangle}{3\langle m^2(T, p; L)\rangle^2}.$$

To determine T_c it is necessary to calculate the cumulant $U_L(T, p) = [U_L(T, p; i)]$, averaged over $i = 1, \ldots, N_L$ samples, for several lattice sizes $(L_1, L_2, \ldots, L_n)$. The critical temperature T_c is determined as the temperature at which the average value of the cumulant is independent of the lattice size $U_{L_1}(T_c, p) = U_{L_2}(T_c, p) = \ldots = U_{L_n}(T_c, p)$.

In [27] the calculation of Binder's cumulant by the transfer- matrix method was checked for a two-dimensional Ising model with random bonds J_{ij} on a strip of finite width for the special case of a self-dual model. For this model the analysis simplifies substantially because the critical temperature is known exactly. It was found that the value of Binder's cumulant seems to be universal. At the same time it was found that the accuracy of determination is inadequate to establish universality reliably. Our preliminary Monte Carlo calculations showed that even for our case of a model with impurities at the sites the computational accuracy is inadequate to establish reliably the universality of Binder's cumulant.

The second method uses the maximum susceptibility. For a fixed lattice size L the temperature $T^*(p, L)$ at which the susceptibility $\chi(p, T; L, i)$ is maximum can be taken as "critical." In this case it is difficult to determine for a specific sample the position of the maximum with an error much less than the smoothing radius τ_r. To determine $T_c(p)$ by this method it is necessary to have data for different lattice sizes and to determine the limit $T_c(p) = \lim_{L \to \infty} T^*(p; L)$.

The temperatures $T^*(p; L, i)$ generally speaking are different for realizations of samples i with different impurity distributions. The average temperature $T^*(p; L)$ can be determined from these realizations by two methods. These approaches are equivalent when self-averaging of the computed quantity occurs [28]. We note that the question of self-averaging of the temperature has not been adequately studied for a system with nonmagnetic impurities [29], especially near the percolation point.

In the first averaging method, the magnetic susceptibility can first be averaged over samples:

$$\chi(T, p; L) = [\chi_i(T, p; L)] \equiv \frac{1}{N_L}\sum_{i=1}^{N_L}\chi_i(T, p; L),$$

and then $T^*(p, L)$ can be determined from the maximum of the average susceptibility $\chi(T, p; L)$.

Table 1. Two variants of a fit to the critical amplitudes of the susceptibility for the Ising model on a square lattice with linear size $L = 256$ without impurities ($p = 1$) and a comparison with exact values

Variant	β_c	Γ'	Γ	$\gamma = \gamma'$	Γ/Γ'
Exact theory	0.440687	0.02554	0.9625	1.75	37.685
Fit of the amplitudes	0.440687*	0.024730(27)	0.9667(30)	1.75*	39.09(15)
Method of the present paper	0.440831(14)	0.02388(8)	0.9596(53)	1.7567(11)	40.19(23)

*The value of the parameter is fixed and does not change in the course of the fit. The value in parentheses is the statistical error to the last significant figures of average quantities, calculated as the standard deviation when averaging over samples.

The second approach consists of finding the critical point $T^*(p; L, i)$ for a given sample i and averaging over samples:

$$T^*(p; L) = [T^*(p; L, i)] \equiv \frac{1}{N_L}\sum_{i}^{N_L} T^*(p; L, i).$$

Simulation of large lattices ($L \sim 256$) requires a large volume of computer calculations for each sample. Consequently, the number of samples is often not very large. It would be attractive to have a method for analyzing computational results that would make it possible to process data for an individual sample. To determine the critical temperature according to the data for an individual configuration of impurities, we proposed the following method for analyzing the data.

3.2. New Method

Our method is based on a single assumption about the critical behavior of the susceptibility. We assume that the critical exponents γ and γ' are the same on both sides of the transition point and that the analytical corrections to the singular part of the susceptibility are of the same nature. This actually reduces to equality of the "effective" exponents γ and γ', since we do not know the corrections to scaling in the presence of impurities. In the approximation using equations (7), we obtained a value of the temperature for each sample $T^*(p; L, i)$ in a manner so as to make the effective exponents γ and γ' equal to one another. In so doing, we have two parameters, for each exponent, for fitting the experimental numerical data in each phase. In the high-temperature phase we have the exponent γ and the amplitude Γ; in the low temperature phase we have γ' and Γ'.

The results of an application of this method to the pure Ising model ($p = 1$) and a comparison of this method with the known exact values are presented in Table 1. The exact numerical values of the critical amplitudes Γ and Γ' [24] and the reciprocal of the critical temperature $\beta_c = 1/T_c$ for an infinite system are given in the first row. The second row gives the values of the critical amplitudes obtained by approximating the numerical data for a finite lattice $L = 256$ using equations (7), where we used the exact values $\beta_c = \ln(\sqrt{2} + 1)/2$ and $\gamma = 7/4$. Finally, the third row gives the results of our method for analyzing the same numerical data. The difference of the value found for the reciprocal of the critical temperature from the exact value is less than the roundoff interval of the singularity $\tau_r \approx 1/L \approx 0.004$. The difference in the ratios of the critical amplitudes in the second and third rows on five standard deviations can be attributed to the systematic error of the method employed. This is due to the difficulty of performing calculations of the critical amplitude Γ', which is small in magnitude, in the ordered phase. As one can see from Table 1, the accuracy of the calculation of the critical amplitude Γ is much higher and is identical, to within the statistical error, to the exact value for an infinite lattice and with the approximate value presented in the second row in Table 1.

In the critical region the plot of the susceptibility (multiplied by T) should be identical to $\Gamma\tau^{-1.75}$ for $\tau > 0$ and $\Gamma'(-\tau)^{-1.75}$ for $\tau < 0$, and in double logarithmic coordinates it should follow two straight lines with the tangents of the slope angles $\gamma = \gamma' = 1.75$. The singular part of the susceptibility $\chi_s(\tau)$ in double logarithmic coordinates is represented by two straight lines:

$$\begin{aligned} \ln\chi_s &= \ln\Gamma - \gamma\ln\tau, \\ \ln\chi_s &= \ln\Gamma' - \gamma'\ln(-\tau). \end{aligned} \qquad (8)$$

The numerical data agree well with the known behavior of the singular part of the susceptibility, $\chi_s \simeq |\tau|^{-1.75}$. The difference of the critical amplitudes from the exact values is explained by the finiteness of the lattice—the exact value equal to $\gamma = 1.75$ and the critical exponents γ and γ' assume values on an infinite lattice, where the dependence (6) is valid as a whole.

The values presented in Table 1 were obtained by analyzing the numerical data from a sequence of 1.5×10^6 Monte Carlo steps divided into five segments. To decrease the correlation between the values at successive steps, the values of the energy and magnetization (4) calculated every three Monte Carlo steps were used to perform the averaging. The first 10^4 Monte Carlo steps were used for relaxation of the system from the initial state and were not used for averaging.

The error of the method described above is discussed in [30]. The numerical values presented for the critical exponents and the preexponential completely correspond to the known values for the pure model. The

Table 2. Number of samples N_L in the simulation of the Ising model with magnetic site concentration p

p	1.0	0.97	0.93	0.90	0.85	0.82	0.80	0.78	0.75
N_L	5	10	13	12	20	13	21	15	18

Table 3. The results of a fit to the critical behavior of the magnetic susceptibility by the method of this paper (Section 3.2)

p	$p - p_c$	Γ'	Γ	γ	β	T	Γ/Γ'
1.00	0.407	0.02388(8)	0.9596(53)	1.7567(11)	0.440831(14)	2.26844(7)	40.19(23)
0.97	0.377	0.02366(15)	0.9456(82)	1.8144(26)	0.462742(36)	2.16103(7)	39.96(19)
0.93	0.337	0.023919(42)	0.955(16)	1.8846(67)	0.496368(63)	2.01463(2)	39.98(38)
0.90	0.307	0.02526(95)	0.999(31)	1.927(13)	0.52544(13)	1.9032(5)	39.65(58)
0.85	0.257	0.03240(98)	1.266(33)	1.954(11)	0.58488(15)	1.7098(5)	39.25(56)
0.82	0.227	0.0369(21)	1.423(64)	1.985(21)	0.62872(29)	1.5905(1)	38.92(85)
0.80	0.207	0.0427(13)	1.703(48)	1.969(13)	0.66214(19)	1.5103(4)	39.98(58)
0.78	0.187	0.0518(26)	1.991(84)	1.959(17)	0.70130(59)	1.4259(1)	39.08(1.4)
0.75	0.157	0.0678(49)	2.611(93)	1.952(21)	0.77060(98)	1.2977(17)	40.55(20)

estimates made in [30] show that the temperature $T(p; L, i)$ depends relatively weakly on the accuracy with which the critical exponents γ and γ' are determined. Conversely, the ratio Γ/Γ' of the critical amplitudes depends very strongly on the accuracy with which the critical exponents γ and γ' are determined and on the accuracy with which the temperature $T(p; L, i)$ is determined.

4. SIMULATION RESULTS

The method proposed in Section 3 was used to process the numerical data which we obtained by simulating a two-dimensional Ising model with random nonmagnetic impurities occupying sites of a square lattice with linear size $L = 256$. We employed periodic boundary conditions, for which, as is well known, the effect of the boundaries is weaker than in other methods [23, 25]. The number of samples for each concentration p of Ising spins from $p = 1$ to $p = 0.75$ is indicated in Table 2. The samples were produced by randomly selecting the coordinates of the nonmagnetic sites, which was done in order to achieve the required concentration of nonmagnetic sites $q = 1 - p$. To this end, a shift register type random number generator was used (see, for example, [15] and the brief discussion of the quality of random numbers in [31]). In the simulation the initial sequence of 10^4 Monte Carlo steps was used for relaxation from the initial ordered state. The energy e_n and magnetization m_n (see equation (4)) as well as the squared magnetization m_n^2 and the fourth power of the magnetization m_n^4 were calculated for $N = 10^5$ states of the system; another 2×10^5 intermediate states were omitted. This guaranteed that there would be no correlations between successive values of the magnetization and its moments; the autocorrelation time for a system of size $L = 256$ is about three Monte Carlo steps for a pure Ising model [21]. In our case, the correlation time for an impurity concentration of less than 30% does not change much.

The susceptibility $\chi(T; L, i)$ was calculated using the expressions (5) in the critical region $0.009 < |\tau| < 0.11$ for magnetic site concentration $p = 1$, in the region $0.01 < |\tau| < 0.1$ for $p = 0.97$, and in the region $0.03 < |\tau| < 0.1$ for all other concentrations presented in Table 2. The simulation for each sample was performed with 44 values of the temperature in each temperature range. Thus, the data analyzed below were obtained by analyzing a total of approximately 5×10^8 configurations of a system of 65526 spin in the equilibrium state. The remaining approximately 10^9 configurations were dropped during averaging.

The functions $\Gamma\tau^{-\gamma}$ in the disordered phase and $\Gamma'(-\tau)^{-\gamma'}$ in the ordered phase, using double logarithmic coordinates (see equation (8)), were fit to the numerical data for each sample. The value of the "critical" temperature $T^*(p; L, i)$ for each sample was determined from the condition that the critical exponents are equal to one another: $\gamma(i) = \gamma'(i)$. We recall that the reduced temperature τ depends on $T^*(p; L, i)$. The values obtained in this manner for the temperature $T^*(p; L, i)$, the critical exponents $\gamma(i)$, and the critical amplitudes $\Gamma(i)$ and $\Gamma'(i)$ were averaged over different realizations of samples with a fixed impurity concentration q. These values are presented in Table 3. It is evident that the critical amplitudes Γ and Γ' vary very strongly, approximately by a factor of 3, as the concentration of magnetic sites varies from $p = 1$ to $p = 0.75$, while the ratio Γ/Γ' of the critical amplitudes remains unchanged to

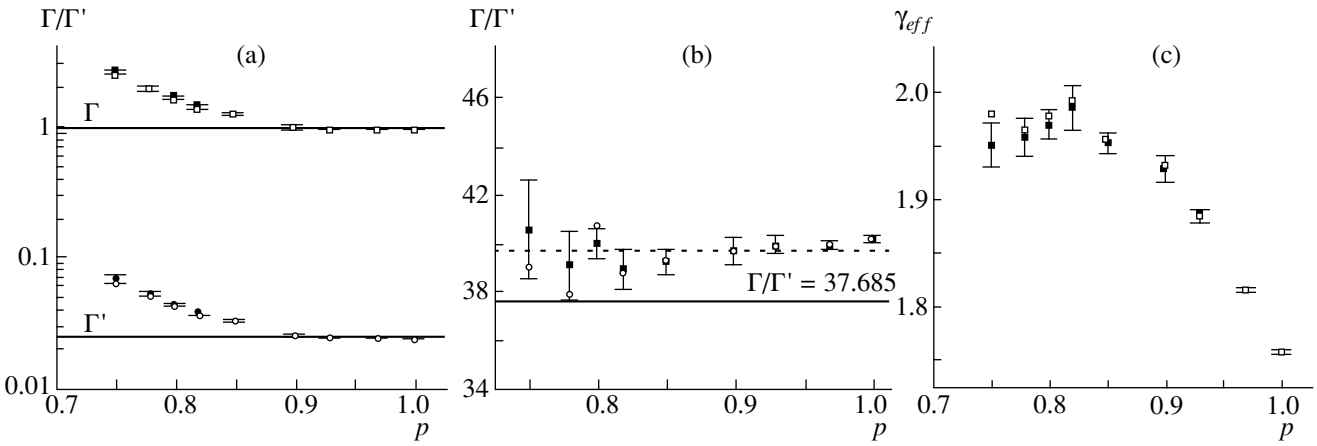


Fig. 4. (a) Critical amplitudes of the susceptibility Γ and Γ', (b) their ratio Γ/Γ', and (c) the effective critical exponent γ as a function of the concentration of the Ising spins. The solid straight lines represent the exact values for the "pure" Ising model; dotted line represents the average value of the amplitude ratio. The filled squares and circles with error bars show the data from Table 3; the open circles show the data from Table 4.

within the statistical error. Table 1 also gives the values Δp of the distances up to the percolation point $p_c = 0.592746$ [14]. As an illustration, the dependences of the critical amplitudes Γ and Γ' are constructed in Fig. 4a, and the dependence of their ratio Γ/Γ' is constructed in Fig. 4b. The conclusion that the critical amplitude ratio Γ/Γ' is independent, which can be drawn from the data presented, is the main physical result of this paper.

We note that while the ratio Γ/Γ' of the critical amplitudes of the susceptibility remains unchanged in a quite wide range of impurity concentrations, the effective critical exponent γ of the susceptibility varies quite appreciably. It is evident from Fig. 4c that the value of this exponent increases continuously, and it assumes its maximum value at an impurity concentration of approximately 20%, after which it decreases. This variation of the effective critical exponent γ agrees qualitatively with the variation noted in [9–11].

To monitor the accuracy we also performed a somewhat different analysis of the same numerical data. Specifically, first we averaged the numerical data for the susceptibility over the samples:

$$\chi(T; L) = \frac{1}{N_L}\sum_{i=1}^{N_L}\chi(T; L, i). \qquad (9)$$

The functions $\Gamma\tau^{-\gamma}$ and $\Gamma'(-\tau)^{-\gamma'}$ were then fit to these values. The values obtained in this manner for the critical exponents γ and the critical amplitudes Γ and Γ' are presented in Table 4. The values are identical to those presented in Table 3. This indicates that the fitting procedure is stable and the results of the analysis of the numerical data are reliable. Moreover, the agreement between the results of the two methods of averaging is an indirect indication of the fact that the susceptibility in the indicated parameter range is a self-averaging quantity.

Table 4. The result of a fit to the average susceptibility $\chi(T; L)$ (see the definition (9))

p	$p - p_c$	Γ'	Γ	γ	β	T_c	Γ/Γ'
1.00	0.407	0.02388	0.959880	1.7567	0.44083	2.26845	40.20
0.97	0.377	0.02364	0.944944	1.8146	0.46274	2.16104	39.97
0.93	0.337	0.02386	0.953161	1.8850	0.49637	2.01463	39.95
0.90	0.307	0.02482	0.984451	1.9323	0.52551	1.90291	39.66
0.85	0.257	0.03202	1.257918	1.9551	0.58492	1.70964	39.28
0.82	0.227	0.03538	1.371543	1.9947	0.62873	1.59051	38.76
0.80	0.207	0.04064	1.653404	1.9796	0.66224	1.51003	40.69
0.78	0.187	0.05037	1.908872	1.9688	0.70111	1.42631	37.90
0.75	0.157	0.06197	2.418006	1.9808	0.77073	1.29747	39.02

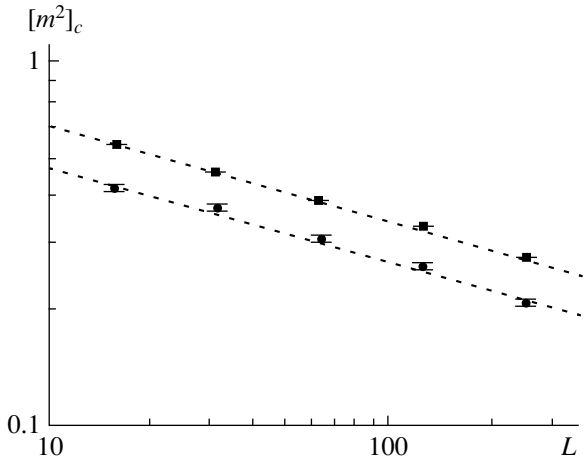


Fig. 5. $[m^2]_c$ versus the lattice size L at the critical point with sing spin concentrations $p = 1.0$ (filled squares) and 0.85 (filled circles). The dotted lines show the result of a fit with he obtained values of the critical exponent ratio γ/ν = 1.7505(6) and 1.747(1), respectively.

5. CONCLUSIONS

We have presented the results of a numerical investigation of the behavior of the magnetic susceptibility in the critical region of the two-dimensional Ising model with nonmagnetic impurities. Our data are based on an analysis of the temperature dependences of thermodynamic quantities. These are the dependences that make it possible to determine the critical amplitudes.

In virtually all previous works (with the exception of [26]) the critical exponents were extracted from an analysis of the dependences of the thermodynamic quantities on the dimensions of the system at the critical point (see the review in [2]). To demonstrate the quality of our numerical data and the possibility of making a direct comparison with the results obtained by other authors, we also performed a similar finite-dimensional analysis.

We shall now investigate the dependence of the susceptibility on the lattice size for the critical point. Our method makes it possible to determine the critical point for each sample. Consequently, the susceptibility can be calculated for each sample i and then averaged over N_L samples:

$$[m^2]_{av1}(L) = \frac{1}{N_L}\sum_{i=1}^{N_L} m_i^2(\beta_i, L),$$

where N_L is the number of samples. The variance β_i is small compared with the radius of "smoothing" of the susceptibility at the critical point, $\tau_r < 1/L \simeq 0.004$ for $L \le 256$. Consequently, we can determine

$$[\beta_c] = \frac{1}{N_L}\sum_{i=1}^{N_L} \beta_i,$$

and then find

$$[m^2]_{av2}(L) = \frac{1}{N_L}\sum_{i=1}^{N_L} m_i^2([\beta_c], L).$$

For $L = 256$ we obtained

$$[m^2]_{av1} = 547(22),$$

$$[m^2]_{av2} = 563(16),$$

i.e., both approaches give consistent values within the limits of error.

For lattices sizes $L < 128$ it is impossible to determine the critical temperature for each sample by the above-described method because the critical region where a power-law fit of the susceptibility is possible vanishes. Consequently, we used the second method for calculating the susceptibility at the critical point: $[m^2]_c(L) = [m^2]_{av2}(L)$. Figure 5 shows the dependence of $[m^2]_c(L)$ for concentrations $p = 1.0$ and $p = 0.85$. The numerical values of the effective exponents are identical to those presented by other authors (compare, for example, [10]).

Thus, the results of this paper can be divided into two parts.

The first part is methodological. We developed and successfully applied a new method for analyzing the critical behavior. This method makes it possible to extract simultaneously from the numerical data the values of the critical temperature, the critical exponents, and the amplitudes. Our determination of the "critical" temperature of a finite-size sample has a definite meaning. Specifically, this is the temperature which determines the equality of the critical exponents of susceptibility in the low- and high-temperature regions. It is obvious that such a conditional determination of the critical temperature is consistent with the well-known fact that a phase transition is impossible in a finite- size sample.

The second part of the results is associated with the investigation of the critical behavior of the Ising model with nonmagnetic impurities. The basic result of this paper is a numerical determination of the fact that the ratio of the critical amplitudes of the susceptibility is independent for concentrations of nonmagnetic impurities q ranging from 0 to 0.25. Hence it can be concluded that in the indicated range of impurities the critical behavior of our system once again lies in the Ising universality class. The variation of the effective critical exponent of the susceptibility, which we observed and confirmed in our investigation, can be attributed to the impurity concentration dependence of corrections, which are unknown to us, to the scaling. At least, this is so for a sufficiently low concentration of impurities, when the Dotsenko–Dotsenko–Shalaev–Shankar–Lüdwig theory is applicable [5, 6]. Specifically, when the known logarithmic behavior is included in the fitting procedure, the critical exponent γ remains equal to the "pure" value 7/4. For large values of the concentrations, a fit of this kind

becomes unstable. This indicates the appearance of new correction terms of a form which is unknown to us.

A very attractive feature of our method is that, as we found, the ratio of the critical amplitudes of the magnetic susceptibility can be calculated to a higher degree of accuracy than, for example, the other ratio of the critical amplitudes: Binder's cumulants [27]. It is of interest to determine the ratio of the critical amplitudes near the percolation point of the lattice. In so doing, it could be possible to investigate a transition from the Ising universality class to the percolation class.

ACKNOWLEDGMENTS

We are grateful to B. Berche, K. Binder, H.W.J. Blöte, P. Butera, B. Derrida, J. Heringa, W. Janke, J.-K. Kim, D.P. Landau, E. Luitjen, M. Novotny, W. Selke, D. Stauffer, R.B. Stinchcombe, M. Weigel, and S. Wiseman for numerous helpful discussions. We thank the NWO Fund (Netherlands) and the Russian Foundation for Basic Research for partial support. O.A.V. thanks the L.D. Landau Grant Committee (Forshungzentrum/KFA Jülich) for support.

REFERENCES

1. R. B. Stinchcombe, in *Phase Transitions and Critical Phenomena*, Ed. by C. Domb and J. L. Lebowitz (Academic, New York, 1983), Vol. 7.
2. W. Selke, L. N. Shchur, and A. L. Talapov, in *Annual Reviews of Computational Physics*, Ed. by D. Stauffer (World Scientific, Singapore, 1995), Vol. 1.
3. K. Binder and W. Heermann, *Monte Carlo Simulation in Statistical Physics* (Springer-Verlag, Berlin, 1998).
4. A. B. Harris, J. Phys. C **7**, 1671 (1974).
5. Vik. Dotsenko and Vl. Dotsenko, Adv. Phys. **32**, 129 (1983).
6. B. N. Shalaev, Phys. Rep. **237**, 129 (1994).
7. H.-O. Heuer, Phys. Rev. B **45**, 5691 (1992).
8. H. G. Ballesteros, L. A. Fernández, V. Martín-Mayor, *et al.*, J. Phys. A **30**, 8379 (1997).
9. R. Kühn, cond-mat/9405047.
10. J.-K. Kim and A. Patrascioiu, Phys. Rev. B **49**, 15764 (1994); Phys. Rev. Lett. **72**, 2785 (1994).
11. S. L. A. de Queiroz and R. B. Stinchcombe, Phys. Rev. B **50**, 9976 (1994).
12. V. Privman, P. C. Hohenberg, and A. Aharony, in *Phase Transitions and Critical Phenomena*, Ed. by C. Domb and J. L. Lebowitz (Academic, New York, 1991), Vol. 14.
13. D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor and Francis, London, 1992).
14. R. M. Ziff, Phys. Rev. Lett. **69**, 2670 (1992).
15. S. Kirkpatrick and E. Stoll, J. Comput. Phys. **40**, 517 (1981).
16. J. R. Heringa, H. W. J. Blöte, and A. Compagner, Int. J. Mod. Phys. C **3**, 561 (1992).
17. L. N. Shchur and H. W. J. Blöte, Phys. Rev. E **55**, R4905 (1997).
18. L. N. Shchur, H. W. J. Blöte, and J. R. Heringa, Physica A (Amsterdam) **241**, 579 (1997).
19. J. Hoshen and R. Kopelman, Phys. Rev. B **14**, 3438 (1976).
20. R. H. Swedsen and J. S. Wang, Phys. Rev. Lett. **58**, 86 (1987).
21. U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).
22. R. G. Edwards and A. D. Sokal, Phys. Rev. D **38**, 2009 (1988).
23. M. E. Fisher and A. E. Ferdinand, Phys. Rev. B **19**, 169 (1967).
24. T. T. Wu, B. M. McCoy, C. A. Tracy, and E. Barouch, Phys. Rev. B **13**, 316 (1976).
25. D. P. Landau, Phys. Rev. B **13**, 2997 (1976).
26. A. L. Talapov and L. N. Shchur, J. Phys.: Condens. Matter **6**, 8295 (1994).
27. B. Derrida, B. W. Southern, and D. Stauffer, J. Phys. (Paris) **48**, 335 (1987).
28. S. Wiseman and E. Domany, Phys. Rev. E **52**, 3469 (1995).
29. W. Selke, L. N. Shchur, and O. A. Vasilyev, Physica A (Amsterdam) **259**, 388 (1998).
30. O. A. Vasilyev, http://www.itp.ac.ru/preprints/vas000201.
31. L. N. Shchur, Comput. Phys. Commun. **121/122**, 83 (1999).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 974–978.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1122–1126.

SOLIDS
Structure

Effect of Phonon Damping on the Resonance Reflection of a Transverse Sound Wave from a Planar Defect in a Crystal

Yu. A. Kosevich*, E. S. Syrkin, D. A. Semagin**, and A. M. Kosevich**, *****

**Moscow State Technological University "STANKIN," Moscow, 101472 Russia*
and Max Planck Institute for the Physics of Complex Systems, Nothnitzer Str. 38, Dresden, D-01187 Germany

***Verkin Physicotechnical Institute for Low Temperatures, National Academy of Sciences of Ukraine, Kharkov, 61164 Ukraine*

****e-mail: kosevich@ilt.kharkov.ua*

Received February 1, 2000

Abstract—Resonance scattering of a transverse sound wave by a planar defect in an elastic isotropic medium is studied in a wide range of values of the ratio of the damping length and the size of the region of localization of longitudinal oscillations. The transition between two limiting cases is described. The character of the transition is demonstrated by typical plots of the dependence of the reflection and transmission coefficients on the parameter relating the wavelength of the incident wave and the strength of the defect. It is shown that renormalization of total reflection into conventional dissipative passage occurs for values of this parameter below a certain critical value. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In the last few years peculiarities of the resonance reflection of particles or waves by defects has been found in scattering theory for cases where the dispersion law possesses more than one branch of stationary states or, in other words, the Green's function of the corresponding equations has at least two components. Examples are the resonance interaction of the transverse and longitudinal components of an acoustic wave at a planar defect in the theory of elasticity [1–4] or in the simplest vector model of a crystal lattice [5], the scattering of waves described by a scalar equation taking into account the highest-order dispersion (i.e., including fourth-order spatial derivatives) [6], two-channel scattering by a point defect [7], and finally, oscillations in waveguide systems, where transverse quantization gives rise to a system of branches in the dispersion law for one-dimensional motion [8]. In all of these examples the same particle energy (or wave frequency) can correspond to two stationary states of a different nature: one state being a wave propagating in all space (or only in one half-space) and the other state being localized at a defect. In the presence of a defect the independent modes of an ideal system interact, and the traveling wave can undergo resonance scattering by a localized oscillation similar to scattering by an internal dynamic mode of a defect.

One feature of resonance reflection, which will be analyzed below, is the predicted total reflection of a wave by a passive defect (a defect with no internal dynamical degrees of freedom) for a certain value of the wave frequency. Specifically, scattering of a transverse sound wave by a planar defect in an elastic isotropic medium is studied. It is assumed that the phase velocity c of the wave along the defect falls between the transverse c_t and longitudinal c_l sound velocities in the bulk. It has been shown [1] that in the limit $c_l - c \ll c_l$ the reflection coefficient for such a wave is exactly 1 along the curve

$$\frac{c_l - c}{c_l} = \text{const}(kH)^2, \quad kH \ll 1 \qquad (1)$$

in the (c, k) plane, where H is the effective thickness of the defect layer, defined taking into account the strength η of the defect: $H = \eta h$, where h is the thickness of the defect layer. A relation determining this curve in a wide range of values of c and k has been found in [2]. It is interesting that although the values of c and k for which the reflection coefficient is 1 and hence the total reflection curve depends on the strength and thickness of the defect, the fact of total reflection itself remains unchanged for arbitrarily small thicknesses of the defect layer.

In [4] it is asserted and confirmed by a calculation that total reflection of a wave by a planar defect with infinitesimal thickness is unphysical or, at least, unstable with respect to the processes that are ordinarily neglected in an idealized theory. One such process is any weak damping of sound waves, which is usually neglected in the theory of elasticity.

The basic idea is as follows. Let τ be the relaxation time determining the damping of a sound wave. It is related with the viscosity ν by the well-known relation $\tau^{-1} = \nu(\omega/c)^2$, where ω is the wave frequency. The

damping of a sound wave is assumed to be weak, the criterion for which is $\omega\tau \gg 1$. To first order in the small parameter $(\omega\tau)^{-1}$ the velocity dispersion is determined by the relation

$$s_l = c_l\left(1 - \frac{i}{2\omega\tau}\right). \quad (2)$$

Consequently, the damping length l_0 of a sound wave is approximately

$$l_0 \sim c\tau = c^3/\nu\omega^2. \quad (3)$$

As noted above, the resonance interaction of transverse sound with a planar defect is due to the interaction of an incident transverse wave with localized longitudinal oscillation. The resonance properties of such an interaction can be manifested in full measure only if the damping length l_0 is much greater than the size of the region of localization of longitudinal oscillations:

$$l_0 \gg 1/\kappa_l, \quad (4)$$

where κ_l is a constant that determines the spatial exponential decay of the amplitude of longitudinal oscillations. By definition

$$\kappa_l = k\sqrt{1 - (c/c_l)^2}. \quad (5)$$

Consequently, using the estimate (1), which was obtained neglecting damping, the condition (4) assumes the form

$$\eta hk \gg \frac{1}{c\tau k} = \frac{\nu\omega}{c^2}. \quad (6)$$

It follows from equation (6) that for $\eta hk \ll 1$ the possibility of describing the resonance properties of a planar defect neglecting damping of sound is limited, at least, to the region (6). It will be shown below that a systematic calculation will give more stringent limitations than the stronger inequality (6), so that the latter can be interpreted as a necessary condition for neglecting dissipative effects when analyzing resonance scattering.

In the opposite limit there should be no resonance scattering. Systematic analytic calculations have confirmed that resonance vanishes in the limit $kH \longrightarrow 0$ [4].

In the present work resonance scattering by a planar defect is studied in a wide range of values of the parameter $l_0\kappa_l$ and the transition between different limiting cases is described. The character of the transition is demonstrated by typical plots of the dependence of the reflection and transmission coefficients on the parameter Hk.

2. RESONANCE SCATTERING WITH DISSIPATION

We shall use the simple model proposed in [3] for a planar defect to study the qualitative physical phenomena associated with the problem under discussion. In [3] the problem of the scattering of a transverse Rayleigh-polarized sound wave

$$\mathbf{u}(x, z) = (u_x, 0, u_z) = \mathbf{u}(z)\exp(ikx - i\omega t)$$

by a flat isotope-defect ($Z = 0$ plane) in an isotropic medium was solved. This model is a particular case of the model examined in [4], if in the latter the coefficients associated with capillary effects (the surface tension and the surface elastic moduli),[1] are set equal to 0 and only the mass defect is taken into account. Such a defect and, hence, the boundary conditions do not fundamentally change the character of the phenomena. The boundary conditions presuppose that the elastic displacements are continuous at the plane of the defect and the jump in the normal components of the stress tensor, which is determined by the defect strength parameter[2] $\eta = (M - m)/m$, where m is the mass of the atoms of the medium and M is the mass of the isotopes in the defect layer.

The solution $\mathbf{u}(z)$ is assume to be a sum of a volume transverse wave and a longitudinal wave localized at the defect. Application of the boundary conditions yields expressions for the reflection coefficient $R = |A|^2$ and transmission coefficient $T = |B|^2$, where A and B are, respectively, the amplitudes of the reflected and transmitted waves [3]:

$$A = \frac{\eta h}{\Delta} \times [(2\kappa_0 c_l^2 - \eta h\omega^2)(k^2 - q^2) - \eta hk^2 c_l^2(\kappa_0^2 + q^2)], \quad (7)$$

$$B = \frac{2iq}{\Delta}(2\kappa_0 c_l^2 - \eta h\omega^2), \quad (8)$$

$$\Delta = \eta h\left[(2\kappa_0 c_l^2 - \eta h\omega^2)\frac{\omega^2}{c_t^2} + \eta hk^2 c_l^2(q^2 - \kappa_0^2)\right] + 2iq[(2\kappa_0 c_l^2 - \eta h\omega^2) - (\eta hk)^2\kappa_0 c_l^2], \quad (9)$$

where $\kappa_0 = \kappa_l$ (see equation (5)) and $q = k\sqrt{(c/c_t)^2 - 1}$ is the z component of the wave vector of the transverse wave.

Analysis of these expressions permits writing the obvious conditions for resonance scattering.

The condition for total reflection ($R = 1$, $T = 0$) is

$$2\kappa_0 c_l^2 = \eta h\omega^2, \quad (10)$$

and the condition of total transmission ($R = 0$, $T = 1$) is

$$2\kappa_0 c_l^2 = \eta h\omega^2\,\frac{k^2(c_l^2 + c_t^2) - \omega^2}{2k^2 c_t^2 - \omega^2}. \quad (11)$$

[1] The coefficients g_1 and h_{11} in equations (6) and (10) in [4].
[2] In [3] this parameter was determined with the opposite sign.

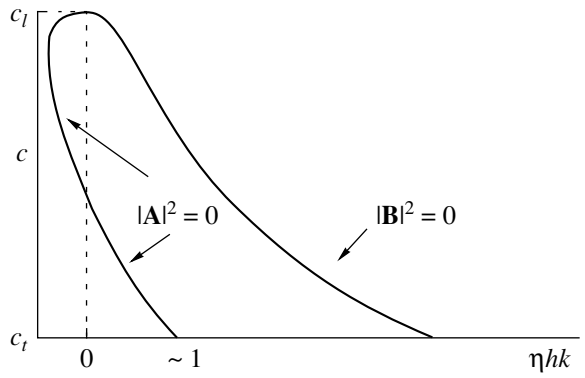


Fig. 1. Curves of the resonance dispersion laws (10) and (11).

The curves of the resonance scattering frequencies (10) and (11), neglecting damping, in the (k, c) plane are displayed in Fig. 1.

Introducing the viscosity into the dynamical equation by taking viscosity into account in the dispersion of the sound velocity (2) does not change the form of the equation itself. This means that to discuss the effect of damping on the effects under consideration it is sufficient to replace in equations (5), (10), and (11) c_l by s_l (and κ_0 by κ).[3]

For convenience, we now introduce the notation $\xi = \nu\omega/2c_l^2$, which by virtue of equation (4) presupposes that $\xi \ll 1$. Then, we obtain the following expression for κ [4] in the leading approximation in ξ:

$$\kappa^2 = \kappa_0^2 - 2i\xi(\omega/c_l)^2.$$

The old formulas (10) and (11) can be used to obtain an expression for κ_0 near resonance scattering and transmission. Consequently, under resonance conditions where $c \longrightarrow c_l$ ($\eta hk \ll 1$) the formula for κ^2 acquires the form

$$\kappa^2 = k^2[\gamma^2(\eta hk)^2 - 2i\xi], \tag{12}$$

where $\gamma = 1/2$ for resonance reflection and $\gamma = (c_l^2 + c_t^2 - c^2)/2(2c_t^2 - c^2)$ for resonance transmission. Since γ is of the order of 1, sound damping in the relation (12) can be neglected only if

$$(\eta hk)^2 \gg \xi, \tag{13}$$

which is a stronger condition than equation (6) (see [4]).

The condition (13) includes almost the entire velocity range under study, $c_t < c < c_l$, except for small neighborhoods near the limits, precisely where the damping is observed.

[3] Since (see the Introduction) the effects under consideration appear in the limit $c \longrightarrow c_l$, the dispersion of the velocity of transverse oscillations can be neglected.

Substituting the expressions (2) and (12) into the expressions (7) and (8) we obtain the coefficients R and T as functions of the phase velocity c and the strength of the defect (the parameter ηhk) on the curves of the frequencies of the "old" resonances (10) and (11):

$$R = \frac{(\eta hk)^2}{D}\left[\left\{(2-w^2)\left(\sqrt{\lambda+1} - \frac{1}{\sqrt{2}\gamma}\right) - \frac{1-\sigma}{\sqrt{2}\gamma\sigma}\right\}^2 + \left\{(2-w^2)\sqrt{\lambda-1} - \frac{\sqrt{2}\xi}{\gamma}\right\}^2\right], \tag{14}$$

$$T = \frac{4(w^2-1)}{D}\left[\left(\sqrt{\lambda+1} - \frac{1}{\sqrt{2}\gamma}\right)^2 + \lambda - 1\right], \tag{15}$$

$$D = \left[\eta hk\left\{w^2\left(\sqrt{\lambda+1} - \frac{1}{\sqrt{2}\gamma}\right) + \frac{1}{\sqrt{2}\gamma}\left(\frac{1+\sigma}{\sigma} - \frac{2}{w^2\sigma}\right)\right\} + \sqrt{(w^2-1)(\lambda-1)}(2-(\eta hk)^2)\right]^2 + \left[\eta hk\left(\frac{\sqrt{2}\xi}{\gamma} - w^2\sqrt{\lambda-1}\right) + \sqrt{(w^2-1)} \times \left(\sqrt{\lambda+1}(2-(\eta hk)^2) - \frac{\sqrt{2}}{\gamma}\right)\right]^2, \tag{16}$$

where $w = c/c_t$, $\sigma = c_t^2/c_l^2$, and $\lambda = \sqrt{1 + (2\xi\sigma w^2/(1-\sigma w^2))^2}$.

We considered it useful to present the relations (14)–(16) here even though they are extremely complicated. They make it possible to calculate the scattering parameters of a wave with any phase velocity in the range studied (i.e., for arbitrary γ) for a defect of arbitrary strength in a slightly viscous medium.

Analysis of equations (14) and (15) shows that, as noted in [4], damping at frequencies of total transmission is weak and conditions stronger than the inequality (6) are not needed to preserve total transmission. Conversely, at the frequencies of total reflection (see Fig. 2) substantial renormalization of the effect occurs at a definite frequency (or for defects with a definite strength), and total reflection gradually vanishes.

The process leading to this qualitative transformation is examined in detail below.

3. APPROXIMATE DESCRIPTION OF THE SCATTERING COEFFICIENTS IN LIMITING CASES

The degree to which damping influences resonance, i.e., the form of the expressions for R and T, is largely determined by the ratio of the real and imaginary parts

in equation (12). The final results are different in the following ranges of values of the parameter ηhk.

1) $(\eta hk)^2 \gg \xi$, the case where the influence of damping is extremely weak (13) (region *a* in Figs. 2 and 3).

In this case, for the old curve of the frequencies of total reflection ($\gamma = 1/2$ in equation (12)) the equality

$$\kappa = k\left[\frac{\eta hk}{2} - \frac{2i\xi}{\eta hk}\right] \tag{17}$$

can be used to a first approximation, and the required substitutions yield approximate expressions for the resonance coefficients (14) and (15), which, naturally, differ from $R_r = 1$ and $T_r = 0$:

$$R_r = \frac{(\eta hk)^6(w^2-1)}{\{(\eta hk)^3\sqrt{w^2-1} + 8\xi\}^2}, \tag{18}$$

$$T_r = \frac{64\xi^2}{\{(\eta hk)^3\sqrt{w^2-1} + 8\xi\}^2}. \tag{19}$$

The formulas (18) and (19) are completely identical to the analogous formulas in [4], provided that the analysis is limited only to an isotope-defect layer.

The fraction of the absorbed sound energy (neglecting other losses) is obtained from equations (18) and (19):

$$E_d = 1 - R_r - T_r = \frac{16\xi(\eta hk)^3\sqrt{w^2-1}}{\{(\eta hk)^3\sqrt{w^2-1} + 8\xi\}^2}. \tag{20}$$

It is evident that for $(\eta hk)^3 = 8\xi/(w^2-1)^{1/2}$ an inflection appears (the region *b* in Figs. 2 and 3): half the energy ($E_d = 1/2$) arriving at the defect is absorbed, and the coefficients R_r and T_r become $R_r = T_r = 1/4$ (Fig. 2). In [4] this phenomenon is called anomalous absorption.

Thus, the form of the denominators in equations (18) and (19) determines a more stringent condition than the condition (13) for the a description neglecting damping (equations (7) and (8)) to be valid, specifically,

$$(\eta hk)^3 \gg \xi, \tag{21}$$

which, as noted in [4], is quite difficult to produce using real materials.

Gradually weakening, the inequality (13) (and, of course, the inequality (21)), i.e., moving along the frequency curves toward longer wavelengths and weaker defects, we obtain the condition $(\eta hk)^2 \sim \xi \ll 1$ and, consequently,

2) the region of strong damping (section *c* in Figs. 2 and 3), where

$$(\eta hk)^2 \ll \xi \ll 1. \tag{22}$$

In this case, a simplification of equation (12) leads to the relation

$$\kappa = k(1-i)\sqrt{\xi}, \tag{23}$$

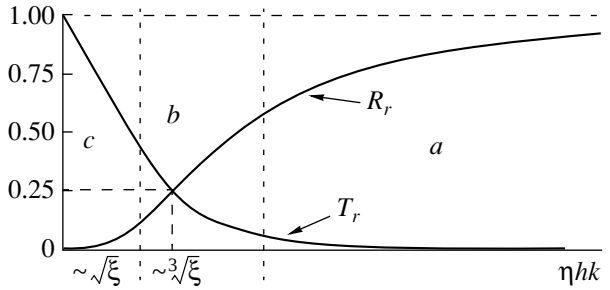


Fig. 2. Reflection *R* and transmission *T* coefficients versus the parameter ηhk at total-reflection frequencies (10).

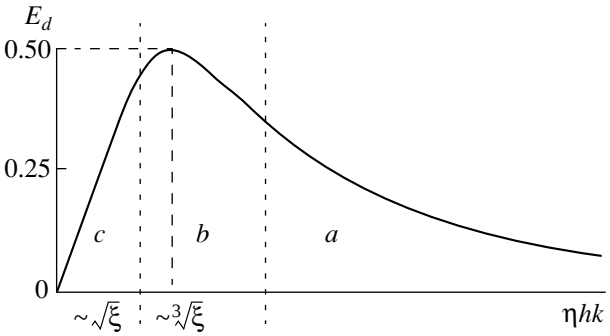


Fig. 3. Dissipation losses versus the parameter ηhk at total-reflection frequencies (10).

which makes it possible to simplify the expressions (14) and (15) for the resonance parameters (for $w \sim 1$):

$$R = \frac{(\eta hk)^2(2-w^2)^2}{4(w^2-1)} \ll 1, \tag{24}$$

$$T = 1 - \frac{(\eta hk)^2 w^4}{4(w^2-1)}. \tag{25}$$

The total dissipation in this case is negligible:

$$E_d = (\eta hk)^2 \ll 1. \tag{26}$$

It is interesting that equations (24), (25), and (26) do not explicitly contain the viscosity coefficient. The viscosity influences the result indirectly via the condition (22).

In the indicated wavelength range, determined by the requirement $(\eta hk)^2 \ll \xi \ll 1$, the interaction of a wave with a defect is not a resonance interaction, and the strength of the interaction is proportional to $(\eta hk)^2$ at any frequency. It follows from equations (24) and (25) that a defect is virtually transparent for such waves.

In summary, our analysis shows that the resonance interaction of acoustic waves with planar defects is extremely sensitive to dissipative losses. Total resonance reflection, which is obtained in a calculation neglecting dissipation processes, is replaced by the transmission of

almost all of the elastic wave through a thin planar defect. The indicated role of dissipation processes agrees with the theoretical and experimental investigations of the Kapitsa jump [9–11], which demonstrate that acoustic absorption increases the transmission coefficient for phonons through a boundary in an elastic medium.

ACKNOWLEDGMENTS

This work was supported in part by INTAS (grant no. 167, 1999).

REFERENCES

1. A. N. Darinskiĭ and G. A. Maugin, Wave Motion **23**, 363 (1996).
2. A. M. Kosevich and A. V. Tutov, Phys. Lett. A **248**, 271 (1998).
3. A. Kosevich and D. Semagin, Cond. Matter Phys. (Lviv) **1**, 409 (1998).
4. Yu. A. Kosevich and E. S. Syrkin, Phys. Lett. A **251**, 378 (1999).
5. A. M. Kosevich, D. V. Matsokin, and S. E. Savotchenko, Fiz. Nizk. Temp. **25**, 63 (1999) [Low Temp. Phys. **25**, 48 (1999)].
6. A. M. Kosevich and S. E. Savotchenko, Fiz. Nizk. Temp. **25**, 737 (1999) [Low Temp. Phys. **25**, 550 (1999)].
7. A. M. Kosevich, Zh. Éksp. Teor. Fiz. **115**, 316 (1999) [JETP **88**, 168 (1999)].
8. C. S. Kim and A. M. Satanin, Zh. Éksp. Teor. Fiz. **115**, 211 (1999) [JETP **88**, 118 (1999)].
9. R. E. Peterson and A. C. Anderson, J. Low Temp. Phys. **11**, 639 (1973).
10. A. F. Andreev, Zh. Éksp. Teor. Fiz. **43**, 358 (1962) [Sov. Phys. JETP **16**, 257 (1963)].
11. K. N. Zinov'eva, Fiz. Nizk. Temp. **23**, 485 (1997) [Low Temp. Phys. **23**, 355 (1997)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 979–986.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1127–1135.

SOLIDS
Structure

Magnetic Phase Diagram of the Intermetallic Compounds $Gd_{1-x}La_xMn_2Ge_2$ and the Effect of a Field on Transitions of the Mn Subsystem from the Antiferromagnetic into the Ferromagnetic State

Guo Guanghua*, R. Z. Levitin*, **, V. V. Snegirev*, D. A. Filippov*, and A. Yu. Sokolov***

**Moscow State University, Moscow, 119899 Russia*
***e-mail: levitin@plms.phys.msu.su*
****Moscow Institute of Radio Engineering, Electronics, and Automatics, Moscow, 117454 Russia*

Received December 17, 1999

Abstract—The magnetic properties of the intermetallic compounds $Gd_{1-x}La_xMn_2Ge_2$ (layered tetragonal crystal structure of the type $ThCr_2Si_2$), investigated in the present work and in [1] A. Sokolov, *et al.*, Sol. State Commun. **105**, 289 (1998), are discussed. It is shown that the basic characteristics of the magnetic ordering of these compounds—magnetic phase transitions from the ferrimagnetic into the antiferromagnetic state and vice versa—observed with increasing temperature in compositions with $x < 0.1$, the effect of a magnetic field on these transitions, the temperature dependences of the magnetization of these intermetallic compounds, as well as their magnetic T–x diagram can be described quantitatively in the Yafet–Kittel model for ferrimagnets with a negative exchange interaction in one of the sublattices using parameters determined in [2] A. Yu. Sokolov *et al.*, JETP **89**, 723 (1999) from investigations of the system $Gd_{1-x}Y_xMn_2Ge_2$, taking into account the dependence of the interplanar exchange interaction Mn–Mn on the crystal lattice parameter a. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The magnetic properties of intermetallic rare-earth compounds RMn_2Ge_2 (tetragonal layered crystal structure of the type $ThCr_2Si_2$, space group $I4/mmm$) are determined by two interacting magnetic subsystems—rare-earth and manganese, which are formed by alternating layers which are perpendicular to the tetragonal axis [3]. Numerous investigations (see, for example, the review [3]) have shown that the positive Mn–Mn exchange interaction in a layer is the strongest interaction. The Mn–Mn and R–Mn interplanar exchange interactions are an order of magnitude weaker. The exchange between the rare earths is another order of magnitude weaker. An important feature is that the interplanar Mn–Mn exchange interaction depends strongly on the distance d between the manganese atoms in a layer (in other words, on the lattice parameter a) and changes sign from negative to positive as the parameter a increases to a critical value a_{cr}. Since the parameters of the crystal structure of the intermetallic compounds RMn_2Ge_2 decrease with increasing number of the rare-earth element because of the lanthanide compression effect, the intrinsic magnetic ordering of the manganese subsystem is ferromagnetic in most of these intermetallic compounds with light rare earths, since for them $a > a_{cr}$, and antiferromagnetic in the intermetallic compounds with heavy rare earths, since in these compounds $a < a_{cr}$. Since exchange between the heavy rare earth and manganese is also negative, intermetallic compounds with heavy rare earths can be treated as two-sublattice ferrimagnets with a negative exchange interaction in the manganese sublattice.

The effect of the interatomic distances on the magnetic properties of intermetallic compounds of the type RMn_2Ge_2 can be clearly seen by comparing the magnetic properties of the system $Gd_{1-x}Y_xMn_2Ge_2$, which we studied in [2], with those of the mixed intermetallic compounds $Gd_{1-x}La_xMn_2Ge_2$, for which a magnetic T–x diagram was constructed in [1] with small ($x < 0.1$) substitutions of lanthanum for gadolinium. Even though in both systems the magnetic gadolinium is replaced by a nonmagnetic element (yttrium or lanthanum), the magnetic properties of the mixed compounds are strongly different.

This seems to be explained by the different dependence of the crystal structure parameter a with substitution of yttrium or lanthanum for gadolinium. Since the atomic radius of yttrium is close to that of gadolinium, in the intermetallic compounds $Gd_{1-x}Y_xMn_2Ge_2$, to a first approximation, the dependence of the interplanar Mn–Mn exchange interaction on the yttrium concentration x can be neglected. At the same time, in the system $Gd_{1-x}La_xMn_2Ge_2$ this effect is very strong, since the atomic radius of lanthanum is much larger than that of gadolinium, and when lanthanum is substituted for gado-

1063-7761/00/9006-0979$20.00

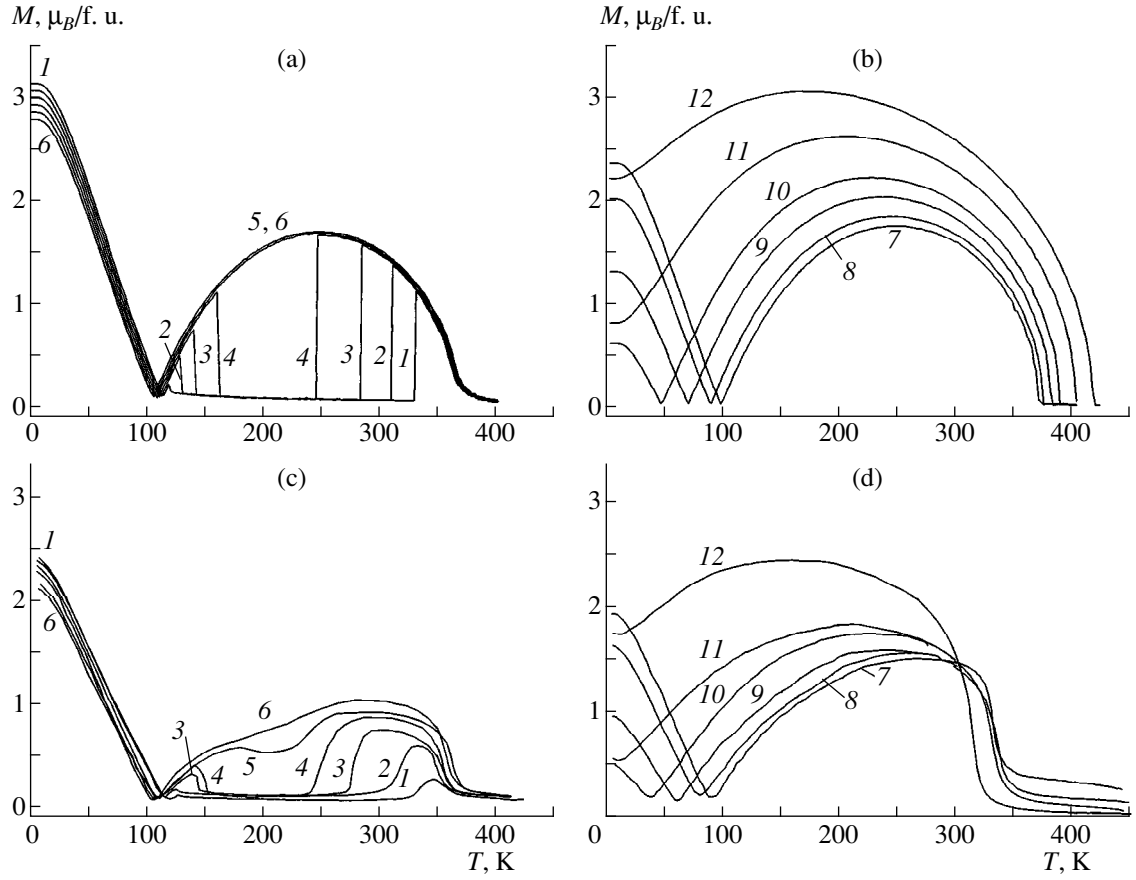


Fig. 1. (a, b) Theoretical and (c, d) experimental temperature dependences of the magnetization of the intermetallic compounds $Gd_{1-x}La_xMn_2Ge_2$ for x = (*1*) 0.04, (*2*) 0.05, (*3*) 0.06, (*4*) 0.07, (*5*) 0.08, (*6*) 0.09, (*7*) 0.15, (*8*) 0.2, (*9*) 0.3, (*10*) 0.4, (*11*) 0.6, and (*12*) 0.8.

linium the sign of the interplanar Mn–Mn exchange interaction changes.

In [2] we showed that the spontaneous and field-induced magnetic phase transitions and other low-temperature magnetic properties of the intermetallic compounds $Gd_{1-x}Y_xMn_2Ge_2$ can be adequately explained by the Yafet–Kittel model [4], modified so as to take account of magnetic anisotropy, for a two-sublattice ferrimagnet with negative exchange in one of the sublattices. In the Yafet–Kittel model it is assumed that this sublattice is divided into two sublattices, whose magnetic moments, depending on the effective field acting on this sublattice, are oriented either parallel or at an angle or antiparallel to one another (see [2] for a detailed description of the possible magnetic structures and magnetic phase transitions in the Yafet–Kittel model). The values of the parameters describing the exchange interactions and the magnetic anisotropy of the compounds $Gd_{1-x}Y_xMn_2Ge_2$ were determined from the experimental data obtained.

The objective of the present work was to determine how the experimental data, obtained in [1], for the system of mixed intermetallide $Gd_{1-x}La_xMn_2Ge_2$ as well as certain experimental results of the present work can be described in the Yafet–Kittel model with parameters determined from investigations of the intermetallic compounds $Gd_{1-x}Y_xMn_2Ge_2$ (taking account of the above-noted dependence of the interlayer Mn–Mn exchange interaction on the lattice parameter a).

2. SAMPLES AND MEASUREMENT PROCEDURE

In the present work polycrystalline samples of the system $Gd_{1-x}La_xMn_2Ge_2$, on some of which the measurements were performed in [1], were used. They were prepared for initial components on a cold hearth in an arc furnace. The samples were homogenized in the course of a week at temperature 800°C in a dynamic vacuum. X-ray diffraction showed that the samples had a single-phase structure.

The magnetic susceptibility of the samples in weak (down to 10^{-3} T) ac magnetic fields was measured. The magnetization of the samples was investigated with a vibrating sample magnetometer in fields up to 0.8 T and in pulsed magnetic fields up to 25 T by the induction method.

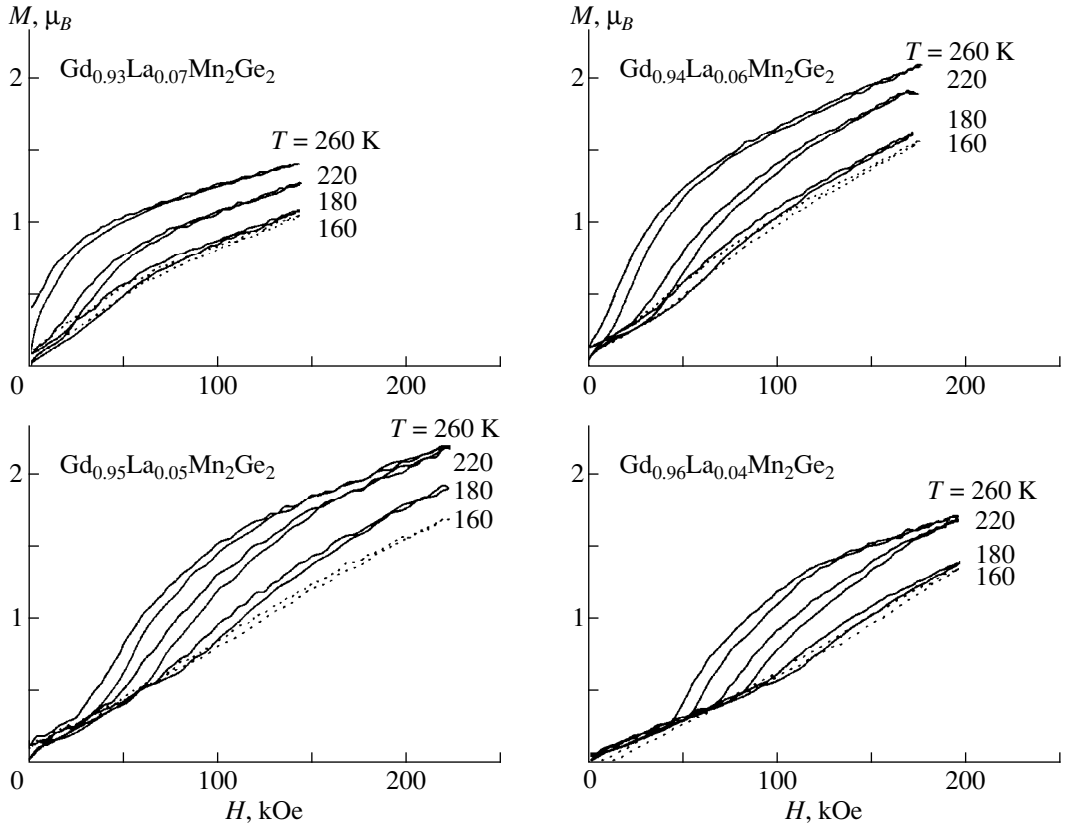


Fig. 2. Magnetization curves of the intermetallic compounds $Gd_{1-x}La_xMn_2Ge_2$.

3. EXPERIMENTAL RESULTS

Figure 1 displays the temperature dependences of the magnetization of the intermetallic compounds $Gd_{1-x}La_xMn_2Ge_2$ in a 0.8 T field (the data for $x < 0.1$ were taken from [1]). As one can see from the figure, the temperature dependences of the magnetization of the intermetallic compounds with lanthanum content greater than $x = 0.1$ are characteristic for ferrimagnets: the magnetic compensation temperature is observed in most of them, and in compositions with lanthanum content $x = 0.6$ and 0.8 the low- temperature magnetization increases with temperature. A different behavior was observed in [1] in compositions with low quantities lanthanum ($x < 0.1$). At low temperatures they are also ferrimagnets, but their magnetization decreases sharply as temperature increases: a first-order transition into an antiferromagnetic state is observed [1]. This transition is similar to the one repeatedly discussed in previous work for pure $GdMn_2Ge_2$ (see, for example, the review [3]) and attributed to a transition of the manganese subsystem from the ferromagnetic into the antiferromagnetic state and magnetic disordering of the gadolinium subsystem. The temperature of this transition increases with the lanthanum concentration. In compositions with $x > 0.04$ a sharp increase of magnetization is observed with a further increase in temperature: a reentrant first-order transition into the ferrimagnetic state occurs as a result of a sign change of the interlayer Mn–Mn exchange interaction [1]. The temperature of this transition decreases as x increases, so that the antiferromagnetic phase does not arise in compositions with $x > 0.08$.

Figure 2 shows the magnetization curves of the intermetallic compounds $Gd_{1-x}La_xMn_2Ge_2$. It is evident that in this temperature range, where according to [1] the samples are antiferromagnetic, there is no spontaneous magnetization and a metamagnetic transition is observed in a field. A similar transition was found previously in the single crystal $GdMn_2Ge_2$ with the field oriented along the tetragonal axis [5] and it was interpreted as a transition of the manganese subsystem from the antiferromagnetic into the ferromagnetic state.

The temperature dependences of the fields of the transition from the antiferromagnetic into the ferromagnetic state are presented in Fig. 3. It is evident that the field of the metamagnetic transition for compositions with $x = 0.04$ and 0.05 increases monotonically with decreasing temperature, and in compositions with high lanthanum concentration ($x = 0.06$ and 0.07) the magnitude of this field passes through a maximum at a certain temperature.

The complete magnetic T–x phase diagram of the system $Gd_{1-x}La_xMn_2Ge_2$ is constructed in Fig. 4. A part of the magnetic phase diagram for low lanthanum concentrations ($x < 0.1$), which was obtained in [1] in a 0.8 T field, is presented in the inset.

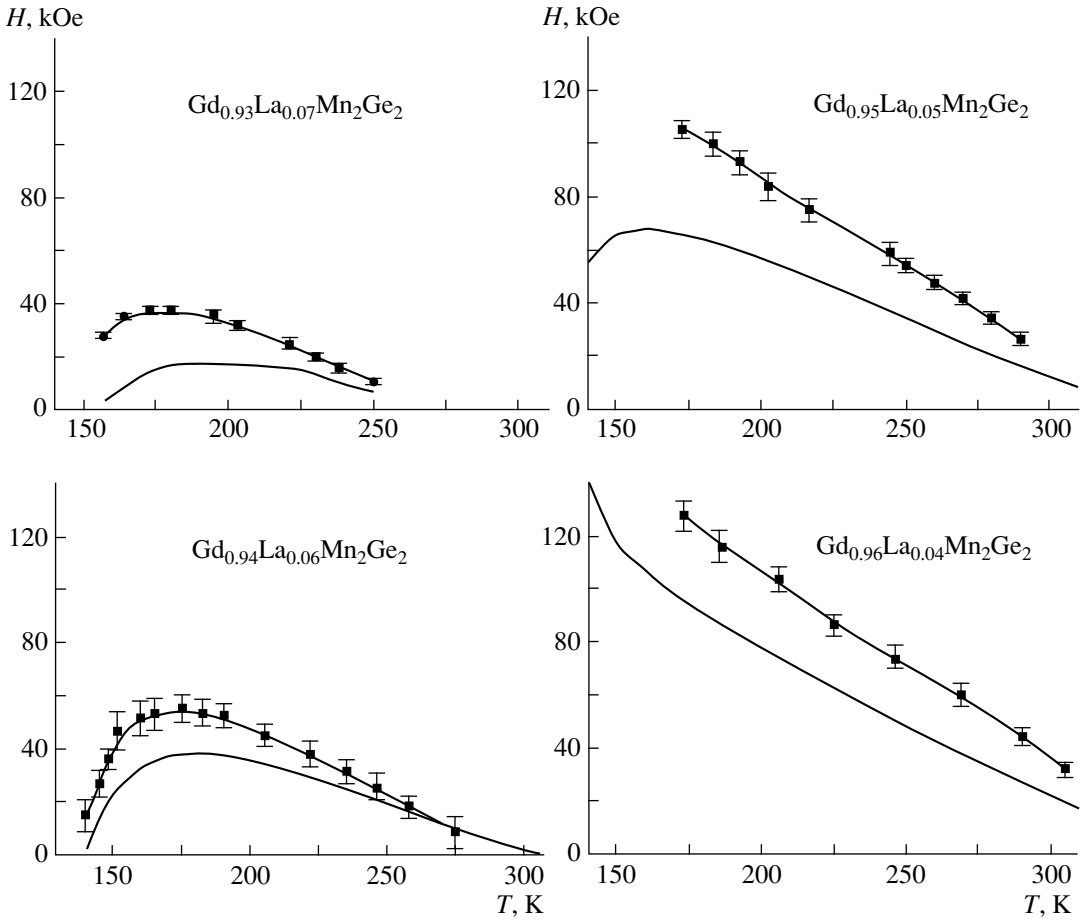


Fig. 3. Temperature dependences of the fields of the transitions from the antiferromagnetic into the ferromagnetic state in the intermetallic compounds $Gd_{1-x}La_xMn_2Ge_2$. The dots represent the experimental data; the vertical bars describe the magnitude of the hysteresis at the transition. The solid lines are the theoretically computed curves.

4. DISCUSSION OF THE EXPERIMENTAL RESULTS

Comparing the magnetic T–x phase diagram of the system $Gd_{1-x}La_xMn_2Ge_2$ with the phase diagram of the system $Gd_{1-x}Y_xMn_2Ge_2$, presented in Fig. 3 of [2], shows that they are different. In the system $Gd_{1-x}La_xMn_2Ge_2$, a ferrimagnetic state occurs for all x at low temperatures, whereas in the system with yttrium the transition from the ferrimagnetic into the antiferromagnetic phase through triangular phases is observed as the yttrium concentration increases. As the temperature in the system with lanthanum increases with low lanthanum concentrations, transitions occur from the ferrimagnetic into antiferromagnetic state and vice versa, while compounds with a low yttrium content transform with increasing temperature from the ferrimagnetic into the triangular phase and then into the antiferromagnetic phase. As already mentioned above, these differences are due to the fact that in a system with yttrium the crystal lattice parameter a is essentially independent of the concentration x, and therefore it can be assumed, to a first approximation, that the interlayer Mn–Mn exchange interaction does not change when yttrium is substituted for gadolinium, while in a system with lanthanum the dependence of this exchange on the lanthanum concentration must be taken into account, since the lattice parameter a increases strongly with the lanthanum content.

We showed in [2] that the magnetic properties of the intermetallic compounds $Gd_{1-x}Y_xMn_2Ge_2$ can be adequately described in a Yafet–Kittel model, modified so as to take account of the magnetic anisotropy, for a two sublattice ferrimagnet with an antiferromagnetic exchange interaction in one of the sublattices. In what follows we shall consider the possibility of describing in this model the magnetic properties of the system $Gd_{1-x}La_xMn_2Ge_2$.

We shall view these intermetallic compounds as two-sublattice ferrimagnets, one sublattice of which is formed by the rare-earth atoms and the other by the manganese atoms.

In accordance with the Yafet–Kittel model, we shall assume that the manganese sublattice divides into two sublattices with moments $\mathbf{M}_2'$ and $\mathbf{M}_2''$ with the same

magnitude ($M_2' = M_2'' = M_2$), which can either be collinear to one another and to the magnetic moment $(1 - x)\mathbf{M}_1$ of the gadolinium subsystem or be oriented at an angle with respect to one another and to the magnetic moment of the gadolinium sublattice (triangular magnetic structure). In addition, depending on the magnitude of the magnetic anisotropy, the magnetic moments can be oriented differently relative to the *c* axis of the crystal.

In the Yafet–Kittel model the magnetic energy of the intermetallic compounds $Gd_{1-x}La_xMn_2Ge_2$ can be written in the form

$$E = -\lambda_{12}(1-x)\mathbf{M}_1(\mathbf{M}_2' + \mathbf{M}_2'') - \lambda_{22}'\mathbf{M}_2^2 - \lambda_{22}\mathbf{M}_2'\mathbf{M}_2'' - (1/2)\lambda_{11}(1-x)^2\mathbf{M}_1^2 + \kappa(M_{2c}'^2 + M_{2c}''^2) - \mathbf{H}[(1-x)\mathbf{M}_1 + \mathbf{M}_2' + \mathbf{M}_2'']. \quad (1)$$

Here the first term describes the exchange interaction of the gadolinium and manganese subsystems ($\lambda_{12} < 0$), the second term describes the Mn–Mn exchange in the layer ($\lambda_{22}' > 0$), the third term describes exchange between the layers of manganese ($\lambda_{22} < 0$ for $a < a_{cr}$, $\lambda_{22} > 0$ for $a > a_{cr}$), the fourth term describes the exchange interaction in the gadolinium subsystem, the fifth term described the magnetic anisotropy of the manganese subsystem (M_{2c} is the component of the magnetization of manganese along the tetragonal axis; $\kappa < 0$ is related with the uniaxial anisotropy constant *K* of the manganese subsystem as $K = -2\kappa M_2^2$), and the last term describes the Zeeman energy. We note that in equation (1), in contrast to the expression presented in [2], we have taken into account the Mn–Mn exchange interaction in a layer, since in the present work we are interested in the high-temperature properties; in this case the temperature dependence of the moment of manganese cannot be neglected.

As shown in [2], for $\lambda_{22} < 0$ four different magnetic phases are possible in our system in a zero external field.

In the first phase the magnetic moments of the two manganese sublattices are parallel to one another and antiparallel to the magnetic moment of the gadolinium sublattice, so that a collinear ferrimagnetic structure with orientation of the resulting magnetization parallel to the tetragonal axis (Fi phase) is formed. In the two other phases the magnetic moments of the manganese sublattices make angles with one another and with the magnetic moment of the gadolinium sublattice, so that triangular magnetic ordering arises. These phases (in [2] they are designated as *T* and *T*') differ by the orientation of the magnetic moment of the gadolinium sublattice relative to the crystallographic axes: they are parallel to the *c* axis of the crystal in the first phase and perpendicular in the second phase. Finally, the fourth phase (AF) corresponds to antiferromagnetic ordering in the manganese subsystem. The energies of these phases depend on the magnitudes of the exchange interactions of the subsystems and the magnetic anisotropy energy.

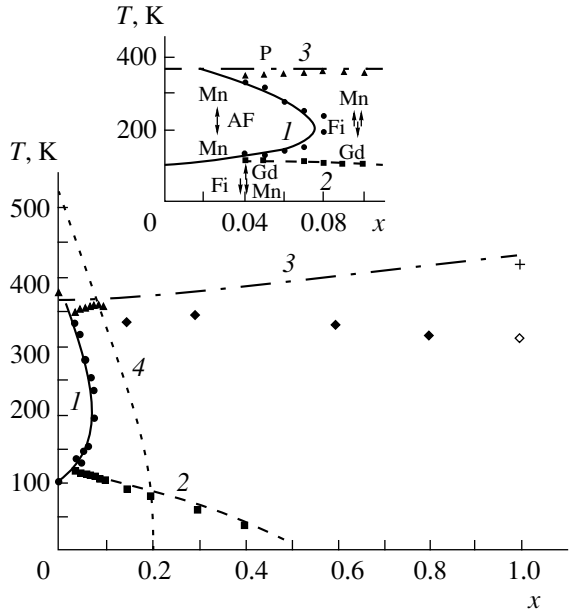


Fig. 4. *T–x* magnetic phase diagram of the intermetallic compounds $Gd_{1-x}La_xMn_2Ge_2$ in a 0.8 T field. The symbols are the experimental data; the lines show the theoretically computed dependences: ●, (*1*) first-order phase transition from the ferrimagnetic (Fi) state into the antiferromagnetic (AF) state; ■, (*2*) magnetic compensation temperature; (*3*) second-order transition into the paramagnetic (P) state (Néel temperature T_N); ▲, ◆—temperature at which ferrimagnetic ordering is destroyed (Curie temperature T_C); (*4*) line showing the variation of the sign of the Mn–Mn interlayer exchange interaction; +, ◇—the temperatures T_N and T_C from [6]. Inset: Experimental part of the phase diagram for $x < 0.1$ from [1] compared with the computational results.

The conditions for the existence of various magnetic phases and phase transitions between them at 0 K are presented in [2]. The magnetic parameters describing the behavior of this system in the Yafet–Kittel model were also determined in [2] from an analysis of the magnetic properties of the intermetallic compounds $Gd_{1-x}Y_xMn_2Ge_2$:

$$\lambda_{12} = -7.7\text{T}/\mu_B \text{ f. u.},$$

$$\lambda_{22} = -10.9\text{T}/\mu_B \text{ f. u.},$$

$$\lambda_{11} = 2.2\text{T}/\mu_B \text{ f. u.},$$

$$K = 15.8\mu_B \text{ T/f. u.}$$

We determined the quantity $\lambda_{22}' = 245\text{T}/\mu_B$ formula units from the Néel temperature of $GdMn_2Ge_2$. The magnetic moment of gadolinium was assumed to be equal to the moment of the trivalent ion ($7\mu_B$); the value

of the magnetic moment of manganese ($1.8\mu_B$) was determined from the data for the single crystal [7].

As already noted above, in calculating the magnetic properties of the system $Gd_{1-x}La_xMn_2Ge_2$ it is necessary to take into account the dependence of the interlayer Mn–Mn exchange interaction on the crystal lattice parameter a and therefore on the temperature and concentration. Assuming that this dependence is linear, we can write for the parameter λ_{22}

$$\lambda_{22} = \rho(a - a_{cr}). \quad (2)$$

According to the Kittel exchange inversion model [8], part of the energy due to the Mn–Mn exchange interaction, taking into account the dependence on the lattice parameter a, can be represented in the form

$$E_{Mn-Mn} = \frac{1}{2}\frac{N(a-a_T)^2}{a_T^2} - \rho(a-a_{cr})\mathbf{M}_2'\mathbf{M}_2''. \quad (3)$$

Here the first term describes the elastic energy, N describes the corresponding elastic modulus, a_T is the lattice parameter at a given temperature without a magnetoelastic contribution (for $\mathbf{M}_2' \perp \mathbf{M}_2''$), and the second term describes the Mn–Mn exchange interaction.

We can find the equilibrium value of the lattice parameter a, taking into account the magnetoelastic energy, from the condition for a minimum of the energy:

$$a = a_T + \frac{\rho a_T^2}{N}\mathbf{M}_2'\mathbf{M}_2''. \quad (4)$$

Substituting this value into the formula for the energy, we find the equilibrium energy of the interplanar Mn–Mn exchange interaction:

$$E_{Mn-Mn} = -\rho(a_T - a_{cr})\mathbf{M}_2'\mathbf{M}_2'' - \frac{\rho^2 a_T^2}{2N}(\mathbf{M}_2'\mathbf{M}_2'')^2. \quad (5)$$

Thus, taking into account the lattice parameter dependence of the exchange interaction leads to the appearance of an additional contribution that is biquadratic in the magnetic moments (biquadratic exchange) in the expression for the exchange energy. We note that, as the estimates showed, in intermetallide systems Cd–La biquadratic exchange is weak (it is about 1 or 2% of the ordinary exchange which is quadratic in the magnetic moments). Moreover, at a phase transition of the manganese subsystem from the ferromagnetic into the antiferromagnetic state, which we are discussing, the biquadratic exchange energy does not change. Consequently, we shall neglect the biquadratic exchange in the calculations below.

To calculate $\lambda_{22} = \rho(a_T - a_{cr})$ it is necessary to known the critical lattice parameter a_{cr}, the lattice parameter a_T at a given temperature, neglecting the magnetoelastic contribution from the Mn–Mn exchange interaction, and the parameter ρ.

We used the value of the critical lattice parameter $a_{cr} = 4.045$ Å, determined in [6] from an analysis of the magnetic properties of the intermetallic compounds $Y_{1-x}La_xMn_2Ge_2$.

The values of a_T in the magnetically order state were found from our measurements of the lattice parameters of the intermetallic compounds $LaMn_2Ge_2$ and $GdMn_2Ge_2$ [9] by extrapolation using the Debye law for the phonon contribution to the thermal expansion from the paramagnetic temperature range. The Debye temperature 415 K was determined by analyzing our experimental data for the temperature dependences of the parameters of the crystal structure of the nonmagnetic intermetallide YCo_2Ge_2. The values of a_T for the mixed intermetallic compounds Gd–La were calculated from the following formula using the data for pure $GdMn_2Ge_2$ and $LaMn_2Ge_2$ assuming Vegard's law to hold:

$$a_T(x) = (1-x)a_T^{Gd} + xa_T^{La}. \quad (6)$$

Knowing the parameter a_T for $GdMn_2Ge_2$ and the value of a_{cr}, we found using equation (2) and the value of the parameter $\lambda_{22} = -10.9T/\mu_B$ formula units for this compound at low temperatures that $\rho = 310T/\mu_B$ Å formula units.

We shall analyze first the properties at low temperatures, where the field dependence of the magnetic moments of the gadolinium and manganese subsystems can be neglected. Analytic expressions for this case were obtained in [2] for the conditions of existence of various magnetic phases. Using these conditions it is possible to explain why, in contrast to the system $Gd_{1-x}La_xMn_2Ge_2$, triangular magnetic phases do not arise in the system $Gd_{1-x}Y_xMn_2Ge_2$.

It is shown in [2] that triangular phases in ferrimagnets with negative intrasublattice exchange are possible only if the anisotropy is small compared with this exchange:

$$K < -2\lambda_{22}M_2^2, \quad (7)$$

and when this condition is not satisfied, they become energetically unfavorable and only ferrimagnetic (Fi) and antiferromagnetic (AF) phases are possible in the system, i.e., the system becomes an Ising system. Since the quantity λ_{22} in the intermetallic compounds $Gd_{1-x}La_xMn_2Ge_2$ decreases in absolute magnitude as the lanthanum content increases (see equation (2)), the condition (7) for the appearance of triangular phases in the system holds only for small values of x.

On the other hand, it is shown in [2] that noncollinear phases appear because of the decrease in the intersublattice exchange interaction with substitutions when

$$(1-x)\lambda_{12}M_1 < 2\lambda_{22}M_2 + K/M_2. \quad (8)$$

A combined analysis of equations (7) and (8), taking into account the dependences of the Mn–Mn interlayer

exchange interaction on the interatomic distances (see equations (3), (4), and (5)), shows that the triangular phase cannot arise in gadolinium–lanthanum intermetallic compounds at low temperatures, since with substitutions the system becomes an Ising system before the condition (8) for the appearance of triangular phases is satisfied.

To construct the temperature and field dependences of the magnetization of the intermetallic compounds $Gd_{1-x}La_xMn_2Ge_2$ and the magnetic phase diagram of this system it is necessary to calculate the equilibrium free energies of various phases taking into account the field and temperature dependences of the magnetizations of the magnetic subsystems and to determine, by comparing them, the magnetic phase with the lowest energy. Such a numerical calculation was performed in the molecular-field approximation using the scheme examined in detail in [7, 10] and briefly described in [2]. The regions of existence of various magnetic phases, the temperatures of magnetic phase transitions, and other characteristics of $Gd_{1-x}La_xMn_2Ge_2$ were calculated as a function of the magnetic field and the lanthanum concentration. The computational results were compared with the experimental data.

We note that approximate analytical formulas can be obtained for the temperatures and fields of a transition of the manganese subsystem from the antiferromagnetic into the ferromagnetic state, if it is assumed that the gadolinium subsystem is paramagnetic and its magnetization is a linear function of the effective field and the magnetic susceptibility of the manganese subsystem is neglected (as estimates show, this approximation "works" in the compound studied in the temperature range 100–250 K). In this case the intrinsic energy of the gadolinium subsystem (fourth term in equation (1)) can be expressed as

$$E_1 = (1-x)M_1^2/2\chi_1, \tag{9}$$

where the intrinsic susceptibility χ_1 of the gadolinium subsystem can be described by the Curie–Weiss law:

$$\chi_1 = C/(T-\Theta). \tag{10}$$

We obtain for the transition field in this approximation

$$H_{cr} = -\lambda_{12}M_2\frac{\chi_1\lambda_{12} + \lambda_{22}/\lambda_{12}}{\chi_1\lambda_{12} + 1}. \tag{11}$$

We note that the calculation performed using the approximate formulas gives qualitatively the same result as an exact computer calculation, though the transition fields and temperatures differ somewhat in magnitude.

We compared the results of the computer calculation with the experimental data. The results of this comparison are displayed in Figs. 1, 3, and 4.

As follows from Fig. 1, satisfactory agreement is observed between the computed dependences $M(T)$ and the experimental data: the computed dependences describe well the transitions Fi⟶AF⟶Fi observed in compositions with $x < 0.1$, the values of the magnetic compensation temperatures, and the character of the temperature dependences of the magnetization in compositions with a high lanthanum content. The somewhat lower experimental values of the magnetization as compared with the computed values can be explained by the fact that the measurements were performed on polycrystalline samples in a 0.8 T field, which is inadequate for technical saturation.

The theoretical dependences of the critical fields of the transitions of the manganese subsystem from the antiferromagnetic into the ferromagnetic state describe qualitatively the experimentally observed results, though the experimental and theoretical values of the transition fields are different in a number of cases (Fig. 3).[1]

In Fig. 4 the experimental T–x phase diagram of the system $Gd_{1-x}La_xMn_2Ge_2$ is compared with the theoretically computed diagram. It is evident that the transitions Fi⟶AF⟶Fi are described well on the basis of the theoretical ideas developed above.

5. CONCLUSIONS

In summary, the magnetic state does not change in the system of intermetallic compounds $Gd_{1-x}La_xMn_2Ge_2$ at low temperatures as the gadolinium concentration decreases: ferrimagnetic ordering occurs in all mixed compounds. This behavior is substantially different from the magnetic behavior of the intermetallic compounds $Gd_{1-x}Y_xMn_2Ge_2$, whereas x increases, the ferrimagnetic (Fi) phase transforms into a phase with triangular magnetic ordering (the phase T'). The magnetic phase transitions observed in both systems with increasing temperature are also different. In contrast to the system $Gd_{1-x}Y_xMn_2Ge_2$, where the phase transitions Fi⟶AF (for $x < 0.3$), Fi⟶T⟶T' ⟶AF (for $0.3 < x < 0.5$), and T'⟶AF (for $x > 0.5$) are observed with increasing temperature, the sequence of phase transitions Fi⟶AF⟶Fi is observed in the intermetallic compounds $Gd_{1-x}La_xMn_2Ge_2$ with $x < 0.09$, as mentioned above, while intermetallic compounds with a high lanthanum content are ordinarily ferrimagnets.

These differences are all observed in a model of a ferrimagnet with a negative exchange interaction in one of the sublattices, if the fact that the parameter λ_{22}, describing the Mn–Mn interplanar exchange interaction, in the system $Gd_{1-x}La_xMn_2Ge_2$ depends on the concentration (while for comparatively low temperatures it also depends on temperature) because of the change in the lattice parameter a. The line $a(T) = a_{cr}$, separating regions with $\lambda_{22} < 0$ and $\lambda_{22} > 0$ (Fig. 4), is shown in the T–x phase diagram of $Gd_{1-x}La_xMn_2Ge_2$. It is evident that Mn–Mn exchange is negative only for small values of x, and even for lanthanum concentrations greater than approximately $x = 0.22$ these inter-

[1] We note that the agreement between the experimental and theoretical dependences $H_{cr}(T)$ can be substantially improved by making small variations in the values of the critical parameter a_{cr}.

metallic compounds are normal two-sublattice ferrimagnets. At the same time $\lambda_{22} < 0$ in the entire magnetically ordered region in the system $Gd_{1-x}Y_xMn_2Ge_2$ where the lattice parameter a is essentially concentration-independent.

We note that, as one can see from Fig. 4, the transition AF⟶Fi at a given temperature occurs at concentrations less than the critical value for which the lattice parameter $a = a_{cr}$. Physically, this is due to the fact that, in addition to manganese, the intermetallic compounds studied contain a second—gadolinium—magnetic subsystem, the interaction with which stabilizes the ferrimagnetic phase Fi, since in this phase the gadolinium subsystem is ferromagnetically ordered, while in the AF phase it is paramagnetic. It is also easy to explain, as shown above, that triangular phases, which are observed in the yttrium-substituted intermetallic compounds, are absent in a system with lanthanum.

In summary, using values of the parameters determined from the experimental investigations of the system $Gd_{1-x}Y_xMn_2Ge_2$, it has been shown that the Yafet–Kittel model for ferrimagnets with negative intrasublattice exchange describes well the transitions Fi⟶AF⟶Fi, the magnetic compensation point, and other characteristic magnetic properties of the intermetallic compounds $Gd_{1-x}La_xMn_2Ge_2$, if it is assumed that the Mn–Mn interlayer exchange interaction depends on the lattice parameter a.

We note that this simple model, however, does not permit describing quantitatively the temperatures at which the ferrimagnetic ordering T_C is destroyed in these compounds. As one can see from Fig. 4, the experimental values of T_C are much lower than the values computed on the basis of the model described above. This seems to be explained by the fact that in the intermetallic compounds $Gd_{1-x}La_xMn_2Ge_2$ with a high lanthanum content the transition at the point T_C is the transition Fi⟶AF, and the transition into the paramagnetic state occurs at a higher temperature T_N. This behavior was recently observed in $LaMn_2Ge_2$ and other intermetallic compounds RMn_2Ge_2 with light rare earths [10]. To describe this behavior it is necessary to take into account, together with the ferromagnetic, antiferromagnetic Mn–Mn exchange interactions in a layer [6]. We note that we observed in the compounds investigated small features in the temperature dependences of the magnetization at temperatures of about 420–430 K, close to T_N for $LaMn_2Ge_2$ [6]. It is possible that these features are due to the destruction of the antiferromagnetic ordering in $Gd_{1-x}La_xMn_2Ge_2$, but it has not been ruled out that they are due to the presence of trace impurities of other phases. This question requires additional study.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 99-02-17358 and project no. 96-15-96429 for support of Scientific Schools), and the Ministry of Science of the Russian Federation (project no. 97-0-7.3-157).

REFERENCES

1. A. Sokolov, H. Wada, M. Shiga, and T. Goto, Solid State Commun. **105**, 289 (1998).
2. A. Yu. Sokolov, Guo Guanghua, S. A. Granovskiĭ, *et al.*, Zh. Éksp. Teor. Fiz. **116**, 1346 (1999) [JETP **89**, 723 (1999)].
3. A. Szytula and J. Leciejewicz, in *Hahdbook Physics and Chemistry of Rare Earths*, Ed. by K. A. Gschneidner, Jr. and L. Eyring (North Holland, Amsterdam, 1989), Vol. 12, p. 133.
4. Y. Yafet and C. Kittel, Phys. Rev. **87**, 290 (1952).
5. H. Kobayashi, H. Onodera, and H. Yamomoto, J. Magn. Magn. Mater. **79**, 76 (1989).
6. G. Venturiny, J. Alloys Compd. **232**, 133 (1996).
7. N. Iwata, K. Hattori, and T. Shigeoka, J. Magn. Magn. Mater. **53**, 318 (1986).
8. C. Kittel, Phys. Rev. **120**, 335 (1960).
9. I. Yu. Gaĭdukova, Guo Guanghua, S. A. Granovskiĭ, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **41**, 2053 (1999) [Phys. Solid State **41**, 1885 (1999)].
10. H. Wada, H. Yamaguchi, and M. Shiga, J. Magn. Magn. Mater. **152**, 165 (1996).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 987–992.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1136–1142.

SOLIDS
Electronic Properties

Long-Time Relaxation of the Magnetization and Tunneling Magnetoresistance of Granular Ferromagnets

E. Z. Meilikhov

Russian Research Centre Kurchatov Institute, 123182 Moscow, Russia
e-mail: meilikhov@imp.kiae.ru

Received November 29, 1999

Abstract—Magnetic anisotropy and orientational variance as well as shape diversity of granules largely determine the magnetic properties of granular ferromagnetic metals. The model of magnetically anisotropic ellipsoidal granules explains the glassy nature of the magnetic state of such systems. The relaxation of the magnetization and the magnetoresistance of granular ferromagnetic metals is examined on the basis of this model. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The model of single-domain superparamagnetic (and not interacting with one another) granules is ordinarily used to describe the properties of granular ferromagnets (see, for example, [1]). It is shown in [2] that the strong geometric magnetic anisotropy of granules makes this model inapplicable. In the present paper we examine a simple model in which the magnetic anisotropy of the granules is due to their aspherical shape and the orientation and shape distributions of the granules in real systems are taken into account statistically. Such a model leads naturally to an explanation of the glassy behavior of granular ferromagnetic metals and, specifically, to a description of the relaxation characteristics of their magnetization.

It is well known that the electric resistance R of nanocomposite materials is determined by intergranular tunneling of electrons. For ferromagnetic granules this probability depends strongly on the magnetic field (the so-called giant magnetoresistance). The relative change in the resistance is found to be related with the average magnetization m of the system by the simple relation [1] $\Delta R/R \propto m^2$. Hence it follows that the relaxation of the magnetization of such a system is always accompanied by a relaxation of its resistance. Consequently, an experimental study of the relaxation processes can be based on magnetic as well as galvanomagnetic measurements. For definiteness, in what follows we shall be concerned with magnetization relaxation.

The most diverse types of glasses can be described, as a rule, on the basis of the so-called two-level systems models—multiple microscopic subsystems with two energy states separated by a barrier [3]. Transitions (tunneling or activation) between these states occur in a subsystem. Often, the specific "arrangement" of two-level systems is unknown (or it is attributed to lattice defects whose origin is unclear), but it is assumed that the corresponding transition times are distributed randomly over exponentially wide limits. A consequence of this assumption is that the relaxation time of the corresponding physical parameter (specific heat, magnetization, and others) is quite long, falling within a quite wide range described by a logarithmic time-dependence of this parameter. The model considered in the present work is, essentially, a model of two-level systems, whose well-known nature makes it possible to describe the properties of individual two-level systems, to determine the statistical properties of their set as a whole, and as a result to calculate the relaxation characteristics of the system.

The present work is a continuation of the analysis initiated in [2, 4] of the magnetic properties of nanocomposite materials with ferromagnetic granules.

2. RELAXATION OF THE MAGNETIC MOMENT OF AN INDIVIDUAL ASPHERICAL FERROMAGNETIC GRANULE

The typical sizes of ferromagnetic granules in nanocomposite materials are very small (10–100 Å). This makes it possible to assume such materials to consist of a single domain. In the absence of an external magnetic field the spontaneous magnetic moment of an aspherical granule (the moment being produced by the intragranular exchange interaction and equal to VI_s, where V is the granule volume and I_s is the saturation magnetization) is always directed parallel to its "easy" magnetization axis. When an external magnetic field is switched on (the direction of the field, generally speaking, is different from that of this axis) the magnetic moment strives to turn and an energy barrier, whose magnitude depends on the relative orientation of the magnetic moment, the external magnetic field, and the "easy" axis, impedes this rotation.

1063-7761/00/9006-0987$20.00

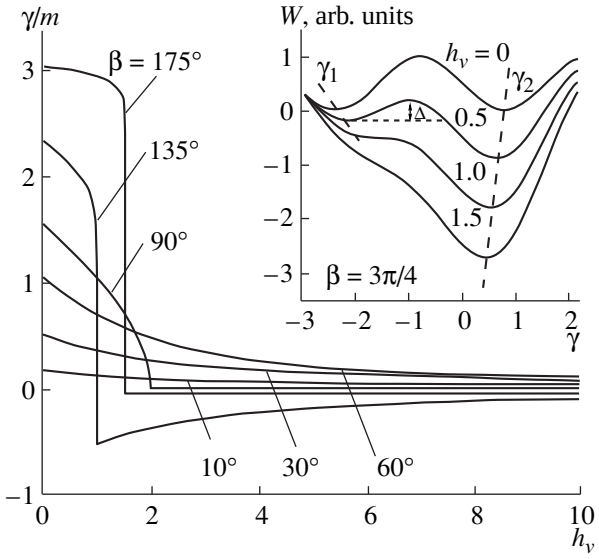


Fig. 1. Field dependences of the equilibrium angles of inclination γ of the magnetic moment of ellipsoidal granules. Inset: Magnetic energy of a granule versus the angle γ between its magnetic moment and the external magnetic field for various values of the latter (the dashed lines γ_1 and γ_2 correspond to the equilibrium state).

We shall assume in what follows that the granules are prolate ellipsoids of revolution with semiaxes $a > b = c$. Then the "easy" magnetization axis of a granule is aligned along the major axis of the ellipsoid. The magnetic energy W of such a granule in an external field H is determined by the relation [2]

$$\frac{W}{0.5I_s^2 \nu V} = \sin^2(\beta - \gamma) - h_\nu \cos\gamma, \quad h_\nu = \frac{2H}{I_s \nu}, \qquad (1)$$

where γ is the angle between the magnetic moment of a granule in the magnetic field, β is the angle between the magnetic field and the major axis of the ellipsoid, and $\nu = N_b - N_a$, where N_b and N_a are the demagnetization factors of an ellipsoid along the corresponding axes.

The ground state of the system, obviously, corresponds to the minimum energy W. A calculation shows (and simple analytic calculations confirm) that for an arbitrary orientation of the granules relative to the direction of the external magnetic field the angular dependence $W(\gamma)$ of the energy of a granule in a strong magnetic field ($h_\nu > 2$) always possesses one minimum (irrespective of the angle β). However, in a weak magnetic field (more accurately, for $h_\nu < 2$) this dependence has either two minima—γ_1 and γ_2—separated by an energy barrier of height $\Delta(h_\nu, \beta)$ or one minimum. This is illustrated in the inset in Fig. 1. Correspondingly, the ground state of the system can evolve according to two qualitatively different scenarios (see Fig. 1):

1) If $\beta < \pi/2$ (the initial spontaneous magnetic moment of a granule makes an acute angle with the "future" magnetic field), then as the magnetic field increases, the angle $\gamma = \gamma_2$ decreases monotonically.

2) If $\pi/2 < \beta < \pi$, then in a certain critical magnetic field $\tilde{h}_\nu(\beta)$ the barrier between the minima vanishes ($\Delta(\tilde{h}_\nu, \beta) = 0$) and an abrupt change in angle $\gamma_1 \longrightarrow \gamma_2$ occurs (orientational magnetic phase transition), after which the equilibrium angle will decrease monotonically as the field increases.

At finite temperature the orientational transition starts in a field below the critical field ($h_\nu < \tilde{h}_\nu$, when $\Delta(h_\nu, \beta) > 0$) and will be manifested as temporal relaxation of the magnetic moment of a granule with a characteristic time τ_ν. The field range where such relaxation can be observed is determined, obviously, by the condition $\tau_{exp} > \tau_\nu$, where τ_{exp} is the duration of the experiment.

The characteristic relaxation time of the magnetic moment $\tau_\nu \sim \tau_{\nu 0} \exp[\Delta(h_\nu, \beta)/kT]$, where $\tau_{\nu 0} \sim mc/eH$ is the precession period of the magnetic moment, depends on the height of the barrier between the two minima of the energy W [5] (the initial minimum lies above the final minimum). The magnetic moment of an individual granule (with angle $\beta > \pi/2$) can assume two values: $m_{\nu 1} = VI_s \cos\gamma_1$ (initial moment) and $m_{\nu 2} = VI_s \cos\gamma_2$ (final moment).[1] The collection of a large number of identical granules (with the same shape, volume, and orientation) can be characterized by a single magnetic moment $m_\nu(t) = p_1(t)m_{\nu 1} + p_2(t)m_{\nu 2}$, where $p_1(t)$ and $p_2(t) = 1 - p_1(t)$ are instantaneous (nonequilibrium) probabilities of finding a granule in the initial (p_1) or final (p_2) state, respectively.

In a constant magnetic field H, the temporal variation of these probabilities is described by the simple equations

$$\frac{dp_2(t)}{dt} = -\frac{dp_1(t)}{dt} = -\frac{p_2(t) - p_2^\infty(H)}{\tau_\nu(H)}, \qquad (2)$$

where $p_2^\infty(H)$ is the corresponding equilibrium probability in a field H.

Let us consider one possible process leading to the relaxation of the average magnetic moment m_ν of such a system accompanying an abrupt change (at the time $t = 0$) of the magnetic field from $H = H_1$ to $H = H_2$, assuming that for $t < 0$ the system was in a thermodynamic equilibrium where $p_2(0) = p_2^\infty(H_1)$. Then the solution of equations (2) has the form

$$p_2(t) = p_2^\infty(H_2) + [p_2^\infty(H_1) - p_2^\infty(H_2)] \exp\left[-\frac{t}{\tau_\nu(H)}\right]. \qquad (3)$$

[1] On account of thermal excitation the energy of the system is determined only to within kT, so that the magnetic moment is "smeared." However, this effect can be neglected, since the depth of the lower minimum of the energy is much greater than kT.

This relation can be used to describe the experimental results on the magnetic relaxation only if the initial state of the system is definitely an equilibrium state. As a rule, such an equilibrium state must be specially "prepared." In our case there are apparently only two simple ways to do this: 1) increase the temperature of the system to a value at which the system "instantaneously" relaxes into the ground state (in practice, room temperature and $H_1 = 0$ are sufficient for this) and 2) apply such a strong magnetic field that the condition $h_\nu > 2$ obtains even for the most prolate granules ($\nu_{max} \sim 10$) (for this, a field $H_1 \gtrsim H^\infty = \nu_{max} I_s \sim 10$ kOe is sufficient at any temperature). Consequently, it is reasonable to examine the following two regimes for studying magnetic relaxation: (1) heating the system up to room temperature with $H_1 = 0$ and then cooling the system to a prescribed temperature in zero field and switching on the required field H_2 (zero field cooling (ZFC) regime); (2) placing the system at a fixed temperature into an "infinite" magnetic field $H_1 > +H^\infty$ and then switching the field, changing its orientation at the same time, to the required value $H = -H_2$ (infinite field (IF) regime).

In an equilibrium initial state of the ZFC regime, half the granules occupy one energy minimum and half occupy the other, i.e., $p_1^\infty(0) = p_2^\infty(0) = 1/2$. In an equilibrium final state all granules occupy the bottom minimum: $p_2^\infty(H_2) = 1$. Therefore it follows from equation (3) that

$$p_2^{ZFC}(t) = \frac{1}{2} + \frac{1}{2}\left\{1 - \exp\left[-\frac{t}{\tau_\nu(H_2)}\right]\right\}. \qquad (4)$$

When a field is switched on, the magnetic moments of half the granules (with $\beta < \pi/2$) "instantaneously" reach the values $m_{\nu 2}(H_2)$ and the average magnetic moment of the remaining granules (with $\beta > \pi/2$) likewise "instantaneously" reaches the value $m_{\nu 1}(H_2)$ and then relaxes to $m_{\nu 2}(H_2)$ according to the law (4). Consequently, the relaxation of the average magnetic moment of all granules of the system is determined by the dependence

$$m_\nu^{ZFC}(t) = \frac{1}{2} m_{\nu 2}(H_2) + \frac{1}{2}\left\{ m_{\nu 1}(H_2) + [m_{\nu 2}(H_2) - m_{\nu 1}(H_2)]\left[1 - \exp\left(-\frac{t}{\tau_\nu(H_2)}\right)\right]\right\}, \qquad (5)$$

where the first term refers to granules with $\beta < \pi/2$ and the second refers to granules with $\beta > \pi/2$.

If the final field is sufficiently strong (so that $2H_2/I_\nu > \tilde{h}_\nu(\beta)$), then in the final state there is no barrier Δ between the two energy minima and, consequently, the relaxation time is short compared with the switching time of the field. The magnetic moment of the system changes almost instantaneously. Long-time (on the order of the characteristic time of the experiment) relaxation of the magnetic moment can be expected only in weak fields. However, to observe such relaxation the characteristic time $\tau_\nu(H)$ still cannot be too long, i.e., the barrier height Δ should not be very large. As shown in the inset in Fig. 2, this condition holds only in a narrow range of fields near $\tilde{h}_\nu(\beta) \sim 0.1$. For $I_s \sim 10^3$ G and $\nu \sim 1$, this corresponds to magnetic fields $H \sim 0.1 I_s \nu \sim 100$ Oe.

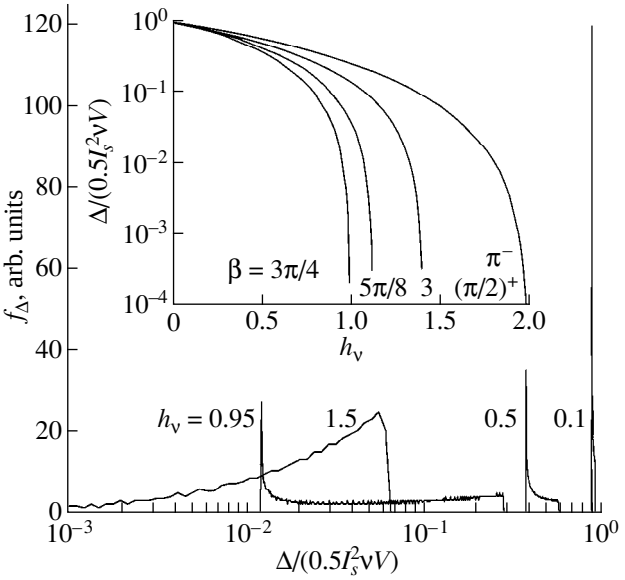


Fig. 2. Distributions $f_\Delta(\Delta)$ of the barrier heights in a system of randomly oriented ellipsoidal granules with the same shape and volume in various magnetic fields. Inset: Field dependences of the barrier height for individual granules with different orientation ($\pi/2 < \beta < \pi$, angles π^-, $(\pi/2)^+$ infinitely close to the boundaries of this range).

A different picture of magnetization relaxation should be observed when the field is switched off (from H_1 to $H_2 = 0$). In this case the magnetic moments of half the granules (with $\beta < \pi/2$) "instantaneously" reach the value $m_{\nu 2}(0)$, and the magnetic moments of the remaining granules (with $\beta > \pi/2$) become "frozen," since the corresponding energy barrier remains high throughout the entire process of switching-off the field. It vanishes only in the process of switching the direction of the field.

In an equilibrium initial state of the IF regime, all granules occupy the same (bottom) energy minimum, i.e., $p_1^\infty(0) = 1$ and $p_2^\infty(0) = 0$. In the equilibrium final state all granules occupy the other minimum: $p_1^\infty(H_2) = 1 - p_2^\infty(H_2) = 0$. Consequently, from equation (3) follows

$$p_2^{IF}(t) = 1 - \exp\left[-\frac{t}{\tau_\nu(H_2)}\right]. \qquad (6)$$

When the field is removed, the magnetic moments of all granules "instantaneously" reach the value $m_{\nu 1}(H_2)$. Then the relaxation of the average magnetic moment of the granules in the system is determined by the dependence

$$m_\nu^{IF}(t) = m_{\nu 2}(H_2) + \left\{ m_{\nu 1}(H_2) + [m_{\nu 2}(H_2) - m_{\nu 1}(H_2)]\left[1 - \exp\left(-\frac{t}{\tau_\nu(H_2)}\right)\right]\right\}, \quad (7)$$

which differs from the dependence (5) only by the absence of the factor 1/2.

3. RELAXATION OF THE MAGNETIC MOMENT OF AN ENSEMBLE OF FERROMAGNETIC GRANULES WITH DIFFERENT SHAPES AND ORIENTATIONS

Real nanocomposite materials consist of granules with different sizes, shapes, and orientations. Since the individual relaxation times τ_ν are exponential functions of the height of the corresponding energy barrier, even a relatively small variance of the parameters leads to an exponentially strong variance of these times. As a rule, relaxation experiments in systems of this type demonstrate a logarithmic relaxation of the magnetic moment ($m(t) \propto \text{const} - \ln t$).[2]

[2] The origin of this logarithmic dependence is associated with the variance of the heights Δ of the energy barriers. Their distribution function $f_\Delta(\Delta)$ determines the distribution function of the relaxation times $f_\tau(\tau) = f_\Delta[\Delta(\tau)](\partial\tau/\partial\Delta)^{-1} = kTf_\Delta[\Delta(\tau)]d\tau/\tau$, where $\Delta(\tau) = kT\ln(\tau/\tau_0)$. Then, we obtain instead of the exponential in equations (5) and (7),

$$G(t) \equiv \int_{\tau_{\min}}^{\tau_{\max}} \exp\left(-\frac{t}{\tau}\right) f_\tau(\tau)d\tau \propto \int_{\tau_{\min}}^{\tau_{\max}} \left[\exp\left(-\frac{t}{\tau}\right)\frac{1}{\tau}\right] f_\Delta[\Delta(\tau)]d\tau,$$

where $\tau_{\min}$ and $\tau_{\max}$ are the minimum and maximum relaxation times. If $f_\Delta = \text{const}$ in the integration range, then

$$G(t) \propto \int_{\tau_{\min}}^{\tau_{\max}} \left[\exp\left(-\frac{t}{\tau}\right)\frac{1}{\tau}\right]d\tau \propto \text{Ei}\left(-\frac{t}{\tau_{\max}}\right) - \text{Ei}\left(\frac{t}{\tau_{\min}}\right),$$

where

$$\text{Ei}(-x) = -\int_x^\infty e^{-z}\frac{dz}{z} = C + \ln x + \int_0^x (e^{-z} - 1)\frac{dz}{z}$$

($C \simeq 0.5772$ is Euler's constant). It is easy to see that if $\tau_{\min} \ll t \ll \tau_{\max}$, then $G(t) \propto \text{const} - \ln(t/\tau_{\max})$. If $f_\Delta \neq \text{const}$, then the function $G(t)$ is no longer logarithmic in the entire exponentially wide range $\tau_{\min} < t < \tau_{\max}$, but it is often close to logarithmic in the actual (experimentally achieved) range of times.

Thus, a simple exponential relaxation of the type (5), which is characteristic for individual granules, does not occur. In order to take into account the above-mentioned diversity of granule parameters, the relaxation processes must be averaged in an ensemble of granules:

$$m(t) = \iiint_{V\nu\beta} m_\nu t)\, f(V, \nu, \beta)dV d\nu d\beta, \quad (8)$$

where m is the ensemble-averaged magnetic moment of the granules and $f(V, \nu, \beta)$ is the distribution function of the corresponding geometric parameters of granules in a nanocomposite material.

At present there are no experimental data that would make it possible to establish the form of this function, and there are no indications of the existence of any correlations between these parameters. There are only a few works in which the granule size distribution $f_a(a)$, where a is the granule size, is found in an indisputable manner (for example, using an electron microscope). Here the size is any quantity obtained by intuitive averaging for an aspherical granule. It is assumed below that all geometric parameters are statistically independent, i.e., the relation $f(V, \beta, \nu) = f_V(V)f_\beta(\beta)f_\nu(\nu)$, where the corresponding partial distribution functions appear on the right-hand side, is valid.

We shall start the averaging over the function (5) by taking into account the variance of the granule orientations determined by the angle β. For simplicity, we shall assume that these angles are distributed uniformly in the range $0 < \beta < \pi$, i.e., $f_\beta(\beta) = 1/\pi$. In this case even granules with the same shape and volume can possess extremely different activation energies Δ; this is what leads to the exponentially strong variance of the relaxation times. This is illustrated in Fig. 2, which displays the results of a numerical calculation of the distribution $f_\Delta(\Delta)$ of the barrier heights in the system under study in various magnetic fields. It is evident that magnetic fields $h_\nu \sim 1$ not only decrease the barriers, as is indicated by the displacement of the right-hand boundary and center of gravity of the distribution $f_\Delta(\Delta)$ leftward, but also produce a strong variance of the barrier height (the relative width of the distribution function increases substantially).[3]

In accordance with the expression (8), the relaxation of the magnetic moment of the system of granules

[3] It can be shown that as the magnetic field increases, the maximum barrier height $\Delta_{\max}$ (corresponding to the angles $\beta = (\pi/2)^+, \pi^-$) decreases as $\Delta_{\max} = (1 - h_\nu/2)^2$ and vanishes for $h_\nu \geq 2$, while the minimum height $\Delta_{\min}$ (corresponding to the angle $\beta = 3\pi/4$) decreases more rapidly: $\Delta_{\min} = 1 - (h_\nu/2)[(8 + h_\nu)^{1/2} - h_\nu]$ and vanishes for $h_\nu \geq 1$. Consequently, for $1 < h_\nu < 2$ the distribution function $f_\Delta(\Delta)$ is different from zero in the entire range $0 < \Delta < \Delta_{\max}$.

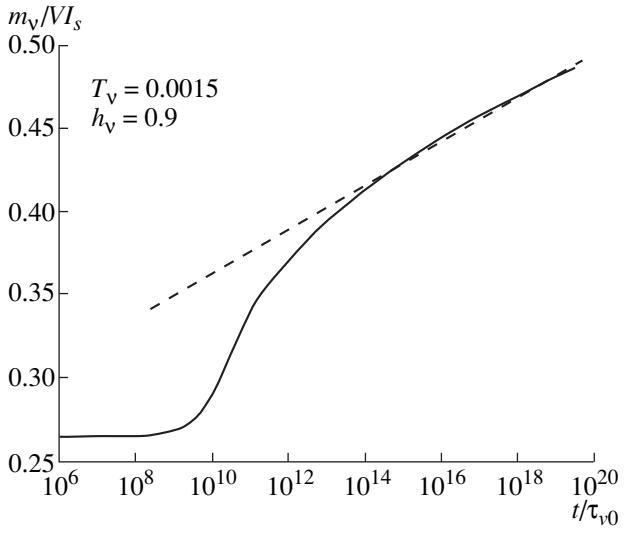


Fig. 3. Relaxation of the magnetization of a system of randomly oriented ellipsoidal granules with the same shapes and volumes.

(after an abrupt change in the field from 0 to h_ν) is described by the relation

$$\frac{m(t)}{VI_s} = \frac{1}{\pi}\int_0^{\pi/2} \cos\gamma_2(h_\nu)\,d\beta$$

$$+\frac{1}{\pi}\int_{\pi/2}^{\pi}\left\{\cos(\gamma_1(h_\nu) + [\cos\gamma_2(h_\nu) - \cos\gamma_1(h_\nu)]\right) \quad (9)$$

$$\times\left[1 - \exp\left\{-\exp\left[-\frac{\Delta(h_\nu, \beta)}{kT}\right]\frac{t}{t_{\nu 0}}\right\}\right]\right\}d\beta.$$

Figure 3 shows a typical relaxation dependence, obtained using the relation (9), of the magnetic moment of a system of randomly oriented granules. For ellipsoidal Fe granules (I_s = 1700 G) with volume $V_1 = 6 \times 10^{-20}$ cm^3 (diameter 50 Å) and $\nu = 1$, the parameter values used in the calculation $h_\nu = 0.9$ and $T_\nu = kT/(0.5I_s^2\,\nu V) = 0.0015$ correspond to a magnetic field $H \approx 800$ Oe and temperature $T \approx 4$ K. It is evident that a logarithmic function cannot describe the initial, relatively rapid, change in the magnetic moment (in the actual time interval $t/\tau_0 \sim 10^{10}$–10^{15}; see below). A logarithmic dependence appears only at a subsequent stage of relaxation.

The next step is to take account of the shape variance of the granules, determined by the parameter ν. It leads to the appearance of a logarithmic relaxation even at the initial stage of relaxation. We shall assume that the values of this parameter are uniformly distributed in the range $\nu_{min} < \nu < \nu_{max}$, where $\nu_{min} = 0.5$ and $\nu_{max} = 2$ (this corresponds to ellipsoidal granules with the ratio of the axes ranging from 1.1 to 1.55), i.e., $f_\nu(\nu) = 1/(\nu_{max} - \nu_{min})$. Then, we have instead of (7)

$$\frac{m(t)}{VI_s} = \frac{1}{\pi(\nu_{max} - \nu_{min})}\left\{\int_{\nu_{min}}^{\nu_{max}} d\nu \int_0^{\pi/2} \cos\gamma_2(h_\nu)\,d\beta\right.$$

$$+\int_{\nu_{min}}^{\nu_{max}} d\nu \int_{\pi/2}^{\pi}\left\{\cos\gamma_1(h_\nu) + [\cos\gamma_2(h_\nu) - \cos\gamma_1(h_\nu)]\right. \quad (10)$$

$$\left.\left.\times\left[1 - \exp\left[-\exp\left[-\frac{\Delta(h_\nu, \beta)}{kT}\right]\frac{t}{\tau_{\nu 0}}\right]\right]\right\}d\beta\right\}.$$

The relaxation dependence of the magnetic moment of the system of granules under study with shape variance

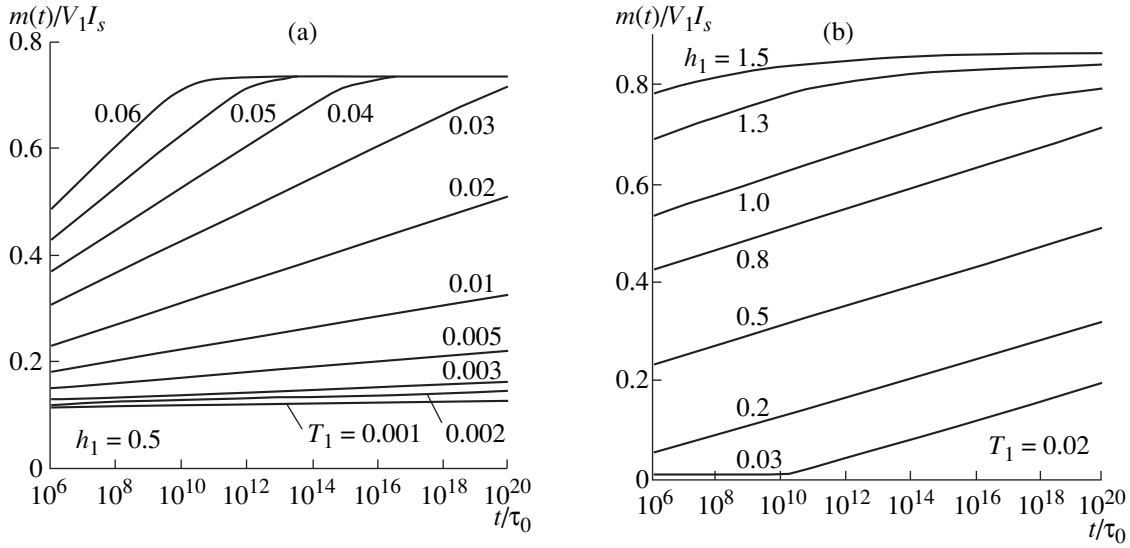


Fig. 4. Relaxation of the magnetization of a system of randomly oriented ellipsoidal granules with the same volume and a random demagnetization factor at various (a) temperatures and (b) fields.

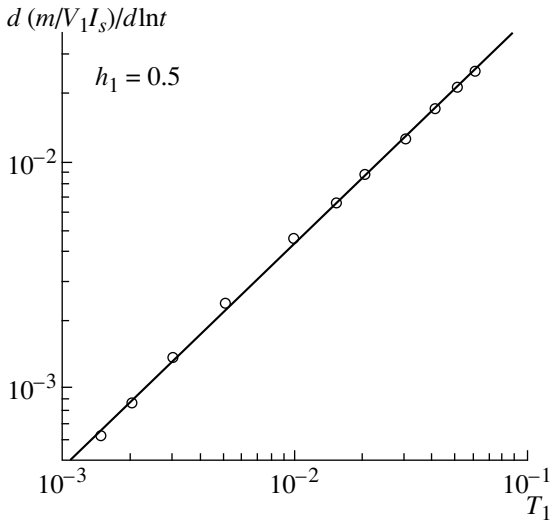


Fig. 5. Temperature dependence of the logarithmic relaxation rate (for the case corresponding to Fig. 4a).

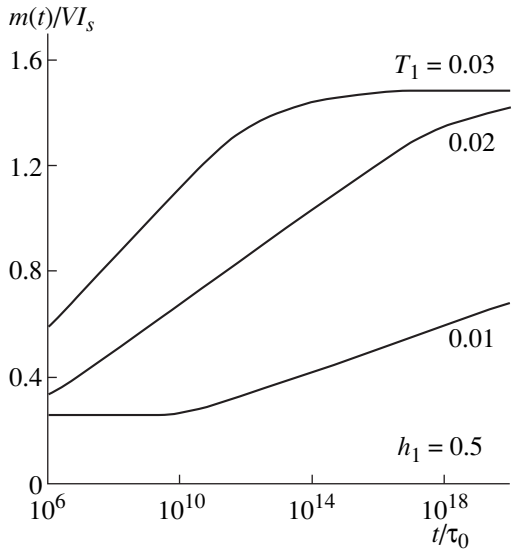


Fig. 6. Relaxation of the magnetization of a system of randomly oriented ellipsoidal granules with a random demagnetization factor and a random volume at various temperatures.

(calculated using equation (10) with the previous values of the parameters V_1, h_v, and T_v) is presented in Fig. 4a. The logarithmic character of the relaxation is now clearly visible. As expected, its rate $S = dm/d\ln t$ increases with temperature. Figure 5 shows that the function $S(T)$ is linear. At the same time, Fig. 4b demonstrates that the logarithmic relaxation rate S in a wide range of fields is independent of the magnitude of the magnetic field.

To use the data presented in Figs. 4a and 4b to assess the situation in a specific experiment, it is necessary to switch to the real time scale. In this connection we note that, for example, for α-Fe granules, $\tau_0 \approx 10^{-10}$ s [6]. In this case, as one can see in Fig. 4a, logarithmic relaxation of the magnetization, lasting for less than 10^2–10^3 s, can be observed only at relatively low temperatures ($T_1 \leq 0.05$ for $h_1 = 0.5$, i.e., $T < 130$ K for $V = 6 \times 10^{-20}$ cm^3).

Finally, we add the variance of the granule volumes to the orientational and shape variances of the granules which were taken into account above. Figure 6 shows the relaxation dependences of the magnetization, which differ from those shown in Fig. 4a in that the corresponding calculation was performed assuming a uniform distribution of granule volumes in the range $[V_{\min}, V_{\max}]$, where $V_{\min} = 0.5V$ and $V_{\max} = 2V$. It is evident that there are no qualitative changes here, and the quantitative changes are small.

It is easy, in principle, to calculate on the basis of the model examined above any magnetic properties of a nanocomposite material, specifically, the often employed characteristics such as the hysteresis curves, the temperature dependences of the magnetization in the *ZFC* and *FC* regimes, and others.

In conclusion, we remind the reader that everywhere above it was assumed that the magnetic field inside an experimental sample is identical to the external magnetic field. The samples of the ferromagnetic nanocomposite materials that are usually studied are thin films (thickness 10^{-5}–10^{-4} cm). Consequently, the external field must be directed along the normal to the plane of the sample. More accurately, the angle α between this normal and the direction of the field must be quite small: $\alpha \ll \alpha_0 \sim H/2\pi I_s$. For $H \sim 100$ Oe this means that $\alpha \ll 1°$.

ACKNOWLEDGMENTS

I am grateful to V.V. Ryl'kov for a helpful discussion. This work was supported by the Russian Foundation for Basic Research (project nos. 98-02-17412 and 00-02-17191) and the PICS-RFFI program (grant no. 98-02-2237).

REFERENCES

1. P. Allia, M. Knobel, P. Tiberto, and F. Vinai, Phys. Rev. B **52**, 15398 (1995); S. Honda, T. Okada, M. Nawate, and M. Tokumoto, Phys. Rev. B **56**, 14566 (1997); F. C. S. da Silva and M. Knobel, Phys. Rev. B **56**, 6086 (1997).
2. E. Z. Meĭlikhov, Zh. Éksp. Teor. Fiz. **116**, 2182 (1999) [JETP **89**, 1184 (1999)].
3. W. A. Phillips, Rep. Prog. Phys. **50**, 1657 (1987).
4. E. Z. Meĭlikhov, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 579 (1999) [JETP Lett. **69**, 623 (1999)].
5. W. J. Schuele, S. Shtrikman, and D. Treves, J. Appl. Phys. **36**, 1010 (1965).
6. F. Bödker, S. Mörup, M. S. Pedersen, *et al.*, J. Magn. Magn. Mater. **177–181**, 925 (1998).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 6, 2000, pp. 993–1009.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 6, 2000, pp. 1143–1160.

SOLIDS
Electronic Properties

Phase Fluctuations and Single-Fermion Spectral Density in 2D Systems with Attraction¶

V. P. Gusynin[a, *], V. M. Loktev[a, **], and S. G. Sharapov[b, c]

[a]*Bogolyubov Institute for Theoretical Physics, National Academy of Sciences of Ukraine, Kiev, 03143 Ukraine*

**e-mail: vgusynin@bitp.kiev.ua*

***e-mail: vloktev@bitp.kiev.ua*

[b]*Department of Physics, University of Pretoria, Pretoria, 0002 South Africa*

[c]*On leave of absence from Bogolyubov Institute for Theoretical Physics, National Academy of Sciences of Ukraine, Kiev, 03143 Ukraine*

Received August 12, 1999

Abstract—The effect of static fluctuations in the phase of the order parameter on the normal and superconducting properties of a 2D system with attractive four-fermion interaction is studied. Analytic expressions for the fermion Green's function, its spectral density, and the density of states are derived in the approximation where the coupling between the spin and charge degrees of freedom is neglected. The resulting single-particle Green's function clearly demonstrates a non-Fermi-liquid behavior. The results show that as the temperature increases through the 2D critical temperature, the widths of the quasiparticle peaks broaden significantly. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

One of the most convincing manifestations of the difference between the BCS scenario and superconductivity in the cuprates is the pseudogap, or the depletion of a single-particle spectral weight around the Fermi level [1]. This is observed mainly in underdoped cuprates where the pseudogap opens in the normal state as the temperature T decreases below the crossover temperature T^* and extends over a wide range of T.

Due to the complex nature of cuprate systems, there are a number of theoretical explanations for the pseudogap behavior. One of them is based on the model of a nearly antiferromagnetic Fermi liquid [2]. Another possible explanation relates the pseudogap to spin- and/or charge-density waves [3]. A third direction, which we take in this paper, argues that precursor superconducting fluctuations may be responsible for the pseudogap phenomena. Indeed, an incoherent pair tunneling experiment [4] proposed recently may allow one to determine whether superconducting fluctuations are really responsible for the pseudogap behavior. Furthermore, one cannot exclude the possibility that the pseudogap is the result of a combination of various mechanisms, e.g., both spin and superconducting fluctuations.

Precursor superconducting fluctuations have recently been extensively studied using different approaches. In most cases, the attractive 2D or 3D Hubbard model was considered. In particular, this model has been studied, both analytically [5–7] and numerically [8–11], in the conserving T-matrix approximation that is "Φ-derivable" in the sense of Baym [12]. The non-"Φ-derivable" T-matrix approximation was considered in [13]. In this approach, the pseudogap is related to the resonant pair scattering of correlated electrons above T_c. The pseudogap was also studied in [14] for d-wave pairing (for a review, see [15]), and Monte Carlo simulations for the 2D attractive Hubbard model were performed in [16].

It is known, however, that while the T-matrix approximation provides an adequate description of 3D systems at all temperatures, including the superconducting state with a long-range order, it fails (see, for example, [9]) to describe the Berezinskii–Kosterlitz–Thouless (BKT) transition into the state with an algebraic order, which is only possible in 2D systems. This is why, in most of the papers cited above, the T-matrix approximation was used to study either 3D systems [5, 6, 10, 13] or 2D systems above T_c [8, 9, 14, 15] in order to avoid the BKT transition, even though it is generally accepted that 2D models are more relevant to the description of cuprates. The effect of interaction between the layers has recently been experimentally studied by intercalating an organic compound into bismuth-based cuprates. Even though the distance between the layers increased remarkably, the value of T_c was nearly the same as that for pristine material (see [17]).

Of course, the superconducting transition itself is not of the BKT type, because even a weak interplanar coupling produces a transition in the $d = 3$ XY univer-

¶This article was submitted by authors in English.

sality class, sufficiently close to the transition temperature. Outside the transition region, however, the low-energy physics is governed by vortex fluctuations [18] and one can expect the 2D model to be especially relevant to the description of the pseudogap phase. This was confirmed for the quasi-2D model [19] (see also [20]).

Regarding the pseudogap, it is sufficient to consider the case where $T > T_c$. However, one definitely needs a different approach from the T-matrix if one wants to study the 2D theory for the entire temperature range and wants to connect the pseudogap to the superconducting gap. An alternative approach, which overcomes the above difficulty, was proposed in [21–23]. For a 2D system, one should rewrite the complex order field $\Phi(x)$ in terms of its modulus $\rho(x)$ and its phase $\theta(x)$ as $\Phi(x) = \rho(x)\exp[i\theta(x)]$, which was originally suggested by Witten in the context of 2D quantum field theory [24]. It is impossible to obtain $\Phi \equiv \langle\Phi(x)\rangle \neq 0$ at finite T, because this would correspond to the formation of symmetry-breaking homogeneous long-range order, which is forbidden by the Coleman–Mermin–Wagner–Hohenberg (CMWH) theorem [25]. However, it is possible to obtain $\rho \equiv \langle\rho(x)\rangle \neq 0$ with $\Phi = \rho\langle\exp[i\theta(x)]\rangle = 0$ at the same time because of random fluctuations of the phase $\theta(x)$ (i.e., because of transverse fluctuations of the order field originating in the modulus conservation principle [26]). We stress that $\rho \neq 0$ does not imply any long-range superconducting order (which is destroyed by phase fluctuations) and, therefore, does not contradict the above-mentioned theorem.

For the simple model studied in [21, 22], there are three regions in the 2D phase diagram. The first one is the superconducting (here, BKT) phase with $\rho \neq 0$ at $T < T_{\mathrm{BKT}}$, where T_{BKT} is the BKT transition temperature, which plays the role of T_c in pure 2D superconducting systems. In this region, there is an algebraic order or a power-law decay of the $\langle\Phi^*\Phi\rangle$ correlations. The second region corresponds to the so-called pseudogap phase ($T_{\mathrm{BKT}} < T < T_\rho$), where T_ρ is the temperature at which ρ is supposed to become zero. In this phase, ρ is still nonzero, but the above correlations decay exponentially. The third is the normal (Fermi-liquid) phase at $T > T_\rho$, where $\rho = 0$. Note that Φ and all the symmetry-violating correlation functions like $\langle\Phi(x)\Phi(0)\rangle$ vanish everywhere.

The proposed description of the phase fluctuations and the BKT transition is very similar to that given by Emery and Kivelson [27]. However, the field $\rho(x)$ does not appear in the phenomenological approach of [27], whereas in the present microscopic approach, it occurs naturally. We also mention here the application of similar ideas to the 3D case [28], where instead of the 2D temperature T_{BKT}, one has the phase-transition temperature in the 3D XY-model, T_c^{XY}.

The main quantity of interest in the present paper is the one-fermion Green's function and the associated spectral function $A(\omega, \mathbf{k}) = -(1/\pi)\mathrm{Im}G(\omega + i0, \mathbf{k})$. The second quantity, being proportional to the intensity of the angle-resolved photoemission spectrum (ARPES) [29], encodes information about the pseudogap and quasiparticles. Following the approach of [21–23], the Green's function for charged (physical) fermions is given by the convolution (in momentum space) of the propagator for neutral fermions (which has a gap $\rho \neq 0$) and the Fourier transform of the phase correlation function $\langle\exp(i\tau_3\theta(x)/2)\exp(-i\tau_3\theta(0)/2)\rangle$.

Thus, the approximation employed here assumes the absence of coupling between spin and charge degrees of freedom; this can be taken into account at the next stage of approximation. We demonstrate that the quasiparticle spectral function broadens considerably when passing from the superconducting to the normal state, as was observed experimentally [29]. More importantly, the phase fluctuations result in a non-Fermi-liquid behavior of the system both below and above T_{BKT}.

We note that the effect of classical phase fluctuations of the order field on the spectral properties of underdoped cuprates has also been analyzed by Franz and Millis [30]. Being experimentally motivated, they were able to show that the corresponding photoemission and tunneling data are well accounted for by a simple model where d-wave charge excitations are coupled to supercurrent fluctuations.

A brief overview of the paper is as follows: In Section 2, we present the modulus–phase formalism for the fermion Green's function and explain why it is so important to use this formalism to describe 2D models. In Section 3, we obtain and discuss the Green's function of phase fluctuations both below and above T_{BKT}. This expression is then used in Section 4 to derive the temperature and retarded fermion Green's functions. We show that this Green's function exhibits a non-Fermi-liquid behavior. In Section 5, we obtain an analytic expression for the spectral density of the fermion Green's function and discuss this result in detail. The density of states (DOS) is considered in Section 6. Appendix A contains technical details on the calculation of the long-distance asymptotic behavior of the phase correlation function. Appendix B contains the derivation of an alternative representation for the fermion Green's function which is useful in calculating the spectral density. The integrals for the DOS are given in Appendix C.

2. THE MODULUS–PHASE REPRESENTATION FOR THE FERMION GREEN'S FUNCTION

Our starting point is a continuum version of the two-dimensional attractive Hubbard model defined by the Hamiltonian density [21–23]

$$\mathcal{H} = \psi_\sigma^\dagger(x)\left(-\frac{\nabla^2}{2m} - \mu\right)\psi_\sigma(x) - V\psi_\uparrow^\dagger(x)\psi_\downarrow^\dagger(x)\psi_\downarrow(x)\psi_\uparrow(x), \tag{2.1}$$

where $x = \mathbf{r}$, τ denotes the space and imaginary time variables, $\psi_\sigma(x)$ is a fermion field with the spin $\sigma = \uparrow, \downarrow$, m is the effective fermion mass, μ is the chemical potential, and V is an effective local attraction constant; we take $\hbar = k_B = 1$. The model with the Hamiltonian density (2.1) is equivalent to the model with an auxiliary BCS-like pairing field, which can be written as

$$\mathcal{H} = \Psi^\dagger(x) \times \left[\tau_3\left(-\frac{\nabla^2}{2m} - \mu\right) - \tau_+(\Phi) - \tau_-\Phi^*(x)\right]\Psi(x) + \frac{|\Phi(x)|^2}{V} \tag{2.2}$$

in terms of Nambu variables

$$\Psi(x) = \begin{pmatrix}\psi_\uparrow(x)\\ \psi_\downarrow^\dagger(x)\end{pmatrix}, \quad \Psi^\dagger(x) = (\psi_\uparrow^\dagger(x)\psi_\downarrow(x)), \tag{2.3}$$

where $\tau_\pm = (\tau_1 \pm i\tau_1)/2$ and τ_3 are the Pauli matrices and $\Phi(x) = V\Psi^\dagger(x)\tau_-\Psi(x) = V\psi_\downarrow\psi_\uparrow$ is the complex order field.

We consider the full fermion Green's function in the Matsubara finite-temperature formalism

$$G(x) = \langle \Psi(x)\Psi^\dagger(0)\rangle. \tag{2.4}$$

For the 3D case of the BCS theory, the frequency–momentum representation for (2.4) in the mean-field approximation is known to be [31]

$$G(i\omega_n, \mathbf{k}) = -\frac{i\omega_n\hat{I} + \tau_3\xi(\mathbf{k}) - \tau_+\Phi - \tau_-\Phi^*}{\omega_n^2 + \xi^2(\mathbf{k}) + |\Phi|^2}, \tag{2.5}$$

where $\omega_n = (2n+1)\pi T$ is the odd (fermion) Matsubara frequency, $\xi(\mathbf{k})$ is the dispersion law of electrons evaluated from the chemical potential μ, and $\Phi \equiv \langle\Phi(x)\rangle$ is the complex order parameter.

A problem arises when one tries to apply equation (2.5) directly to 2D systems, since it has been proved (see [25]) that nonzero Φ values are forbidden. Nevertheless, one can assume that the modulus of the order parameter $\rho = |\Phi|$ has a nonzero value, while its phase $\theta(x)$, defined by

$$\Phi(x) = \rho(x)\exp[i\theta(x)] \tag{2.6}$$

is a random quantity. To be consistent with (2.6), one should also introduce the spin–charge variables for the Nambu spinors

$$\Psi(x) = \Upsilon(x)\exp\left[\frac{i\tau_3\theta(x)}{2}\right], \quad \Psi^\dagger(x) = \Upsilon^\dagger(x)\exp\left[-\frac{i\tau_3\theta(x)}{2}\right], \tag{2.7}$$

where Υ is the neutral fermion-field operator. The strategy of treating charge and spin (neutral) degrees of freedom as independent seems to be quite useful and, at the same time, a very general feature of 2D systems [24, 32].

Applying (2.7), we thus split the Green's function (2.4) into spin and charge parts

$$G_{\alpha\beta}(x) = \sum_{\alpha',\beta'}\mathcal{G}_{\alpha'\beta'}(x)\left\langle (e^{i\tau_3\theta(x)/2})_{\alpha\alpha'}(e^{-i\tau_3\theta(0)/2})_{\beta'\beta}\right\rangle, \tag{2.8}$$

where

$$\mathcal{G}_{\alpha\beta}(x) = \langle \Upsilon_\alpha(x)\Upsilon_\beta^\dagger(0)\rangle \tag{2.9}$$

is the Green's function for neutral fermions. Introducing the projectors $P_\pm = (1/2)(\hat{I} \pm \tau_3)$, we obtain

$$e^{i\tau_3\theta/2} = P_+e^{i\theta/2} + P_-e^{-i\theta/2}, \quad e^{-i\tau_3\theta/2} = P_-e^{i\theta/2} + P_+e^{-i\theta/2}, \tag{2.10}$$

so that (2.8) can be rewritten as

$$G(x) = \sum_{\alpha,\beta=\pm} P_\alpha\mathcal{G}(x)P_\beta\left\langle \exp\left[\frac{i\alpha\theta(x)}{2}\right]\exp\left[-\frac{i\beta\theta(0)}{2}\right]\right\rangle, \tag{2.11}$$

where $\alpha = \beta$ and $\alpha = -\beta$ correspond to the diagonal and nondiagonal parts of the Green's function, respectively.

For the frequency–momentum representation of (2.11), we have

$$G(i\omega_n, \mathbf{k}) = T\sum_{m=-\infty}^{\infty}\int\frac{d^2p}{(2\pi)^2} \times \sum_{\alpha,\beta=\pm} P_\alpha\mathcal{G}(i\omega_m, \mathbf{p})P_\beta D_{\alpha\beta}(i\omega_n - i\omega_m, \mathbf{k}-\mathbf{p}), \tag{2.12}$$

where

$$\mathcal{G}(i\omega_m, \mathbf{k}) = \int_0^{1/T} d\tau\int d^2r\,\mathcal{G}(\tau, \mathbf{r})\exp[i\omega_m\tau - i\mathbf{k}\cdot\mathbf{r}] \tag{2.13}$$

and

$$D_{\alpha\beta}(i\Omega_n, \mathbf{q}) = \int_0^{1/T} d\tau\int d^2r\exp(i\Omega_n\tau - i\mathbf{q}\cdot\mathbf{r}) \times\left\langle \exp\left[\frac{i\alpha\theta(\tau, \mathbf{r})}{2}\right]\exp\left[-\frac{i\beta\theta(0)}{2}\right]\right\rangle \tag{2.14}$$

is the correlation function of phase fluctuations with even (boson) frequencies $\Omega_n = 2\pi nT$.

There is a good reason to believe (see [22]) that for T close to T_{BKT}, the fluctuations of the order-parameter modulus ρ (the so-called longitudinal fluctuations,

which in fact correspond to carrier-density fluctuations and undoubtedly must be taken into account in the very underdoped region)[1] are irrelevant and one can safely use the Green's function (2.13) of neutral fermions in the mean-field approximation [compare with (2.5)]

$$\mathcal{G}(i\omega_n, \mathbf{k}) = -\frac{i\omega_n\hat{I} + \tau_3\xi(\mathbf{k}) - \tau_1\rho}{\omega_n^2 + \xi^2(\mathbf{k}) + \rho^2}. \qquad (2.15)$$

Here, $\xi(\mathbf{k}) = \mathbf{k}^2/2m - \mu$, with $\mathbf{k}$ being a 2D vector and $\rho \equiv \langle\rho(x)\rangle$. Note that in [21, 22], $\rho(x)$ was treated only in the mean-field approximation, which means that fluctuations in both $\rho(x)$ and $\theta(x)$ were neglected, and therefore a second-order phase transition was obtained at T_ρ. However, as stressed in the Introduction, experimentally, the formation of the pseudogap phase does not display any sharp transition and the temperature T^* observed in various experiments is to be considered as a characteristic energy scale rather than as a temperature where the pseudogap is reduced to zero [33]. We believe that taking the $\rho(x)$ fluctuations into account may resolve the discrepancy between the experimental behavior of T^* and the temperature T_ρ introduced in the theory.

3. THE CORRELATION FUNCTION FOR THE PHASE FLUCTUATIONS

As stated above, we expect the phase fluctuations to be responsible for the difference between properties of the charged and neutral fermions defined above. The latter are described by the Green's function (2.15), which coincides with the BCS Green's function (2.5) only under the assumption that the phase θ of the order parameter $\Phi = \rho\exp(i\theta)$ is a constant and can be chosen to vanish. This is not the case for the 2D model, where there is a decay of the phase correlations and the Green's functions of charged and neutral fermions are nontrivially related via equation (2.12). To establish their relationship, one must know the correlation function for the phase fluctuations. Its calculation is quite straightforward for $T < T_{BKT}$, while for $T > T_{BKT}$ one can apply the results of the BKT transition theory [34].

3A. The Correlation Function for $T < T_{BKT}$

In the superconducting phase, the free vortex excitations are absent and the exponential correlation function is easily expressed in terms of the Green's function

$$D_\theta(x) = \langle\theta(x)\theta(0)\rangle \qquad (3.1)$$

(here, as above, $x \equiv \tau, \mathbf{r}$) via the Gaussian functional integral

$$D_{\alpha\beta}(x) = \int\mathcal{D}\theta(x)\exp\left\{-\int_0^{1/T} d\tau_1\int d^2r \times\left[\frac{1}{2}\theta(x_1)D_\theta^{-1}(x_1)\theta(x_1) + I(x_1)\theta(x_1)\right]\right\}$$
$$= \exp\left[-\frac{1}{2}\int_0^{1/T} d\tau_1\int_0^{1/T} d\tau_2\int d^2r_1\int d^2r_2 \times I(\tau_1, \mathbf{r}_1)D_\theta(\tau_1 - \tau_2, \mathbf{r}_1 - \mathbf{r}_2)I(\tau_2, \mathbf{r}_2)\right], \qquad (3.2)$$

with the source

$$I(x_1) = -i\frac{\alpha}{2}\delta(\tau_1 - \tau)\delta(\mathbf{r}_1 - \mathbf{r}) + i\frac{\beta}{2}\delta(\tau_1)\delta(\mathbf{r}_1), \quad \alpha, \beta = \pm. \qquad (3.3)$$

The Green's function

$$D_\theta^{-1}(x) = -J(\mu, T, \rho)\nabla_r^2 - K(\mu, T, \rho)(\partial_\tau)^2 \qquad (3.4)$$

for this model was found in [22]. Note that, here, the superfluid stiffness J and compressibility K are functions of μ, T, and ρ and also that the Green's function (3.4) includes only the lowest derivatives of the phase θ. The higher terms are also present in the expansion, but we neglect them. In the simplest case, $J(\mu, T, \rho) \sim n_f$, the density of carriers, and $K(\mu, T, \rho) \sim$ const [22].

Substituting (3.4) into (3.2), we obtain

$$D_{\alpha\beta}(x) = \exp\left[-\frac{T}{4}\sum_{n=-\infty}^{\infty}\int\frac{d^2q}{(2\pi)^2}\frac{1 - \alpha\beta\cos(\mathbf{q}\cdot\mathbf{r} - \Omega_n\tau)}{Jq^2 + K\Omega_n^2}\right]. \qquad (3.5)$$

It is easy to see that for zero frequency $\Omega_n = 0$, the integral in equation (3.5) is divergent at $\mathbf{q} = 0$ unless $\alpha = \beta$, and therefore only two terms survive in the sum over α and β in equation (2.12), namely

$$P_-\mathcal{G}(i\omega_n, \mathbf{k})P_- + P_+\mathcal{G}(i\omega_n, \mathbf{k})P_+ = -\frac{i\omega_n\hat{I} + \tau_3\xi(\mathbf{k})}{\omega_n^2 + \xi^2(\mathbf{k}) + \rho^2}. \qquad (3.6)$$

It is important that the terms like $P_\pm\mathcal{G}(i\omega_n, \mathbf{k})P_\mp$, which are proportional to τ_1 and thus violate the gauge symmetry, do not contribute to equation (2.12) due to vanishing of the corresponding D_{+-} and D_{-+} correlation functions standing after them. This explicitly demonstrates that the nondiagonal part of the 2D Green's function vanishes at all finite temperatures. Thus, making use of the Gor'kov equations to calculate its diagonal part and the

[1] This in turn means that the approximation used proves to be justified not far from optimal doping.

gap function is questionable. For nonzero correlation functions, we have

$$D(x) \equiv D_{++}(x) = D_{--}(x) = \exp\left[-\frac{T}{4}\sum_{n=-\infty}^{\infty}\int\frac{qdqd\varphi}{(2\pi)^2}\frac{1-\cos(qr\cos\varphi)\cos\Omega_n\tau}{Jq^2+K\Omega_n^2}\right]. \quad (3.7)$$

In what follows, we consider in detail only the static case $\tau = 0$. The restriction to this case is one of the few main assumptions we use throughout the paper.

The summation over n and the integration with respect to φ in (3.7) can readily be done, yielding the following exponent of (3.7):

$$-\frac{1}{16\pi\sqrt{JK}}\int_0^\infty dqe^{-q/\Lambda}[1-J_0(qr)]\tanh\frac{qr_0}{4}, \quad (3.8)$$

where we introduced the scale

$$r_0 = \frac{2}{T}\sqrt{\frac{J}{K}}, \quad (3.9)$$

which is a function of the variables used (in the simplest case, $r_0 \sim \sqrt{n_f}/T$). In (3.8), we introduced the cutoff Λ by means of the exponential function. This cutoff represents the maximum possible momentum in the theory, i.e., the Brillouin momentum.

One can derive from (3.8) (see Appendix A) the following asymptotic expressions:

$$D(0,\mathbf{r}) \sim \begin{cases}\left(\frac{r}{r_0}\right)^{-T/8\pi J}, & r \gg r_0 \gg \Lambda^{-1},\\ \left(\frac{\Lambda r}{2}\right)^{-T/8\pi J}, & r \gg \Lambda^{-1} \gg r_0.\end{cases} \quad (3.10)$$

This long-distance behavior governs the physics of θ fluctuations that we intend to study in what follows.

We now discuss the meaning of the value r_0. Again using the phase stiffness $J(T=0)$ and compressibility K from [22], we readily obtain $r_0 = 2\sqrt{\epsilon_F/m}/T$, which is the single-particle thermal de Broglie wavelength ($\epsilon_F = \pi n_f/m$ is the Fermi energy). Then, assuming that $T \sim T_{BKT}$ and taking $T_{BKT} \simeq \epsilon_F/8$ [21, 22], we can estimate

$$r_0 \sim \frac{16}{\sqrt{\epsilon_F m}} = \frac{16\sqrt{2}}{k_F}, \quad (3.11)$$

where k_F is the Fermi momentum. The value of k_F for cuprates is less than the Brillouin momentum Λ, which is why the first case in (3.10) seems to be more relevant.

There is another way to estimate r_0: we can use the value $2\Delta/T_c$; and hence,

$$r_0 \sim \sqrt{2}\pi\frac{2\Delta}{T_c}\xi_0, \quad (3.12)$$

where $\xi_0 = v_F/(\pi\Delta)$ is the BCS coherence length. This shows that r_0 has the meaning of a coherence length, which appears to be rather natural, since the minimum size of the phase coherence region should be of the order of ξ_0. Since the coherence length in cuprates is larger than the lattice spacing Λ^{-1}, the first case in (3.10) is applicable. Therefore, for $T < T_{BKT}$ and for static fluctuations, we have

$$D(\mathbf{r}) = \left(\frac{r}{r_0}\right)^{-T/8\pi J}, \quad (3.13)$$

where $r_0 = 16/\sqrt{\epsilon_F m}$.

3B. The Correlation Function for $T > T_{BKT}$

For $T > T_{BKT}$, the expression for the static correlation function (3.13) can be generalized using the well-known results of the BKT transition theory [34, 35]:

$$D(\mathbf{r}) = \left(\frac{r}{r_0}\right)^{-T/8\pi J}\exp\left(-\frac{r}{\xi_+(T)}\right), \quad (3.14)$$

where

$$\xi_+(T) = C\exp\sqrt{\frac{T_\rho - T}{T - T_{BKT}}} \quad (3.15)$$

is the BKT coherence length and C is a constant whose value is discussed later. One can consider equation (3.14) as a general representation for $D(\mathbf{r})$ for both $T > T_{BKT}$ and $T < T_{BKT}$ if the coherence length $\xi_+(T)$ is considered to be infinite for $T < T_{BKT}$. The prefactor in equation (3.14) is related to the longitudinal (spin-wave) phase fluctuations, while the exponent is responsible for the transverse (vortex) excitations, which are present only above T_{BKT}. The prefactor appears to be important for the non-Fermi-liquid behavior discussed in what follows. Note, however, that the longitudinal phase fluctuations can be suppressed by the Coulomb interaction [30], which is not included in the present simple model. One further comment is that, while the approximation used to study the vortex fluctuations in [30] is good for T well above T_{BKT}, the form of the correlation function D is appropriate for T close to T_{BKT}.

The constant C can be estimated from the condition that $\xi_+(T)$ cannot be much less than the parameter r_0, which is a natural cutoff in the theory, and we thus take $C = r_0/4$ in our numerical calculations. In any case, for $T \gtrsim T_{BKT}$, where (3.15) is valid, the value $\xi_+(T)$ is large and not very sensitive to the initial value of C.

There also exists a dynamical generalization of (3.14) proposed from phenomenological backgrounds in [36]:

$$D(t,\mathbf{r}) = \exp(-\gamma t)\left(\frac{r}{r_0}\right)^{-T/8\pi J}\exp\left(-\frac{r}{\xi_+(T)}\right). \quad (3.16)$$

Note that t is the real time and γ is the decay constant, and therefore (3.16) is the retarded Green's function. We hope to consider the more general case of dynamical phase fluctuations (3.16) elsewhere.

3C. The Fourier Transform of $D(\mathbf{r})$

For the Fourier transform (2.14) of (3.14), we have

$$D(i\Omega_n, \mathbf{q}) = \int_0^{1/T} d\tau \int d^2 r \exp(i\Omega_n \tau - i\mathbf{q}\cdot\mathbf{r}) \times \left(\frac{r}{r_0}\right)^{-T/8\pi J} \exp\left(-\frac{r}{\xi_+(T)}\right) = 2\pi \frac{\delta_{n,0}}{T} r_0^{T/8\pi J} \times \int_0^{\infty} dr r^{1-T/8\pi J} J_0(qr) \exp\left(-\frac{r}{\xi_+(T)}\right). \quad (3.17)$$

The integral (3.17) can be calculated (see, for example, [37]) with the result

$$D(i\Omega_n, \mathbf{q}) = \frac{\delta_{n,0}}{T} \frac{2\pi r_0^{2(1-\alpha)} \Gamma(2\alpha)}{[q^2 + (1/\xi_+)^2]^{\alpha}} \times {}_2F_1\left(\alpha, -\alpha + \frac{1}{2}; 1; \frac{q^2}{q^2 + (1/\xi_+)^2}\right). \quad (3.18)$$

The hypergeometric function $F(a, b; c; z)$ can be well approximated by a constant, since it is slowly varying at all values of $\mathbf{q}$. We can take the value of the hypergeometric function at $q = \infty$ for this constant. Thus, we have

$$D(i\Omega_n, \mathbf{q}) = \frac{\delta_{n,0}}{T} A \left[q^2 + \left(\frac{1}{\xi_+}\right)^2\right]^{-\alpha}, \quad (3.19)$$

where

$$A \equiv \frac{4\pi\Gamma(\alpha)}{\Gamma(1-\alpha)} \left(\frac{2}{r_0}\right)^{2(\alpha-1)}, \quad \alpha \equiv 1 - \frac{T}{16\pi J}. \quad (3.20)$$

It should be stressed that for $T > T_{\mathrm{BKT}}$, the parameter α quickly deviates from unity as ϵ_{F} decreases; in other words, the underdoped region has to reveal highly nonstandard properties in comparison with the overdoped one.

Note that for $\xi_+^{-1} = 0$ ($T < T_{\mathrm{BKT}}$), equation (3.19) is an exact Fourier transform of the correlation function (3.13).

One should take into account that even for $T < T_{\mathrm{BKT}}$, propagator (3.19) does not have the canonical behavior $1/q^2$, which is typical, for example, for the Bogolyubov mode in dimensions $d > 2$. In 2D, the modes with a propagator $1/q^2$ would lead to severe infrared singularities [25]; to avoid them, these modes transform into softer ones ($1/q^{2\alpha}$, $\alpha < 1$).

Finally, substituting (3.6) and (3.19) in (2.12), we obtain

$$G(i\omega_n, \mathbf{k}) = -A \int \frac{d^2 q}{(2\pi)^2} \times \frac{i\omega_n + \tau_3 \xi(\mathbf{q})}{\omega_n^2 + \xi^2(\mathbf{q}) + \rho^2} \frac{1}{[(\mathbf{k}-\mathbf{q})^2 + (1/\xi_+)^2]^{\alpha}}. \quad (3.21)$$

The coincidence of the Matsubara frequency on the left- and right-hand sides of equation (3.21) is evidently related to the static approximation used in this paper. As we will see in the next sections, the Green's function (3.21), spectral density, and the density of states can be evaluated exactly.

4. THE DERIVATION OF THE FERMION GREEN'S FUNCTION

The fermion Green's function can be calculated analytically if we split the fermion part of (3.21) as

$$\frac{i\omega_n \hat{I} + \tau_3 \xi(\mathbf{k})}{\omega_n^2 + \xi^2(\mathbf{k}) + \rho^2} = \frac{A_1}{\xi(\mathbf{k}) + i\sqrt{\omega_n^2 + \rho^2}} + \frac{A_2}{\xi(\mathbf{k}) - i\sqrt{\omega_n^2 + \rho^2}}, \quad (4.1)$$

where

$$A_1 = \frac{1}{2}\left(\tau_3 - \frac{\omega_n}{\sqrt{\omega_n^2 + \rho^2}}\right), \quad A_2 = \frac{1}{2}\left(\tau_3 + \frac{\omega_n}{\sqrt{\omega_n^2 + \rho^2}}\right). \quad (4.2)$$

Using the representations

$$\frac{1}{a \pm ib} = \mp i \int_0^{\infty} ds \exp[\pm is(a \pm ib)], \quad (4.3)$$

$$\frac{1}{c^{\alpha}} = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} dt t^{\alpha-1} e^{-ct} \quad (4.4)$$

and taking (4.1) into account, we can rewrite (3.21) as

$$G(i\omega_n, \mathbf{k}) = \frac{iA}{\Gamma(\alpha)} \int_0^{\infty} ds \int_0^{\infty} dt t^{\alpha-1} \exp[-\xi_+^{-2} t - s\sqrt{\omega_n^2 + \rho^2}] \times \int \frac{d^2 q}{(2\pi)^2} \left\{ A_1 \exp\left[is\frac{q^2}{2m} - i\mu s - (\mathbf{k}-\mathbf{q})^2 t\right] - A_2 \exp\left[-is\frac{q^2}{2m} + i\mu s - (\mathbf{k}-\mathbf{q})^2 t\right] \right\}. \quad (4.5)$$

Note that the special form of the integral representation (4.3) [compare with representation (4.4)] guarantees that the Gaussian integral with respect to $\mathbf{q}$ is well defined independently of the sign of $\xi(\mathbf{q}) = \mathbf{q}^2/2m - \mu$. Now the Gaussian integration with respect to momenta $\mathbf{q}$ in (4.5) can be done explicitly:

$$G(i\omega_n, \mathbf{k}) = \frac{iA}{4\pi\Gamma(\alpha)}\int_0^\infty ds \int_0^\infty dt t^{\alpha-1} \times \exp[-\xi_+^{-2}t - s\sqrt{\omega_n^2 + \rho^2}] \times \left[\frac{A_1}{t - is/2m}\exp\left(i\frac{\mathbf{k}^2}{2m}\frac{st}{t - is/2m} - i\mu s\right) - \frac{A_1}{t + is/2m}\exp\left(-i\frac{\mathbf{k}^2}{2m}\frac{st}{t + is/2m} + i\mu s\right)\right]. \quad (4.6)$$

Changing the variables as $s \longrightarrow 2ms$ and further as $t \longrightarrow st$, we can integrate with respect to s with the result

$$G(i\omega_n, \mathbf{k}) = \frac{imA}{2\pi}\int_0^\infty dt t^{\alpha-1}\left\{[A_1(t-i)^{\alpha-1}][\xi_+^{-2}t(t-i) + 2m\sqrt{\omega_n^2 + \rho^2}(t-i) - it\mathbf{k}^2 + 2im\mu(t-i)]^{-\alpha} - [A_2(t+i)^{\alpha-1}][\xi_+^{-2}t(t+i) + 2m\sqrt{\omega_n^2 + \rho^2}(t+i) + it\mathbf{k}^2 - 2im\mu(t+i)]^{-\alpha}\right\}. \quad (4.7)$$

In the general case where $\xi_+^{-1} \neq 0$, the denominator of (4.7) is quadratic in t and some further transformations are needed. Replacing $t \longrightarrow -iu$ and expanding the quadratic polynomial in the denominator, we have

$$G(i\omega_n, \mathbf{k}) = -\frac{Am\xi_+^{2\alpha}}{2\pi}\left\{\int_0^{i\infty} du \frac{A_1 u^{\alpha-1}(u+1)^{\alpha-1}}{[(u+u_1)(u+u_2)]^\alpha} + \int_0^{-i\infty} du \frac{A_2 u^{\alpha-1}(u+1)^{\alpha-1}}{[(u+\tilde{u}_1)(u+\tilde{u}_2)]^\alpha}\right\}, \quad (4.8)$$

where

$$u_1 = m\xi_+^2\left(\frac{k^2\xi_+^2 + 1}{2m\xi_+^2} - \mu + i\sqrt{\omega_n^2 + \rho^2} + \sqrt{D}\right),$$
$$u_2 = m\xi_+^2\left(\frac{k^2\xi_+^2 + 1}{2m\xi_+^2} - \mu + i\sqrt{\omega_n^2 + \rho^2} - \sqrt{D}\right), \quad (4.9)$$

with

$$D \equiv \left(\frac{k^2\xi_+^2 + 1}{2m\xi_+^2} - \mu + i\sqrt{\omega_n^2 + \rho^2}\right)^2 + \frac{2}{m\xi_+^2}(\mu - i\sqrt{\omega_n^2 + \rho^2}) \quad (4.10)$$

and

$$\tilde{u}_i = \left(u_i\sqrt{\omega_n^2 + \rho^2} \longrightarrow -\sqrt{\omega_n^2 + \rho^2}\right). \quad (4.11)$$

We can verify from (4.9) that $\mathrm{Re}\, u_i > 0$ for $\mu < 0$, and therefore we can rotate the integration contour to the real axis:

$$G(i\omega_n, \mathbf{k}) = -\frac{Am\xi_+^{2\alpha}}{2\pi}\left\{\int_0^\infty du \frac{A_1 u^{\alpha-1}(u+1)^{\alpha-1}}{[(u+u_1)(u+u_2)]^2} + \left(\sqrt{\omega_n^2 + \rho^2} \longrightarrow -\sqrt{\omega_n^2 + \rho^2}\right)\right\}. \quad (4.12)$$

The integral representation (4.12) can then be analytically continued to $\mu > 0$. The change of the variable $z = u/(u+1)$ allows equation (4.12) to be expressed in terms of Appell's function [38]

$$F_1(\alpha, \beta, \beta', \gamma; x, y) = \frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\gamma - \alpha)} \times \int_0^1 \frac{z^{\alpha-1}(1-z)^{\gamma-\alpha-1}}{(1-zx)^\beta(1-zy)^{\beta'}}dz, \quad (4.13)$$

and hence,

$$G(i\omega_n, \mathbf{k}) = -\frac{Am\xi_+^{2\alpha}}{2\pi\alpha}\left[\frac{A_1}{(u_1u_2)^\alpha} \times F\left(\alpha, \alpha, \alpha; \alpha+1; \frac{u_1 - 1}{u_1}, \frac{u_2 - 1}{u_2}\right) + \left(\sqrt{\omega_n^2 + \rho^2} \longrightarrow -\sqrt{\omega_n^2 + \rho^2}\right)\right]. \quad (4.14)$$

For $T < T_{\mathrm{BKT}}$, the BKT coherence length is infinite ($\xi_+^{-1} = 0$), which means $(u_1 - 1)/u = 1$ in the first argument of the Appell's function. This allows us to apply the reduction formula [38]

$$F(\alpha, \beta, \beta', \gamma; x, 1) = \frac{\Gamma(\gamma)\Gamma(\gamma - \alpha - \beta')}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta')}F_1(\alpha, \beta; \gamma - \beta'; x) \quad (4.15)$$

and express the result via the hypergeometric function

$$G(i\omega_n, \mathbf{k}) = -\Gamma^2(\alpha)\left(\frac{2}{mr_0^2}\right)^{\alpha-1} \times \left\{ \frac{A_1}{[-(\mu - i\sqrt{\omega_n^2+\rho^2})]^\alpha} \times {}_2F_1\left(\alpha, \alpha; 1; \frac{k^2/2m}{\mu - i\sqrt{\omega_n^2+\rho^2}}\right) + \frac{A_2}{[-(\mu + i\sqrt{\omega_n^2+\rho^2})]^\alpha} {}_2F_1\left(\alpha, \alpha; 1; \frac{k^2/2m}{\mu + i\sqrt{\omega_n^2+\rho^2}}\right)\right\}, \tag{4.16}$$

where we inserted the value of A from (3.20).

This completes our derivation of the temperature-dependent fermion Green's function.

4A. The Retarded Fermion Green's Function

To obtain the spectral density, we need to obtain the retarded real-time Green's function from the temperature-dependent Green's function by means of analytic continuation $i\omega_n \longrightarrow \omega + i0$, and where $\sqrt{\omega_n^2+\rho^2} \longrightarrow i\sqrt{\omega^2-\rho^2}$. This results in the following rules (compare with (4.2), (4 9), and (4.10)):

$$A_1 \longrightarrow \mathscr{A}_1 = \frac{1}{2}\left(\tau_3 + \frac{\omega}{\sqrt{\omega^2-\rho^2}}\right), \quad A_2 \longrightarrow \mathscr{A}_2 = \frac{1}{2}\left(\tau_3 - \frac{\omega}{\sqrt{\omega^2-\rho^2}}\right); \tag{4.17}$$

$$u_1 \longrightarrow v_1 = m\xi_+^2\left(\frac{k^2\xi_+^2+1}{2m\xi_+^2} - \mu - \sqrt{\omega^2-\rho^2} + \sqrt{\mathscr{D}}\right), \quad u_2 \longrightarrow v_2 = m\xi_+^2\left(\frac{k^2\xi_+^2+1}{2m\xi_+^2} - \mu - \sqrt{\omega^2-\rho^2} - \sqrt{\mathscr{D}}\right); \tag{4.18}$$

with

$$D \longrightarrow \mathscr{D} = \left(\frac{k^2\xi_+^2+1}{2m\xi_+^2} - \mu - \sqrt{\omega^2-\rho^2}\right)^2 + \frac{2}{m\xi_+^2}\left(\mu + \sqrt{\omega^2-\rho^2}\right) \tag{4.19}$$

and

$$\tilde{v}_i = v_i\left(\sqrt{\omega^2-\rho^2} \longrightarrow -\sqrt{\omega^2-\rho^2}\right). \tag{4.20}$$

For the retarded Green's function, we thus have

$$G(\omega, \mathbf{k}) = -\frac{Am\xi_+^{2\alpha}}{2\pi\alpha}\left\{ \frac{\mathscr{A}_1}{(v_1 v_2)^\alpha} \times F_1\left(\alpha, \alpha, \alpha; \alpha+1; \frac{v_1-1}{v_1}, \frac{v_2-1}{v_2}\right) + \left(\sqrt{\omega^2-\rho^2} \longrightarrow -\sqrt{\omega^2-\rho^2}\right)\right\}. \tag{4.21}$$

It is easy to see that

$$v_1 v_2 = -2m\xi_+^2(\mu + \sqrt{\omega^2-\rho^2}). \tag{4.22}$$

We now discuss the condition under which the imaginary part of $G(\omega + i0, \mathbf{k})$ is nonvanishing.

For $|\omega| < \rho$, we can see that $\tilde{v}_1 = v_1^*$, $\tilde{v}_2 = v_2^*$, and therefore $G(\omega, \mathbf{k})$ is real and $\mathrm{Im}\,G(\omega + i0, \mathbf{k}) = 0$. The case where $|\omega| > \rho$ is more complicated. It follows from the Appel function transformation property [38]

$$F_1(\alpha, \beta, \beta', \gamma; x, y) = (1-x)^\alpha \times F_1\left(\alpha, \gamma-\beta-\beta', \beta', \gamma; \frac{x}{x-1}, \frac{y-x}{1-x}\right), \tag{4.23}$$

that for real x and y, the function F_1 becomes complex if $x > 1$ or/and $y > 1$. This implies that $G(\omega, \mathbf{k})$ has an imaginary part if $v_1 < 0$ or/and $v_2 < 0$. Looking at the expressions (4.18) for v_1 and v_2, we can see that v_1 is always positive, whereas v_2 may be negative. This means that $G(\omega, \mathbf{k})$ has a nonvanishing imaginary part if $v_1 v_2 < 0$. Using (4.22), the condition for the existence of a nonzero imaginary part of $G(\omega, \mathbf{k})$ can then be written as $\mu + \sqrt{\omega^2-\rho^2} > 0$.

4B. The Branch Cut Structure of $G(\omega, \mathbf{k})$ and a Non-Fermi-Liquid Behavior

We now consider the retarded fermion Green's function (4.16) for $T < T_{\mathrm{BKT}}$. Applying the analytic continuation rules from the previous subsection to equation (4.16), we obtain

$$G(\omega, \mathbf{k}) = -\Gamma^2(\alpha)\left(\frac{2}{mr_0^2}\right)^{\alpha-1}\left[\frac{\mathscr{A}_1}{[-(\mu + i\sqrt{\omega^2+\rho^2})]^\alpha} \times {}_2F_1\left(\alpha, \alpha; 1; \frac{k^2/2m}{\mu + \sqrt{\omega^2-\rho^2}}\right) + \frac{\mathscr{A}_2}{[-(\mu - \sqrt{\omega^2-\rho^2})]^\alpha} {}_2F_1\left(\alpha, \alpha; 1; \frac{k^2/2m}{\mu - \sqrt{\omega^2-\rho^2}}\right)\right]. \tag{4.24}$$

Near the quasiparticle peaks where $\omega \simeq \pm E(\mathbf{k})$, the arguments of the hypergeometric function in (4.24) are

close to 1. One can consider, for instance, the first hypergeometric function; then

$$z_1 \equiv \frac{k^2/2m}{\mu + \sqrt{\omega^2 - \rho^2}} \simeq 1. \tag{4.25}$$

Using the relation between the hypergeometric functions [38]

$$\begin{aligned} {}_2F_1(a, b; c; z) &= \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} \\ &\times {}_2F_1(a, b; a+b+1-c; 1-z) \\ &+ \frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)}(1-z)^{c-a-b} \\ &\times {}_2F_1(ca, cb; c+1-a-b; 1-z), \end{aligned} \tag{4.26}$$

we obtain that, near $z_1 \simeq 1$,

$$\begin{aligned} G(\omega, \mathbf{k}) \sim -\Gamma^2(\alpha)\left(\frac{2}{mr_0^2}\right)^{\alpha-1} &\frac{\mathcal{A}_1}{[-(\mu + \sqrt{\omega^2 - \rho^2})]^{\alpha}} \\ \times \left\{ \frac{\Gamma(1-2\alpha)}{\Gamma^2(1-\alpha)} + \frac{\Gamma(2\alpha-1)}{\Gamma^2(\alpha)} \right. &\left. \frac{1}{(1-z_1)^{2\alpha-1}} \right\}. \end{aligned} \tag{4.27}$$

It can be seen that the expression for this Green's function is evidently a nonstandard one: besides containing a branch cut, it clearly displays its nonpole character. The latter in turn corresponds to non-Fermi-liquid behavior of the system as a whole. It must be stressed that non-Fermi-liquid peculiarities are tightly related to the charge (i.e., observable) fermions only, because the Green's function (2.15) of neutral fermions has a typical (pole-type) BCS form. In addition, it follows from (4.27) that new properties appear as a consequence of the θ-particle presence (leading to $\alpha \neq 1$); and because the parameter α is a function of T [see (3.20)], the non-Fermi-liquid behavior is developed with increasing temperature and is preserved until ρ vanishes.

It is interesting that in Anderson's theory [39], it was postulated that the Fermi liquid theory is broken down in the normal state as a result of strong correlations. Here, we started from the Fermi liquid theory and found that it is broken down due to strong phase fluctuations. As suggested in [39], the non-Fermi-liquid behavior may lead to suppression of the coherent tunneling between layers, which in turn confines carriers in the layers and leads to strong phase fluctuations. In contrast to [39], however, our model predicts the restoration of the Fermi liquid behavior as T decreases, since $\alpha \longrightarrow 1$ as $T \longrightarrow 0$ (see the discussion in Section *5C*, item 4).

The $T = 0$ limit can also be obtained as follows. Strictly speaking, one cannot estimate the value of r_0 as $T \longrightarrow 0$ in (4.24) via equation (3.11), because the substitution of $T_{BKT} \simeq \epsilon_F/8$ in (3.9) is not valid in this case. However, this is not essential, because $T/8\pi J \longrightarrow 0$, so that correlation function (3.13), $D(\mathbf{r}) \longrightarrow 1$, which evidently means the formation of a long-range order in the system. Furthermore, the value of α in (3.20) goes to unity as $T \longrightarrow 0$ and the hypergeometric function in (4.24) reduces to the geometrical series

$${}_2F_1(1, 1; 1; z) = \frac{1}{1-z}. \tag{4.28}$$

Therefore, inserting (4.28) in (4.24), we obtain the standard BCS expression

$$G_{11}(\omega, \mathbf{k}) = \frac{\omega + \xi(\mathbf{k})}{\omega^2 - \xi^2(\mathbf{k}) - \rho^2} \tag{4.29}$$

for the diagonal component $G_{11}(\omega, \mathbf{k})$ of the Nambu–Gor'kov Green's function $G(\omega, \mathbf{k})$.

Evidently, equation (4.29) results in the standard BCS spectral density [31] with two δ-function peaks

$$\begin{aligned} A(\omega, \mathbf{k}) &= \frac{1}{2}\left[1 + \frac{\xi(\mathbf{k})}{E(\mathbf{k})}\right]\delta(\omega - E(\mathbf{k})) \\ &+ \frac{1}{2}\left[1 - \frac{\xi(\mathbf{k})}{E(\mathbf{k})}\right]\delta(\omega + E(\mathbf{k})), \end{aligned} \tag{4.30}$$

where $E(\mathbf{k}) = \sqrt{\xi^2(\mathbf{k}) + \rho^2}$. To recover the nondiagonal components of G, one has to restore the correlation functions $D_{-+}(\mathbf{r})$ and $D_{+-}(\mathbf{r})$ that were omitted in Section 3A.

5. THE SPECTRAL DENSITY OF THE FERMION GREEN'S FUNCTION

As is well known, [31], the spectral features of any system are entirely controlled by its spectral density

$$A(\omega, \mathbf{k}) = -\frac{1}{\pi}\mathrm{Im}G_{11}(\omega + i0, \mathbf{k}), \tag{5.1}$$

which, for example, for cuprates is measured in ARPES experiments (see [29]). This function defines the spectrum anisotropy, the presence of a gap, the DOS, etc. In what follows, we calculate $A(\omega, \mathbf{k})$ for the Green's function obtained above.

5A. Analytic Expression for the Spectral Density

For $v_1 > 0$ and $v_2 < 0$, the retarded fermion Green's function (4.21) can be rewritten (see Appendix B) as

$$\begin{aligned} G(\omega, \mathbf{k}) &= -\frac{Am\xi_+^{2\alpha}}{2\pi} \\ \times \left\{ \mathcal{A}_1 \right. &\left[\frac{(-1)^{\alpha}\Gamma(\alpha)\Gamma(1-\alpha)}{[v_1(1-v_2)]^{\alpha}} {}_2F_1\left(\alpha, \alpha; 1; \frac{v_2(1-v_1)}{v_1(1-v_2)}\right) \right. \\ &+ \left. \frac{1}{|v_2|^{\alpha}} \frac{\Gamma(1-\alpha)}{\Gamma(2-\alpha)} F_1\left(1, \alpha, 1-\alpha; 2-\alpha; \frac{v_1}{v_2}, \frac{1}{u_2}\right) \right] \\ &+ \left. \left(\sqrt{\omega^2 - \rho^2} \longrightarrow -\sqrt{\omega^2 - \rho^2}\right) \right\}. \end{aligned} \tag{5.2}$$

Then, according to (5.1) the spectral density for the Green's function (5.2) has the form

$$A(\omega, \mathbf{k}) = \frac{Am\xi_+^{2\alpha}\sin(\pi\alpha)}{2\pi^2}\operatorname{sgn}\omega\theta(\omega^2-\rho^2)$$
$$\times\left[(\mathcal{A}_1)_{11}\frac{\Gamma(\alpha)\Gamma(1-\alpha)}{[v_1(1-v_2)]^\alpha}{}_2F_1\left(\alpha,\alpha;1;\frac{v_2(1-v_1)}{v_1(1-v_2)}\right)\right. \quad (5.3)$$
$$\left.\times\theta(\mu+\sqrt{\omega^2-\rho^2})-(\sqrt{\omega^2-\rho^2}\longrightarrow-\sqrt{\omega^2-\rho^2})\right].$$

Using the quadratic transformation for the hypergeometric function [38]

$${}_2F_1(a,b;a-b+1;z) = (1-z)^{-a}$$
$$\times{}_2F_1\left(\frac{a}{2},-b+\frac{a+1}{2};1+a-b;-\frac{4z}{(1-z)^2}\right), \quad (5.4)$$

the expression (3.20) for A, and equations (4.18) and (4.19), we finally obtain

$$A(\omega, \mathbf{k}) = \frac{\Gamma(\alpha)}{\Gamma(1-\alpha)}\left(\frac{2}{mr_0^2}\right)^{\alpha-1}\operatorname{sgn}\omega\theta(\omega^2-\rho^2)$$
$$\times\left[\frac{(\mathcal{A}_1)_{11}}{\mathcal{D}^{\alpha/2}}{}_2F_1\left(\frac{\alpha}{2},\frac{1-\alpha}{2};1;-4\frac{\frac{k^2}{2m}(\mu+\sqrt{\omega^2-\rho^2})}{\mathcal{D}}\right)\right. \quad (5.5)$$
$$\left.\times\theta(\mu+\sqrt{\omega^2-\rho^2})-(\sqrt{\omega^2-\rho^2}\longrightarrow-\sqrt{\omega^2-\rho^2})\right].$$

where the chemical potential μ can be, in principle, determined from the equation that fixes the carrier density [22]. Here, however, we assume that the carrier density is sufficiently high and $\mu = \epsilon_F$.

In the BCS theory, $A(\omega, \mathbf{k})$ given by equation (4.30) consists of two pieces that are the spectral weights of adding and removing a fermion from the system, respectively. Note that our splitting of $A(\omega, \mathbf{k})$ is different, since each term in (5.5) corresponds to both the addition and the removal of a fermion.

In the next subsections, we verify the sum rule for (5.5), plot it for different temperatures, and discuss the results.

5B. The Sum Rule for the Spectral Density

It is well known that for the exact Green's function $G(\omega, \mathbf{k})$, the spectral function (5.1) must satisfy the sum rule

$$\int_{-\infty}^{\infty} d\omega A(\omega, \mathbf{k}) = 1. \quad (5.6)$$

The Green's function calculated in (4.21) is, of course, approximate. This is related to the use of the long-distance asymptotic behavior (3.10) of the phase correlation function (3.7). This means that its Fourier transform (3.19) is, strictly speaking, valid for small $\mathbf{k}$ only, whereas we have integrated our expressions to infinity. Another approximation that we have made is the restriction to static phase fluctuations. Thus, it is important to check whether the sum rule (5.6) is satisfied with sufficient accuracy.

It is remarkable that for (5.5), the sum rule (5.6) can be tested analytically with the help of the techniques used in calculating $N(\omega)$ in Appendix C. We obtain

$$\int_{-\infty}^{\infty} d\omega A(\omega, \mathbf{k}) = \frac{\Gamma(\alpha)}{\Gamma(2-\alpha)}. \quad (5.7)$$

The numerical value of the integral at the temperatures of interest can be estimated as follows. Taking the phase stiffness $J = 2/\pi T_{BKT}$ at $T = T_{BKT}$, the value α from (3.20) is given by

$$\alpha \simeq 1 - \frac{1}{32}\frac{T}{T_{BKT}}, \quad T \sim T_{BKT} \quad (5.8)$$

for T close to T_{BKT}. In particular, $\alpha(T = T_{BKT}) = 31/32$ gives the following estimate for the right-hand side of (5.7): $\Gamma(\alpha)/\Gamma(2-\alpha) \simeq 1.037$. This shows that for $T \sim T_{BKT}$, the spectral density (5.5) is reasonably good at the temperatures of interest.

The parameter α can, however, differ strongly from unity at $T > T_{BKT}$ and in the underdoped regime.

5C. Results for the Spectral Density

The plots of the spectral density $A(\omega, \mathbf{k})$ given by (5.5) at temperatures below and above T_{BKT} are presented in Figs. 1–3. To draw these plots, we used the value of α from equation (5.8) and the mean-field value of ρ obtained from the corresponding equation in [21, 22]. From these figures and our analytic expressions, we can infer the following results:

(1) For $T < T_{BKT}$ (the case presented in Fig. 1), there are two highly pronounced quasiparticle peaks at $\omega = \pm E(\mathbf{k})$. They are simply related to the contribution of zeros of $\mathcal{D}$ [see equation (4.19)] to $A(\omega, \mathbf{k})$.

(2) We also observe two peaks at $\omega = \pm\rho$ when $k \neq k_F$ (for $k = k_F$, the value $E(\mathbf{k}_F) = \rho$, so that the two sets of peaks coincide). One can check that the divergence at these points is weaker than at the former peaks at $\omega = \pm E(\mathbf{k})$. In fact, these peaks are the result of the static and noninteracting approximation for the phase fluctuations used here. They are essential to satisfy the sum rule (5.6).

If the dynamical fluctuations are taken into account, it is clear that the "external" frequency ω in $A(\omega, \mathbf{k})$ is different from the "internal" frequency in $\mathcal{A}_1$ and $\mathcal{A}_2$ [see the discussion after equation (3.21) and compare it with equation (2.12)]. We believe that this additional summation over the "internal" frequency (which is

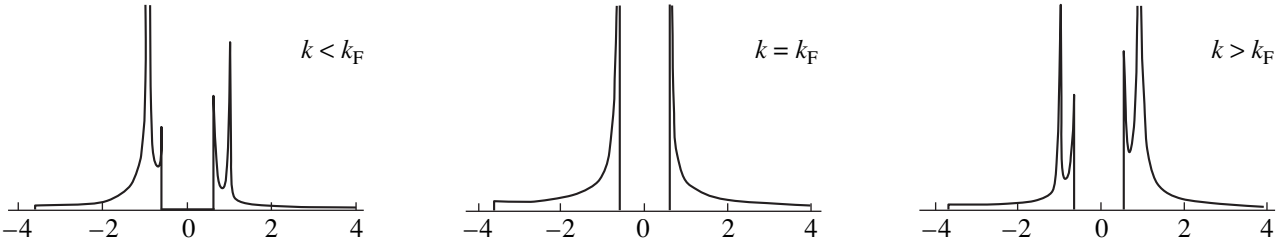


Fig. 1. Spectral function $A(\omega, \mathbf{k})$ versus ω in units of the zero-temperature gap Δ for $k < k_F$, $k = k_F$, and $k > k_F$ at $T = 0.99\,T_{BKT}$.

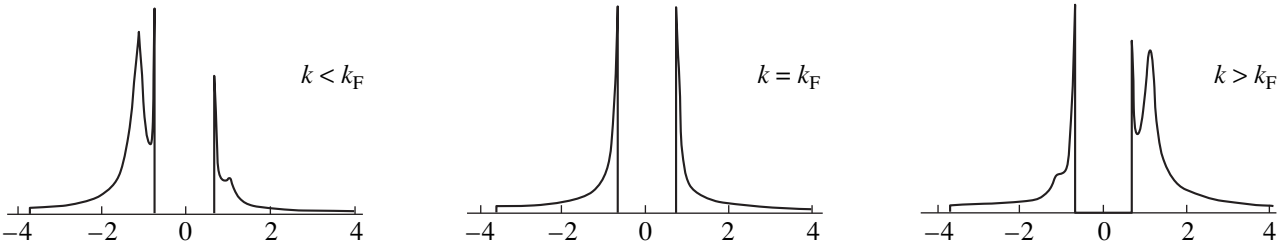


Fig. 2. Spectral function $A(\omega, \mathbf{k})$ versus ω in units of the zero-temperature gap Δ for $k < k_F$, $k = k_F$, and $k > k_F$ at $T = 1.043\,T_{BKT}$.

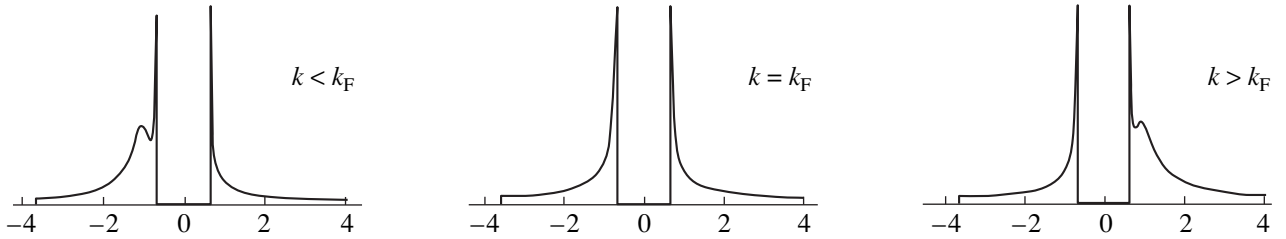


Fig. 3. Spectral function $A(\omega, \mathbf{k})$ versus ω in units of the zero-temperature gap Δ for $k < k_F$, $k = k_F$, and $k > k_F$ at $T = 1.088\,T_{BKT}$.

present if the dynamical fluctuations are considered) would considerably smear these peaks, moving the excess spectral weight inside the gap. This assumption is supported by the results of [36] (see item 3 below). The same effect can also be reached when the interaction between the charge and spin degrees of freedom is taken into account [30]. Note also that full cancellation of these peaks occurs in the $T = 0$ case given by equation (4.29).

(3) For $\omega < |\rho|$, we have $A(\omega, \mathbf{k}) = 0$ and a gap exists at all T (including $T > T_{BKT}$). This result is also a consequence of the static approximation used above. The dynamical fluctuations should fill the empty region, resulting in pseudogap formation in the normal state. Indeed, filling of the gap was obtained in a related calculation [36], where the correlation function $\langle \exp(i\theta(\mathbf{r}, t))\exp(-i\theta(0))\rangle$ (which differs from (3.16) only by the factor 1/2 multiplying the phase), which includes the dynamical phase fluctuations, was used in the numerical calculation of the self-energy of fermions and in the subsequent extraction of the spectral function from the fermion Green's function.

In the approximation used in the present paper, the spin and charge degrees of freedom are decoupled [see (2.8)]. However, this coupling can be included at the next stage of the approximation and also leads to pseudogap filling. Indeed, using the special form of the scattering rate proposed in [40]. It was obtained in [30] that $A(\omega, \mathbf{k}) \neq 0$, even for $\omega < |\rho|$. On the other hand, as stated above, there are also indications [36] that filling of the gap can be obtained by considering the dynamical phase fluctuations only. At present, it is not clear which of these gap-filling mechanisms plays the main role; this is the subject of our current investigations.

(4) The main peaks at $\omega = \pm E(\mathbf{k})$ have a finite temperature-dependent width which is, of course, related to the spin-wave (longitudinal) phase fluctuations. As $T \longrightarrow 0$, the width goes to zero, but this limit cannot

be correctly derived from (5.5), because this is an ordinary function, while the BCS spectral density (4.30) is a distribution. The correct limit can, however, be obtained for the integral of $A(\omega, \mathbf{k})$ (see Section 6, where the density of states is discussed). This sharpening of the peaks with decreasing T in the superconducting state was experimentally observed [29] and represents a striking difference from the BCS "pileup" (4.30) which is present for all $T < T_c$.

It was pointed out in [30] that the broadening of the spectral function caused by these fluctuations can be greater than the experimental data permit. This leads [30] to the conclusion that the spin-wave phase fluctuations are probably suppressed by the Coulomb interaction.

(5) For $T > T_{\mathrm{BKT}}$ (see Figs. 2, 3), one can see that the quasiparticle peaks at $\omega \simeq \pm E(\mathbf{k})$ are less pronounced as the temperature increases. Indeed, the value of $A(\omega, \mathbf{k})$ at $\omega = \pm E(\mathbf{k})$ is, in contrast to the case where $T < T_{\mathrm{BKT}}$, already finite. This is caused by the fact that $\mathscr{D} \neq 0$, since ξ_+ is already finite due to the influence of the vortex fluctuations. As the temperature is increased further, ξ_+ decreases, so that the quasiparticle peaks disappear (cf. Figs. 2, 3). This behavior qualitatively reproduces the ARPES studies of cuprates for the antinode direction [29] (see also [41]), which show that the quasiparticle spectral function broadens dramatically when passing from the superconducting to the normal state.

(6) It is important to stress that due to a very smooth dependence of ξ_+^{-1} on T [see (3.15)] as the temperature varies from $T < T_{\mathrm{BKT}}$ to $T > T_{\mathrm{BKT}}$, there is no sharp transition at the point $T = T_{\mathrm{BKT}}$. There is a smooth evolution of the superconducting (excitation) gap $\Delta_{SC} = \rho$ into the gap Δ_{PG}, which is also equal to ρ and in fact can be called a pseudogap, because the system is not superconducting at $T > T_{\mathrm{BKT}}$. This qualitatively fits the experiment [29, 33, 41] and appears to be completely different from the BCS theory [31], where the gap vanishes at $T = T_c$. As was already mentioned, the gap obtained at $T > T_{\mathrm{BKT}}$ occurs in the static approximation only and begins to be filled after dynamical fluctuations are taken into account (see, for example, [36]).

(7) Again, for $T > T_{\mathrm{BKT}}$, one has $A(\omega, \mathbf{k}) = 0$ when $|\omega| < \rho$ and we expect the gapped region to be filled by the dynamical phase fluctuations [36]. We predict, however, an essential difference between filling of the gap at $T > T_{\mathrm{BKT}}$ and $T < T_{\mathrm{BKT}}$. Indeed, due to the presence of the vortices above T_{BKT}, the value of the decay constant γ in equation (3.16) should be much larger than for $T < T_{\mathrm{BKT}}$. This and a nonzero value of ξ_+^{-1} above T_{BKT} may explain the break at $T = T_c$ in the scattering rate Γ_1 introduced in [40]. In general, it is interesting to establish a correspondence between the phenomenological parameters Γ_1 and Γ_0 introduced in [40] and the vortex parameters ξ_+ and γ used here. Note, however, that this correspondence cannot be simple because of the nonpole character of the Green's function derived here.

As mentioned above (see item 3), filling of the gap due to dynamical phase fluctuations is not the only possible mechanism for filling and the presence of vortices above T_{BKT} can be taken into account via coupling the spin and charge degrees of freedom [30]. It could also be that both these mechanisms are physically equivalent, since they relate gap filling to the presence of vortices in the system.

(8) Since we used the mean-field dependence $\rho(T)$, it is clear that the distance between the quasiparticle peaks (which is approximately equal to 2ρ) diminishes as T increases. This process of pseudogap closing is accompanied by the destruction of the quasiparticle peaks. It is also evident that for $\rho = 0$, the normal Fermi liquid behavior is immediately restored, because $J(\rho = 0) = 0$ [21, 22]. Recall, however, that the description proposed here cannot be applied when ρ is rather small, because, as already mentioned, the fluctuations of $\rho(x)$ also have to be taken into account in this region.

6. THE DENSITY OF STATES

The density of states can be found from the formula

$$N(\omega) = \int \frac{d^2k}{(2\pi)^2} A(\omega, \mathbf{k}) = N_0 \int_0^W d\frac{k^2}{2m} A(\omega, \mathbf{k}), \quad (6.1)$$

where $N_0 \equiv m/2\pi$ is the density of 2D states in the normal state (W is the bandwidth).

This integral can be calculated analytically (see Appendix C), which gives

$$N(\omega) = N_0 \frac{\Gamma(\alpha)}{\Gamma(2-\alpha)} \left(\frac{2}{mr_0^2}\right)^{\alpha-1} \operatorname{sgn}\omega\, \theta(\omega^2 - \rho^2)$$
$$\times \left\{ (\mathscr{A}_1)_{11} \left[\left(\frac{1}{2m\xi_+^2} + W - \mu - \sqrt{\omega^2 - \rho^2} \right)^{1-\alpha} \right.\right.$$
$$\left. - \left(\frac{1}{2m\xi_+^2} \right)^{1-\alpha} \right] \theta(\mu + \sqrt{\omega^2 - \rho^2}) \quad (6.2)$$
$$\left. - \left(\sqrt{\omega^2 - \rho^2} \longrightarrow -\sqrt{\omega^2 - \rho^2} \right) \right\}.$$

Again, for $T = 0$ and large $\mu \gg \rho$, equation (6.2) reduces to the BCS result [31]

$$N(\omega) = N_0 \frac{|\omega|}{\sqrt{\omega^2 - \rho^2}}. \quad (6.3)$$

The plots for the DOS (6.2) are presented in Fig. 4 ($T < T_{\mathrm{BKT}}$) and Figs. 5 and 6 for $T > T_{\mathrm{BKT}}$, respectively. If one does not pay attention to a small difference in the curves shown in these figures, it is clearly seen that the

form of the DOS does not differ qualitatively from standard BCS curves. Moreover, similarly to the spectral function, the DOS in the static approximation has a gap both above and below T_{BKT} and does not reveal any change when the temperature crosses the phase-transition point. This confirms once more the crossover character of the latter, although, as was already pointed out, a 2D system is superconducting below T_{BKT} only. According to generally accepted views, the existence of an empty gap above the critical temperature is impossible. The reasons for its persistence were discussed in the previous section. Recall only that dynamical fluctuations or fluctuations of the modulus ρ undoubtedly result in gap filling above T_{BKT}. One must also take the dependence of the decay constant γ into account [see (3.16)], which for $T > T_{BKT}$, can be considerably greater than in the region $T < T_{BKT}$ due to the presence of the vortices.

From the physical point of view, the filling of a gap (transforming it into a pseudogap) above T_{BKT} (or T_c in the quasi-2D case) has to continue up to T^* (or T_ρ if there is a point where $\rho = 0$). However, taking ρ-fluctuations into account (i.e., $\rho(x) \longrightarrow \rho + \Delta\rho$) will cause the appearance of the self-energy, in addition to ρ^2, in the denominator of the mean-field Green's function (2.15); it is proportional to the quantity $\langle\Delta\rho(x)\Delta\rho(0)\rangle$, whose contribution could persist at all T. In this case, the beginning of the pseudogap opening will be defined by the experimental technique sensitivity of the spectral function or DOS measurements.

7. CONCLUSION

To summarize, we have derived analytic expressions for the fermion Green's function, its spectral density, and the density of states in the modulus–phase representation for the simplest 2D attractive Hubbard model with the *s*-wave nonretarded attractive interaction.

While there is still no generally accepted microscopic theory of HTSC compounds and their basic features (including the pairing mechanism), it seems that this approach, although in a sense phenomenological, is of great interest, since it enables one to propose a reasonable interpretation for the pseudogap phenomena related to vortex fluctuations. The results presented here are entirely analytic, which allows a deeper understanding than in the case of a numerical investigation. In particular, the analytic investigation of the Green's function structure revealed that the phase fluctuations lead to a non-Fermi-liquid behavior below and above T_{BKT}.

Evidently, there are a number of important open questions. The main question is whether the pseudogap is related to some kind of superconducting (in our case, phase) fluctuations. Hopefully, the experiment proposed in [4] may answer this question. It seems plausible from the theoretical point of view that superconducting fluctuations should contribute to the pseudogap

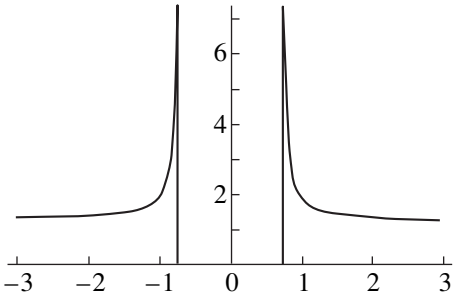


Fig. 4. The density of states $N(\omega)/N_0$ at $T = 0.99\ T_{BKT}$.

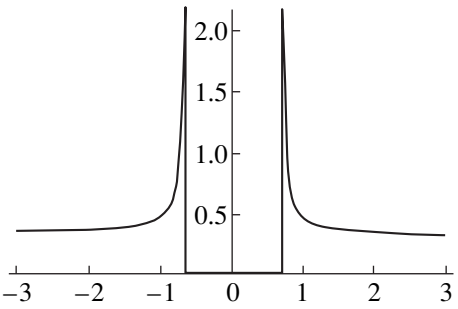


Fig. 5. The density of states $N(\omega)/N_0$ at $T = 1.043\ T_{BKT}$.

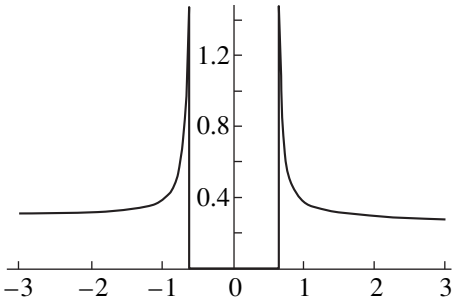


Fig. 6. The density of states $N(\omega)/N_0$ at $T = 1.088\ T_{BKT}$.

(see, however, [11]). Nevertheless, one cannot exclude the possibility that the superconducting contribution may be neither the only, nor the main, contribution.

Another open question is which approach allows one to obtain the pseudogap from the attractive Hubbard model. The schemes used in [30] and in our paper are very different from those of [13]. In particular, our approach allowed us to establish a direct relationship between the superconducting fluctuations and the non-Fermi-liquid behavior in a very natural and transparent way. Also, it relates the pseudogap to the "soup" of fluctuating vortices (see also [30, 42]), while [13] emphasizes the existence of metastable pairs above T_c. It is possible that both these pictures capture some of the physics, but in different temperature regions. When T is high and close to T^*, the value of ρ is small, so that

ρ-fluctuations or metastable pairs dominate. Then, as the temperature approaches T_{BKT}, the values of ρ and the phase stiffness J are growing larger, so that the vortex excitations dominate and ρ-fluctuations become less important. We stress once more that the vortex excitations cannot be adequately described within the T-matrix approximation [9].

Recently, the last part of this picture was supported experimentally [43] by measurements of the screening and dissipation of a high-frequency electromagnetic field in bismuth cuprate films. These measurements provide evidence for a phase-fluctuation-driven transition from the superconducting to normal state.

Finally, there remains the problem of a more complete treatment of the pseudogap in the modulus–phase variables. In particular, the effects of dynamical phase fluctuations and the fluctuations of the order-field modulus must be considered. The latter are especially important for a *d*-wave superconductor, since the modulus can be arbitrarily small in the nodal directions. In this case, it will again be important to check the complete structure of the Green's function, especially its nonpole structure. Another important question that has to be addressed is which factor is more important for gap filling, the spin–charge coupling proposed in [30] or dynamical phase fluctuations, which, as was shown in [36], also result in filling.

ACKNOWLEDGMENTS

We thank Prof. R.M. Quick for many thoughtful comments on this manuscript. One of us (S.G.S.) is grateful to the members of the Department of Physics of the University of Pretoria, especially Prof. R.M. Quick and Dr. N.J. Davidson, for very useful points and hospitality. S.G.S. thanks Prof. A. Sudbø for a useful discussion. S.G.S. also acknowledges the financial support of the Foundation for Research Development, Pretoria.

APPENDIX

A. The Asymptotic Behavior of the Phase Correlation Function

To calculate the integral in equation (3.8), we first write it as

$$I \equiv \int_0^\infty e^{-q/\Lambda}[1 - J_0(qr)]\coth aq$$

$$= \frac{1}{a}\int_0^\infty dt e^{-t/\Lambda a}\left(\coth t - \frac{1}{t}\right)\left[1 - J_0\left(\frac{r}{a}t\right)\right] \qquad \text{(A.1)}$$

$$+ \frac{1}{a}\int_0^\infty \frac{dt}{t} e^{-t/\Lambda}[1 - J_0(rt)].$$

The following formulas are used when calculating I:

$$\int_0^\infty dt e^{-\beta t}\left(\coth t - \frac{1}{t}\right) = \ln\frac{\beta}{2} + \frac{1}{\beta} - \psi\left(1 + \frac{\beta}{2}\right),$$

$$\int_0^\infty \frac{dt}{t} e^{-pt}[1 - J_0(ct)] = \ln\frac{p + \sqrt{p^2 + c^2}}{2p}. \qquad \text{(A.2)}$$

Hence, we obtain

$$I = \frac{4}{r_0}\left[\ln\frac{1 + \sqrt{1 + (\Lambda r)^2}}{\Lambda r_0} + \frac{\Lambda r_0}{4} - \psi\left(1 + \frac{2}{\Lambda r_0}\right)\right]$$

$$- \frac{1}{r}\int_0^\infty dt e^{-t/\Lambda r}\left(\coth\frac{r_0 t}{4r} - \frac{4r}{r_0 t}\right)J_0(t) \qquad \text{(A.3)}$$

$$\sim \frac{4}{r_0}\left[\ln\frac{r}{r_0} + \frac{\Lambda r_0}{4} - \psi\left(1 + \frac{2}{\Lambda r_0}\right)\right] - \frac{1}{r}\frac{1}{\sqrt{1 + 1/(\Lambda r)^2}},$$

$$r \gg r_0, \Lambda^{-1}.$$

Now, depending on the relationship between Λ and r_0, we obtain

$$I \sim \begin{cases} \dfrac{4}{r_0}\ln\dfrac{r}{R_0} + \Lambda, & r \gg r_0 \gg \Lambda^{-1} \\ \dfrac{4}{r_0}\ln\dfrac{\Lambda r}{2}, & r \gg \Lambda^{-1} \gg r_0, \end{cases} \qquad \text{(A.4)}$$

which gives equation (3.10).

B. Another Representation for the Retarded Green's Function

Here, we obtain another representation for the retarded fermion Green's function that is more convenient for the derivation of the spectral density. Recall that when the imaginary part of $G(\omega, \mathbf{k})$ is nonzero, $\mu + \sqrt{\omega^2 - \rho^2} > 0$ and $\nu_1 > 0$, $\nu_2 < 0$. This allows one to transform the analytically continued [by means of equation (4.18)] integral

$$L \equiv \int_0^\infty du \frac{[u(u+1)]^{\alpha-1}}{[(u+\nu_1)(u+\nu_2)]^\alpha} \qquad \text{(B.1)}$$

from equation (4.14) as (for $\alpha < 1$)

$$L = (-1)^\alpha \int_0^{|\nu_2|} du \frac{[u(u+1)]^{\alpha-1}}{[(u+\nu_1)(|\nu_2| - u)]^\alpha}$$

$$+\int_{|v_2|}^{\infty} du \frac{[u(u+1)]^{\alpha-1}}{[(u+v_1)(u-|v_2|)]^{\alpha}} = \frac{(-1)^{\alpha}}{u_1^{\alpha}}\Gamma(\alpha) \times \Gamma(1-\alpha) F_1\left(\alpha, \alpha, 1-\alpha; 1; \frac{v_2}{v_1}, u_2\right) + \frac{1}{|v_2|}\frac{\Gamma(1-\alpha)}{\Gamma(1-\alpha)} F_1\left(1, \alpha, 1-\alpha; 2-\alpha; \frac{v_1}{v_2}, \frac{1}{v_2}\right). \quad \text{(B.2)}$$

The first Appel function in (B.2) can be reduced to the hypergeometric function using the identity [38] that is valid for $\gamma = \beta + \beta'$:

$$F_1(\alpha, \beta, \beta', \beta+\beta'; x, y) = (1-y)^{\alpha} {}_2F_1\left(\alpha, \beta; \beta+\beta'; \frac{x-y}{1-y}\right). \quad \text{(B.3)}$$

Thus, one obtains

$$L = \frac{(-1)^{\alpha}\Gamma(\alpha)\Gamma(1-\alpha)}{[u_1(1-u_2)]^{\alpha}} {}_2F_1\left(\alpha, \alpha; 1; \frac{u_2(1-u_1)}{u_1(1-u_2)}\right) + \frac{1}{|u_2|}\frac{\Gamma(1-\alpha)}{\Gamma(2-\alpha)} F_1\left(1, \alpha, 1-\alpha; 2-\alpha; \frac{u_1}{u_2}, \frac{1}{u_2}\right); \quad \text{(B.4)}$$

$$\frac{u_2(1-u_1)}{u_1(1-u_2)} < 1, \quad \frac{u_1}{u_2} < 0, \quad \frac{1}{u_2} < 0.$$

This completes the derivation of equation (5.2).

C. The Calculation of the Density of States

Introducing

$$y = \frac{k^2/2m}{\mu + \sqrt{\omega^2 - \rho^2}}, \quad b = \frac{1}{2m\xi_+^2}\frac{1}{\mu + \sqrt{\omega^2-\rho^2}}, \quad y_0 = \frac{W}{\mu + \sqrt{\omega^2-\rho^2}}, \quad \text{(C.1)}$$

and substituting (5.5) in (6.1), we can write

$$N(\omega) = N_0 \frac{\Gamma(\alpha)}{\Gamma(1-\alpha)}\left(\frac{2}{mr_0^2}\right)^{\alpha-1} \operatorname{sgn}\omega\,\theta(\omega^2-\rho^2) \times \left[(\mathcal{A}_1)_{11}(\mu + \sqrt{\omega^2-\rho^2})^{1-\alpha}\int_0^{y_0} \frac{dy}{[(y+b-1)^2+4b]^{\alpha/2}} \times {}_2F_1\left(\frac{\alpha}{2}, \frac{1-\alpha}{2}; 1; -\frac{4y}{(y+b-1)^2+4b}\right) \times \theta(\mu + \sqrt{\omega^2-\rho^2}) - (\sqrt{\omega^2-\rho^2} \longrightarrow -\sqrt{\omega^2-\rho^2})\right]. \quad \text{(C.2)}$$

We now consider the integral from (C.2),

$$I = \int_0^{y_0} \frac{dy}{[(y+b-1)^2+4b]^{\alpha/2}} \times {}_2F_1\left(\frac{\alpha}{2}, \frac{1-\alpha}{2}; 1; -\frac{4y}{(y+b-1)^2+4b}\right). \quad \text{(C.3)}$$

Using the relation [38]

$${}_2F_1(a, b; c; z) = (1-z)^{-a} {}_2F_1\left(a, c-b; c; \frac{z}{z-1}\right), \quad \text{(C.4)}$$

the integral can then be rewritten as

$$I = \int_0^{y_0} \frac{dy}{(y+b+1)^{\alpha}} \times {}_2F_1\left(\frac{\alpha}{2}, \frac{1+\alpha}{2}; 1; \frac{4y}{(y+b+1)^2}\right). \quad \text{(C.5)}$$

Replacing $x = \dfrac{b+1}{y+b+1}$ in (C.5), we obtain

$$I = (b+1)^{1-\alpha}\int_{x_0}^{1} dx x^{\alpha-2} \times {}_2F_1\left(\frac{\alpha}{2}, \frac{1+\alpha}{2}; 1; \frac{4x(1-x)}{b+1}\right), \quad \text{(C.6)}$$

$$x_0 = \frac{b+1}{y_0+b+1}.$$

Integral (C.5) diverges at the lower limit as $x_0 \longrightarrow 0$, or equivalently, as $y_0 \longrightarrow \infty$. To handle this, we can write

$$I = (b+1)^{1-\alpha}\int_{x_0}^{1} dx x^{\alpha-2} \times \left[{}_2F_1\left(\frac{\alpha}{2}, \frac{1+\alpha}{2}; 1; \frac{4x(1-x)}{b+1}\right) - 1 + 1\right] = (b+1)^{1-\alpha}\left\{\frac{1-x_0^{\alpha-1}}{\alpha-1} + \int_{x_0}^{1} dx x^{\alpha-2}\left[{}_2F_1\left(\frac{\alpha}{2}, \frac{1+\alpha}{2}; 1; \frac{4x(1-x)}{b+1}\right) - 1\right]\right\}. \quad \text{(C.7)}$$

To calculate the last integral in (C.7), we rewrite it as

$$E = \lim_{\gamma \to \alpha - 1} \int_0^1 dx x^{\gamma - 1} \times \left[{}_2F_1\left(\frac{\alpha}{2}, \frac{1+\alpha}{2}; 1; \frac{4x(1-x)}{b+1}\right) - 1 \right]. \tag{C.8}$$

For $\gamma > 0$, we can compute the integral with the help of the formula (2.21.29) [44],

$$\int_0^y x^{\alpha-1}(y-x)^{\beta-1} {}_2F_1(a, b; c; \omega x(y-x))dx = y^{\alpha+\beta-1} B(\alpha, \beta) \times {}_4F_3\left(a, b, \alpha, \beta; c, \frac{\alpha+\beta}{2}, \frac{\alpha+\beta+1}{2}; \frac{\omega y^4}{4}\right), \tag{C.9}$$

so that

$$E = \lim_{\gamma \to \alpha - 1} \left\{ B(\gamma, 1) \times {}_4F_3\left(\frac{\alpha}{2}, \frac{1+\alpha}{2}, \gamma, 1; 1, \frac{\gamma+1}{2}, \frac{\gamma+2}{2}; \frac{1}{b+1}\right) - \frac{1}{\gamma} \right\}$$

$$= B(\alpha - 1, 1) \times {}_4F_3\left(\frac{\alpha}{2}, \frac{1+\alpha}{2}, \alpha - 1, 1; 1, \frac{\alpha}{2}, \frac{1+\alpha}{2}; \frac{1}{b+1}\right) \tag{C.10}$$

$$-\frac{1}{\alpha - 1} = \frac{1}{\alpha - 1} {}_1F_0\left(\alpha - 1; \frac{1}{b+1}\right) - \frac{1}{\alpha - 1} = \frac{1}{1-\alpha}\left[1 - \left(\frac{b}{b+1}\right)^{1-\alpha}\right] > 0.$$

Thus, for integral (C.3), we find

$$I = \frac{1}{1-\alpha}[(y_0 + b + 1)^{1-\alpha} - b^{1-\alpha}]. \tag{C.11}$$

Now substituting (C.11) into (C.2), we obtain

$$N(\omega) = N_0 \frac{\Gamma(\alpha)}{\Gamma(2-\alpha)} \left(\frac{2}{m r_0^2}\right)^{\alpha-1} \operatorname{sgn}\omega \theta(\omega^2 - \rho^2) \times \left\{ (\mathcal{A}_1)_{11} (\mu + \sqrt{\omega^2 - \rho^2})^{1-\alpha} \times \left[(y_0 + b + 1)^{1-\alpha} - b^{1-\alpha} \right] \theta(\mu + \sqrt{\omega^2 - \rho^2}) - (\sqrt{\omega^2 - \rho^2} \to -\sqrt{\omega^2 - \rho^2}) \right\}. \tag{C.12}$$

Finally, replacing y_0 and b in (C.12) by expressions from (C.1), we arrive at equation (6.2).

REFERENCES

1. B. G. Levi, Phys. Today **49**, 17 (1996).
2. D. Pines, Turk. J. Phys. **20**, 535 (1996); A. V. Chubukov and A. J. Schmalian, Phys. Rev. B **57**, R11085 (1998).
3. R. A. Klemm, in *Proceedings of the First International Conference on New Theories, Discoveries, and Applications of Superconductors and Related Materials, Baton Rouge, 1998*; Int. J. Mod. Phys. B **12**, 2920 (1998).
4. B. Jankó, I. Kosztin, K. Levin, *et al.*, Phys. Rev. Lett. **82**, 4304 (1999).
5. R. Haussmann, Z. Phys. B **91**, 291 (1993).
6. O. Tchernyshyov, Phys. Rev. B **56**, 3372 (1997).
7. M. Yu. Kagan *et al.*, Phys. Rev. B **57**, 5995 (1998).
8. J. Serene, Phys. Rev. B **40**, 10873 (1989); J. J. Deisz, D. W. Hess, and J. W. Serene, Phys. Rev. Lett. **80**, 373 (1998).
9. R. Micnas, M. H. Pedersen, S. Schafroth, *et al.*, Phys. Rev. B **52**, 16223 (1995).
10. R. Haussmann, Phys. Rev. B **49**, 12975 (1994).
11. M. Letz and R. J. Gooding, J. Phys.: Condens. Matter **10**, 6931 (1998).
12. G. Baym, Phys. Rev. **127**, 1391 (1962).
13. B. Jankó, J. Maly, and K. Levin, Phys. Rev. B **56**, R11407 (1997); **59**, 1354 (1999); Physica C (Amsterdam) **321**, 113 (1999); I. Kosztin, Q. Chen, B. Jankó, and K. Levin, Phys. Rev. B **58**, R5936 (1998); Phys. Rev. Lett. **81**, 4708 (1998).
14. J. R. Engelbrecht, A. Nazarenko, M. Randeria, and E. Dagotto, Phys. Rev. B **57**, 13406 (1998).
15. M. Randeria, in *Proceedings of International School of Physics, "Enrico Fermi," Varenna, 1997* (North-Holland, New York) (in press); cond-mat/9710223 (1997).
16. J. M. Singer, M. H. Pedersen, T. Schneider, *et al.*, Phys. Rev. B **54**, 1286 (1996).
17. J. H. Choy, S. J. Kwon, and G. S. Park, Science (Washington) **280**, 1589 (1998).
18. A. K. Nguyen and A. Sudbø, Phys. Rev. B **57**, 3123 (1998).
19. G. Preosti, Y. M. Vilk, and M. R. Norman, Phys. Rev. B **59**, 1474 (1999).
20. E. V. Gorbar, V. M. Loktev, and S. G. Sharapov, Physica C (Amsterdam) **257**, 355 (1996); R. M. Quick and S. G. Sharapov, Physica C (Amsterdam) **301**, 262 (1998); V. M. Loktev, R. M. Quick, and S. G. Sharapov, Physica C (Amsterdam) **314**, 233 (1999); Q. Chen, I. Kosztin, B. Jankó, and K. Levin, Phys. Rev. B **59**, 7083 (1999).
21. V. P. Gusynin, V. M. Loktev, and S. G. Sharapov, JETP Lett. **65**, 182 (1997).
22. V. P. Gusynin, V. M. Loktev, and S. G. Sharapov, Zh. Éksp. Teor. Fiz. **115**, 1243 (1999) [JETP **88**, 685 (1999)].
23. V. P. Gusynin, V. M. Loktev, and S. G. Sharapov, JETP Lett. **69**, 141 (1999).
24. E. Witten, Nucl. Phys. B **145**, 110 (1978).

25. N. D. Mermin and H. Wagner, Phys. Rev. Lett. **17**, 1113 (1966); P. C. Hohenberg, Phys. Rev. **158**, 383 (1967); S. Coleman, Commun. Math. Phys. **31**, 259 (1973).
26. A. Z. Patashinskiĭ and V. L. Pokrovskiĭ, *Fluctuation Theory of Phase Transitions* (Nauka, Moscow, 1982; Pergamon, Oxford, 1979).
27. V. Emery and S. A. Kivelson, Nature (London) **374**, 434 (1995); Phys. Rev. Lett. **74**, 3253 (1995); cond-mat/9710059 (1997).
28. E. Babaev and H. Kleinert, Phys. Rev. B **59**, 12083 (1999).
29. M. Randeria and J. C. Campuzano, in *Proceedings of International School of Physics, "Enrico Fermi," Varenna, 1997* (North-Holland, New York) (in press); cond-mat/9709107 (1997).
30. M. Franz and A. J. Millis, Phys. Rev. B **58**, 14572 (1998).
31. R. Schrieffer, *Theory of Superconductivity* (Benjamin, New York, 1964).
32. P. B. Wiegmann, in *Proceedings of International School of Physics, "Enrico Fermi," Varenna, 1997* (North-Holland, New York) (in press); cond-mat/9808004 (1998).
33. Ch. Renner *et al.*, Phys. Rev. Lett. **80**, 149 (1998); **80**, 3606 (1998).
34. M. Plischke and B. Bergersen, in *Equilibrium Statistical Physics* (Prentice-Hall, New Jersey, 1989), p. 167.
35. S. W. Pierson, Philos. Mag. B **76**, 715 (1997).
36. M. Capezzali and H. Beck, Physica B (Amsterdam) **259/261**, 501 (1999).
37. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Nauka, Moscow, 1971; Academic, New York, 1980).
38. H. Bateman and A. Erdélyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953; Nauka, Moscow, 1965).
39. S. Chakravarty and P. W. Anderson, Phys. Rev. Lett. **72**, 3859 (1994); A. Sudbø, Phys. Rev. Lett. **74**, 2575 (1995).
40. M. R. Norman, M. Randeria, H. Ding, and J. C. Campuzano, Phys. Rev. B **57**, R11093 (1998).
41. M. R. Norman, H. Ding, *et al.*, Phys. Rev. B **60**, 7585 (1999).
42. H. J. Kwon and A. T. Dorsey, cond-mat/9809225 (1998); H. J. Kwon, cond-mat/9811059 (1998).
43. J. Corson *et al.*, Nature (London) **398**, 221 (1999).
44. A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marychev, *Integrals and Series* (Nauka, Moscow, 1986; Gordon & Breach, New York, 1989), Vol. III.