Journal of Experimental and Theoretical Physics, Vol. 100, No. 4, 2005, pp. 637–644.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 4, 2005, pp. 723–731.

NUCLEI, PARTICLES, FIELDS, GRAVITATION, AND ASTROPHYSICS

Redshifts in Space Caused by Stimulated Raman Scattering in Cold Intergalactic Rydberg Matter with Experimental Verification¶

L. Holmlid

Atmospheric Science, Department of Chemistry, Göteborg University, SE-412 96 Göteborg, Sweden
e-mail: holmlid@chem.gu.se

Received September 23, 2004

Abstract—The quantized redshifts observed from galaxies in the local supercluster have recently been shown to be well described by stimulated Stokes Raman processes in intergalactic Rydberg matter (RM). The size of the quanta corresponds to transitions in the planar clusters forming the RM, of the order of 6×10^{-6} cm^{-1}. A stimulated Stokes Raman process gives redshifts that are independent of the wavelength of the radiation, and it allows the radiation to proceed without deflection, in agreement with observation. Such redshifts must also be additive during the passage through space. Rydberg matter is common in space and explains the observed Faraday rotation in intergalactic space and the spectroscopic signatures called unidentified infrared bands (UIBs) and diffuse interstellar bands (DIBs). Rydberg matter was also recently proposed to be baryonic dark matter. Experiments now show directly that IR light is redshifted by a Stokes stimulated Raman process in cold RM. Shifts of 0.02 cm^{-1} are regularly observed. It is shown by detailed calculations based on the experimental results that the redshifts due to Stokes scattering are of at least the same magnitude as observations. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The redshifts observed from distant extragalactic sources in space are quite intriguing, and their origin has been debated by many authors. The important reports that the redshifts are quantized [1, 2] are at variance with the accepted interpretation of the redshifts. The quantization of galactic redshifts was recently shown to agree with the expected stimulated Raman process in intergalactic Rydberg matter (RM) [3]. That intergalactic matter exists at nonnegligible densities is shown, for example, by the observation of a Faraday rotation effect at radiofrequencies in intergalactic space [4]. This effect is well described by very low densities of RM, with its high electron density and inherent magnetic field [5]. It is also proposed that RM is the (baryonic) dark matter in space [6]. The properties of RM are such that it is an excellent candidate for the missing dark matter. For example, a hydrogen atom in RM takes up a volume 5×10^{12} times larger than a ground-state hydrogen atom.

Rydberg matter is a special form of matter, which is built up of weakly interacting, highly excited atoms or molecules in circular, metastable Rydberg states. It has metallic properties due to the delocalized electrons and a very low density under the conditions in space. We have shown that low-density RM can be conveniently studied by laser fragmentation and time-of-flight methods [7–9]. Rydberg matter can be built up from alkali atoms and also from small gas molecules like H_2 and N_2 [8, 10, 11]. It can also be formed from H atoms [12, 13]. The existence and structure of RM was predicted by Manykin, Ozhovan and Poluéktov [14, 15] more than 20 years ago, and by 1991, the first experimental studies of a macroscopic Cs Rydberg matter phase were performed [16, 17]. These early studies of RM have recently been independently confirmed [18]. Improved quantum mechanical calculations of the properties of RM built up by Cs were also published [19, 20]. Later, studies by time-of-flight methods [7, 10, 21] identified the special planar cluster shapes predicted by theory [22]. The distance between the ions in the RM can be measured by the repulsion energy released in Coulomb explosions in the matter [8, 9, 11, 12]. This bond distance is a few nanometers at the relatively low excitation levels studied by this technique.

Due to the extremely large polarizability of RM, Raman spectroscopic studies will be very useful to find information on the properties of RM. A blueshift of single-mode IR laser light passing through RM was observed and interpreted as an anti-Stokes stimulated electronic Raman effect (ASERS) [23]. Experiments have also directly observed anti-Stokes Raman scattering from RM surface layers. Transitions in K Rydberg atoms [24] and bands from H_2 and other small molecules, and bands from transitions in the RM [25] were observed. The stimulated Raman effect was also used in a micro-Raman spectroscopy study of the interaction of

¶ This article was submitted by author in English.

water molecules with one of the solid materials used to emit RM in laboratory [26]. A recent study interpreted the interaction observed between laser modes observed in transmission and reflection of IR lasers in RM as being due to the stimulated Raman effect [27]. Another method recently employed for the study of RM is stimulated emission, which works well due to extremely long lifetimes of the electronically excited states in RM. In fact, the first thermally excited laser was recently demonstrated in the IR range [28, 29]. This CW laser is tunable over the range 800–14000 nm. The numerous bands observed in the stimulated transitions agree with the transitions expected in RM.

The lifetime of RM is very long in space, which allows its rate of formation to be very low. Extrapolation from the theoretical results in [20] gives a radiative lifetime of undisturbed RM of the order of the lifetime of the universe. That RM is a common type of matter in space is shown by the interpretation [30] of the unidentified infrared bands (UIR, UIB) as being due to transitions in RM, in fact, the same transitions experimentally studied by stimulated emission [29] and by a few stimulated Raman experiments [24, 25]. Further evidence comes from the Faraday rotation observations in intergalactic space and their interpretation as being due to RM [5]. More evidence also exists. A large number (at least 60) of the so-called diffuse interstellar bands (DIBs) have been calculated accurately using a theory based on the RM concept [31].

Because the conduction band in RM is not filled, almost continuous electronic excitations are possible. In a stimulated electronic Raman process, a photon gains or loses some energy from the interaction with matter, and RM may then change the frequency of radiation passing through it by very small amounts, in the form of a red or blue shift (Stokes or anti-Stokes shifts) of the radiation. A blueshift was observed in the experiments in [23], where the RM used was electronically excited by its formation process and by the high temperature of the surrounding equipment. This shift is expected to change to a redshift as in Stokes scattering in cold RM in space. Further experiments using the same technique have recently been published [32]. In that study, redshifts were observed in reflection from a layer of cold RM deposited on a window. We now demonstrate directly that a redshift is observed in cold RM even in transmission. Redshifts of the radiation from distant galaxies are found from the interaction between radiation and RM, as explained in [3].

2. THEORY

In what follows, the term light is used for simplicity in many places where the term electromagnetic radiation could be used instead.

2.1. Estimated Gain

An important process for the interaction between light and RM is the stimulated electronic Raman scattering of the Stokes or anti-Stokes type (SERS, ASERS) [33]. Anti-Stokes Raman scattering and similar processes in RM have been observed in different types of experiments [24, 25, 34], even at low light intensities [23, 26, 27]. For cold ground-state RM, Stokes scattering should be observed instead. There is no phase matching condition for this type of process [33], and the Raman-scattered light proceeds in the same direction as the incident light.

It is necessary to estimate the magnitude of the stimulated Raman effect in the case of an RM material. For this, we use the ordinary classical steady-state gain factor derived for molecular vibrational transitions. The formulas are given for the Stokes scattering, which is the case of interest for astrophysical processes. The equations are identical in form for the anti-Stokes (*as*) and Stokes (*s*) cases. The incident light frequency is ω_L and the generated Stokes wave has a frequency ω_s. The difference

$$\omega_L - \omega_s = \omega_E$$

is the resulting electronic excitation in the RM material. For the Stokes wave, the steady-state gain factor G_{ss} at its maximum is given by [35]

$$G_{ss} = \frac{N_d k_s}{8m\epsilon\gamma(\omega_L - \omega_s)}\left(\frac{\partial\alpha}{\partial q}\right)_0^2 |E_L|^2. \quad (1)$$

In this expression N_d is the density of the dipoles created by the incident light wave, here chosen to correspond to the number of electrons in the RM; k_s is the wavenumber for the Stokes wave; m is the mass of the driven oscillator, in this case, one electron mass; $\epsilon = \epsilon_r\epsilon_0$ is the permittivity of the medium, which is unknown but probably has ϵ_r on the order of unity; γ is the coupling constant for the electronic excitation in the intermediate state to other degrees of freedom; $\partial\alpha/\partial q$ is the variation of the polarizability with the coordinate describing the excited motion, here the electronic motion; and E_L is the electric field strength of the incident light.

The gain factor G_{ss} can be large even at low light intensities due to the very large polarizability of the RM and an arbitrarily small value of the difference $\omega_L - \omega_s$. The reason for this small value is the almost continuous nature of the energy levels in RM. Another factor of great importance is the coupling (dephasing) constant γ, which is much smaller than in the case of ordinary matter because the coupling of the electronic motion in RM to other modes of motion is weak. From the experiments with IR lasers, it was concluded that γ is on the order of 10^3 s^{-1} or smaller [23]. This would give a lower

limit on the magnitude of the electronic excitations in terms of wavenumbers from

$$c\Delta\tilde{\nu} = \omega_E \gg \gamma \text{ as } \Delta\tilde{\nu} \gg 3\times10^{-8}\ \text{cm}^{-1}.$$

The quantized redshifts interpreted in [3] give

$$\Delta\tilde{\nu} \approx 6\times10^{-6}\ \text{cm}^{-1}$$

in agreement with the experimentally found lower limit for γ. We use this value as a reasonable quantum magnitude of the electronic excitation in RM above each metastable state with a certain excitation level n. Lifetime measurements of RM both at 77 K [36] and at 800 K [16, 37] and calculations at low excitation levels [20] seem to indicate a much smaller value of γ, maybe on the order of 10^{-3} s^{-1} [23, 34]. However, these lifetimes probably correspond to deexcitation from an excitation level n down to the fully dissociated ground state consisting of separate atoms or molecules.

To estimate reasonable values of the quantities in Eq. (1), we use a phase of RM with the excitation state $n = 80$, which is the average value deduced from the study in [30]. Each Rydberg atom is then considered a polarizable particle. With interatomic distances of 0.5 μm at $n = 80$, the density is found to be approximately 10^{18} m^{-3}. The wavenumber k_s is on the order of 2×10^4 cm^{-1} or 2×10^6 m^{-1} in the visible range, and the difference of the frequencies $\omega_L - \omega_s$ is assumed to be 10^{-6} cm^{-1} or 10^{-4} m^{-1}. In fact, with RM, this value is probably even smaller. The mass m is the electron mass and ϵ is the dielectric constant ϵ_r times ϵ_0. The polarizability variation $\partial\alpha/\partial q$ is estimated as the volume added for a change in the radial distance for the Rydberg electron (times ϵ_0 to give the right dimension) and becomes 4.4×10^{-24} As m V^{-1}. Finally, the field strength due to light is estimated for the power density 1 mW cm^{-2}, giving $E = 90$ V m^{-1}. This gives

$$G_{ss} = \frac{3\times10^{17}}{\epsilon_r\gamma}\ \text{m}^{-1} \qquad (2)$$

with γ in s^{-1} and ϵ_r on the order of unity. Even for rather large values of γ, the G_{ss} factor is large, which means that the stimulated Raman effect is strong and an efficient conversion to the Stokes wave takes place in a short distance. It may be assumed that γ is of the order of 10^3 s^{-1}. This means that the G_{ss} factor is very large and gives rise to a strong stimulated effect by the equation for the Stokes field strength [33, 35]

$$E_s^f = E_s(0)\exp(G_{ss}x), \qquad (3)$$

where the index f indicates a forward wave, $E_s(0)$ is the Stokes wave at position zero (noise photons), and x is the distance along the laser beam. Thus, it is obvious that stimulated Raman effects should be observable even for low light intensities, especially if x is very large as in intergalactic space.

2.2. Continuous Excitations

The form of Eq. (1) is best suited for molecular vibrational Raman problems. Other forms of this equation exist (as, e.g., in [38]) which in general are better suited for electronic excitations. However, the theory of RM shows that the delocalized electrons, as in ordinary Rydberg states, are close to the classical limit. For example, in [22], the bond energies and electronic levels of RM are calculated from electrostatic formulas, only with the addition of the electron correlation as a quantum mechanical effect. The transitions of interest here, between translational states of the RM electrons in the RM clusters, have very small quanta and are almost classical in nature. This means that a description of the interaction between the light wave and the molecular system in terms of motion of point particles, like in the vibrational Raman transitions, is quite well adapted to RM electronic excitations.

In the case of stimulated Raman scattering, well-defined Raman transitions are normally studied, and not a continuous range of wavenumbers ω_E. Because one well-defined Stokes (or anti-Stokes) frequency usually dominates and gives the Raman wave, it is not directly clear what happens if a range of transitions is possible. In the present case, consecutive Stokes components are formed along the laser beam, as shown, for example, in [33, 39]. If a range of frequencies is possible, this could be thought to lead to a broadening of the Stokes wavevector and a less well-defined stimulated Raman appearance. From Eq. (3), however, one can see that the switch from one Stokes component to the next occurs when the exponent $G_{ss}x$ is sufficiently large. Because

$$G_{ss}x \propto \frac{x}{\omega_E},$$

a smaller value of ω_E gives the same value of the exponent in a proportionally shorter distance x. This means that if a smaller ω_E takes over, it switches over to the next Stokes component faster, keeping the resulting shift after a larger distance constant, independently of which ω_E actually dominates. This shows that the detailed process of the Stokes component switching with a continuous range of transitions is of no great concern, because the result is the same.

2.3. Intensity Dependence

The small value of γ means that the form of G_{ss} given by Eq. (1) is not strictly valid. Assuming that this effect does not change the theoretical formulas completely, one can still use the treatment by Shen and

Bloembergen [39] and Shen [38] to find a more complete description of the Raman signal. For a small wave vector momentum mismatch along the light direction (the *x*-direction)

$$\Delta k = 2k_{L(x)} - k_{s(x)} - k_{as(x)},$$

the gain is observed to decrease linearly towards zero as $\Delta k \longrightarrow 0$. The plots for this effect are given as reduced plots in [39] and also in [23], where the scales of both axes are in reduced quantities, relative to G_{ss}. Because G_{ss} is proportional to $|E_L|^2$, an increase in the light intensity increases G_{ss} and thus moves a point on the curve downwards, for constant Δk. This means that the relative gain decreases by the same factor, keeping the real gain G constant. Thus, the gain G varies linearly with the mismatch Δk. The gain G is then independent of the light intensity for constant Δk in the region of small momentum mismatch, which probably is the range of interest for space. Because the value of G_{ss} is extremely large, as described above, the resulting gain can still be substantial. The analysis in [38, 39] shows further that the anti-Stokes wave formed together with the Stokes wave has a lower intensity at relatively small values of Δk.

3. RESULTS

3.1. Cold Rydberg Matter in the Laboratory

Rydberg matter can be produced by several techniques in the laboratory in different surroundings and from different starting materials [40]. Usually, the formation process involves relatively high temperatures, from 300 K up to 1500 K depending on the technique used. This means that the RM produced is in an electronically excited state with electron translation in the RM cluster above its ground state, which is characterized by a certain principal quantum number *n*. The reason for the excitation is that the RM is formed in the experiments by condensation of excited clusters, molecules, and atoms in Rydberg states. The excess energy in the RM clusters comes from the condensation energy and from collisions with an excitation energy transfer from the Rydberg species. However, a cooling process can be applied to reach lower electronic temperatures. At RM densities of interest in space, the condensed phase consists mainly of hexagonal planar clusters with the magic number (number of members in a stable cluster) N = 7, 19, 37, 61, and 91 as observed experimentally [7, 11].

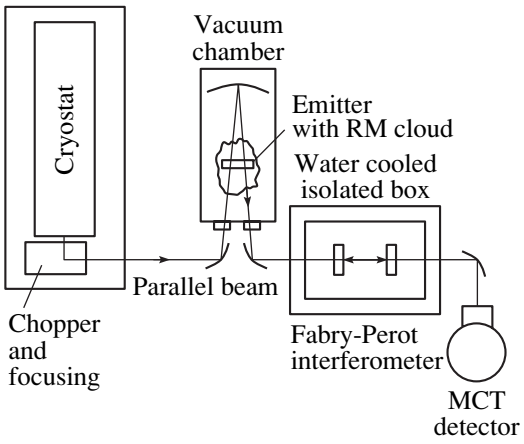


Fig. 1. Block diagram of the experimental setup for a transmission measurement. The RM cloud around the hot emitter is indicated.

In the present stimulated Raman experiment, an almost single-mode CW laser with mW power is used in the IR range, with ω_L close to 1100 cm^{-1}. The slope of the laser modes is approximately 7×10^{-3} cm^{-1} mA^{-1}. The setup is similar to the one used in [23, 32] with an air-spaced Fabry–Perot interferometer (FPI) with ZnSe mirrors between the RM chamber and the detector. A mirror is used inside the RM chamber to reflect the laser beam passing below the RM emitter. See Fig. 1. After formation, the RM cloud is allowed to cool itself with the RM emitter turned off. The lifetime of RM in the laboratory under these conditions is of the order of a few hours. As observed directly in laser fragmentation experiments, the cold RM gives translational temperatures of a few K for large clusters [9, 11]. This cooling is probably driven by stimulated emission in the IR range [29]. The result of this cooling process is that a redshift of the light passing through the RM cloud can be detected with the interferometer.

A typical result with blueshifting (leftshift) at a high emitter temperature due to the formation of a cloud of warm RM is shown in Fig. 2. At the end of the experiment, a redshift (rightshift) is observed after cooling of the RM. A complete run with less blueshifting due to colder RM (exhausted RM emitter) is shown in Fig. 3. Time and FPI temperature are given in the figure. A temperature-stable (often water-cooled) FPI with an Invar base, usually in a well insulated box was used, with a free spectral range of 0.19 cm^{-1}. The temperature coefficient of the FPI is determined to be less than 1.5×10^{-2} cm^{-1} K^{-1}. This gives an elongation of the cavity with increasing temperature and a drift of the peaks (fringes) to the left in the figure. Thus, the true redshift in the figures (rightshift) is slightly larger than directly observed because the temperature of the FPI increased during the experiment. The redshift observed is of the order of 0.02 cm^{-1} for a passage with length of the order of 25 cm through RM in a surrounding air pressure of 10^{-4} mbar. It is estimated that the density of RM is close to the gas density in the chamber. This gives a value of 10^{18} m^{-3} if RM fills the volume almost completely, with the excitation level n = 80. Due to the planar cluster structure of RM, a filling factor of 10% is more likely, giving the density 10^{17} m^{-3} in the experiments.

A redshifting of the interferometer fringes can also be observed with cold RM deposited on the inner surface of the ZnSe window used in the RM chamber. The window is antireflection-coated on the outer surface, and the reflection on the inner surface is observed. This effect was demonstrated in [32]. In this configuration, only the evanescent wave interacts with the cold RM on the inner surface. A similar configuration was shown to work well in stimulated emission experiments using RM [41], where the coupling to the external laser cavity took place via the evanescent wave. The results from a reflection experiment are shown in Fig. 4. The noise in the signal is relatively high because just a fraction of the laser beam intensity is reflected. The time for the scans and the temperature of the mirror holder in the FPI are shown. A redshift (rightshift) is seen, caused by the deposition of RM on the window during heating of the RM emitter. After return to room temperature, the redshift gradually disappears and after two hours, the shift is zero. We note that the FPI temperature is still higher than initially and that the drift of the fringes due to temperature changes of the FPI is thus small. A more complete discussion of the influence of temperature changes of the FPI is given in [32].

3.2. Cold Rydberg Matter in Space

In space, Rydberg matter can exist both at low temperatures and very low densities, and at higher temperatures of the order of 300–600 K and higher local densities, for example, surrounding particles in interstellar space. The excitation levels observed from the so-called unidentified infrared bands [30] have the most probable value close to $n = 80$. A comparison with the parameter values appropriate for the laboratory studies of cold RM shows that the density N_d is several orders of magnitude smaller in space than in laboratory experiments. On the other hand, for visible light, the wavenumber k_s is approximately 50 times larger than for the IR studies described here.

Equation (3) shows that a shift due to the stimulated Raman effect exists if the product $G_{ss}x$ is larger than unity. With Eq. (1) used for G_{ss} and with reasonable values also used for estimation of this factor above, it is possible to find a condition on the electric field strength of the light field. The distance covered by light in space is denoted by l. Then the inequality

$$\frac{k_s}{8m\epsilon\gamma c}\left(\frac{\partial\alpha}{\partial q}\right)_0^2 \geq \frac{\Delta\tilde{\nu}}{N_d l|E_L|^2} \qquad (4)$$

should be valid for the effect to exist. In this inequality, the space-related quantities are collected in the right hand side. The difference $\omega_L - \omega_s$ was replaced by $c\Delta\tilde{\nu}$, where $\Delta\tilde{\nu}$ is in wavenumbers. Assuming conservatively that γ is as large as the upper experimental limit found, 10^3 s^{-1}, we find that the left-hand side is equal to 2×10^{-12} m^3 V^{-2}, with the other parameter values estimated as before. With the experimental values from the study of the cold RM described here, $\Delta\tilde{\nu} = 0.02$ cm^{-1}, $N_d = 10^{17}$ m^{-3}, $l = 0.25$ m, and $E_L = 90$ V m^{-1}, the right-

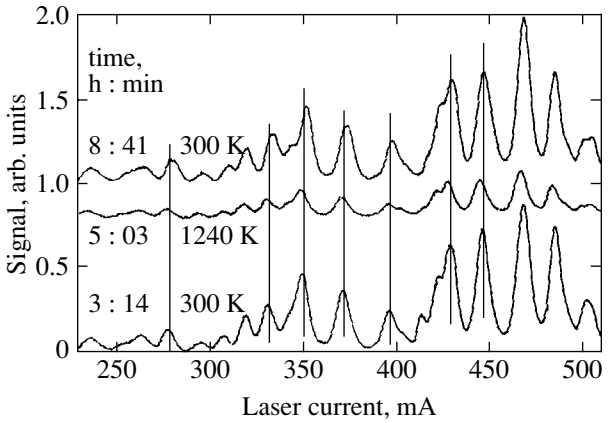


Fig. 2. Initial blueshift due to warm RM and final redshift due to cold RM of light from a diode laser at 1100 cm^{-1} passing through the RM cloud. The lower trace is taken with the RM emitter at room temperature, while the top one is after heating to 1240 K and subsequent cooling, at an air pressure of 10^{-4} mbar in the RM chamber. Note the blueshift at 1240 K and the redshift (rightshift) in the top curve. The same data as in [32].

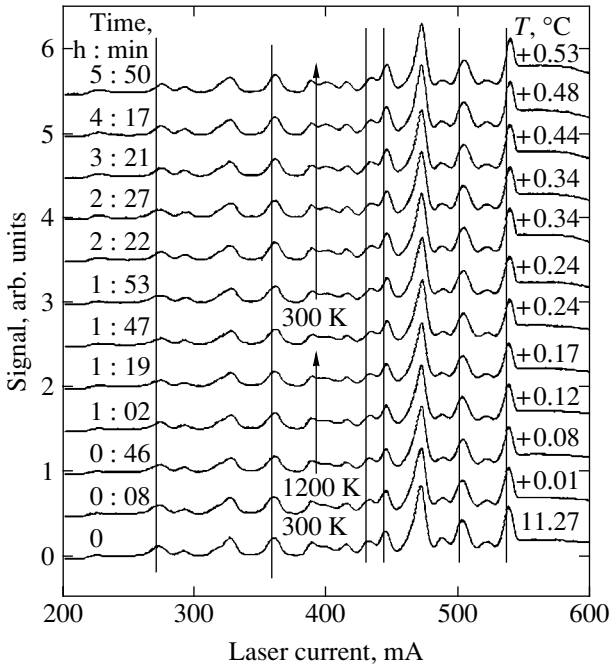


Fig. 3. Redshift of light from a diode laser at 1100 cm^{-1} passing through cold RM (in the top part of the figure). Time and the temperature of the water cooled interferometer base as well as the RM emitter temperature are given. The experiment was done at the air pressure 10^{-4} mbar in the RM chamber. The curves are consecutively shifted upwards to increase visibility.

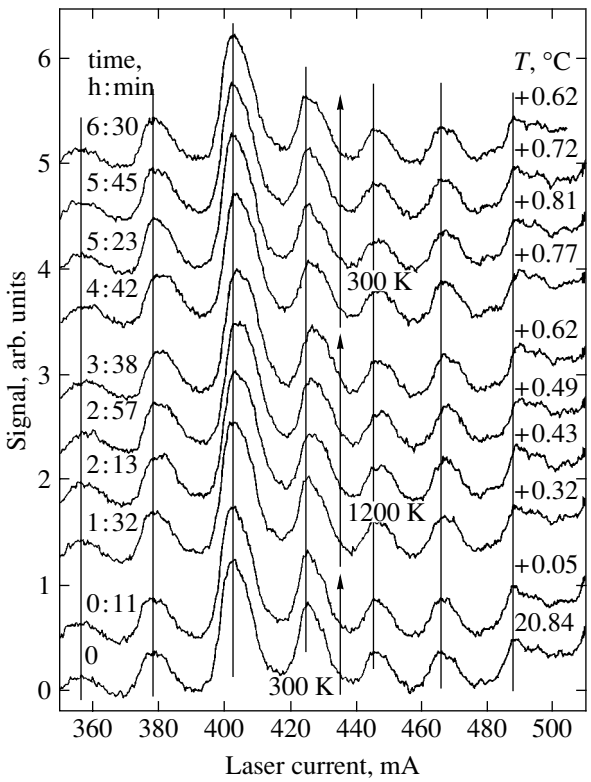


Fig. 4. Redshift of light from a diode laser at 1100 cm^{-1} reflected from a layer of RM on the inner face of the vacuum window. Time and the temperature of the FPI mirror holder and the RM emitter temperature are given. The experiment was conducted at an air pressure of 10^{-5} mbar in the RM chamber. The curves are consecutively shifted upwards to increase visibility. RM is slowly deexcited on the window when RM deposition is cancelled (at room temperature).

hand side becomes 1×10^{-20} m^3 V^{-2}. This is many orders of magnitude smaller than the left-hand side, as required by Eq. (4). Because some of the quantities are anyway somewhat uncertain, we may assume that the right-hand side of Eq. (4) in space should be as small as this value in experiments. (This is a very conservative estimate.) Then $\Delta\tilde{\nu} \approx 10^4$ cm^{-1} (the typical total summed shift in observations), $N_d = 10^6$ m^{-3} [42], and $l = 8 \times 10^8$ pc $= 2.5 \times 10^{25}$ m may be used, giving $E_L = 3 \times 10^{-3}$ V m^{-1}. The distance l used is smaller than the radius of the observable universe by about a factor of 10. This gives a light intensity of 2×10^{-8} W m^{-2} as the minimum intensity required to make the stimulated Raman process work, corresponding to the intensity of light from the Sun at our nearest star α Centauri. This is a very conservative estimate, based on the experimental results. If we instead use the condition that the right-hand side in Eq. (4) should only be smaller than the estimated value 2×10^{-12} m^3 V^{-2} of the left-hand side, the required light intensity may even be a factor of 10^9 smaller, thus larger than 2×10^{-17} W m^{-2} corresponding to the field strength $E_L = 10^{-7}$ V m^{-1}. This is the same as the intensity of the light from the Sun at the distance 4×10^4 pc, i.e., on the other side of our galaxy. Thus, a redshift due to stimulated Raman scattering develops even at very large distances from the source.

3.3. Redshifts

It is possible that the stimulated Raman process discussed here operates at the very low intensity levels in space, due to the extreme properties of intergalactic RM. It is of course interesting to investigate whether such an effect gives the behavior of ordinary redshifts, for example, whether the shift varies correctly with wavelength as

$$z = \frac{\lambda_{\text{obs}} - \lambda_{\text{em}}}{\lambda_{\text{em}}}, \tag{5}$$

which should be constant for a certain astronomical object independently of which spectral line is used for the determination. Thus, it is required that relation (5) is a constant, or that $\Delta\lambda/\lambda_L$ is a constant. This may be compared with the predictions from stimulated Raman theory. From Eq. (1), it follows that

$$G_{ss} \propto \frac{k_s}{\omega_L - \omega_s} \propto \frac{\lambda_s^{-1}}{\lambda_L^{-1} - \lambda_s^{-1}} = \frac{\lambda_L}{\lambda_s - \lambda_L} = \frac{\lambda_L}{\Delta\lambda} \tag{6}$$

holds, or that the observed behavior of the redshift is consistent with the stimulated Raman effect with a constant gain factor. The G_{ss} factor is not the real gain, as shown in [39] and cited above, but it is modified to the gain G that is nevertheless proportional to G_{ss}. This gain is constant independently of the shifting light intensities, e.g., for different spectral lines with different emitted intensities. Thus, the stimulated Raman effect gives the correct observed behavior.

From the description given here, it is also clear that the redshifting by the stimulated Raman process in RM is an additive process, given by the distance covered by light in the RM phase.

4. DISCUSSION

4.1. Quantum Effects

The main effect that may prevent the stimulated Raman effect from shifting the frequency of light continuously, as described above, is the quantal nature of excitations in the RM and the quantal nature of light. A finite lower bound to the size of the possible excitations in RM tends to prevent G_{ss} from reaching extremely large values, as can be seen from Eq. (1). The experimental redshifts or blueshifts are not observed to be

quantized at a level of 10^{-2} cm^{-1} or larger. This means that the quantum size is small. It is relevant to note that quantized redshifts are in fact observed, both for galaxies [1] and for quasars [2].

From the description in the theoretical section, a lower limit of the size of the quantized shifts could be 3×10^{-8} cm^{-1}. The quantized redshifts for nearby galaxies [1] instead suggest a quantum size of 6×10^{-6} cm^{-1} [3]. We use this value to make conservative estimates of the quantum effects. With the total summed shift 10^4 cm^{-1} in the visible range, the total number of shifts is then on the order of 10^9. If the distance traveled by the light is on the order of 10^{25} m, this means a shift of 6×10^{-6} cm^{-1} per 6×10^{15} m on average. Therefore, the shifts are very uncommon events. If the number of shifts is 10^9, the statistical variation in this number, the square root of this, is 3×10^4 on average. The statistical nature of the shifts is then not observed, but all light is shifted the same amount with an uncertainty on the order of 3×10^{-5}. This gives 0.3 cm^{-1} in a total shift of 10^4 cm^{-1}. Typical uncertainties in optical redshifts are ± 0.001 in z [43], giving ± 20 cm^{-1} for a typical optical line. Thus, the width of the observed lines are much larger than the widths resulting from the statistics of the redshift quanta.

4.2. Cosmological Arguments

In a recent book on cosmology [44], the four general cases of possible explanations of the redshifts of distant astronomical objects are summarized, with the fourth possibility "interaction (scattering, absorption)" being most relevant to the present discussion. It should be noted that stimulated Raman scattering was not considered explicitly, but only more "normal" scattering events were considered. The main arguments against the interaction processes given in [4] are as follows: (a) the resulting redshift should have an exponential dependence on distance, (b) the redshift $\Delta\lambda/\lambda$ would not be independent of frequency, and (c) a scattering process would smear out the light from distant sources, which is at variance with observations. The present model, based on the known physics of stimulated Raman processes, gives not an exponential but a linear dependence of the redshift on distance. As shown above, the redshift $\Delta\lambda/\lambda$ should be frequency-independent for the stimulated Raman process. The stimulated Raman process does not change the direction of light, and hence the final argument against a scattering mechanism of the stimulated Raman type is not valid either.

5. CONCLUSIONS

We conclude that the stimulated Raman mechanism for redshifting radiation in RM in space is a possible process to explain at least a part of the redshifts observed from distant astronomical objects. It gives the correct behavior, for example, concerning the frequency variation, the distance variation, and the direction of scattering. Further, the observed redshifts in cold RM in the laboratory are recalculated to astronomical distances and found to easily cover the range of the observed redshifts.

REFERENCES

1. B. N. G. Guthrie and W. M. Napier, Astron. Astrophys. **310**, 353 (1996).
2. G. Burbridge and W. M. Napier, Astron. J. **121**, 21 (2001).
3. L. Holmlid, Astrophys. Space Sci. **291**, 99 (2004).
4. T. E. Clarke, P. P. Kronberg, and H. Böhringer, Astrophys. J. **547**, L111 (2001).
5. S. Badiei and L. Holmlid, Mon. Not. R. Astron. Soc. **335**, L94 (2002).
6. S. Badiei and L. Holmlid, Mon. Not. R. Astron. Soc. **333**, 360 (2002).
7. J. Wang and L. Holmlid, Chem. Phys. Lett. **295**, 500 (1998).
8. S. Badiei and L. Holmlid, Int. J. Mass Spectrom. **220**, 127 (2002).
9. S. Badiei and L. Holmlid, Chem. Phys. **282**, 137 (2002).
10. J. Wang and L. Holmlid, Chem. Phys. **261**, 481 (2000).
11. J. Wang and L. Holmlid, Chem. Phys. **277**, 201 (2002).
12. S. Badiei and L. Holmlid, Phys. Lett. A **327**, 186 (2004).
13. S. Badiei and L. Holmlid, J. Phys.: Condens. Matter **16**, 7017 (2004).
14. É. A. Manykin, M. I. Ozhovan, and P. P. Poluéktov, Sov. Tech. Phys. Lett. **6**, 95 (1980).
15. É. A. Manykin, M. I. Ozhovan, and P. P. Poluéktov, Sov. Phys. Dokl. **26**, 974 (1981).
16. R. Svensson, L. Holmlid, and L. Lundgren, J. Appl. Phys. **70**, 1489 (1991).
17. R. Svensson and L. Holmlid, Surf. Sci. **269–270**, 695 (1992).
18. V. I. Yarygin, V. N. Sidelnikov, I. I. Kasikov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 330 (2003) [JETP Lett. **77**, 280 (2003)].
19. É. A. Manykin, M. I. Ozhovan, and P. P. Poluéktov, Zh. Éksp. Teor. Fiz. **102**, 804 (1992) [Sov. Phys. JETP **75**, 440 (1992)].
20. É. A. Manykin, M. I. Ozhovan, and P. P. Poluéktov, Zh. Éksp. Teor. Fiz. **102**, 1109 (1992) [Sov. Phys. JETP **75**, 602 (1992)].
21. J. Wang and L. Holmlid, Chem. Phys. Lett. **325**, 264 (2000).
22. L. Holmlid, Chem. Phys. **237**, 11 (1998).
23. L. Holmlid, Phys. Rev. A **63**, 013817 (2001).
24. L. Holmlid, Langmuir **17**, 268 (2001).
25. L. Holmlid, Astrophys. J. **548**, L249 (2001).
26. F. Olofson, S. Badiei, and L. Holmlid, Langmuir **19**, 5756 (2003).
27. L. Holmlid, Eur. Phys. J.: Appl. Phys. **26**, 103 (2004).

28. S. Badiei and L. Holmlid, Chem. Phys. Lett. **376**, 812 (2003).
29. L. Holmlid, J. Phys. B: At. Mol. Opt. Phys. **37**, 357 (2004).
30. L. Holmlid, Astron. Astrophys. **358**, 276 (2000).
31. L. Holmlid, Phys. Chem. Chem. Phys. **6**, 2048 (2004).
32. L. Holmlid, Appl. Phys. B **79**, 871 (2004).
33. J. C. White, in *Tunable Lasers*, Ed. by L. F. Mollenauer, J. C. White, and C. R. Pollock, 2nd ed. (Springer, Berlin, 1992), Topics in Applied Physics, Vol. 59.
34. R. Svensson and L. Holmlid, Phys. Rev. Lett. **83**, 1739 (1999).
35. A. Yariv, *Quantum Electronics*, 3rd ed. (Wiley, New York, 1989; Sovetskoe Radio, Moscow, 1973).
36. C. Åman, J. B. C. Pettersson, H. Lindroth, and L. Holmlid, J. Mater. Res. **7**, 100 (1992).
37. L. Holmlid and É. A. Manykin, Zh. Éksp. Teor. Fiz. **111**, 1601 (1997) [JETP **84**, 875 (1997)].
38. Y. R. Shen, *The Principles of Nonlinear Optics* (Wiley, New York, 1984; Nauka, Moscow, 1989).
39. Y. R. Shen and N. Bloembergen, Phys. Rev. A **137**, 1787 (1965).
40. L. Holmlid, J. Phys.: Condens. Matter **14**, 13469 (2002).
41. L. Holmlid, Chem. Phys. Lett. **367**, 556 (2003).
42. J. E. Dyson and D. A. Williams, *The Physics of the Interstellar Medium* (Inst. of Physics, Bristol, 1997).
43. R. C. Vermeulen, Y. M. Pihlström, W. Tschager, *et al.*, Astron. Astrophys. **404**, 861 (2003).
44. E. V. Linder, *First Principles of Cosmology* (Addison-Wesley, Harlow, 1997).

Journal of Experimental and Theoretical Physics, Vol. 100, No. 4, 2005, pp. 645–655.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 4, 2005, pp. 732–743.

NUCLEI, PARTICLES, FIELDS, GRAVITATION, AND ASTROPHYSICS

Multiple Exchanges in Lepton Pair Production in High-Energy Heavy Ion Collisions¶

E. Bartoš[a,c,*], S. R. Gevorkyan[a,d,], E. A. Kuraev[a,***], and N. N. Nikolaev[b,e,****]**

[a] *Joint Institute for Nuclear Research, Dubna, 141980 Russia*
**e-mail: bartos@thsun1.jinr.ru*
***e-mail: gevs@nusun.jinr.ru*
****e-mail: kuraev@thsun1.jinr.ru*
[b] *Landau Institute for Theoretical Physics, Chernogolovka, Moscow oblast, 142432 Russia*
[c] *Department of Theoretical Physics, Comenius University, 84248, Bratislava, Slovakia*
[d] *Yerevan Physics Institute, 375036, Yerevan, Armenia*
[e] *Institut für Kernphysik, Forschungszentrum Jülich, D-52425, Jülich, Germany*
*****e-mail: n.nikolaev@fz-juelich.de*

Received October 21, 2004

Abstract—As an archetype reaction for pQCD multigluon hard processes in collisions of ultrarelativistic nuclei, we analyze generic features of lepton pair production via multiphoton processes in peripheral heavy ion scattering. We report explicit results for collisions of two photons from one nucleus with two photons from the other nucleus, $2\gamma + 2\gamma \longrightarrow l^+l^-$. The results suggest that the familiar eikonalization of Coulomb distortions breaks down for oppositely moving Coulomb centers. The breaking of eikonalization in QED suggests that multigluon pQCD processes cannot be described in terms of collective nuclear gluon distributions. We discuss a logarithmic enhancement of the contribution from the $2\gamma + 2\gamma \longrightarrow l^+l^-$ process to production of lepton pairs with large transverse momentum; similar enhancement is absent for the $n\gamma + m\gamma \longrightarrow l^+l^-$ processes with m, $n > 2$. We comment on the general structure of multiphoton collisions and properties of higher-order terms that cannot be eikonalized. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The exact theory of Coulomb distortions of the spectrum of ultrarelativistic lepton pairs photoproduced in the Coulomb field of a nucleus has been developed by Bethe and Maximon [1]. It is based on the description of leptons by exact solutions of the Dirac equation in the Coulomb field (see, e.g., [2]). In Feynman diagram language, one has to sum multiphoton exchanges between the produced electrons and positrons and the target nucleus. For ultrarelativistic leptons, the result of this summation is the eikonal factors in the impact parameter representation. In the momentum space, the same eikonal form leads to simple recursive relations between the $(n + 1)$- and n-photon exchange amplitudes [3], where the incoming photon can be either real or virtual. There are two fundamental points behind these simple results.

(i) The lightcone momenta of ultrarelativistic leptons are conserved in a multiple scattering process (i.e., if the nucleus moves along the n_--lightcone and the produced leptons move along the n_+-lightcone, then the p_+-components of the lepton momenta are conserved).

(ii) The s-channel helicity of leptons is conserved in high-energy QED (see [2]). It is the latter property by which distortions reduce to a simple eikonal factor.

The same properties allow one to express the pair production cross section in the dipole representation [4]. They also underlie the color dipole perturbative Quantum Chromo Dynamics (pQCD) analysis of nuclear distortions and the derivation of nonlinear $k_\perp$-factorization for multijet hard processes in DIS off nuclei [5].

As shown in [6], the so-called Abelianization takes place in certain cases of practical interest. Specifically, the hard dijet production in a hadron-nucleus collision is dominated by a hard collision of an isolated parton from the beam hadron simultaneously with many gluons from the nucleus, which belong to different nucleons of a target nucleus. Nevertheless, at least for single-particle spectra, the interaction with a large number of nuclear gluons can be reduced to that with a single gluon from the collective gluon field of a nucleus; i.e., the nonlinear $k_\perp$ factorization reduces to the linear one, and in terms of the collective glue, one only needs to evaluate the familiar Born cross sections. Extending the nonlinear $k_\perp$ factorization for hard processes from hadron-nucleus collisions to collisions of ultrarelativistic nuclei is a formidable task that has not been properly

¶ This article was submitted by the authors in English.

addressed so far. The lightcone QED and QCD share many properties, and we here address a much simpler, Abelian problem of Coulomb distortions of lepton pairs produced in peripheral collisions of relativistic nuclei.

The process of lepton pair production in the Coulomb fields of two colliding ultrarelativistic heavy ions was intensely investigated recently [7–14]. Such an activity is mainly connected with new practical interest in pair production opened with operation of the facilities such as the RHIC and the LHC. Despite the high activity in this area, the issue of correct allowance for the final-state interaction of produced leptons with the colliding ion Coulomb field remains open. The main results obtained so far in this direction are as follows.

(i) The produced high-energy lepton pair interacts strongly with the Coulomb field of heavy ions, and the corresponding corrections have a noticeable impact on the cross section of the process [10].

(ii) The perturbation series corresponding to multiple interaction of a produced pair with Coulomb fields can be summed and the result can be expressed in an eikonal-like form [14] if one restricts oneself to terms growing with energy in the cross section [12]. In QED, such an approximation can be considered satisfactory, but it is not warranted in QCD, and the problem of higher-order corrections in pair production requires further investigation.

In our paper [12], we cited the amplitude $M_{(2)}^{(2)}$, which is irrelevant in the leading and next-to-leading logarithmic approximations in QED. Nevertheless, knowledge of contributions of this type becomes important for similar processes in QCD with multigluon exchanges between the color constituents of each of the colliding hadrons and the created quark–antiquark pair. This is the main motivation for our interest in multiple exchanges and their impact on the lepton pair yield in the ultrarelativistic heavy ion collisions, an issue that is not only useful in understanding the electromagnetic processes but also broadly applicable in QCD.

We skip the previously studied case where one of the ions radiates a single photon and the other radiates an arbitrary number of photons absorbed by the created pair [14]. The photon exchanges between the ions were not taken into account either [13].

This paper is organized as follows. In Section 2, we consider the case where each of the colliding ions radiates two photons, which create a lepton pair. We derive the relevant amplitude $M_{(2)}^{(2)}$ using the powerful Sudakov technique, well suited for calculations of processes at high energies. In Section 3, we study the wide-angle limit in pair production kinematics corresponding to the case of large transverse momenta of pair components. In this limit, the results are much more transparent than in the general case, as can be seen from the form of the differential cross section given below. In Section 4, we discuss the generalization of the process under consideration to the case where the number of photons exchanged by each ion exceeds two.

2. LEPTON PAIR PRODUCTION

We are interested in the process of lepton pair production in the collision of two relativistic nuclei A and B with the charge numbers Z_1 and Z_2,

$$\begin{aligned} \mathrm{A}(p_1) + \mathrm{B}(p_2) &\longrightarrow l^-(q_-) + l^+(q_+) \\ &+ \mathrm{A}(p_1') + \mathrm{B}(p_2'), \end{aligned} \tag{1}$$

with the kinematical invariants

$$\begin{gathered} s = (p_1 + p_2)^2, \quad q_1^2 = (p_1 - p_1')^2, \\ q_2^2 = (p_2 - p_2')^2, \quad s_1 = (q_+ + q_-)^2, \\ p_1^2 = p_1'^2 = M_1^2, \quad p_2^2 = p_2'^2 = M_2^2, \quad q_\pm^2 = m^2. \end{gathered} \tag{2}$$

We are interested in peripheral kinematics, i.e.,

$$s \gg M_1^2, M_2^2, |q_1^2|, |q_2^2| \gg m^2, \tag{3}$$

which corresponds to small scattering angles of ions A and B.

It is convenient to use the Sudakov parameterization for all 4-momenta entering process (1),

$$\begin{aligned} q_1 &= a_1\tilde{p}_2 + b_1\tilde{p}_1 + q_{1\perp}, \\ q_2 &= a_2\tilde{p}_2 + b_2\tilde{p}_1 + q_{2\perp}, \\ k_1 &= \alpha_1\tilde{p}_2 + \beta_1\tilde{p}_1 + k_{1\perp}, \\ k_2 &= \alpha_2\tilde{p}_2 + \beta_2\tilde{p}_1 + k_{2\perp}, \\ q_\pm &= \alpha_\pm\tilde{p}_2 + \beta_\pm\tilde{p}_1 + q_{\pm\perp}, \end{aligned} \tag{4}$$

with lightcone 4- vectors $\tilde{p}_{1,2}$ obeying the conditions

$$\tilde{p}_1^2 = \tilde{p}_2^2 = 0, \quad \tilde{p}_{1,2}\cdot q_\perp = 0, \quad 2\tilde{p}_1\cdot\tilde{p}_2 = s.$$

2.1. The Pair Production by 4-Photons

We consider the creation of a lepton pair from four virtual photons (Fig. 1). The photons with momenta k_1 and $q_1 - k_1$ (referred to as photons 1 and 2 hereafter) are emitted by ion A and the photons with momenta k_2 and $q_2 - k_2$ (referred as the photons 3 and 4) by ion B. The

leading contribution to the cross section comes from the following regions of the Sudakov variables:

$$\alpha_1 \ll \beta_1 \sim b_1, \quad \beta_+ + \beta_- = b_1,$$

$$\beta_2 \ll \alpha_2 \sim \alpha_2, \quad \alpha_+ + \alpha_- = a_2,$$

$$|a_1| \ll a_2, \quad |b_2| \ll b_1, \quad q_{i\perp} = \mathbf{q}_i, \tag{5}$$

$$\mathbf{q}_1 + \mathbf{q}_2 = \mathbf{q}_+ + \mathbf{q}_-,$$

$$\alpha_\pm = \frac{\mathbf{q}_\pm^2}{s\beta_\pm}, \quad \mathbf{q}_\pm^2 \gg m^2.$$

Hereinafter, the boldface $\mathbf{q}_i$, denotes the two-dimensional transverse part of any considered 4-momentum. For definiteness, we assume $\beta_+, \beta_- > 0$, which corresponds to the situation where the pair moves along the momentum of ion A (the momentum p_1). With a possible extension to pQCD in mind, we neglect the lepton masses whenever appropriate.

The contribution to the matrix element of such a set of the Feynman diagrams (FD) is given by

$$M_{(2)}^{(2)} = is\frac{(Z_1Z_2)^2(4\pi\alpha)^4}{(2\pi)^8}$$
$$\times\int\frac{d^4k_1d^4k_2}{k_1^2k_2^2(q_1-k_1)^2(q_2-k_2)^2}$$
$$\times\frac{1}{s}\bar u^\eta(p_1')O_1^{\mu_1\nu_1}u^\eta(p_1)\bar u^\lambda(p_2')O_2^{\rho_1\sigma_1}u^\lambda(p_2) \tag{6}$$
$$\times\bar u(q_-)T^{\mu\nu\rho\sigma}v(q_+)g_{\mu\mu_1}g_{\nu\nu_1}g_{\rho\rho_1}g_{\sigma\sigma_1},$$

where u and v are the leptonic Dirac bispinors and O_1, O_2, and T are the corresponding tensors of the upper, down, and pair production blocks. To see the proportionality of matrix element (6) to the invariant energy s, we use the Gribov representation for the virtual photon Green functions

$$g_{\mu\mu_1}g_{\nu\nu_1}g_{\rho\rho_1}g_{\sigma\sigma_1}$$
$$\approx\left(\frac{2}{s}\right)^4 p_{1\mu}p_{1\nu}p_{1\rho_1}p_{1\sigma_1}p_{2\mu_1}p_{2\nu_1}p_{2\rho}p_{2\sigma}. \tag{7}$$

The numerators of the Green functions of nucleus A can be written as s^2N_1 with

$$N_1 = \frac{1}{s}\bar u^\eta(p_1')\hat p_2u^\eta(p_1), \quad \sum_\eta|N_1|^2 = 2,$$

and a similar expression exists for nucleus B. The denominators of the virtual photon Green functions in the considered kinematics depend only on the transverse components of the corresponding 4-vectors, and therefore

$$k_1^2k_2^2(q_1-k_1)^2(q_2-k_2)^2 = \mathbf{k}_1^2\mathbf{k}_2^2(\mathbf{q}_1-\mathbf{k}_1)^2(\mathbf{q}_2-\mathbf{k}_2)^2.$$

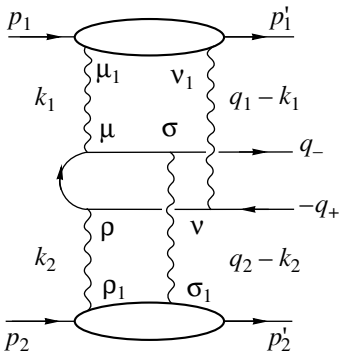


Fig. 1. Typical Feynman diagram for amplitude $M_{(2)}^{(2)}$.

There are 24 FD contributing to $M_{(2)}^{(2)}$. Instead of them, it is convenient to consider $24 \cdot 2 \cdot 2 = 96$ FD with all possible permutations of emission and absorption points of the photons exchanged by the nuclei (Fig. 2). Then the result must be divided by $(2!)^2$. This trick [15] provides the convergence of integrals over β_2,

$$\frac{1}{2\pi i}\int_{-\infty}^{\infty}d\beta_2\left[\frac{s}{s\beta_2-c+i0}+\frac{s}{-s\beta_2-(d+i0)}\right] = -1, \tag{8}$$

and of a similar integral over α_1. After all operations, we can write the matrix element as

$$M_{(2)}^{(2)} = is\frac{(16\pi\alpha^2Z_1Z_2)^2N_1N_2}{(2!)^2}$$
$$\times\int\frac{d^2k_1d^2k_2}{\pi^2}\frac{\bar u(q_-)Rv(q_+)}{\mathbf{k}_1^2\mathbf{k}_2^2(\mathbf{q}_1-\mathbf{k}_1)^2(\mathbf{q}_2-\mathbf{k}_2)^2}, \tag{9}$$

where

$$R = \frac{1}{s}\int\frac{d\beta_1d\alpha_2}{(2\pi i)^2}p_{1\mu}p_{1\nu}p_{2\rho}p_{2\sigma}T^{\mu\nu\rho\sigma}.$$

2.2. Classification of Feynman Diagrams

It is convenient to classify FDs by the ordering of the exchanged photons absorbed by the lepton world line (Fig. 3). We label them as R_{ijkl}, $R = \sum R_{ijkl}$, with pairwise distinct integers i, j, k, l from one to four, counting from a negative lepton emission point.

(a) We first consider the set of four FDs (Fig. 4a), labeled R_{1234}, R_{2134}, R_{1243}, and R_{2143}, in which the interactions with two nuclei are ordered consecutively against the lepton line direction. The sum of the relevant contributions provides the convergence of the β_1

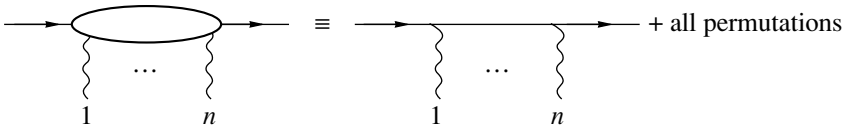


Fig. 2. Notation for permutations of n virtual photons emitted by a heavy ion.

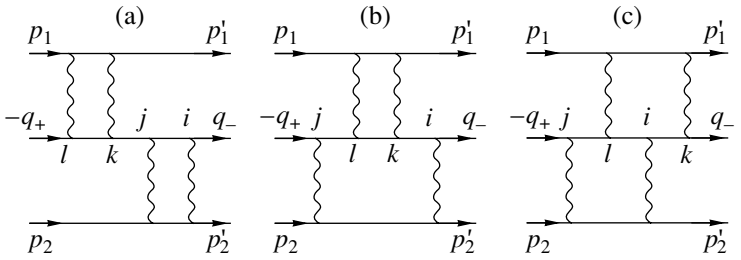


Fig. 3. Set of basic Feynman diagrams for amplitude $M_{(2)}^{(2)}$.

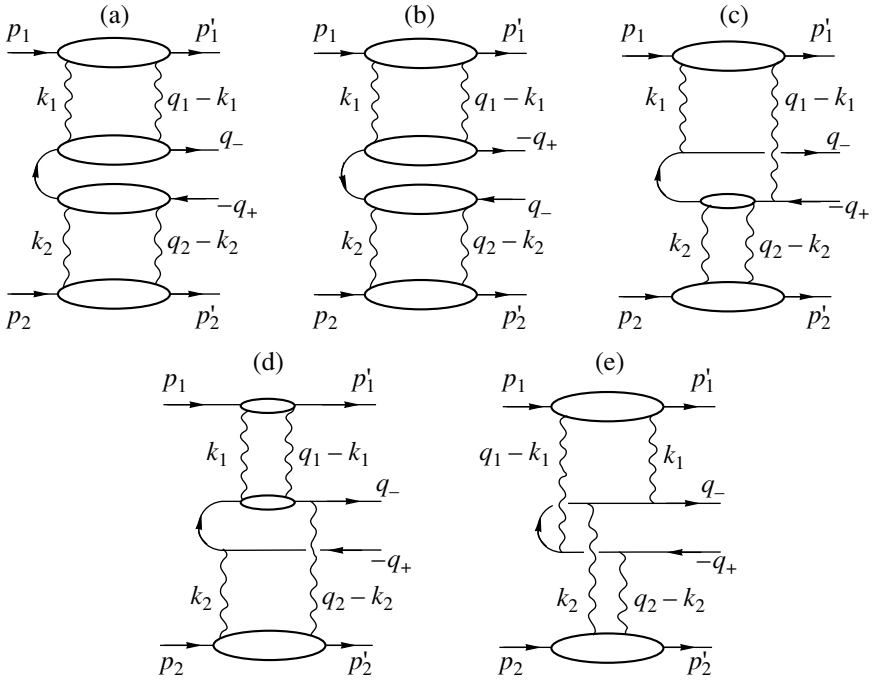


Fig. 4. Feynman diagrams for amplitude $M_{(2)}^{(2)}$.

and α_2 integrations. After a standard calculation, we obtain

$$\begin{aligned} &R_{1234} + R_{2134} + R_{1243} + R_{2143} \\ &= \frac{\beta_- \hat{p}_1(\hat{q}_- - \hat{q}_1)_\perp}{\beta_+ \mathbf{q}_-^2 + \beta_-(\mathbf{q}_- - \mathbf{q}_1)^2}\frac{\hat{p}_2}{s} = -B\frac{\hat{p}_2}{s}, \\ &B = \frac{\hat{q}_{-\perp}(\hat{q}_- - \hat{q}_1)_\perp}{\beta_+ \mathbf{q}_-^2 + \beta_-(\mathbf{q}_- - \mathbf{q}_1)^2}. \end{aligned} \tag{10}$$

The last equality in Eq. (10) is the result of the Dirac equation for massless particles,

$$\bar{u}(q_-)\beta_-\hat{p}_1\hat{p}_2 = -\bar{u}(q_-)\hat{q}_{-\perp}\hat{p}_2. \tag{11}$$

A result similar to Eq. (10) is obtained for the set of the crossing diagrams (Fig. 4b) corresponding to the R_{3412}, R_{3421}, R_{4312}, and R_{4321} terms in the amplitude, with

only the replacement $B \longrightarrow \tilde{B}$, where

$$\tilde{B} = \frac{(-\hat{q}_+ + \hat{q}_1)_\perp \hat{q}_{+\perp}}{\beta_- \mathbf{q}_+^2 + \beta_+ (\mathbf{q}_1 - \mathbf{q}_+)^2}. \tag{12}$$

(b) We next consider the set of diagrams R_{1342}, R_{1432}, R_{2341}, R_{2431} (Fig. 4c) and R_{3124}, R_{3214}, R_{4123}, R_{4213} (Fig. 4d), where exchanges with ion B (A) are attached to the lepton line between the interactions with ion A (B).

For definiteness, we consider the sum $R_{1342} + R_{1432}$. Using the relevant denominators of the lepton line, we obtain the following integrals over β_1 and α_2:

$$\begin{aligned}
&\int \frac{d\beta_1}{2\pi i} \frac{1}{s\alpha_-(\beta_- - \beta_1) - (\mathbf{q}_- - \mathbf{k}_1)^2 + i0} \\
&\times \frac{1}{-s\alpha_+(\beta_- - \beta_1) - (-\mathbf{q}_+ + \mathbf{q}_1 - \mathbf{k}_1)^2 + i0} \\
&\times \int \frac{d\alpha_2}{2\pi i} \left[\frac{s(\beta_- - \beta_1)}{s(\beta_- - \beta_1)(\alpha_- - \alpha_2) - (\mathbf{q}_- - \mathbf{k}_1)^2 + i0} \right. \\
&\left. + \frac{s(\beta_- - \beta_1)}{s(\beta_- - \beta_1)(-\alpha_+ + \alpha_2) - (-\mathbf{q}_+ + \mathbf{q}_1 - \mathbf{k}_1)^2 + i0} \right].
\end{aligned} \tag{13}$$

The second integral after closing the integration contour in the lower half-plane gives the function $\mathrm{sgn}(\beta_- - \beta_1)$, and hence Eq. (13) becomes

$$\begin{aligned}
&\int \frac{d\beta_1}{2\pi i} \frac{\mathrm{sgn}(\beta_1 - \beta_-)}{s\alpha_-(\beta_- - \beta_1) - (\mathbf{q}_- - \mathbf{k}_1)^2 + i0} \\
&\times \frac{1}{-s\alpha_+(\beta_- - \beta_1) - (-\mathbf{q}_+ + \mathbf{q}_1 - \mathbf{k}_1)^2 + i0}.
\end{aligned} \tag{14}$$

Using the relation

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{dx}{2\pi i} \frac{\mathrm{sgn}\, x}{(-ax - b + i0)(cx - d + i0)} \\
&= \frac{1}{\pi i (ad + bc)} \ln \frac{ad}{bc},
\end{aligned} \tag{15}$$

we obtain the result

$$\begin{aligned}
&R_{1342} + R_{1432} + R_{2341} + R_{2431} \\
&= \frac{\hat{p}_1}{i\pi s} \left[\frac{(\hat{q}_- - \hat{k}_1)_\perp (-\hat{q}_+ + \hat{q}_1 - \hat{k}_1)_\perp}{\alpha_+ (\mathbf{q}_- - \mathbf{k}_1)^2 + \alpha_- (-\mathbf{q}_+ + \mathbf{q}_1 - \mathbf{k}_1)^2} \right. \\
&\times \ln \frac{\alpha_+ (\mathbf{q}_- - \mathbf{k}_1)^2}{\alpha_- (-\mathbf{q}_+ + \mathbf{q}_1 - \mathbf{k}_1)^2} \\
&+ \frac{(\hat{q}_- - \hat{q}_1 + \hat{k}_1)_\perp (-\hat{q}_+ + \hat{k}_1)_\perp}{\alpha_+ (\mathbf{q}_- - \mathbf{q}_1 + \mathbf{k}_1)^2 + \alpha_- (-\mathbf{q}_+ + \mathbf{k}_1)^2} \\
&\left. \times \ln \frac{\alpha_+ (\mathbf{q}_- - \mathbf{q}_1 + \mathbf{k}_1)^2}{\alpha_- (-\mathbf{q}_+ + \mathbf{k}_1)^2} \right],
\end{aligned} \tag{16}$$

$$\begin{aligned}
&R_{3124} + R_{3214} + R_{4123} + R_{4213} \\
&= \frac{\hat{p}_2}{i\pi s} \left[\frac{(\hat{q}_- - \hat{k}_2)_\perp (-\hat{q}_+ + \hat{q}_2 - \hat{k}_2)_\perp}{\beta_+ (\mathbf{q}_- - \mathbf{k}_2)^2 + \beta_- (-\mathbf{q}_+ + \mathbf{q}_2 - \mathbf{k}_2)^2} \right. \\
&\times \ln \frac{\beta_- (-\mathbf{q}_+ + \mathbf{q}_2 - \mathbf{k}_2)^2}{\beta_+ (\mathbf{q}_- - \mathbf{k}_2)^2} \\
&+ \frac{(\hat{q}_- - \hat{q}_2 + \hat{k}_2)_\perp (-\hat{q}_+ + \hat{k}_2)_\perp}{\beta_+ (\mathbf{q}_- - \mathbf{q}_2 + \mathbf{k}_2)^2 + \beta_- (-\mathbf{q}_+ + \mathbf{k}_2)^2} \\
&\left. \times \ln \frac{\beta_- (-\mathbf{q}_+ + \mathbf{k}_2)^2}{\beta_+ (\mathbf{q}_- - \mathbf{q}_2 + \mathbf{k}_2)^2} \right].
\end{aligned}$$

We note that the expressions (16) are purely imaginary, and therefore their interference with the Born term in the cross section is zero.

(c) We now consider the case of interactions with different nuclei alternating along the lepton line, for instance, the amplitude R_{1324} (Fig. 4e). After some algebra, we obtain the relevant numerator

$$\begin{aligned}
N_{1324} &= s\hat{p}_1 \hat{p}_2 (\hat{q}_- - \hat{k}_1)_\perp \\
&\times (\hat{q}_- - \hat{k}_1 - \hat{k}_2)_\perp (\hat{q}_- - \hat{q}_1 - \hat{k}_2)_\perp,
\end{aligned} \tag{17}$$

which is very different from the numerators of Born-like amplitudes. Specifically, it is a higher-order term in the running transverse momenta $\mathbf{k}_i$.

The relevant denominators are given by

$$\begin{aligned}
\{1\} &\equiv (q_- - k_1)^2 + i0 \\
&= s(\beta_- - \beta_1)\alpha_- - (\mathbf{q}_- - \mathbf{k}_1)^2 + i0, \\
\{2\} &\equiv (q_- - k_1 - k_2)^2 + i0 \\
&= s(\beta_- - \beta_1)(\alpha_- - \alpha_2) - (\mathbf{q}_- - \mathbf{k}_1 - \mathbf{k}_2)^2 + i0, \\
\{3\} &\equiv (-q_+ + q_2 - k_2)^2 + i0 \\
&= s(-\beta_+)(\alpha_- - \alpha_2) - (-\mathbf{q}_+ + \mathbf{q}_2 - \mathbf{k}_2)^2 + i0.
\end{aligned} \tag{18}$$

The nonvanishing contribution only emerges if the poles are located in different α_2 half-planes, which

Coefficients in formula (23). The brackets denote index permutation, e.g., (12) ≡ 12 + 21

n	R_{ijkl}	a_n	b_n	c_n	d_n
1	$R_{(12)(34)}$	q_-	$q_- - q_1$	–	–
2	$R_{(34)(12)}$	$q_1 - q_+$	q_+	–	–
3	R_{1324}	q_-	$q_- - k_1$	$q_- - k_1 - k_2$	$q_- - q_1 - k_2$
4	R_{1423}	q_-	$q_- - k_1$	$q_- - q_2 + k_2 - k_1$	$-q_+ + k_2$
5	R_{2314}	q_-	$q_- - q_1 + k_1$	$q_- - q_1 + k_1 - k_2$	$-q_+ + q_2 - k_2$
6	R_{2413}	q_-	$q_- - q_1 + k_1$	$-q_+ + k_1 + k_2$	$-q_+ + k_2$
7	R_{4231}	$q_- - q_2 + k_2$	$-q_+ + k_1 + k_2$	$-q_+ + k_1$	q_+
8	R_{3241}	$q_- - k_2$	$q_- - q_1 + k_1 - k_2$	$-q_+ + k_1$	q_+
9	R_{4132}	$q_- - q_2 + k_2$	$q_- - q_2 + k_2 - k_1$	$-q_+ + q_1 - k_1$	q_+
10	R_{3142}	$q_- - k_2$	$q_- - k_1 - k_2$	$-q_+ + q_1 - k_1$	q_+
11	$R_{3(12)4}$	$q_- - k_2$	$-q_+ + q_2 - k_2$	–	–
12	$R_{4(12)3}$	$q_- - q_2 + k_2$	$-q_+ + k_2$	–	–

takes place only if $\beta_1 < \beta_-$ ($\beta_\pm > 0$). Taking the residue at pole $\{2\}$, we find

$$\int \frac{s d\alpha_2}{2\pi i}\frac{1}{\{2\}\{3\}} = -\frac{\theta(\beta_- - \beta_1)}{(\beta_1 - \beta_-)(-\mathbf{q}_+ + \mathbf{q}_2 - \mathbf{k}_2)^2 - \beta_+(\mathbf{q}_- - \mathbf{k}_1 - \mathbf{k}_2)^2}. \tag{19}$$

Further integration over β_1 can be performed using the relation

$$\int_{-\infty}^{\infty}\frac{dx}{2\pi i}\frac{\theta(x)}{(ax - b + i0)(cx + d + i0)} = -\frac{1}{2(ad + bc)}\left(1 + \frac{i}{\pi}\ln\frac{ad}{bc}\right), \tag{20}$$

with the result

$$R_{1324} = -\frac{\beta_- N_{1324}}{2sD_{1324}}\left(1 + \frac{i}{\pi}\ln\frac{ad}{bc}\right),$$

$$D_{1324} = \beta_-(\mathbf{q}_- - \mathbf{k}_1)^2(-\mathbf{q}_+ + \mathbf{q}_2 - \mathbf{k}_2)^2 + \beta_+\mathbf{q}_-^2(\mathbf{q}_- - \mathbf{k}_1 - \mathbf{k}_2)^2 = ad + bc. \tag{21}$$

The highly nonlinear denominator given by Eq. (21) makes the contribution of the considered case dramatically different from the Born amplitude and corrections to it from the higher-order processes in which only one photon is emitted by one of the ions [12]. Technically, the nonlinearity is not surprising because of the related nonlinearity of the numerator. The principal difference from the Born-like amplitude is that with the alternating ordering of interactions, we have the situation in which the p_+ component of the lightcone momentum-is conserved in the scattering on one ion but is not conserved in the scattering on the second ion. Depending on the ordering of interaction vertices and the order of integrations, we encounter a sequence of vertices with conservation and nonconservation of the p_--component of the lightcone momentum.

Similar results can be obtained for other contributions of these types.

(d) The final result is given by (see table)

$$M_{(2)}^{(2)} = \frac{is}{(2!)^2}(16\pi\alpha^2 Z_1 Z_2)^2 N_1 N_2 \times \int \frac{d^2k_1}{\pi}\frac{d^2k_2}{\pi}\frac{\bar{u}(q_-)R_{(2)}^{(2)}\hat{p}_2 v(q_+)}{s\mathbf{k}_1^2\mathbf{k}_2^2(\mathbf{q}_1 - \mathbf{k}_1)^2(\mathbf{q}_2 - \mathbf{k}_2)^2}, \tag{22}$$

$$R_{(2)}^{(2)} = \sum_{n=1}^{2}\frac{[\hat{a}_n\hat{b}_n]_\perp}{\beta_-\mathbf{b}_n^2 + \beta_+\mathbf{a}_n^2} - \sum_{n=3}^{10}\frac{[\hat{a}_n\hat{b}_n\hat{c}_n\hat{d}_n]_\perp}{2[\beta_-\mathbf{b}_n^2\mathbf{d}_n^2 + \beta_+\mathbf{a}_n^2\mathbf{c}_n^2]} \times\left(1 + i\frac{(-1)^{n+1}}{\pi}\ln\frac{\beta_-\mathbf{b}_n^2\mathbf{d}_n^2}{\beta_+\mathbf{a}_n^2\mathbf{c}_n^2}\right) + \sum_{n=11}^{12} i\frac{(-1)^{n+1}}{\pi}\frac{[\hat{a}_n\hat{b}_n]_\perp}{\beta_-\mathbf{b}_n^2 + \beta_+\mathbf{a}_n^2}\ln\frac{\beta_-\mathbf{b}_n^2}{\beta_+\mathbf{a}_n^2}. \tag{23}$$

To verify gauge invariance, we give the explicit form for the real part of the amplitude:

$$\mathrm{Re}R_{(2)}^{(2)} = \frac{[\hat{q}_-(\hat{q}_- - \hat{q}_1)]_\perp}{\beta_+\mathbf{q}_-^2 + \beta_-(\mathbf{q}_- - \mathbf{q}_1)^2} + \frac{[(-\hat{q}_+ + \hat{q}_1)\hat{q}_+]_\perp}{\beta_-\mathbf{q}_+^2 + \beta_+(\mathbf{q}_+ - \mathbf{q}_1)^2}$$

$$- \frac{[\hat{q}_-(\hat{q}_- - \hat{k}_1)(\hat{q}_- - \hat{k}_1 - \hat{k}_2)(\hat{q}_- - \hat{q}_1 - \hat{k}_2)]_\perp}{2[\beta_-(\mathbf{q}_- - \mathbf{k}_1)^2(-\mathbf{q}_+ + \mathbf{q}_2 - \mathbf{k}_2)^2 + \beta_+\mathbf{q}_-^2(\mathbf{q}_- - \mathbf{k}_1 - \mathbf{k}_2)^2]}$$

$$- \frac{[\hat{q}_-(\hat{q}_- - \hat{k}_1)(\hat{q}_- - \hat{q}_2 + \hat{k}_2 - \hat{k}_1)(-\hat{q}_+ + \hat{k}_2)]_\perp}{2[\beta_-(\mathbf{q}_- - \mathbf{k}_1)^2(-\mathbf{q}_+ + \mathbf{k}_2)^2 + \beta_+\mathbf{q}_-^2(\mathbf{q}_- - \mathbf{q}_2 + \mathbf{k}_2 - \mathbf{k}_1)^2]}$$

$$- \frac{[\hat{q}_-(\hat{q}_- - \hat{q}_1 + \hat{k}_1)(\hat{q}_- - \hat{q}_1 + \hat{k}_1 - \hat{k}_2)(-\hat{q}_+ + \hat{q}_2 - \hat{k}_2)]_\perp}{2[\beta_-(\mathbf{q}_- - \mathbf{q}_1 + \mathbf{k}_1)^2(-\mathbf{q}_+ + \mathbf{q}_2 - \mathbf{k}_2)^2 + \beta_+\mathbf{q}_-^2(\mathbf{q}_- - \mathbf{q}_1 + \mathbf{k}_1 - \mathbf{k}_2)^2]}$$

$$- \frac{[\hat{q}_-(\hat{q}_- - \hat{q}_1 + \hat{k}_1)(-\hat{q}_+ + \hat{k}_1 + \hat{k}_2)(-\hat{q}_+ + \hat{k}_2)]_\perp}{2[\beta_-(\mathbf{q}_- - \mathbf{q}_1 + \mathbf{k}_1)^2(-\mathbf{q}_+ + \mathbf{k}_2)^2 + \beta_+\mathbf{q}_-^2(-\mathbf{q}_+ + \mathbf{k}_1 + \mathbf{k}_2)^2]}$$

$$- \frac{[(\hat{q}_- - \hat{q}_2 + \hat{k}_2)(-\hat{q}_+ + \hat{k}_1 + \hat{k}_2)(-\hat{q}_+ + \hat{k}_1)\hat{q}_+]_\perp}{2[\beta_-\mathbf{q}_+^2(-\mathbf{q}_+ + \mathbf{k}_1 + \mathbf{k}_2)^2 + \beta_+(-\mathbf{q}_+ + \mathbf{k}_1)^2(\mathbf{q}_- - \mathbf{q}_2 + \mathbf{k}_2)^2]}$$

$$- \frac{[(\hat{q}_- - \hat{k}_2)(\hat{q}_- - \hat{q}_1 + \hat{k}_1 - \hat{k}_2)(-\hat{q}_+ + \hat{k}_1)\hat{q}_+]_\perp}{2[\beta_-\mathbf{q}_+^2(\mathbf{q}_- - \mathbf{q}_1 + \mathbf{k}_1 - \mathbf{k}_2)^2 + \beta_+(-\mathbf{q}_+ + \mathbf{k}_1)^2(\mathbf{q}_- - \mathbf{k}_2)^2]}$$

$$- \frac{[(\hat{q}_- - \hat{q}_2 + \hat{k}_2)(\hat{q}_- - \hat{q}_2 + \hat{k}_2 - \hat{k}_1)(-\hat{q}_+ + \hat{q}_1 - \hat{k}_1)\hat{q}_+]_\perp}{2[\beta_-\mathbf{q}_+^2(\mathbf{q}_- - \mathbf{q}_2 + \mathbf{k}_2 - \mathbf{k}_1)^2 + \beta_+(\mathbf{q}_- - \mathbf{q}_2 + \mathbf{k}_2)^2(-\mathbf{q}_+ + \mathbf{q}_1 - \mathbf{k}_1)^2]}$$

$$- \frac{[(\hat{q}_- - \hat{k}_2)(\hat{q}_- - \hat{k}_1 - \hat{k}_2)(-\hat{q}_+ + \hat{q}_1 - \hat{k}_1)\hat{q}_+]_\perp}{2[\beta_-\mathbf{q}_+^2(\mathbf{q}_- - \mathbf{k}_1 - \mathbf{k}_2)^2 + \beta_+(-\mathbf{q}_+ + \mathbf{q}_1 - \mathbf{k}_1)^2(\mathbf{q}_- - \mathbf{k}_2)^2]}.$$

We can then verify that the following condition is satisfied:

$$\mathrm{Re}R_{(2)}^{(2)} = 0 \text{ if } \mathbf{k}_1 = 0 \text{ or } \mathbf{k}_2 = 0 \text{ or } \mathbf{k}_1 = \mathbf{q}_1 \text{ or } \mathbf{k}_2 = \mathbf{q}_2. \quad (24)$$

This fact is also correct for the whole amplitude (23). This property (24) is crucial for the gauge invariance and infrared convergence of integrations over d^2k_i.

In the loop integration, we can shift the integration variable as $\mathbf{k}_i \longrightarrow \mathbf{q}_i - \mathbf{k}_i$. Then expression (23) for $\mathrm{Re}R_{(2)}^{(2)}$ can be simplified to

$$\mathrm{Re}R_{(2)}^{(2)} = \frac{\hat{q}_{-\perp}(\hat{q}_- - \hat{q}_1)_\perp}{\beta_+\mathbf{q}_-^2 + \beta_-(\mathbf{q}_- - \mathbf{q}_1)^2} + \frac{(-\hat{q}_+ + \hat{q}_1)_\perp\hat{q}_{+\perp}}{\beta_-\mathbf{q}_+^2 + \beta_+(\mathbf{q}_1 - \mathbf{q}_+)^2}$$

$$- 2\frac{[\hat{q}_-(\hat{q}_- - \hat{k}_1)(\hat{q}_- - \hat{k}_1 - \hat{k}_2)(\hat{q}_- - \hat{q}_1 - \hat{k}_2)]_\perp}{\beta_-(\mathbf{q}_- - \mathbf{k}_1)^2(-\mathbf{q}_+ + \mathbf{q}_2 - \mathbf{k}_2)^2 + \beta_+\mathbf{q}_-^2(\mathbf{q}_- - \mathbf{k}_1 - \mathbf{k}_2)^2} \quad (25)$$

$$- 2\frac{[(-\hat{q}_+ + \hat{q}_1 + \hat{k}_2)(-\hat{q}_+ + \hat{k}_1 + \hat{k}_2)(-\hat{q}_+ + \hat{k}_1)\hat{q}_+]_\perp}{\beta_-\mathbf{q}_+^2(-\mathbf{q}_+ + \mathbf{k}_1 + \mathbf{k}_2)^2 + \beta_+(-\mathbf{q}_+ + \mathbf{k}_1)^2(\mathbf{q}_- - \mathbf{q}_2 + \mathbf{k}_2)^2}.$$

Although the gauge invariance property is not manifested here, as in the previous case, the final results after integration over k_i coincide.

3. THE WIDE-ANGLE LIMIT OF THE $M_{(2)}^{(2)}$ AMPLITUDE

We consider the behavior of expression (25) in the case where the transverse component of lepton momenta is large compared to the momenta transferred to the ions,

$$\mathbf{q}_- \approx -\mathbf{q}_+ = \mathbf{q}, \quad |\mathbf{q}| \gg |\mathbf{q}_{1,2}|. \tag{26}$$

The main contribution to the matrix element is then given by the region

$$|\mathbf{q}_i| \ll |\mathbf{k}_i| \ll |\mathbf{q}|. \tag{27}$$

The amplitude $M_{(1)}^{(1)}$ is

$$M_{(1)}^{(1)} = -is\frac{(8\pi\alpha)^2 N_1 N_2 Z_1 Z_2}{\mathbf{q}_1^2\mathbf{q}_2^2}\bar{u}(q_-)\frac{R_{(1)}^{(1)}}{s}v(q_+),$$

$$R_{(1)}^{(1)} = \hat{p}_1\frac{\hat{q}_- - \hat{q}_1}{(q_- - q_1)^2}\hat{p}_2 + \hat{p}_2\frac{\hat{q}_1 - \hat{q}_+}{(q_1 - q_+)^2}\hat{p}_1 \tag{28}$$

$$= (B - \tilde{B})\hat{p}_2.$$

For wide-angle kinematics, we have

$$\frac{1}{s}R_{(1)}^{(1)} = \frac{\hat{p}_2}{s}\frac{1}{b_1^2(\mathbf{q}^2)^2}[2\mathbf{q}\cdot\mathbf{q}_2(b_1\hat{q}\hat{q}_1 + 2\beta_-\mathbf{q}\cdot\mathbf{q}_1) + \mathbf{q}^2(b_1\hat{q}_1\hat{q}_2 + 2\beta_+\mathbf{q}_1\cdot\mathbf{q}_2)], \tag{29}$$

where $\beta_1 = \beta_- + \beta_+$, $\mathbf{q} = \mathbf{q}_- \approx -\mathbf{q}_+$, and $\mathbf{q}_{1,2}$ are the momenta transferred to ions.

For the matrix element $M_{(2)}^{(1)}$ we have (in agreement with the result obtained in paper [16])

$$M_{(2)}^{(1)} = -s\frac{2^7\pi^2\alpha^3 Z_1 Z_2^2 N_1 N_2}{\mathbf{q}_1^2} \times \int\frac{d^2k_2}{\pi}\frac{\bar{u}(q_-)R_{(2)}^{(1)}\hat{p}_2 v(q_+)}{s\mathbf{k}_2^2(\mathbf{q}_2 - \mathbf{k}_2)^2}, \tag{30}$$

where

$$R_{(2)}^{(1)} = B + \tilde{B} - \frac{(\hat{q}_- - \hat{k}_2)_\perp(\hat{q}_- - \hat{q}_1 - \hat{k}_2)_\perp}{\beta_-(\mathbf{q}_- - \mathbf{q}_1 - \mathbf{k}_2)^2 + \beta_+(\mathbf{q}_- - \mathbf{k}_2)^2} - \frac{(\hat{q}_+ - \hat{k}_2 - \hat{q}_1)_\perp(\hat{q}_+ - \hat{k}_2)_\perp}{\beta_+(\mathbf{q}_- - \mathbf{q}_2 + \mathbf{k}_2)^2 + \beta_-(\mathbf{q}_+ - \mathbf{k}_2)^2}. \tag{31}$$

In the considered limit, this expression becomes

$$R_{(2)}^{(1)} \sim \frac{1}{b_1\mathbf{q}^2}\Bigg[(2\beta_-\mathbf{q}_-\cdot\mathbf{q}_1 + \hat{q}_-\hat{q}_1) \times\left(\frac{4(\mathbf{q}_-\cdot\mathbf{k}_2)^2}{(\mathbf{q}^2)^2} - \frac{\mathbf{k}_2^2}{\mathbf{q}^2}\right) - \frac{2\mathbf{q}_-\cdot\mathbf{k}_2}{\mathbf{q}^2}(\hat{k}_2\hat{q}_1 + 2\beta_-\mathbf{k}_2\cdot\mathbf{q}_1)\Bigg] + (\beta_- \longrightarrow \beta_+), \quad |\mathbf{k}_2| \gg |\mathbf{q}_2|. \tag{32}$$

This expression vanishes after angular averaging. It can be shown that the quantity $M_{(3)}^{(1)}$ also vanishes in the limit of wide-angle pair production and is proportional to $|\mathbf{q}_2|/|\mathbf{q}| \ll 1$, which is in agreement with [3].

For the amplitude $M_{(2)}^{(2)}$ in Eq. (22), the quantity $R_{(2)}^{(2)}$ plays the role of a cut-off parameter in the region $|\mathbf{k}_i| > |\mathbf{q}|$. From very general arguments, it can be written in the form

$$\mathrm{Re}R_{(2)}^{(2)} \approx \frac{[k_1^\mu(q_1 - k_1)^\nu k_2^\alpha(q_2 - k_2)^\beta]_\perp}{(\mathbf{q}^2)^2}R_{\mu\nu\alpha\beta}, \tag{33}$$

with some dimensionless tensor matrix $R_{\mu\nu\alpha\beta}$ independent of $\mathbf{k}_i$ and $\mathbf{q}_i$. Expanding expression (25), we obtain

$$\int\frac{d^2k_1 d^2k_2}{\pi^2}\frac{\mathrm{Re}R_{(2)}^{(2)}}{\mathbf{k}_1^2\mathbf{k}_2^2(\mathbf{q}_1 - \mathbf{k}_1)^2 q(\mathbf{q}_2 - \mathbf{k}_2)^2} \approx \frac{I}{(\mathbf{q}^2)^2}\frac{4(\beta_+ - \beta_-)}{(\beta_- + \beta_+)^2}\ln\frac{\mathbf{q}_{\max}^2}{\mathbf{q}_1^2}\ln\frac{\mathbf{q}_{\max}^2}{\mathbf{q}_2^2}, \tag{34}$$

where I is the unit matrix and $q_{\max} \approx 1/R$ is the upper integration limit, with R being the nucleus radius. Such a logarithmic enhancement is absent if the number of the exchanged photons from each ion exceeds two (Fig. 5). In fact, the amplitudes $M_{(n)}^{(2)}$, $M_{(2)}^{(n)}$, $n > 2$ contain only the first power of the large logarithm, whereas $M_{(n)}^{(m)}$, m, $n > 2$ do not contain such a factor at all because the corresponding loop momenta integrals are

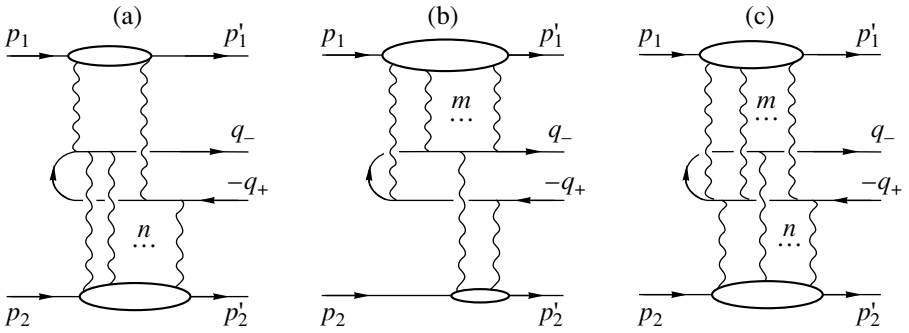


Fig. 5. Some Feynman diagrams for amplitudes of the type $M_{(n)}^{(2)}$ (a), $M_{(2)}^{(n)}$ (b), and $M_{(n)}^{(m)}$ (c) with $m, n \geq 2$.

convergent in both infrared and ultraviolet regions and we can safely put $|\mathbf{q}_{1(2)}| = 0$ in loop integrations.

Thus, the differential cross section for the considered kinematics is determined by the interference term $(M_{(1)}^{(1)})^* M_{(2)}^{(2)}$, which has the form (for comparison, we also present the Born term)

$$\frac{d\sigma_0}{db_1 dx} = \frac{16(Z_1 Z_2 \alpha^2)^2}{\pi^4} \times \frac{x^2 + (1-x)^2}{\mathbf{q}_1^2 \mathbf{q}_2^2 (\mathbf{q}^2)^2 b_1} d^2q_1 d^2q_2 d^2q, \tag{35}$$

$$\frac{d\sigma_{\text{int}}}{db_1 dx} = \frac{16(Z_1 Z_2 \alpha^2)^3}{\mathbf{q}_1^2 \mathbf{q}_2^2 \mathbf{q}_+^2 \mathbf{q}_-^2} \frac{1-2x}{b_1} \times \ln\frac{\mathbf{q}_{\max}^2}{\mathbf{q}_1^2} \ln\frac{\mathbf{q}_{\max}^2}{\mathbf{q}_2^2} Q d^2q_1 d^2q_2 d^2q_-, \tag{36}$$

where

$$Q = \frac{\mathbf{q}_- \cdot (\mathbf{q}_1 - \mathbf{q}_-)}{(1-x)\mathbf{q}_-^2 + x(\mathbf{q}_- - \mathbf{q}_1)^2} + \frac{\mathbf{q}_+ \cdot (\mathbf{q}_+ - \mathbf{q}_1)}{x\mathbf{q}_+^2 + (1-x)(\mathbf{q}_1 - \mathbf{q}_+)^2},$$

$$x = \frac{\beta_-}{b_1}, \quad \epsilon < x, \quad b_1 < 1 - \epsilon, \quad \epsilon = \frac{4m^2 x(1-x)}{\mathbf{q}_\pm^2}.$$

We note that expression (36) is symmetric under simultaneous substitutions $\mathbf{q}_+ \longleftrightarrow \mathbf{q}_-$ and $\beta_+ \longleftrightarrow \beta_-$ due to the *C*-even nature of the interference.

Finally, from a very straightforward generalization of (33), it can be shown that the set of amplitudes with an odd number of exchanges with one or both nuclei is suppressed in the limit of wide-angle production:

$$M_{(2n+1)}^{(2m)} \sim O\left(\frac{|\mathbf{q}_1|}{|\mathbf{q}|}\right), \quad M_{(2n)}^{(2m+1)} \sim O\left(\frac{|\mathbf{q}_2|}{|\mathbf{q}|}\right), \quad M_{(2n+1)}^{(2m+1)} \sim O\left(\frac{|\mathbf{q}_1||\mathbf{q}_2|}{|\mathbf{q}^2|}\right). \tag{37}$$

4. MULTIPHOTON EXCHANGE

We now generalize the above picture to the case of multiple photon exchanges ($m, n > 2$). Using the relation

$$I_n = \frac{1}{\pi^{n-1}} \times \int \frac{d^2k_1 \ldots d^2k_{n-1}}{(\mathbf{k}_1^2 + \lambda^2)\ldots(\mathbf{k}_{n-1}^2 + \lambda^2)[(\mathbf{q} - \mathbf{k}_1 - \ldots - \mathbf{k}_{n-1})^2 + \lambda^2]} = \frac{n \ln^{n-1}(\mathbf{q}^2/\lambda^2)}{\mathbf{q}^2} \tag{38}$$

and taking the combinatorial factor $1/n!$ coming from the symmetric integration over α_i and β_i into account, we have to replace any single photon exchange with an infinite set of photons by multiplying the amplitude by eikonal factors of the type $\exp\{i\varphi_i(\mathbf{q}^2)\}$ with the phase $\varphi_i(\mathbf{q}^2) = \pm\alpha Z_i \ln(\mathbf{q}^2/\lambda^2)$. The scattering amplitudes of an electron and a positron differ only by the sign of the phase (which is positive for electrons) [9]. This replacement is shown in Fig. 6, where the double photon line corresponds to infinitely many photons.

Using the same technique as in [17], we can see that the amplitude relevant to Fig. 7a and Fig. 7b can be written as

$$\tilde{R}_{(1)}^{(1)} = B\exp\{-i[\varphi_1(\mathbf{q}_1^2) - \varphi_2(\mathbf{q}_2^2)]\} + \tilde{B}\exp\{i[\varphi_1(\mathbf{q}_1^2) - \varphi_2(\mathbf{q}_2^2)]\}. \tag{39}$$

The interactions of the electron and the positron with the Coulomb field differ only by signs. Although this expres-

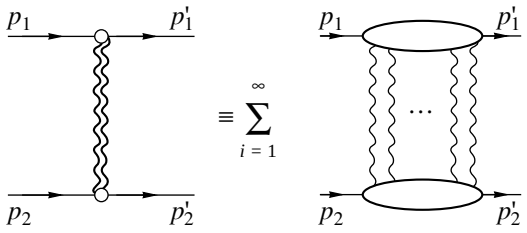


Fig. 6. Representation of all eikonal exchanges.

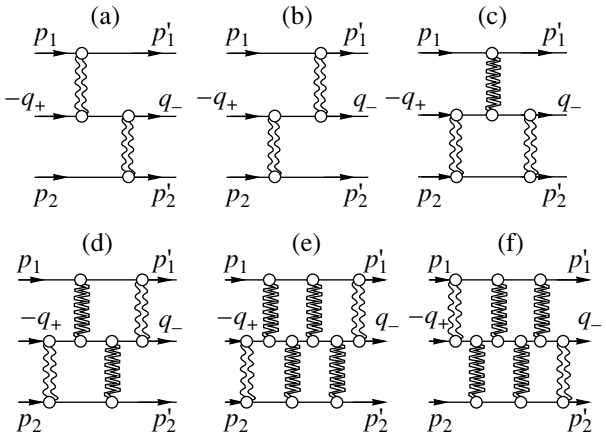


Fig. 7. Feynman diagrams for amplitudes with many photon exchanges. The double photon line represents any number of exchanged photons, the double zigzag line represents only an odd number of exchanged photons.

sion is infrared-unstable in the case where $Z_1 \neq Z_2$ the regularization parameter λ enters it in a standard way.

We now consider the class of diagrams shown in Fig. 7c. In Subsection 2.2, we obtained expressions (16) in the case where $m = n = 2$, with $\mathrm{Re}R_{1(34)2} = 0$. It can be shown that higher-order terms with any even number of photons from the same nucleus attached to the lepton world line between two photons from other nuclei do not contribute to the amplitude of the process under consideration. This follows from the relation $(\mathrm{sgn}\alpha)^{2k+1} = \mathrm{sgn}\alpha$.

The general structure of the amplitude corresponding to Fig. 7c can be constructed using the lowest-order truncated amplitude (without single-photon propagators) $R_{(2)}^{(1)}$,

$$\tilde{R}_{(2)}^{(1)} = \frac{\cos(\varphi_1(\mathbf{q}_1^2))}{q_1^2} R_{(2)}^{(1)} \times \exp\{i[\varphi_2(\mathbf{k}^2) - \varphi_2((\mathbf{q}_2 - \mathbf{k})^2)]\},$$
$$R_{(2)}^{(1)} = \frac{1}{i\pi} \frac{(\hat{q}_- - \hat{q}_2 + \hat{k})_\perp (-\hat{q}_+ + \hat{k})_\perp}{\beta_-(\mathbf{q}_+ - \mathbf{k})^2 + \beta_+(\mathbf{q}_- - \mathbf{q}_2 + \mathbf{k})^2} \times \ln \frac{\beta_+(\mathbf{q}_- - \mathbf{q}_2 + \mathbf{k})^2}{\beta_-(\mathbf{q}_+ - \mathbf{k})^2}. \quad (40)$$

The subsequent generalization is obvious. For instance, we give the expression corresponding to the diagram in Fig. 7d,

$$\tilde{R}_{(2)}^{(2)} = \cos(\varphi_1(\mathbf{k}_1^2))\exp[-i\varphi_1((\mathbf{q}_1 - \mathbf{k}_1)^2)] \times \cos(\varphi_2(\mathbf{k}_2^2))\exp[i\varphi_2((\mathbf{q}_2 - \mathbf{k}_2)^2)]R_{1324}. \quad (41)$$

From the above consideration, we conclude that it can therefore be written as the general structure of the matrix element $M_{(n)}^{(m)}$ corresponding to m photon exchanges from one ion (with 4-momenta k_i) and n exchanges from the other (with 4-momenta κ_i) can be schematically written as

$$M_{(n)}^{(m)} = isN_1N_2(Z_1\alpha)^m(Z_2\alpha)^n \frac{\pi^2}{16n!m!} \times \int \frac{d^2k_1}{\pi} \cdots \frac{d^2k_{m-1}}{\pi} \frac{d^2\kappa_1}{\pi} \cdots \frac{d^2\kappa_{n-1}}{\pi} \frac{1}{\mathbf{k}_1^2 \ldots \mathbf{k}_m^2} \times \frac{1}{\boldsymbol{\kappa}_1^2 \ldots \boldsymbol{\kappa}_n^2} \bar{u}(q_-)\bar{R}_{(n)}^{(m)} \frac{\hat{p}_2}{s} v(q_+), \quad (42)$$

where m and n satisfy the condition $|m - n| \leq 1$. At this stage, we omit phase factors in the structure $R_{(n)}^{(m)}$ (in order to understand the problem clearly), and it can therefore be written as

$$\bar{R}_{(n)}^{(m)} = \bar{R}_{(1)}^{(1)} + \bar{R}_{(2)}^{(1)} + \bar{R}_{(1)}^{(2)} + \bar{R}_{(2)}^{(2)} + \bar{R}_{(3)}^{(2)} + \bar{R}_{(2)}^{(3)} + \bar{R}_{(3)}^{(3)R} + \bar{R}_{(3)}^{(3)L} \ldots, \quad (43)$$

where

$$\bar{R}_{(1)}^{(2)} = \frac{1}{i\pi} \frac{(\hat{q}_- - \hat{k})_\perp (-\hat{q}_+ + \hat{q}_1 - \hat{k})_\perp}{\alpha_-(-\mathbf{q}_+ + \mathbf{q}_1 - \mathbf{k})^2 + \alpha_+(\mathbf{q}_- - \mathbf{k})^2} \times \ln \frac{\alpha_+(\mathbf{q}_- - \mathbf{k})^2}{\alpha_-(-\mathbf{q}_+ + \mathbf{q}_1 - \mathbf{k})^2},$$
$$\bar{R}_{(3)}^{(2)} = \bar{R}_{(2)}^{(3)} = 0,$$
$$\bar{R}_{(3)}^{(3)R} = \frac{1}{c_1 + c_2}\left[\frac{\pi^2}{2} + \frac{1}{2}\ln^2\frac{c_1}{c_2}\right],$$
$$c_1 = \beta_-(\mathbf{q}_- - \mathbf{k}_1)^2(\mathbf{q}_- - \mathbf{k}_1 - \mathbf{k}_2 - \boldsymbol{\kappa}_1)^2 \times (-\mathbf{q}_+ + \mathbf{q}_2 - \boldsymbol{\kappa}_1 - \boldsymbol{\kappa}_2)^2,$$
$$c_2 = \beta_+\mathbf{q}_-^2(\mathbf{q}_- - \mathbf{k}_1 - \boldsymbol{\kappa}_1)^2 \times (\mathbf{q}_- - \mathbf{k}_1 - \mathbf{k}_2 - \boldsymbol{\kappa}_1 - \boldsymbol{\kappa}_2)^2, \quad (44)$$
$$\bar{R}_{(4)}^{(3)} = \frac{1}{d_1 + d_2}\left[\frac{\pi^2}{2} + \frac{1}{2}\ln^2\frac{d_1}{d_2}\right],$$

$$d_1 = \beta_+(\mathbf{q}_- - \boldsymbol{\kappa}_1)^2(\mathbf{q}_- - \boldsymbol{\kappa}_1 - \boldsymbol{\kappa}_2 - \mathbf{k}_1)^2$$
$$\times(\mathbf{q}_- - \boldsymbol{\kappa}_1 - \boldsymbol{\kappa}_2 - \boldsymbol{\kappa}_3 - \mathbf{k}_1 - \mathbf{k}_2)^2,$$
$$d_2 = \beta_-(\mathbf{q}_- - \boldsymbol{\kappa}_1 - \mathbf{k}_1)^2$$
$$\times(\mathbf{q}_- - \boldsymbol{\kappa}_1 - \boldsymbol{\kappa}_2 - \mathbf{k}_1 - \mathbf{k}_2)^2$$
$$\times(-\mathbf{q}_+ + \mathbf{q}_2 - \boldsymbol{\kappa}_1 - \boldsymbol{\kappa}_2 - \boldsymbol{\kappa}_3)^2.$$

Here, $\bar{R}^{(2)}_{(2)}$ is only the second term in the right-hand side in Eq. (23) and the index $R(L)$ denotes two possible configurations of photons for $\bar{R}^{(3)R}_{(3)}$ (Fig. 7e) and $\bar{R}^{(3)L}_{(3)}$ (Fig. 7f).

Thus, the general algorithm for constructing an arbitrary term is transparent. Unfortunately, we cannot obtain a compact expression for the whole amplitude. The reason is the increasing nonlinearity of the propagators with the order of interaction. The behavior of the above denominators is very different from the Born-like case, where the simplicity of propagators allows one to obtain eikonal-like expressions.

The result of partial summation like (41) suffers from infrared divergences and cannot be considered final. On the other hand, the final result (44) implies the summation over the classes $R^{(m)}_{(n)}$ of FD and must contain all the dependence on the "photon mass" λ in the form of a general phase factor, proving the infrared stability of the cross section. We believe that this question will be the subject of a separate investigation.

5. CONCLUSIONS

The wide-angle lepton pair production in peripheral interactions of ultrarelativistic heavy ions is an archetype reaction for hard processes in central hadronic hard collisions of heavy nuclei. In the electromagnetic case, the expansion parameter $Z_{1,2}\alpha \sim 1$ makes the multiple photon collisions $m\gamma + n\gamma \longrightarrow l^+l^-$ potentially important, and similarly, the effect of multiple gluon collisions in central collisions is enhanced by a large number of nucleons at the same impact parameter. The crucial issue is whether such multiple photon collisions can be described by the Born cross section in terms of the collective photon fields of colliding nuclei. We have obtained the expression for the amplitude for the $2\gamma + 2\gamma \longrightarrow l^+l^-$ process and have shown that its contribution is dominant in the wide-angle limit. Our principal finding is that the amplitude is manifestly of a non-Born nature, which is suggestive of complete failure of the linear $k_\perp$ factorization even in the Abelian case.

The leading term of the multiphoton collision contribution to the amplitude of the production of high transverse momentum leptons, $2\gamma + 2\gamma \longrightarrow l^+l^-$, is found to have a logarithmic enhancement, while such an enhancement is absent in higher-order terms. We presented the algorithm that allows constructing the full amplitude in all orders. The results can be useful in application to the QCD process of producing high-$k_\perp$ jets as well as the bound state creation (positronium, charmonium), the issue which will be investigated elsewhere.

ACKNOWLEDGMENTS

We are grateful to the participants of the seminars at BLTP JINR, Dubna, and INP, Novosibirsk for critical comments and discussions.

Two of the authors (E. K. and E. B.) acknowledge support of INTAS (grant no. 00366), the Russian Foundation for Basic Research (project no. 03-02-17077), and the Grant Program of Plenipotentiary of Slovak Republic at JINR (grant no. 02-0-1025-98/2005).

REFERENCES

1. H. Bethe and L. Maximon, Phys. Rev. **93**, 768 (1954); H. Davies, H. Bethe, and L. Maximon, Phys. Rev. **93**, 788 (1954).
2. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory*, 4th ed. (Nauka, Moscow, 1989; Pergamon, New York, 1977).
3. D. Yu. Ivanov and K. Melnikov, Phys. Rev. D **57**, 4025 (1998).
4. S. R. Gevorkyan and S. S. Grigoryan, Phys. Rev. A **65**, 022505 (2002).
5. N. N. Nikolaev, W. Schafer, B. G. Zakharov, and V. R. Zoller, Zh. Éksp. Teor. Fiz. **124**, 491 (2003) [JETP **97**, 441 (2003)].
6. N. N. Nikolaev, W. Schafer, B. G. Zakharov, and V. R. Zoller, hep-ph/0408054; *Proc. of 12th Int. Workshop on Deep Inelastic Scattering (DIS2004), Štrbske Pleso, Slovakia.*
7. A. J. Baltz and L. McLerran, Phys. Rev. C **58**, 1679 (1998).
8. B. Segev and J. C. Wells, Phys. Rev. A **57**, 1849 (1998).
9. U. Eichmann, J. Reinhardt, and W. Greiner, Phys. Rev. A **61**, 062710 (2000).
10. D. Yu. Ivanov, A. Schiller, and V. G. Serbo, Phys. Lett. B **457**, 155 (1999).
11. R. N. Lee and A. I. Milstein, Phys. Rev. A **61**, 032103 (2000).
12. E. Bartoš, S. R. Gevorkyan, N. N. Nikolaev, and E. A. Kuraev, Phys. Rev. A **66**, 042720 (2002).
13. E. Bartoš, S. R. Gevorkyan, N. N. Nikolaev, and E. A. Kuraev, Phys. Lett. B **538**, 45 (2002).
14. S. R. Gevorkyan and E. A. Kuraev, J. Phys. G **29**, 1227 (2003).
15. V. N. Gribov, L. N. Lipatov, and G. V. Frolov, Yad. Fiz. **12**, 994 (1970) [Sov. J. Nucl. Phys. **12**, 543 (1971)].
16. I. F. Ginzburg, S. L. Panfil, and V. G. Serbo, Nucl. Phys. B **284**, 685 (1987).
17. E. Bartoš, S. R. Gevorkyan, and E. A. Kuraev, Yad. Fiz. **67**, 1945 (2004) [Phys. At. Nucl. **67**, 1923 (2004)].

Journal of Experimental and Theoretical Physics, Vol. 100, No. 4, 2005, pp. 656–662.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 4, 2005, pp. 744–751.

NUCLEI, PARTICLES, FIELDS, GRAVITATION, AND ASTROPHYSICS

TeV Signatures of Compact UHECR Accelerators¶

A. Neronov[a], P. Tinyakov[b, d], and I. Tkachev[c, d,*]

[a]*Institute of Theoretical Physics, EPFL BSP, CH-1015 Lausanne, Switzerland*
[b]*Service de Physique Théorique, CP 225 Université Libre de Bruxelles, B-1050, Brussels, Belgium*
[c]*Department of Physics, Theory Division, CERN 1211 Geneva 23, Switzerland*
[d]*Institute for Nuclear Research of the Russian Academy of Sciences, Moscow, 117312 Russia*
**e-mail: Igor.tkachev@cern.ch*

Received October 15, 2004

Abstract—We numerically study particle acceleration by the electric field induced near the horizon of a rotating supermassive ($M \sim 10^9$–$10^{10} M_\odot$) black hole embedded in the magnetic field B. We find that acceleration of protons to the energy $E \sim 10^{20}$ eV is possible only at extreme values of M and B. We also find that the acceleration is very inefficient and is accompanied by a broad-band MeV-TeV radiation whose total power exceeds the total power emitted in ultrahigh energy cosmic rays (UHECRs) at least by a factor of 1000. This implies that if $O(10)$ nearby quasar remnants were sources of proton events with an energy $E > 10^{20}$ eV, then each quasar remnant would, e.g., overshine the Crab Nebula by more than two orders of magnitude in the TeV energy band. Recent TeV observations exclude this possibility. A model in which $O(100)$ sources are situated at 100–1000 Mpc is not ruled out and can be experimentally tested by present TeV γ-ray telescopes. Such a model can explain the observed UHECR flux at moderate energies $E \approx (4\text{–}5) \times 10^{19}$ eV.

1. INTRODUCTION

The conventional hypothesis of ultrahigh-energy cosmic ray (UHECR) acceleration in extragalactic astrophysical objects has two important consequences. First, it predicts the Greisen–Zatsepin–Kuzmin (GZK) cutoff [1] in the spectrum of UHECRs at an energy on the order of 5×10^{19} eV. Whether such a cutoff indeed exists in nature is currently an open question [2, 3]. Second, it implies that the observed highest-energy cosmic rays with $E > 10^{20}$ eV should come from within a GZK distance of ~50 Mpc. Moreover, under plausible assumptions about extragalactic magnetic fields supported by recent simulations [4], the propagation of UHE protons over the GZK distance is rectilinear and the observed events should point back to their sources. While sub-GZK UHECRs were found to correlate with BL Lacertae objects [5, 6], no significant correlations of cosmic rays with energies of $E \gtrsim 10^{20}$ eV with nearby sources were found [7].

In view of the last problem, a question arises whether there exist UHECR accelerators that can produce super-GZK protons and are quiet in the electromagnetic (EM) channel. If such quiet accelerators existed, they could explain the apparent absence of sources within ~50 Mpc in the direction of the highest-energy events. This idea was advocated, e.g., in [8], where sources of UHE protons were associated with supermassive black holes in quiet galactic nuclei (so-called "dead quasars"). However, it was pointed out in [9] that most of the energy available for particle acceleration in such an environment is spent for EM radiation by the accelerated particles. As a consequence, the flux of TeV γ-rays produced by such an accelerator may be at a detectable level.

Recent observations by HEGRA/AIROBICC [10], MILAGRO [11] and TIBET [12] arrays substantially improved the upper limits on the flux of 7-rays above 10 TeV from point sources in the Northern Hemisphere. This may completely exclude the possibility of explaining observed super-GZK cosmic rays by the acceleration near supermassive black holes. The purpose of this paper is to analyze this question quantitatively. For this, we numerically study particle acceleration near the black hole horizon. Following [8, 9], we restrict ourselves to the case of protons. The case of heavy nuclei acceleration, propagation, and detection is phenomenologically very different and requires separate consideration. In particular, heavy nuclei can easily desintegrate as early as the acceleration stage.

We stress that our purpose is not to construct a realistic model of a compact UHE proton accelerator, but to find whether quiet compact accelerators can exist, even if the most favorable conditions for the acceleration are provided. For this, we minimize the energy losses of accelerated particles by considering acceleration in the ordered electromagnetic field and neglect all possible losses related to scattering of the accelerated particles on matter and radiation present in the acceleration site. However, we self-consistently take into account the synchrotron–curvature radiation losses, which are

¶ This article was submitted by the authors in English.

intrinsic to the acceleration process. Clearly, this approximation corresponds to the most favorable conditions for particle acceleration. In realistic models, the resulting particle energy must be smaller and the emitted EM power larger. Therefore, our results should be considered as a lower bound on the ratio of the electromagnetic to UHECR power of a cosmic ray accelerator based on a rotating supermassive black hole.

We find that the flux produced by a nearby UHE proton accelerator of super-GZK cosmic rays in the energy band $E_\gamma > 10$ TeV should be at least 100–1000 times larger than that of the Crab Nebula. The existence of such sources is indeed excluded by recent observations [10, 11]. At the same time, the constraints on the sources of sub-GZK cosmic rays are weaker or absent (see Section 7 for details).

This paper is organized as follows. In Section 2, we describe our minimum-loss model in more detail. In Section 3, we present an analytical estimate and the numerical calculation of the maximum particle energy. In Section 4, we consider the self-consistency constraints on the parameters of this model that arise from the requirement on the absence of on-site e^+e^- pair production caused by emitted radiation. In Section 5, the calculation of the EM luminosity of the accelerator is presented. In Section 6, observational constraints are derived. Section 7 contains a discussion of the results and concluding remarks.

2. THE MODEL

The model that we consider is based on a rotating supermassive black hole embedded in a uniform magnetic field. Because of the rotational drag of magnetic field lines, an electric field is generated, leading to acceleration of particles. In the absence of matter, the corresponding solution of the Einstein–Maxwell equations is known analytically at an arbitrary inclination angle of the black hole rotation axis with respect to the magnetic field [13, 14]. We assume low accretion rate and small matter and radiation density near the black hole, and neglect their back reaction on the EM and gravitational fields. We also neglect the effect of matter on propagation of the accelerated protons. This corresponds to the most favorable conditions for particle acceleration, and therefore leads to a maximum proton energy and minimum EM radiation.

The model has three parameters: the black hole mass M, the strength of the magnetic field B, and the inclination angle χ. We consider a maximally rotating black hole with a rotation moment per unit mass $a = M$. This maximizes the strength of the rotation-induced electric field. For a given injection rate and geometry, the above parameters completely determine the trajectories of accelerated particles and, therefore, their final energies and the emitted radiation. We reconstruct particle trajectories numerically, keeping track of the emitted radiation and taking its back reaction onto particle propagation into account.

We assume that protons flow into the acceleration volume from the accretion disk that is situated at larger radii. We model this accretion by injecting nonrelativistic particles uniformly over the sphere of the Schwarzschild radius $R_S = 2GM$, which is two times larger than the horizon of the maximally rotating black hole. We follow the trajectories of particles that propagate toward the horizon and that are then expelled from the vicinity of the black hole with high energies. It turns out that such trajectories exist only if the inclination angle of the magnetic field with respect to the rotation axis is sufficiently large, $\chi \gtrsim 10°$. For smaller inclination angles, all particles that propagate toward the horizon are finally absorbed by the black hole. This means that the stationary regime in which particles accreted onto the black hole are subsequently accelerated and ejected with high energies exists only at $\chi \gtrsim 10°$. In this regime, changes of the inclination angle and the injection radius do not strongly affect the maximum energies of particles.

3. MAXIMUM PARTICLE ENERGIES IN THE STATIONARY REGIME

In the absence of matter and radiation backgrounds, particle energies are limited by the radiation loss intrinsic to the acceleration process. For a general electric and magnetic field configuration, the energy loss in the ultrarelativistic limit is given by [15]

$$\frac{d\mathscr{E}}{dt} = -\frac{2e^4\mathscr{E}^2}{3m^4}[(\mathbf{E} + \mathbf{v}\times\mathbf{B})^2 - (\mathbf{E}\cdot\mathbf{v})^2], \qquad (1)$$

where m, e, and $\mathbf{v}$ are particle mass, charge, and velocity, respectively. We use this equation in our numerical modeling to calculate the electromagnetic radiation produced by the accelerated particles and to account for the back reaction of this radiation on particle trajectories.

Before presenting the numerical results, it is useful to summarize some simple qualitative estimates (see, e.g., [9, 16, 17]). We consider particle acceleration by a generic electromagnetic field obeying $|\mathbf{E}| \sim |\mathbf{B}|$. If the energy losses can be neglected, energies of accelerated particles are estimated as

$$\mathscr{E} = eBR \approx 10^{22}\frac{B}{10^4\ \mathrm{G}}\frac{M}{10^{10}M_\odot}\ \mathrm{eV}, \qquad (2)$$

where we assume that the size R of the acceleration region is of the order of the gravitational radius of the black hole, $R \approx 2GM$. But if the magnetic field strength is high, the synchrotron/curvature energy losses cannot be neglected.

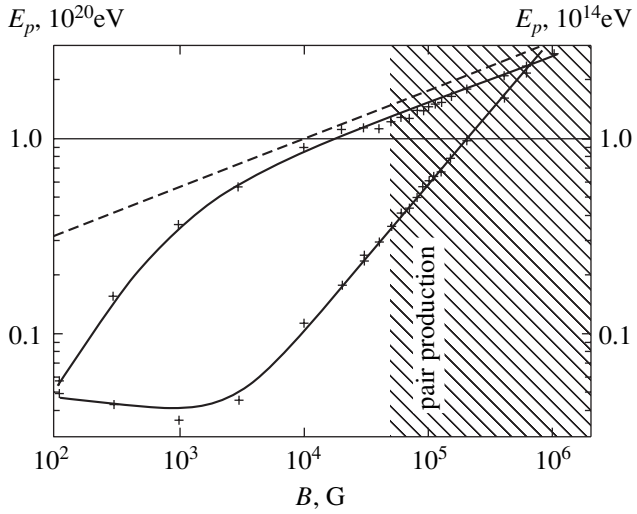


Fig. 1. The results of numerical calculation of maximum energies of accelerated protons and accompanying γ-rays are shown by crosses (solid lines are fits to numerical data). The dashed line is estimate (9) of the proton energies in the curvature-loss dominated regime. The shaded region corresponds to the magnetic field strength exceeding the pair production threshold by the curvature γ-rays. The black hole mass is $M = 10^{10}M_\odot$.

If no special relative orientation of the three vectors **E**, **B**, and **v** is assumed, Eq. (1) becomes

$$\frac{d\mathscr{E}}{dt} \approx -\frac{2e^4B^2\mathscr{E}^2}{3m^4}, \tag{3}$$

which is the standard formula for the synchrotron energy loss. Equating the rate of energy gain $d\mathscr{E}/dt = eE \sim eB$ to the rate of energy loss, we find that in the synchrotron-loss-saturated regime, the maximum energy is given by (see, e.g., [18])

$$\mathscr{E}_{\rm syn} = \left(\frac{3m^4}{2e^3B_0}\right)^{1/2} \approx 1.6\times 10^{18}\left(\frac{B}{10^4\ {\rm G}}\right)^{-1/2}\ {\rm eV}. \tag{4}$$

Here, we assume that accelerated particles are protons; for electrons, the maximum energy is much smaller.

The critical magnetic field strength at which the synchrotron energy loss becomes important can be found from the condition that estimates (2) and (4) give the same result,

$$B_{\rm crit} = \left(\frac{3m^4}{2e^5R^2}\right)^{1/3} \approx 30\left(\frac{M}{10^{10}M_\odot}\right)^{-2/3}\ {\rm G}. \tag{5}$$

Here, it is assumed that $R \sim R_S$. This critical field corresponds to the particle energy

$$\mathscr{E}_{\rm crit} \approx 3\times 10^{19}\left(\frac{M}{10^{10}M_\odot}\right)^{1/3}\ {\rm eV}, \tag{6}$$

which is the maximum energy attainable in the synchrotron-loss-saturated regime for a given black hole mass.

Acceleration is more efficient (loss (1) can be orders of magnitude smaller) in the special case where **E**, **B**, and **v** are nearly aligned. These conditions may be approximately satisfied in some regions around the black hole. In this case, particles closely follow the curved field lines and the curvature radiation loss,

$$\frac{d\mathscr{E}}{dt} = -\frac{2e^2\mathscr{E}^4}{3m^4R^2}, \tag{7}$$

becomes the main energy loss channel for high-energy particles. For an order-of-magnitude estimate, we can assume that the curvature scale of the magnetic field lines is on the order of the size of the acceleration region, $R \approx 2GM$. This translates into the maximum energy

$$\mathscr{E}_{\rm cur} = \left(\frac{3m^4R^2B}{2e}\right)^{1/4}, \tag{8}$$

which gives

$$\mathscr{E}_{\rm cur} = 1.1\times 10^{20}\left(\frac{M}{10^{10}M_\odot}\right)^{1/2}\left(\frac{B}{10^4\ {\rm G}}\right)^{1/4}\ {\rm eV} \tag{9}$$

for protons, where we again set $R = R_S$. The range of applicability of Eqs. (8) and (9) is given by the same condition $B > B_{\rm crit}$.

In the numerical simulations, we injected protons uniformly over the sphere surrounding the black hole. We disregarded trajectories that start at the injection sphere and move outward. Among protons that approach the horizon and are then expelled to infinity, we selected those which have the maximum final energy. For a black hole mass $M = 10^{10}\ M_\odot$, the dependence of this maximum energy on the magnetic field strength is shown in Fig. 1 (upper curve). For energies on the order 10^{20} eV and higher, the numerically calculated curve approaches limit (9), which corresponds to the curvature-loss-saturated regime. The acceleration to these energies requires magnetic fields in excess of 10^4 G. The necessary magnetic field is even stronger for smaller black hole masses, cf. Eq. (9). The maximum energies of protons do not depend strongly on the inclination angle in a wide range of χ.

4. CONSTRAINTS FROM PAIR PRODUCTION

There is an important self-consistency constraint that does not allow increasing *B* and *M* independently in order to reach higher energies. The reason is as follows. In our model, it was assumed that the acceleration proceeds in the vacuum. However, at a sufficiently strong magnetic field, photons of curvature radiation may produce e^+e^- pairs. Electrons and positrons are in turn accelerated and produce more photons, which

again produce e^+e^- pairs, etc. The plasma created by this cascade then neutralizes the electric field and prevents further acceleration of particles. For consistency of the model, we have to require that the cascade does not develop.

We consider this process in more detail. The energy ϵ_γ of the curvature photons in the regime when particle energies are limited by curvature losses is estimated as

$$\epsilon_\gamma = \frac{3\mathscr{E}_{\rm cur}^3}{2m^3R} \propto B^{3/4}M^{1/2}$$

(cf. Eq. (8)). Remarkably, the photon energy is independent of the particle mass. This means that proton-originated and electron-originated photons have the same energy. Numerically, we have

$$\epsilon_\gamma \approx 14\left(\frac{B}{10^4\ {\rm G}}\right)^{3/4}\left(\frac{M}{10^{10}M_\odot}\right)^{1/2}\ {\rm TeV}. \qquad (10)$$

If this energy is enough to produce more than one e^+e^- pair within the acceleration site, the instability may develop.

Therefore, for the stationary operating accelerator, the mean free path d of a γ-ray in the background of a strong magnetic field (see [19]) has to be larger than the size of the acceleration region

$$d \approx 100\frac{10^4\ {\rm G}}{B}\exp\left(\frac{8m_e^3}{3eB\epsilon_\gamma}\right)\ {\rm cm} > R_S. \qquad (11)$$

This requirement leads to the condition

$$B < 3.6\times 10^4\left(\frac{10^{10}M_\odot}{M}\right)^{2/7}\ {\rm G} \qquad (12)$$

on the magnetic field in the vicinity of the horizon.

In the numerical calculation of proton trajectories, we kept track of the emitted photons. For given parameters of the accelerator, we determined the maximum photon energy. The dependence of this energy on the magnetic field strength is shown in Fig. 1 (the lower curve). Substituting the calculated photon energy in Eq. (11), we can check whether the accelerator is in the stationary regime. The shaded region in Fig. 1 corresponds to nonnegligible pair production. The results of numerical calculation are in good agreement with Eq. (12).

From Fig. 1, we conclude that acceleration of protons to energies higher than 10^{20} eV is marginally possible in a small region of the parameter space (M, B). The magnetic field strength B must be close to the pair production threshold. The black hole mass M must be larger than $10^{10}M_\odot$. Such black holes are rare. For example, in [20], it is found that supermassive black holes in AGNs range within $10^{6.5}$–$10^{10.2}M_\odot$, with the mean mass being $10^{8.9}M_\odot$. The list of nearby (within 40 Mpc) candidates for quasar remnants [21] does not contain black holes with masses above $5\times 10^8M_\odot$.

Under reasonable assumptions about the black hole mass, the acceleration to energies above 10^{20} eV is impossible in the stationary regime discussed above (no particle production in the acceleration volume). However, although the accelerator cannot operate permanently, it is possible that UHECR are produced during "flares," or short episodes of activity of the accelerator, interrupted by discharges. The natural duration of one flare is about the time needed for the charge redistribution and neutralization of the electric field in the acceleration volume to establish. This can be roughly estimated as the light-crossing time $T_{\rm flare} \sim R_S/c \approx 10$ hours for the $3\times 10^9M_\odot$ black hole. During such flares, the electromagnetic luminosity of the accelerator must be much higher than the luminosity produced in the stationary regime, because the electromagnetic flux is dominated by the radiation produced by e^+e^- pairs whose number density is much higher than the density of the initial protons. Because we are interested in the possibility of having a "quiet" UHE proton accelerator, we concentrate in the next section on the case of the stationary regime, with the parameters of the model tuned to $B \approx 3\times 10^4$ G, $M \approx 10^{10}M_\odot$.

5. ELECTROMAGNETIC LUMINOSITY OF THE ACCELERATOR

It is clear from Fig. 1 that the acceleration of protons to energies above 10^{20} eV proceeds in the curvature-loss-saturated regime. In this regime, most of the work done by the accelerating electric field is spent on the emission of curvature radiation rather than on the increase in particle energy. The ratio of the dissipated energy to the final energy of a proton is

$$\mathscr{R} \approx \frac{eBR}{\mathscr{E}_{\rm cur}} \approx 2\times 10^2\left(\frac{M}{10^{10}M_\odot}\right)^{1/2}\left(\frac{B}{3\times 10^4\ {\rm G}}\right)^{3/4}. \qquad (13)$$

Thus, the energy carried away by photons is at least a hundred times higher than the energy carried by cosmic rays. Because only a small fraction of the accelerated protons reaches the UHECR energies $\mathscr{E} \geq 10^{20}$ eV, the ratio of the electromagnetic luminosity of the accelerator to its luminosity in UHECR with $E > 10^{20}$ eV is even higher.

Numerically, we calculated this ratio as follows. We summed energies of those protons which were accelerated above 10^{20} eV, and summed the energy emitted in synchrotron–curvature radiation (including the radiation emitted by protons that did not acquire sufficient energy while being expelled to infinity). We then took the ratio of the two sums.

The results of numerical calculation of the ratio of the electromagnetic and UHECR luminosities in the stationary regime are shown in Fig. 2 by crosses. Variations are due to fluctuations in the precise positions of

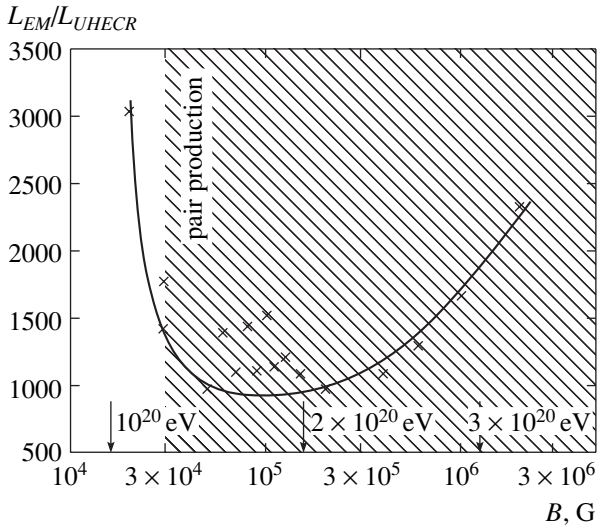


Fig. 2. Numerically calculated ratio of the electromagnetic luminosity of the accelerator to the luminosity emitted in particles with energies $\mathscr{E} \geq 10^{20}$ eV is shown by crosses, with the solid line being the fit to the numerical results. The shaded region corresponds to the magnetic field strength above the threshold of pair production by the curvature γ-rays. The black hole mass is $M = 10^{10} M_{\odot}$.

the injection points.[1] The qualitative behavior of the numerical results is easy to understand. Close to the "threshold" $\mathscr{E} = 10^{20}$ eV, the accelerator emits a finite power L_{EM} but does not produce UHECR with $\mathscr{E} > 10^{20}$ eV. Therefore, the ratio L_{EM}/L_{UHECR} diverges as $\mathscr{E} \longrightarrow 10^{20}$ eV. If the magnetic field is large, the maximum energies of particles increase as well. But the ratio $\mathscr{R}$ also increases for each particle according to Eq. (13), and so does L_{EM}/L_{UHECR}. The minimal value of L_{EM}/L_{UHECR} is reached at $\mathscr{E} \approx 1.5 \times 10^{20}$ eV. The numerically calculated minimum of L_{EM}/L_{UHECR} is larger than estimate (13) by a factor of 10.

In obtaining the results in Fig. 2, we have taken only the curvature radiation produced by protons into account. For the magnetic field strength above the pair production threshold $\sim 3 \times 10^4$ G (and, correspondingly, $\mathscr{E}_{\max} > 1.3 \times 10^{20}$ eV, see Fig. 1), our results give the lower bound on the electromagnetic luminosity of the compact accelerator.

We can see from Fig. 2 that the electromagnetic luminosity of the UHE proton accelerator based on a rotating supermassive black hole is

$$L_{EM} \gtrsim 10^3 L_{UHECR}. \tag{14}$$

Because the typical energy of photons of curvature radiation is about 10 TeV (see Eq. (10)), the above relation implies that such a source of UHE protons would be much more powerful in the 10 TeV band than in the UHECR channel.

6. OBSERVATIONAL CONSTRAINTS

The fact that production of UHE protons in a "quiet" accelerator is accompanied by the emission of the TeV γ-ray flux enables us to put strong constraints on the possibility of existence of such accelerators in the nearby Universe. Following [8], we assume that there are about $N \sim 10$ nearby UHECR accelerators not farther than $D_{GZK} \sim 50$ Mpc from the Earth. If these sources give the major contribution to the flux of cosmic rays above 10^{20} eV [22],

$$F^{\rm tot}_{UHECR} \sim (2\text{–}6) \times 10^{-11} \frac{\rm erg}{\rm cm^2\ s},$$

the mean energy flux produced by each source is

$$F_{UHECR} \sim \frac{F^{\rm tot}_{UHECR}}{N} \sim (2\text{–}6) \times 10^{-12} \frac{\rm erg}{\rm cm^2\ s}. \tag{15}$$

We have seen in the previous section that the lower bound on the ratio of the electromagnetic and UHECR luminosities of such a source is $\mathscr{R} > 10^3$. This means that the electromagnetic flux from the source must be

$$F_{EM} > (2\text{–}6) \times 10^{-9} \frac{10}{N} \frac{\rm erg}{\rm cm^2\ s}, \tag{16}$$

which implies the total luminosity larger than

$$L_{EM} \sim 10^{44}\ {\rm erg/s} \tag{17}$$

at a distance of D_{GZK}. This must be an extremely powerful source of TeV radiation. For comparison, the flux of the Crab Nebula at energies above 15 TeV is $F_{\rm Crab} \sim 10^{-11}$ erg/(cm^2 s) [10]. Thus, the hypothetical "quiet" cosmic ray sources, which would explain the observed UHECR flux, should be 100–1000 times brighter in the TeV band than the Crab Nebula.

The possibility of the existence of persistent point sources of this type in the Northern Hemisphere is excluded by the measurements of the HEGRA AIROBICC Array [10] and by the MILAGRO experiment [11, 23]. The upper limit on the energy flux from an undetected point source of ~15 TeV γ-rays provided by HEGRA/AIROBICC group [10] is at a level of $F_{HEGRA} \lesssim (2\text{–}3) F_{\rm Crab}$. A much tighter upper limit was published recently by the MILAGRO Collaboration, $F_{MILAGRO} \lesssim (0.3\text{–}0.6) F_{\rm Crab}$ [23].

7. DISCUSSION

The above model of particle acceleration near the horizon of a supermassive black hole is based on a number of assumptions: the maximum rotation moment of the black hole, a low matter and radiation density in the acceleration volume, the absence of back reaction

[1] For this calculation, we performed injection in 10^3 randomly chosen points uniformly distributed over the sphere.

of the accelerated particles and their radiation on the EM field, and the uniform magnetic field at large distance from the black hole. These assumptions have one common feature: they facilitate acceleration to higher energies and minimize losses (and, therefore, the radiation produced). We have found that even under these idealized conditions, the acceleration of protons to the energy $E = 10^{20}$ eV requires extreme values of parameters, $M \approx 10^{10} M_\odot$ and $B \approx 3 \times 10^4$ G. Moreover, the acceleration is very inefficient: the total power emitted in TeV gamma rays is 100–1000 times larger than in UHECR. In view of recent TeV observations, this rules out some UHECR models based on this acceleration mechanism, e.g., the model of several nearby dormant galactic nuclei (dead quasars), the goal of which was to explain the observed UHECR flux having an energy of $E > 10^{20}$ eV.

In a more realistic case, the above conditions may not be satisfied completely, and the acceleration of protons to an energy $E \sim 10^{20}$ eV in the continuous regime may not be possible. The synchrotron losses due to the presence of a random component $B_{\rm rand}$ of the magnetic field can be neglected if

$$B_{\rm rand} \ll \frac{B}{\mathcal{R}} \frac{m}{m_p} \approx 10^{-2} \frac{m}{m_p} B, \qquad (18)$$

where $\mathcal{R}$ is given by (13). This means that the presence of a tiny (1% level) random magnetic field leads to a decrease in the maximum energies of accelerated protons and an increase in the electromagnetic luminosity of the accelerator. We note that the synchrotron radiation is emitted in this case at the energies

$$\epsilon_{\rm synch} \le \frac{m}{e^2} \frac{B}{B_{\rm rand}} \approx 0.1 \frac{m}{m_p} \frac{B}{B_{\rm rand}} \text{ TeV}. \qquad (19)$$

The power is still given by Eq. (14).

Even if the strength of the random component of the magnetic field is as small as $10^{-5} B$, for electrons, which are inevitably present in the accelerator, the synchrotron losses dominate over the curvature losses. The electromagnetic power emitted by electrons is then in the 100 MeV–10 TeV energy band (see Eq. (19). Assuming that the density of electrons is on the same order as the density of protons, we obtain the same estimate (14) for the 100 MeV luminosity of the accelerator. This means that such an accelerator is not only a powerful TeV source, but also an extremely powerful EGRET source.

Even if the idealized conditions are realized in nature, the corresponding objects must be extremely rare. Thus, only a very small fraction of (active or quiet) galactic nuclei could be stationary sources of UHE protons with energies above 10^{20} eV.

If the parameters of the model are not precisely tuned to their optimal values, it is expected that the maximum energies of accelerated protons should be somewhat below 10^{20} eV. It is therefore interesting to note that most of the correlations of UHECRs with BL Lacertae objects come from the energy range $(4–5) \times 10^{19}$ eV. The central engine of BL Lacs is thought to consist of a supermassive black hole; it is possible that the acceleration mechanism considered above operates in these objects.[2] This mechanism may also operate in the centers of other galaxies that may have (super)massive black holes, including our own Galaxy, where it may be responsible for the production of cosmic rays of energies up to $\sim 10^{18}$ eV [13, 25].

The constraints from TeV observations are different in this case. First, cosmic rays of lower energies propagate over cosmological distances; hence the UHECR flux is collected from a much larger volume and the number of sources may be larger. Correspondingly, the TeV luminosity of each source is smaller. Second, the TeV radiation attenuates substantially over several hundred megaparsecs. Third, at $E < 10^{20}$ eV, the ratio L_{EM}/L_{UHECR} is smaller. For example, we consider the case of $O(100)$ sources located at $z \approx 0.1$ with a typical maximal energy at the accelerator $E \approx 5 \times 10^{19}$ eV. According to Fig. 1, the typical energy of produced γ-rays is then $\approx$4 TeV. The flux of γ-rays in this energy range is attenuated by a factor 10–100, while according to Eq. (13), $L_{EM}/L_{UHECR} \approx 50$. Therefore, we may expect $F_{EM} \approx (0.01–0.1) F_{\rm Crab}$ for the TeV flux from each of these sources. This is within the range of accessibility of modern telescopes. For example, the TeV flux from the nearby ($z = 0.047$) BL Lac 1ES 1959 + 650, which correlates with the arrival directions of UHECRs [6, 26], is at a level of $0.06 F_{\rm Crab}$ during the quiet phase and rises up to $2.9 F_{\rm Crab}$ during flares. Several other BL Lacs, which are confirmed TeV sources, have fluxes of $\approx 0.03 F_{\rm Crab}$ (see, e.g., [27]).

This paper mainly concerns the stationary regime of acceleration when the acceleration volume is not polluted by the creation of e^+e^- pairs. To ensure this condition, we required that the magnetic field not exceed the critical value (12). If the magnetic field is larger, the acceleration by the mechanism considered here can only occur during flares, which are interrupted by the creation of e^+e^- plasma and neutralization of the electric field as discussed at the end of Section 4. Although we do not have a quantitative model of a flare, some features of this regime and its consequences for the UHECR production can be understood qualitatively. Because there is no constraint on the magnetic field in this regime, the maximum energies of the accelerated protons may exceed 10^{20} eV. However, the efficiency of the acceleration during flares must be much lower than in the stationary case. First, as follows from Fig. 2, the L_{EM}/L_{UHECR} ratio is larger at large B. Second, the dominant part of the EM radiation is produced by the created electrons and positrons, whose number density by far

[2] We note that if the accelerated particles interact with the photon background outside the central engine, the same mechanism may be responsible for "photon jets" discussed in [24].

exceeds the number density of protons. Thus, we expect that the L_{EM}/L_{UHECR} ratio for this sources is much larger than in Eq. (14).

UHECR accelerators operating in the flaring regime would produce an approximately constant UHECR flux at the Earth. The reason is the time delay of protons due to random deflections in the extragalactic magnetic fields. This delay is on the order of $\sim 10^5[\alpha/1^\circ]^2$ yr for a source at 100 Mpc, where α is the deflection angle. Because the time scale of flares (light crossing time) is on the order of more than a day, the variations of UHECR flux would disappear upon averaging. On the contrary, the TeV radiation from such a source would be highly variable, with powerful TeV bursts and an average energy flux in the TeV band exceeding that in UHECRs by a factor of 10^4 or higher. We note that there exist tight constraints on transient TeV sources: the energy flux of a TeV burst having a duration of 10^5 s has to be less than 10^{-10} erg/(cm^2 s) $\sim 10F_{\rm Crab}$ [11, 12]. As in the case of a stationary accelerator, this constraint excludes the possibility of explaining the observed UHECR flux by a few nearby proton accelerators operating in the flaring regime. The hypothesis of several hundred remote sources is not constrained by TeV observations.

To summarize, the model of compact UHE proton accelerators that operate near the horizons of supermassive black holes in galactic nuclei can explain only the sub-GZK flux. A large number (several hundreds) of sources situated at cosmological distances. Production of UHECRs in such sources may be associated with blazar-type activity, TeV γ-radiation being an important signature of the model, testable by existing γ-ray telescopes.

ACKNOWLEDGMENTS

We are grateful to F. Aharonian, M. Kachelriess, D. Semikoz, M. Shaposhnikov, and M. Teshima for many useful discussions during various stages of this project. The work of P. T. is supported in part by the Swiss Science Foundation (grant no. 20-67958.02).

REFERENCES

1. K. Greisen, Phys. Rev. Lett. **16**, 748 (1966); G. T. Zatsepin and V. A. Kuz'min, Pis'ma Zh. Éksp. Teor. Fiz. **4**, 114 (1966) [JETP Lett. **4**, 78 (1966)].
2. M. Takeda *et al.* (AGASA Collab.), Astropart. Phys. **19**, 447 (2003); astro-ph/0209422.
3. T. Abu-Zayyad *et al.* (High Resolution Fly's Eye Collab.), astro-ph/0208301.
4. K. Dolag, D. Grasso, V. Springel, and I. Tkachev, astro-ph/0310902.
5. P. G. Tinyakov and I. I. Tkachev, Pis'ma Zh. Éksp. Teor. Fiz. **74**, 499 (2001) [JETP Lett. **74**, 445 (2001)]; astro-ph/0301336; Astropart. Phys. **18**, 165 (2002).
6. D. S. Gorbunov, P. G. Tinyakov, I. I. Tkachev, and S. V. Troitsky, Astrophys. J. **577**, L93 (2002).
7. M. Takeda *et al.* (AGASA Collab.), Astrophys. J. **522**, 225 (1999).
8. E. Boldt and P. Ghosh, Mon. Not. R. Astron. Soc. **307**, 491 (1999).
9. A. Levinson, Phys. Rev. Lett. **85**, 912 (2000).
10. F. A. Aharonian *et al.* (HEGRA Collab.), Astron. Astrophys. **390**, 39 (2002); astro-ph/0209402.
11. R. Atkins *et al.* (MILAGRO Collab.), astro-ph/0311389.
12. L. K. Ding *et al.*, High Energy Phys. Nucl. Phys. **21**, 9 (1997); M. Amenomiri *et al.*, in *Proceedings of ICRC 2001* (Hamburg, 2001), Vol. 6, p. 2544.
13. R. M. Wald, Phys. Rev. D **10**, 1680 (1974).
14. J. Bičak and L. Dvořak, Gen. Relativ. Gravit. **7**, 959 (1976); J. Bičak and V. Janis, Mon. Not. R. Astron. Soc. **212**, 899 (1985).
15. L. D. Landau and E. M. Lifschitz, *Course of Theoretical Physics*, Vol. 2: *The Classical Theory of Fields*, 7th ed. (Nauka, Moscow, 1988; Pergamon, Oxford, 1975).
16. A. M. Hillas, Annu. Rev. Astron. Astrophys. **22**, 425 (1984).
17. F. A. Aharonian, A. A. Belyanin, E. V. Derishev, *et al.*, Phys. Rev. D **66**, 023005 (2002); astro-ph/0202229.
18. M. V. Medvedev, Phys. Rev. E **67**, 045401(R) (2003); astro-ph/0303271.
19. T. Erber, Rev. Mod. Phys. **38**, 626 (1966).
20. E. W. Liang and H. T. Liu, Mon. Not. R. Astron. Soc. **340**, 632 (2003).
21. D. F. Torres, E. Boldt, T. Hamilton, and M. Loewenstein, Phys. Rev. D **66**, 023001 (2002); astro-ph/0204419.
22. M. Nagano and A. A. Watson, Rev. Mod. Phys. **72**, 689 (2000).
23. A. Atkins *et al.* (MILAGRO Collab.), astro-ph/0403097.
24. A. Neronov, D. Semikoz, F. Aharonian, and O. Kalashev, Phys. Rev. Lett. **89**, 051101 (2002); astro-ph/0201410; A. Neronov and D. V. Semikoz, Phys. Rev. D **66**, 123003 (2002); hep-ph/0208248; New Astron. Rev. **47**, 693 (2003).
25. A. Levinson and E. Boldt, astro-ph/0012314.
26. D. S. Gorbunov, P. G. Tinyakov, I. I. Tkachev, and S. V. Troitsky, Pis'ma Zh. Éksp. Teor. Fiz. **80**, 167 (2004).
27. F. Aharonian *et al.* (HEGRA Collab.), astro-ph/0401301.

Journal of Experimental and Theoretical Physics, Vol. 100, No. 4, 2005, pp. 663–687.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 4, 2005, pp. 752–779.

Experimental Investigation of Muon-catalyzed *dt* Fusion in Wide Ranges of D/T Mixture Conditions¶

V. R. Bom[a], A. M. Demin[b], D. L. Demin[c], C. W. E. van Eijk[a], M. P. Faifman[d], V. V. Filchenkov[c], A. N. Golubkov[b], N. N. Grafov[c], S. K. Grishechkin[b], K. I. Gritsaj[c], V. G. Klevtsov[b], A. D. Konin[c], A. V. Kuryakin[b], S. V. Medved'[c], R. K. Musyaev[b], V. V. Perevozchikov[b], A. I. Rudenko[c], S. M. Sadetsky[e], Yu. I. Vinogradov[b], A. A. Yukhimchuk[b], S. A. Yukhimchuk[c], V. G. Zinov[c], and S. V. Zlatoustovskii[b]

[a]*Delft University of Technology, 2629 JB Delft, the Netherlands*
[b]*Russian Federal Nuclear Center All-Russian Research Institute of Experimental Physics, Sarov, Nizhni Novgorod oblast, 607200 Russia*
[c]*Dzhelepov Laboratory of Nuclear Problems, Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia*
[d]*Russian Research Center Kurchatov Institute, Moscow, 123182 Russia*
[e]*St. Petersburg Nuclear Physics Institute, Gatchina, Leningrad oblast, 188350 Russia*
e-mail: grafov@nu.jinr.ru

Received September 6, 2004

Abstract—A vast program of the experimental investigation of muon-catalyzed *dt* fusion was performed on the Joint Institute for Nuclear Research phasotron. Parameters of the *dt* cycle were obtained in a wide range of the D/T mixture conditions: temperatures of 20–800 K, densities of 0.2–1.2 of the liquid hydrogen density (LHD), and tritium concentrations of 15–86%. In this paper, the results obtained are summarized. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Investigation of the muon-catalyzed fusion (MCF) process is a unique independent direction in modern physics relevant to molecular, atomic, and nuclear physics [1–6] and to astrophysics [7]. Study of nuclear fusion reactions from the bound states of a muonic molecule is of great importance for determining properties of the lightest nuclei, including various exotic nuclear systems. In addition, the high neutron yield of MCF can be effectively used for solving different practical problems such as the construction of an intense 14-MeV neutron source [8] and a nuclear fuel breeder [9].

That is why the process of MCF in hydrogen isotope mixtures has been under active study in many laboratories worldwide over the last several decades. Within this period, many experimental results were obtained by investigating muon-induced processes in different mixtures of hydrogen, deuterium, and tritium, as well as in pure isotopes; most of these experimental results are in good agreement with theory. The most impressive achievement is the precise agreement between experiment and theory in the temperature dependence of the *dd*μ-molecule formation rate in gaseous deuterium [4, 10, 11]. This allowed the binding energy of the loosely bound state of *dd*μ to be determined with a very high accuracy, $\epsilon_{11}^{\text{exp}} = -1962.56^{+32}_{-47}$ meV [10], which should be compared with the theoretical value $\epsilon_{11}^{\text{theor}} = -1964.83$ meV [4].

In comparison with pure deuterium, the MCF process in a D/T mixture manifests much richer physical phenomena (the muon transfer $d \longrightarrow t$ from the ground and excited states of the *d*μ-atom, epithermal, and many-body effects in the *dt*μ-molecule formation). Theory predicts a significant increase of the *dt*μ-molecule formation rate on the D_2 and DT molecule with the rise of temperature and density of the mixture [12, 13]. Therefore, complete theoretical analysis requires measuring temperature and density dependences of the $d + t$ cycle parameters in ranges that are as large as possible. Finally, the results of an experimental and theoretic study of MCF processes in a double D/T mixture will be rather helpful for investigation and explanation of the most difficult case of the triple H/D/T mixture.

Previously, a truly systematic experimental study of MCF process was performed at PSI only for a low-density ($\varphi \approx 0.1$ LHD, with a liquid hydrogen nuclei density of LHD = 4.25×10^{22} cm^{-3}), low-temperature ($T <$ 300 K) gaseous D/T mixture [14]. The same group, as

¶ This article was submitted by the authors in English.

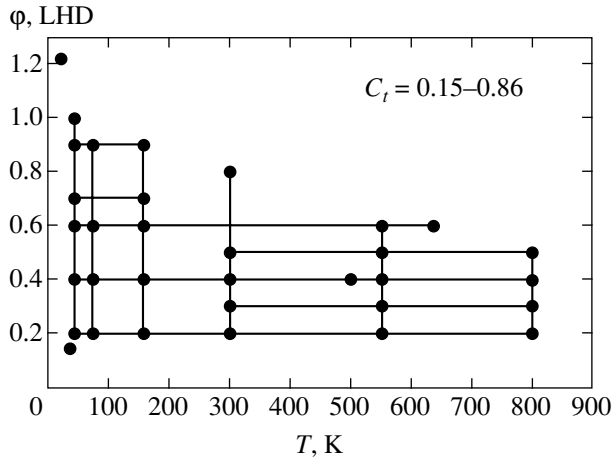


Fig. 1. Experimental conditions (density and temperature) for the MCF process study in the D/T mixtures.

well as the RIKEN-RAL team, performed measurements with liquid and solid D/T mixtures [15, 16]. The only group that investigated MCF in the high-density ($\varphi \approx 1$ LHD), high-temperature ($T \le 600$ K) mixtures was the LAMPF team [17, 18]. But its measurements had a hasty character and caused a lot of questions concerning the analysis. We therefore decided to conduct a full set of measurements in a wide region of the experimental conditions spanning a density range of $\varphi = 0.2$–1.2 LHD and temperature range of $T = 20$–800 K.

The Dzhelepov Laboratory of Nuclear Problems made a prominent contribution to the MCF experimental study. The Dubna group discovered the phenomenon of $dd\mu$-molecule resonance formation [19] and later directly confirmed its existence by measurements of the temperature dependence of the $dd\mu$-molecule formation rate [20]. This group was the first to experimentally investigate [21] muon-catalyzed fusion,

$$dt\mu \longrightarrow {}^4\mathrm{He}(3.5\ \mathrm{MeV}) + n(14.1\ \mathrm{MeV}) + \mu,$$

and to confirm the theoretical predictions [22] of the high intensity of this process, which induced the activity in the study of MCF worldwide.

Since 1997, our collaboration has been carrying out a large program in the investigation of MCF processes in D/T on the JINR phasotron. The distinctive characteristic of our study is the use of novel methods both in measurements and in experimental data analysis, which allows us to obtain accurate and reliable data no worse than those obtained at meson facilities. The experimental method that we used made it possible to measure the MCF cycle parameters in the D/T mixture under a wide variety of mixture conditions [23].

This paper is a report on the most comprehensive measurements of the MCF parameters in the D/T mixture. The preliminary data were published in [23–26]. Figure 1 shows the condition ranges of the experiments conducted up to now. The accumulated data and the MCF cycle parameters cover wide ranges of D/T mixture conditions:

(1) temperatures of 20–800 K;

(2) tritium concentrations of 15–86%;

(3) densities of 0.2–1.2 LHD.

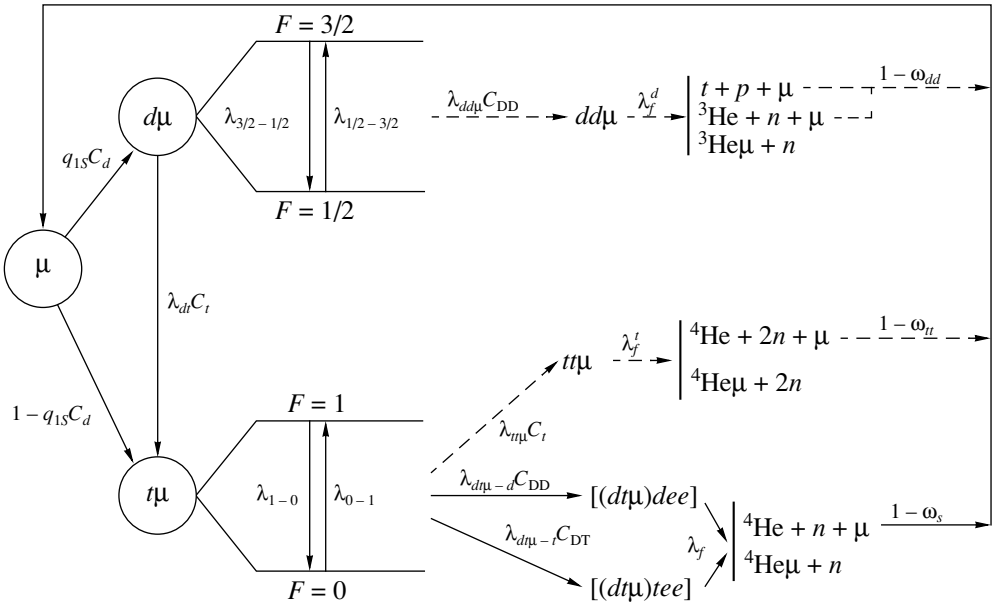


Fig. 2. Scheme of the MCF kinetics in the double D/T mixture.

2. GENERAL DESCRIPTION OF THE PROCESS

The simplified scheme of MCF kinetics in a double D/T mixture is shown in Fig. 2. Muons stopped in the mixture form *d*μ- and *t*μ-atoms in their ground states with the respective probabilities

$$w_{d\mu} = C_d q_{1S}$$

and

$$w_{t\mu} = C_t + C_d(1 - q_{1S}) = 1 - C_d q_{1S},$$

where C_d and C_t are relative hydrogen isotope concentrations and q_{1S} is the fraction of *d*μ-atoms in the ground state after muon cascade processes, with the muon transfer $(d\mu)_n \longrightarrow (t\mu)_n$ from *d*μ to *t*μ during the deexcitation cascade taken into account [27–29]:

$$q_{1S} = \frac{\lambda_{\text{dex}}}{\lambda_{\text{dex}} + \lambda_{\text{tr}}}. \quad (1)$$

Here, λ_{dex} and λ_{tr} are the rates of deexcitation and muon transfer averaged over the *d*μ-atom excited states. One should expect strong dependence of q_{1S} on C_t and φ [30, 31].

The "standard" cascade model, in which the initial μ-atom energies are distributed around E_0 = 1–2 eV, is apparently valid only at very low densities $\varphi \le 10^{-3}$ LHD. Now, it is known that, during the cascade, muonic atoms can be both thermalized and accelerated, obtaining an energy as high as tens of eV (see, e.g., [32–34]). But, until now, the problem of determining the initial energy distribution of muonic atoms after the cascade has not been solved definitely.

Being in the *d*μ-atom ground state, the muon can be transferred to tritium in the collisional process

$$d\mu + t \longrightarrow t\mu + d \quad (2)$$

with a rate of $\lambda_{dt} = 2.8 \times 10^8\ \text{s}^{-1} \cdot \varphi$ [18, 21, 35, 36]. In transfer process (2), the *t*μ-atom acquires an energy of 19 eV. The atoms *t*μ are formed in two hyperfine states with a total spin of $F = 1$ (weight 0.75) and $F = 0$ (weight 0.25) and can take part in the spin-flip processes

$$(t\mu)_{F=1} + t \longleftrightarrow (t\mu)_{F=0} + t. \quad (3)$$

Muonic atoms *t*μ can form *dt*μ- and *tt*μ-molecules, and *d*μ-atoms can form *dd*μ-molecules. In these μ-molecules, fusion reactions occur in which the muon can be either released and stimulate the next MCF cycle or stick to helium produced in the reactions. The notation for the rates of muonic formation and fusion reactions, as well as for the sticking probabilities, is introduced in Fig. 2. Being bound in a μ-atom or a μ-molecule or being free, the muon disappears at a rate of $\lambda_0 = 4.55 \times 10^5\ \text{s}^{-1}$.

The specific feature of the *dd*μ- and *dt*μ-molecule formation processes is their resonance character; that is, the muonic molecular formation rates $\lambda_{dd\mu}$ and $\lambda_{dt\mu}$ turn out to depend on the μ-atom kinetic energy [3]. The MCF $d + d$ cycle has been studied very well. The measured temperature dependence $\lambda_{dd\mu}(T)$ is in excellent agreement with theory [13].

Quite a different situation occurs for the MCF $d + t$ cycle. In fact, this process has been studied in detail in the parameter region (low temperatures), where the "standard" theory predicts its relatively low intensity. It follows from experiment that, just in this region, the MCF process is very effective. Modern theory explains this only qualitatively.

It follows from the original Vesman consideration [37] that the resonance *dt*μ-molecule formation occurs in the interaction of the *t*μ-atom with D_2, DT, or HD molecules according to the scheme [22]

$$t\mu + (\text{DX})_{K_i} \longrightarrow [(dt\mu), x, 2e]_{\nu, K_f}, \quad \text{X} = \text{H, D, T}, \quad x = p, d, t, \quad (4)$$

where the energy released under *dt*μ formation together with the *t*μ-atom kinetic energy $E_{t\mu}$ is transferred to excite the vibration–rotational state of the molecular complex $[(dt\mu), x, 2e]$. Here, K_i and K_f are the respective rotational quantum numbers of the "initial" molecule DX and the "final" complex. The set of the resonance *t*μ-atom energies

$$E^r_{t\mu} = \Delta E_{\nu, K}$$

corresponds to different transitions $\nu = 3, 4, 5$; $K_i \longrightarrow K_f$. Indeed, the spin states of the *t*μ-atom and the *dt*μ-molecule should be taken into account for determination of $E^r_{t\mu}$. In addition, the position and intensity of the resonances depend on the type of the molecule (D_2, DT, and HD) and the temperature of the mixture influencing the population of the rotational states of these molecules.

Once formed, the complex $[(dt\mu), x, 2e]$ either undergoes back decay

$$[(dt\mu), x, 2e] \longrightarrow t\mu + \text{DX},$$

or the fusion reactions

$$dt\mu \longrightarrow {}^4\text{He} + n + \mu, \quad (5)$$

$$dt\mu \longrightarrow {}^4\text{He}\mu + n \quad (6)$$

take place in it (with the high rate $\lambda_f \approx 10^{12}\ \text{s}^{-1}$ [38]). The muon-to-helium sticking probability is $\omega_s \approx 0.5\%$.

The resonance dependences $\lambda_{dt\mu-p,d,t}(E_{t\mu})$ for a *t*μ-atom of spin $F = 0$ are shown in Fig. 3; the calculations presented there are based on the evaluation scheme developed in [13]. The following remarkable features are evident from this figure.

(1) Resonance formation of the *dt*μ-molecule on HD molecules is the most intensive.

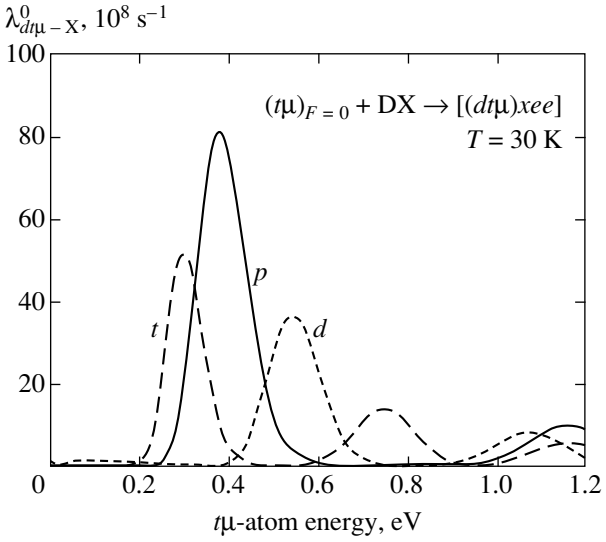


Fig. 3. The $dt\mu$-molecule formation rates on D_2, DT, and HD molecules for the $t\mu$-atom spin $F = 0$ as a function of $E_{t\mu}$ for T = 30 K (calculations based on [13]).

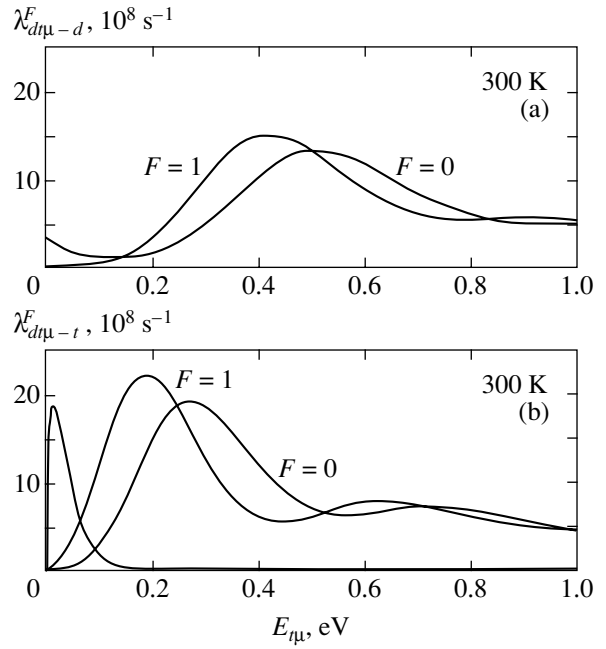


Fig. 4. The dependence of $\lambda_{dt\mu-d}$ (a) and $\lambda_{dt\mu-t}$ (b) on the $t\mu$-atom energy for T = 300 K. The Maxwell distribution is shown in the bottom picture.

(2) The resonance positions correspond to relatively high $t\mu$-atom energies, that is, to high temperatures ($T \sim 10^3$ K) for the thermalized muonic atoms.

(3) The positions of the resonances of each type correspond to various vibration levels of the complex ν = 3, 4, 5. The nearest resonance for $\lambda^0_{dt\mu-d}$ is placed at $E_{t\mu} \approx 0.5$ eV. This means that the nearest "subthreshold" resonance (corresponding to $\nu = 2$) is close to zero at a negative $t\mu$-atom energy of $E_{t\mu} \approx -(10–12)$ meV. Negative energy implies that, for the most intensive dipole transitions $|K_f - K_i| = 1$, an excess energy arises that cannot be transferred in two-particle reaction (4).

For the highest multiple transitions ($\nu = 2$; $K_i = 0$, $1 \longrightarrow K_f = 2, 3, 4$), process (4) becomes possible, but its intensity is two orders of magnitude lower than that of the main transitions. However, in contradiction to the standard theory, experiment manifests high MCF intensity in the low-temperature D/T mixture and reveals a nontrivial density dependence of its cycling rate. We note that the highest value of the cycling rate $\Lambda_c = 185 \pm 13\ \mu s^{-1}$ was measured at PSI [15] in a solid D/T mixture.

This can be qualitatively explained in the modern theory [39], according to which the influence of the subthreshold resonance turns out to be much stronger due to the mechanism of triple collisions. According to the theory, the resonance $dt\mu$ formation at low temperatures occurs at subthreshold resonance in the triple collision process

$$t\mu + D_2 + M \longrightarrow [(dt\mu), d, 2e] + M', \quad M = D_2, DT, T_2. \tag{7}$$

The "additional" second molecule M plays the role of a spectator that carries away the excess energy. Because (7) is a three-particle process, it must depend on the density of molecules M.

Qualitatively, the scheme in (7) explains both the high values of $\lambda_{dt\mu-d}$ and its density dependence observed in experiment. However, despite many attempts undertaken to calculate its intensity (see, e.g., [40, 41]), the quantitative explanation is not yet obtained.

With increasing temperature, the resonance pictures are modified due to the change in the population of the DX molecule rotational states and the thermal motion of the molecule. The calculated rates $\lambda_{dt\mu-d}$ and $\lambda_{dt\mu-t}$ as functions of $E_{t\mu}$ for T = 300 K are presented in Fig. 4. The Maxwell distribution for the thermalized $t\mu$-atoms is shown in one of them ($\lambda_{dt\mu-t}$). As is seen, this distribution only slightly overlaps the nearest resonance.

The resonances for $\lambda_{dt\mu}$ at T = 1000 K are presented in Fig. 5. In this case, the Maxwell distribution considerably overlaps the most intensive resonances for the $dt\mu$ formation on D_2, DT, and HD molecules. Unfortunately, this high temperature is not yet achieved in experiment. The temperature T = 800 K is the highest at which the measurements were made (in Dubna).

As we have mentioned, a substantial part of $t\mu$ has an initial (after cascade) energy of $E_{t\mu} > 1$ eV. In elastic collisions $t\mu + t$, $t\mu + d$, these atoms are quickly thermalized. The thermalization time is approximately equal to *ns* for the 1 LHD of a mixture. Accordingly, the time distribution of the fusion reaction products (neutrons) should have two components: a quick "spike"

corresponding to the first pass through the resonances and a much slower "steady-state" component.

Due to shortness of the epithermal spike and ambiguity in the initial energy of the *t*μ-atom, it is difficult to interpret this effect. That is why the main efforts of different experimental groups were concentrated on the steady-state study, for which the *t*μ-atom energy spectrum is a Maxwell distribution.

For convenience, the comparison of the measurements with the theoretical calculations is performed for the so-called "effective" *dt*μ-molecule formation rate as a function of temperature. It is obtained by integrating over all possible initial states, averaging over all final states, and convolving with the Maxwell spectrum $W(E_{t\mu}, T)$. Such calculations for the D/T mixture were made in [13].

For the steady state, the time distribution of fusion neutrons has the form

$$\frac{dN_n}{dt} = N_\mu \epsilon_n \Lambda_c \exp(-\lambda_n t), \quad \lambda_n = \lambda_0 + \omega\Lambda_c, \qquad (8)$$

where $\Lambda_c = \lambda_c \varphi$, ϵ_n is the neutron detection efficiency, N_μ is the number of muons stopped in the D/T mixture, and ω is the muon loss in the cycle, which is the probability of muon sticking to helium in fusion reactions, mainly in $d + t$ (ω_s), and also, with lower weight, in the accompanying reactions $d + d$ and $t + t$. The cycling rate Λ_c means the inverse of the average time between the closest cycles. It involves mainly the time of $d\mu \longrightarrow t\mu$ transfer (2), *t*μ-atom spin-flip process (3), and *dt*μ-molecule formation (4). The neutron yield Y_n is limited by ω and λ_0:

$$Y_n^{-1} = \omega + \frac{\lambda_0}{\Lambda_c}. \qquad (9)$$

The expression for λ_c, corresponding to the kinetic scheme of Fig. 2, is

$$\frac{1}{\lambda_c} \approx \frac{q_{1S}C_d}{\lambda_{dt}C_t} + \frac{0.75}{\lambda_{1-0}C_t} + \frac{1}{\lambda_{dt\mu-d}C_{\mathrm{DD}} + \lambda_{dt\mu-t}C_{\mathrm{DT}}}. \qquad (10)$$

To extract the values $\lambda_{dt\mu-d}$ and $\lambda_{dt\mu-t}$, one should use formula (10) to analyze the experimental values of λ_c measured at different tritium concentrations changing the relative population of D_2 and DT molecules.

The expression for ω is

$$\omega \approx \omega_s + \frac{\lambda_{tt\mu}C_t\omega_{tt}}{\lambda_{dt\mu-d}C_{\mathrm{DD}} + \lambda_{dt\mu-t}C_{\mathrm{DT}} + \lambda_{tt\mu}C_t} + \frac{q_{1S}C_d\frac{2}{3}\lambda_{dd\mu}^{3/2}C_{\mathrm{DD}}\omega_{dd}r}{\lambda_{dt}C_t + \lambda_{dd\mu}^{3/2}C_{\mathrm{DD}} + \lambda_{3/2-1/2}C_d} + \frac{\lambda_Z C_Z}{\lambda_c}, \qquad (11)$$

where λ_Z is the rate of muon transfer to possible admixtures with $Z > 1$, having concentration C_Z, $\lambda_{dd\mu}^{3/2}$ is the rate of $dd\mu$-molecule formation from the $d\mu$-atom state with spin $F = 3/2$, and r is the branching ratio of the dd fusion channels (^{3}He + n) and ($t + p$); the other variables are defined in Fig. 2. It follows from Eq. (11) that the minimum value of ω is achieved at highest λ_c (large λ_{dt} and $\lambda_{dt\mu-d,t}$), where ω is close to its natural limit $\omega_s \approx 0.5\%$.

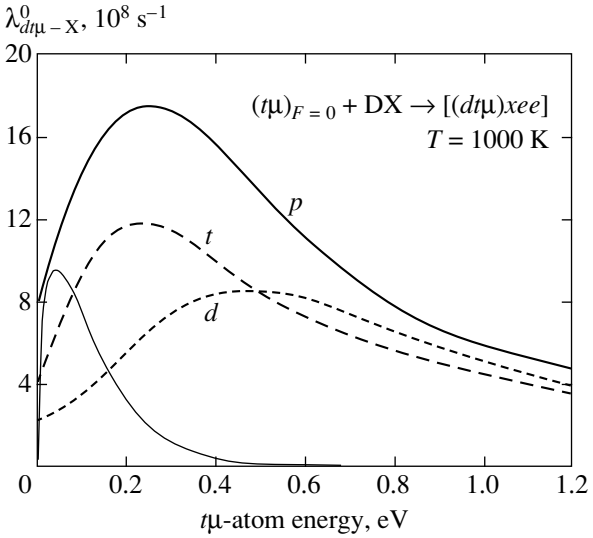


Fig. 5. The *dt*μ-molecule formation rates on D_2, DT, and HD molecules for the *t*μ-atom spin $F = 0$ as a function of $E_{t\mu}$ for $T = 1000$ K (calculations based on [13]) and the Maxwell distribution.

We note that, in expressions (10) and (11) and in what follows, the cycling rate and all collisional rates are normalized to the nuclear density φ of the D/T mixture.

3. EXPERIMENTAL METHOD

All experimental runs were made at the installation "Triton" mounted on the muon channel [42] of the JINR phasotron. The experimental setup is schematically shown in Fig. 6. The novel experimental method in [23] was used. Based on measurements of the total charge produced by the fusion neutrons in a detector, it allowed us to avoid the distortions in the neutron time spectra caused by the pileup and thus to use a high-efficiency detection system.

Incoming muons are detected by scintillation counters *1*, *2*, *3*, a proportional wire counter *4* and stopped in the target. Neutrons from the *d–t* reaction are detected by two full-absorption neutron detectors *ND1* and *ND*2. Electrons from the decay of muons stopped in the target are registered by the proportional wire counter *5* and scintillation detectors *1*-e and *2*-e.

3.1. The Specific Features of the Method

The following important features characterize the method used.

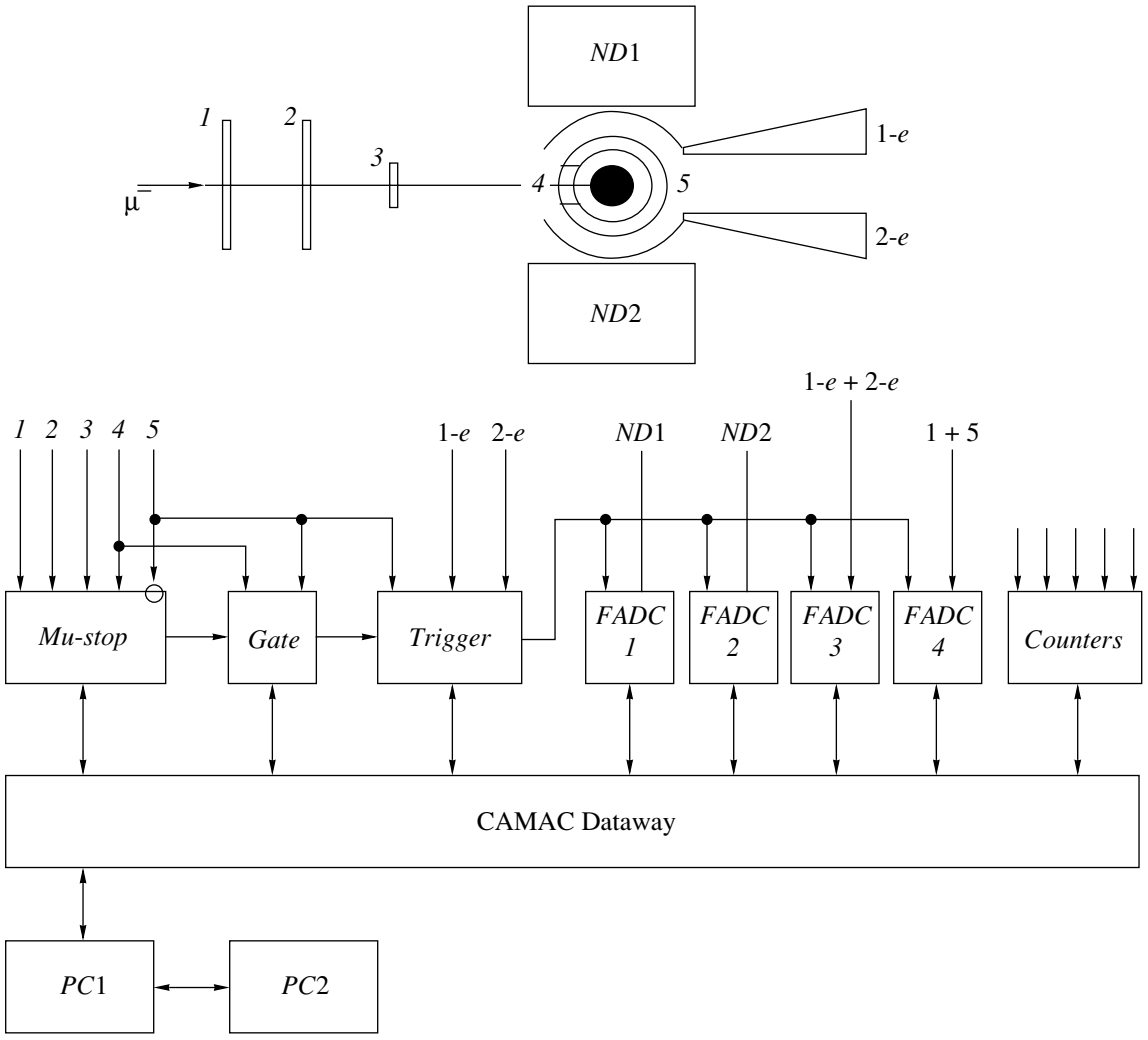


Fig. 6. Experimental layout.

(1) Unique targets and tritium handling system were used, which allowed measurements in a wide range of the D/T mixture densities and temperatures.

(2) A high-efficiency neutron detection system was used in the geometry close to 4π. It provided a high counting rate and low accidental background.

(3) A specially designed proportional counter was used for muon and electron detection; having a low sensitivity to neutrons, it allowed reliable electron identification.

(4) Time distributions of charge were measured instead of the usually registered time spectra of the number of events. Flash ADC were used for this aim. This allowed us to avoid distortions in the neutron time spectra and thus to use a high-efficiency detection system.

(5) The novel analysis methods were used, which turned out to be most effective for the high neutron multiplicity realized in the experiment. In addition to the usually measured neutron time distribution, we measured and analyzed the neutron multiplicity distribution and the spectra of the time between the μ-decay electron and the last neutron in the series. This allowed us to decrease systematic errors and to obtain reliable data.

3.2. Targets and Gas Handling System

A set of targets [43–45] with the working volume 10–18 cm^3 depending on the tritium content was used in the experiments. The targets allowed the following measurements:

(1) with liquid D/T (liquid tritium target (LTT) [43] of 18 cm^3, working temperature 20–40 K, pressure up to 20 bar);

(2) with hot gaseous D/T (high-pressure tritium target (HPTT) [44] of 16 cm^3, working temperature 300–800 K, pressure up to 1600 bar);

(3) with cold gaseous D/T (two high-pressure tritium targets (HPTT) [45] of 8 and 16 cm^3, working temperature 40–200 K, pressure up to 2500 bar).

The special cryogenic system [43] (for the LTT and HPTT) and the cryorefrigerator (for the HPTT) were used at low temperatures ($T < 300$ K), and the system of special heaters was used at high temperatures to maintain the needed temperature regime. Cryogenic filling was used for all targets.

A special preparation system based on palladium filters [46] provided the gas of the required composition and purified of impurities at the level less than 10^{-7} of volume parts. The molecular composition of the mixtures was monitored with the aid of chromatography.

3.3. Detectors and Electronics

The target was surrounded by a set of detectors. Scintillation counters *1–3* detected incoming muons. A cylinder-shaped proportional counter (*PC*, analogous to [47]) with wires grouped in two parts (*4*, *5*) served to select muon stops in the target (signal $1 \cdot 2 \cdot 3 \cdot 4 \cdot \bar{5}$) and to detect electrons from the muon decay. Specially designed cylinder-shaped scintillation counters (*SC*) *1*-e and *2*-e were used to detect μ-decay electrons in coincidence with counter *5* (signals $5 \cdot 1$-e and $5 \cdot 2$-e were considered as a μ-decay electron). The full-absorption neutron spectrometer [48, 49] consisting of two large detectors (*ND*1 and *ND*2) with the volume 12.5 l each was the basis of the detection system. It was aimed at detecting neutrons from reactions (5), (6). A plastic scintillator with dimensions ($\varnothing 31 \times 17$) cm was used in each detector. It was viewed by four PMs XP 2040. The direct contact of the PMs with the scintillator and Teflon used as an optical reflector provided excellent spectrometric properties of the detector. Its energy resolution was

$$\sigma_{FWHM} = 0.09(1 + 1/\sqrt{E_e\ [\mathrm{MeV}]}).$$

The total solid angle covered by two detectors was $\Omega \approx 70\%$, which corresponded to the total neutron detection efficiency $\epsilon_n \approx 2 \times 15\%$. The time resolution of *ND* was dictated by the light collection process and electronics and was $\Delta t = 6$–7 ns.

The trigger [50] allows recording only those events for detection that are connected to electron detection. Because the intensity of the process under study was high, these events were accompanied by neutron detection in nearly every case.

The trigger requirements included the presence of muon stop signals ($1, 2, 3, 4, \bar{5}$) and electron signals (*5*, *1*-e or *5*, *2*-e) during a time window of 20 μs, set off by the incoming muon signal (*1*, *2*). Insertion of the electron signal in the trigger makes it possible to radically suppress the background connected with the muon stops in the target walls, where a muon undergoes predominately (90%) nuclear capture without electron escape. Additional suppression of this background is achieved under the condition that only delayed electrons (later than 0.2 μs after the gate start) are permitted.

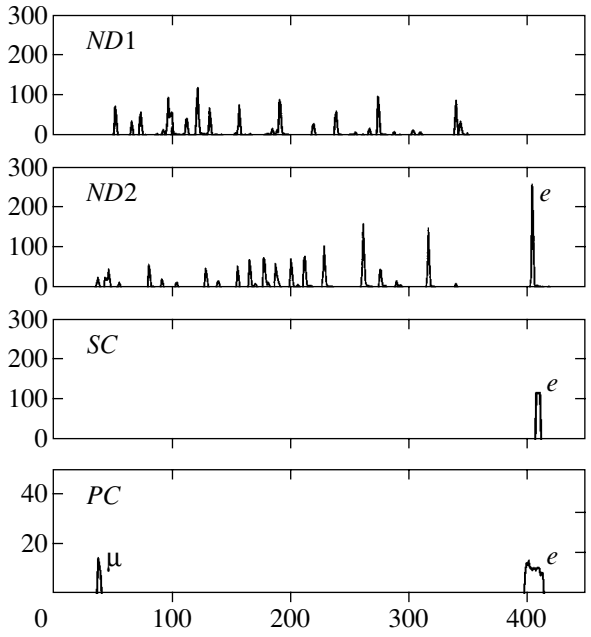


Fig. 7. Flash ADC signals for a single muon.

Another important advantage of this is that direct normalization to the electron number becomes possible without the necessity to determine the number of muon stops in hydrogen. This method was first employed by us in the first experiment on the MCF $d + t$ reaction [21] and allows successful accomplishment of this fundamental work.

Pulses from the neutron spectrometer are registered by the flash ADC (8 bits × 2048 samples, 100 Mc/s) producing a time distribution of the *ND*1, *ND*2 signal amplitude for each single muon. To provide correct time measurements, the signals of the detector for incoming muons and the electron counter are also analyzed by the flash ADC. An example of "oscillograms" observed at the flash ADC is shown in Fig. 7. During each run, on-line monitoring of data accumulation was conducted.

4. EXPERIMENTAL CONDITIONS

A total of 81 exposures with D/T mixtures were carried out. The conditions (density, temperature, and tritium concentration) of each run are presented in Table 1. In each exposure (duration of 6–10 h), at least 20000 electrons from the decay of muons stopped in the target were accumulated. In practically all cases, the neutron statistics was sufficiently large. The special exposures with empty targets were carried out to measure the background of electrons from muons stopped in the target walls.

Table 1. Normalized cycling rates λ_c, muon loss probabilities ω, and neutron yields per muon Y_n. For λ_c, the uncertainties due to statistics, density ($\Delta\varphi$), charge calibration procedure (Δq), and the total uncertainty are indicated. For all parameters, the total errors include the systematic uncertainty in the determination of neutron detection efficiency ($\Delta\epsilon_n = 6\%$)

No.	Conditions		ω, %	Y_n	λ_c, μs^{-1}	Error contributions, %			
	φ, LHD	C_t, %				stat.	$\Delta\varphi$	Δq	total
$T = 22.0 \pm 0.7$ K (liquid)									
1	1.19	18.1(1.5)	0.77(0.07)	68.9(5.9)	61.5	0.4	2.1	3.1	7.7
2	1.20	33.4(1.0)	0.72(0.06)	96.6(7.9)	117.7	0.2	2.1	3.1	7.7
3	1.19	35.2(1.0)	0.63(0.05)	102.9(8.6)	117.9	0.3	2.1	3.1	7.7
4	1.23	63.5(2.0)	0.76(0.07)	82.5(6.9)	84.6	0.4	2.4	3.1	7.8
5	1.24	85.5(2.5)	1.40(0.11)	34.2(2.8)	20.8	0.6	2.4	3.1	7.8
$T = 38.5 \pm 2.0$ K									
6	0.143	57.0(2)	1.03(0.08)	10.1(0.8)	31.1	1.5	3.3	3.1	8.2
$T = 45 \pm 2$ K									
7	0.237	31.4(0.5)	0.77(0.06)	32.0(2.6)	80.5	1.0	3.4	3.4	8.3
8	0.449	16.7(0.5)	0.81(0.06)	39.4(3.2)	59.9	1.0	3.0	3.2	8.1
9	0.450	31.4(0.5)	0.68(0.05)	54.2(4.5)	88.7	0.9	3.3	3.4	8.3
10	0.448	50.6(0.5)	0.81(0.06)	43.8(3.6)	73.1	1.3	3.0	3.2	8.1
11	0.445	71.1(0.5)	1.46(0.11)	19.0(1.6)	30.6	1.3	2.9	3.2	8.1
12	0.689	16.3(0.5)	1.12(0.09)	46.1(3.7)	64.2	0.9	3.0	3.1	8.0
13	0.643	31.1(0.5)	0.74(0.06)	69.9(5.8)	101.2	0.7	3.3	3.4	8.2
14	0.704	52.7(0.5)	1.11(0.09)	53.4(4.2)	84.0	0.7	3.0	3.1	8.0
15	0.766	71.2(0.5)	1.89(0.15)	27.5(2.2)	34.7	1.3	3.0	3.1	8.1
16	1.022	16.3(0.5)	1.55(0.12)	45.0(3.5)	65.2	1.0	3.0	3.1	8.0
17	0.912	31.1(0.5)	0.89(0.07)	76.0(6.3)	118.5	0.8	3.2	3.4	8.2
18	1.024	52.7(0.5)	1.12(0.09)	64.2(5.0)	97.7	0.8	3.0	3.1	8.0
19	1.018	71.2(0.5)	1.83(0.14)	34.2(2.7)	40.6	1.1	3.0	3.1	8.1
$T = 75 \pm 2$ K									
20	0.234	31.4(0.5)	0.85(0.07)	29.8(2.5)	81.2	1.0	3.0	3.4	8.2
21	0.445	31.4(0.5)	0.87(0.07)	50.4(4.2)	92.1	0.9	2.9	3.4	8.1
22	0.635	31.4(0.5)	0.94(0.07)	69.9(5.8)	101.6	0.7	3.0	3.4	8.1
23	0.897	31.1(0.5)	0.91(0.07)	75.5(6.2)	119.5	0.8	3.0	3.4	8.1
$T = 158 \pm 2$ K									
24	0.230	31.4(0.5)	0.94(0.07)	28.6(2.4)	79.3	1.1	3.0	3.4	8.2
25	0.438	16.7(0.5)	1.41(0.11)	31.0(2.5)	58.7	1.1	3.0	3.2	8.1
26	0.424	31.0(0.5)	0.99(0.08)	45.4(3.7)	88.8	1.0	3.1	3.2	8.1
27	0.436	31.4(0.5)	0.88(0.07)	48.1(4.0)	90.9	0.9	3.0	3.4	8.1
28	0.433	50.6(0.5)	1.00(0.08)	39.2(3.2)	74.5	1.3	3.0	3.2	8.1
29	0.430	71.1(0.5)	2.01(0.15)	17.3(1.4)	29.5	1.3	3.0	3.2	8.1
30	0.607	16.3(0.5)	1.94(0.15)	31.7(2.5)	63.2	1.4	3.0	3.1	8.1
31	0.620	31.1(0.5)	0.98(0.08)	57.9(4.8)	100.1	0.9	3.1	3.4	8.2
32	0.621	52.7(0.5)	1.14(0.09)	48.9(3.9)	82.9	0.8	3.1	3.1	8.1
33	0.688	71.2(0.5)	1.64(0.13)	27.1(2.1)	35.4	1.3	3.1	3.1	8.1
34	0.905	16.3(0.5)	1.89(0.15)	36.9(2.9)	64.7	1.3	3.0	3.1	8.1
35	0.876	31.1(0.5)	0.90(0.07)	72.4(6.0)	119.6	0.8	3.0	3.4	8.1
36	0.907	52.7(0.5)	1.09(0.08)	66.3(5.2)	101.7	0.8	3.0	3.1	8.0
37	0.902	71.2(0.5)	1.62(0.12)	34.9(2.8)	40.6	1.1	3.0	3.1	8.1

Table 1. (Contd.)

No.	Conditions		ω, %	Y_n	λ_c, μs^{-1}	Error contributions, %			
	φ, LHD	C_t, %				stat.	Δφ	Δ*q*	total
				$T = 300 \pm 3$ K					
38	0.204	31.4(0.5)	1.23(0.10)	28.2(2.3)	91.4	1.7	3.4	3.5	8.5
39	0.303	17.9(0.5)	2.13(0.16)	21.6(1.7)	67.2	1.1	3.3	3.1	8.1
40	0.302	36.1(0.5)	1.14(0.09)	36.7(3.0)	101.1	0.9	3.4	3.1	8.2
41	0.312	52.0(0.5)	1.27(0.10)	30.3(2.5)	78.7	1.0	3.2	3.1	8.1
42	0.312	68.8(0.5)	1.25(0.10)	21.4(1.7)	47.6	1.0	3.2	3.1	8.1
43	0.434	15.4(0.5)	0.97(0.07)	35.4(2.9)	59.3	0.9	3.2	3.1	8.1
44	0.411	31.0(0.5)	1.08(0.08)	43.7(3.6)	96.0	1.0	3.2	3.2	8.1
45	0.425	32.7(0.5)	0.95(0.07)	49.4(4.0)	99.9	0.7	3.3	3.1	8.1
46	0.443	35.0(1.0)	0.89(0.07)	53.6(4.4)	104.4	0.5	3.4	3.2	8.2
47	0.409	47.7(0.8)	0.97(0.07)	44.0(3.6)	89.3	0.7	3.4	3.1	8.1
48	0.411	68.5(0.5)	1.21(0.09)	27.7(2.2)	50.3	0.8	3.2	3.1	8.1
49	0.515	18.2(0.5)	1.95(0.15)	30.2(2.4)	74.5	1.1	3.7	3.1	8.3
50	0.518	35.2(0.5)	1.38(0.10)	46.8(3.8)	109.2	0.8	3.7	3.1	8.3
51	0.532	52.8(0.5)	1.01(0.08)	50.0(4.0)	92.8	0.7	3.6	3.1	8.2
52	0.787	33.0(1.0)	0.80(0.06)	76.8(6.3)	123.2	0.5	3.0	3.2	8.1
53	0.781	33.7(0.5)	1.19(0.09)	57.1(4.7)	118.4	1.1	3.1	3.5	8.3
				$T = 500 \pm 6$ K					
54	0.425	35.0(1.0)	0.88(0.07)	58.9(4.8)	130.0	0.6	3.1	3.2	8.2
				$T = 550 \pm 6$ K					
55	0.201	33.7(0.5)	1.25(0.10)	30.2(2.5)	113.3	1.6	3.5	3.5	8.5
56	0.293	17.9(0.5)	1.92(0.15)	23.1(1.9)	73.5	1.1	3.5	3.1	8.2
57	0.285	36.1(0.5)	1.14(0.09)	42.8(3.5)	130.1	0.9	3.6	3.1	8.2
58	0.287	52.0(0.5)	1.07(0.08)	43.3(3.5)	135.6	0.9	3.5	3.1	8.2
59	0.292	68.8(0.5)	1.06(0.08)	38.7(3.1)	104.3	0.8	3.5	3.1	8.2
60	0.407	15.4(0.5)	0.93(0.07)	37.4(3.0)	66.1	0.9	3.4	3.2	8.2
61	0.399	32.7(0.5)	0.97(0.07)	35.4(2.9)	128.7	0.7	3.5	3.1	8.2
62	0.383	47.7(0.8)	0.87(0.07)	56.3(4.6)	133.1	0.7	3.6	3.1	8.2
63	0.390	68.5(0.5)	1.00(0.08)	45.8(3.7)	103.5	0.9	3.4	3.1	8.2
64	0.505	18.2(0.5)	1.81(0.14)	32.4(2.6)	79.8	1.1	3.8	3.1	8.4
65	0.490	35.2(0.5)	1.25(0.09)	50.3(4.1)	138.0	0.7	3.9	3.1	8.4
66	0.502	52.8(0.5)	0.93(0.07)	62.4(5.1)	141.8	1.2	3.8	3.1	8.4
67	0.604	51.5(0.5)	0.93(0.07)	68.0(5.5)	142.1	1.1	3.8	3.1	8.4
				$T = 635 \pm 6$ K					
68	0.597	51.5(0.5)	0.94(0.07)	68.5(5.5)	155.5	0.5	4.0	3.1	8.4
				$T = 800 \pm 10$ K					
69	0.191	33.7(0.5)	1.28(0.10)	36.2(3.0)	134.9	1.8	3.9	3.5	8.7
70	0.279	17.9(0.5)	1.88(0.14)	23.6(1.9)	78.8	1.2	4.3	3.1	8.6
71	0.275	36.1(0.5)	1.13(0.09)	40.2(3.3)	150.1	0.9	4.4	3.1	8.6
72	0.278	52.0(0.5)	1.16(0.09)	46.8(3.8)	165.2	0.8	4.0	3.1	8.4
73	0.278	68.8(0.5)	1.24(0.09)	39.7(3.2)	139.9	0.9	4.0	3.1	8.4
74	0.410	18.2(0.5)	1.93(0.15)	29.8(2.4)	84.5	1.0	4.0	3.1	8.5
75	0.400	35.0(0.5)	0.92(0.07)	60.2(4.9)	152.0	0.6	4.0	3.2	8.5
76	0.385	35.2(0.5)	1.50(0.11)	45.6(3.7)	150.3	1.0	4.3	3.1	8.6
77	0.405	51.5(0.5)	1.23(0.09)	51.9(4.2)	164.8	0.6	4.0	3.1	8.4
78	0.375	68.5(0.5)	1.25(0.09)	47.0(3.8)	145.0	0.8	3.9	3.1	8.4
79	0.484	18.2(0.5)	1.84(0.14)	32.7(2.6)	84.2	0.9	4.3	3.1	8.6
80	0.484	35.2(0.5)	1.29(0.10)	50.2(4.1)	155.3	0.7	4.3	3.1	8.6
81	0.491	51.5(0.5)	1.14(0.09)	59.4(4.8)	173.0	0.6	4.3	3.1	8.6

4.1. Temperature and Pressure Control

Temperature of liquid D/T was determined by measuring the D/T vapor pressure with tensometric gauges having an accuracy of 0.5%. Hence, the D/T temperature was determined with an accuracy of 0.1 K. The temperature of gaseous D/T was measured by special thermocouples. During the experimental runs, a small temperature gradient existed in the D/T mixture, which was taken into account in the determination of temperature and its error. The accuracy of determining the temperature was 3–10 K in the 40–800 K range. Pressure was measured with the use of strain pressure gauges having a calibration error of 3%.

4.2. D/T Mixture Density

Nuclear density of liquid D/T was determined using the cryogenic data on deuterium and tritium [51] taking the mixture content into account. Errors in density were 2%.

Nuclear density of gaseous D/T was determined by two ways. The first was the use of the deuterium (tritium) equation of state [52], with the gas temperature and pressure known from measurements. Some corrections for the presence of the buffer volume being at room temperature were made. The second way was density determination via the quantity of gas in the target of a known volume. Both methods gave identical results within an accuracy of 4%. The final error of gas density was 3–4%.

In addition, we have another way to check the mixture density. If the muon beam intensity is stable, the number of muon stops in the mixture per time unit is proportional to the mixture density. In several cases, we made some corrections (a few percent) to the mixture density based on this method.

4.3. Measurements of Isotope and Molecular Gas Composition

The chromatographical method [53] was used to control the isotope and molecular composition of the mixtures. In addition, an ionization chamber was used to obtain the D/T and T_2 content. Measurements were made before filling of the target and after evacuation of the mixture from it.

The chromatographical analysis showed the molecular compositions to be very close to the equilibrium ones,

$$C_{\mathrm{DD}} : C_{\mathrm{DT}} : C_{\mathrm{TT}} = C_d^2 : 2C_dC_t : C_t^2,$$
$$C_t + C_d = 1,$$

for each gaseous mixture exposed to a muon beam. However, for liquid mixtures, the molecular content can differ from equilibrium due to the dynamic effects in evaporation of a multicomponent liquid, which was investigated by us under the conditions of our target in [54]. The deviation from the equilibrium state becomes noticeable for a high tritium concentration of $C_t > 50\%$. Appropriate corrections to the molecular and isotope concentrations of the liquid mixture were made in [55]. We note that the quantity of protium in D/T mixtures did not exceed 1%.

4.4. D/T Mixture Purity and ^{3}He Accumulation

As follows from Eq. (10), the expression for the cycling rate is independent of the muon loss, including the effect of impurities with $Z > 1$. Contrary to this, the muon losses depend on the cycling rate ($\lambda_{dt\mu}$ and λ_{dt}). The larger the cycling rate, the closer the muon losses are to their natural limit, which is equal to the probability ω_s.

As is seen from Eq. (11), the muon transfer to the possible impurities affects the value of ω. That is why the impurity level must be made as small as possible. Indeed, the condition

$$\lambda_Z C_Z \ll \omega_s \lambda_c \tag{12}$$

must be ensured.

It is necessary to distinguish two sorts of impurities: impurities with $Z > 2$ and He admixtures.

4.4.1. Impurities with $Z > 2$ and ^{4}He. These impurities are predominantly helium-4, carbon, oxygen, and nitrogen originating from imperfect purification of the mixture before filling the target and removal of residual gaseous elements from the target walls during the exposure.

The special preparation system based on palladium filters [46] provides filling of a target with gas purified at a level of $C_Z < 10^{-7}$ of volume parts. As the outgassing effect increased with temperature, the mixture purity varied from $C_Z < 10^{-7}$ for $T = 20$ K to $C_Z = 10^{-5}$–10^{-6} for $T = 800$ K. The rate of muon transfer from the $t\mu$-atom to the pointed admixtures is $\lambda_Z \sim 10^{11}$ s^{-1} for nuclei with $Z > 2$ [56] and $\lambda_{^4\mathrm{He}} \sim (1\text{–}5) \times 10^9$ s^{-1} for ^{4}He [57, 58]. Therefore, condition (12) is satisfied only for a liquid D/T mixture where the cycling rate is rather high ($\lambda_c = 50$–120 μs^{-1} depending on the tritium concentration) and most impurities (excluding helium) are solid and freezed out on the target walls.

4.4.2. ^{3}He admixture. The tritium handling system provides an initial concentration of C_{He} of ^{3}He in the mixture before being poured into a target at a level of 10^{-7}. However, due to the tritium β-decay, ^{3}He is accumulated in a target according to relation

$$C_{\mathrm{He}}(\tau) = C_t[1 - \exp(-\lambda_{\mathrm{trit}}\tau)],$$

where $\lambda_{\mathrm{trit}} = 6.4 \times 10^{-6}$ h^{-1} is the tritium decay rate. Hence, the process of muon transfer from the $t\mu$-atom to ^{3}He (with the rate $\lambda_{^3\mathrm{He}} \approx 2 \times 10^8$ s^{-1} [57]) can substantially affect the muon losses.

The ^{3}He accumulation effect is quite different for liquid and gaseous D/T. It was shown in experiment [58] that ^{3}He in liquid D/T diffuses and exits to the vapor gas. Our cooling system of the LTT [43] ensured passage of the entire D/T mixture through the vapor phase in approximately 1 h, which led to ^{3}He escape from liquid D/T. In experiments with liquid D/T, we therefore had no problem with D/T purity.

In experiments with gaseous D/T, we were forced to refill the target every 10–40 h (depending on the tritium content in D/T) to avoid accumulation of ^{3}He larger than the "critical value" $C_{^3\mathrm{He}} \approx 10^{-5}$.

5. DATA TREATMENT

The data processing included the following stages.

(1) Selection of events. By an event, we mean an occurrence of the processes caused by a single muon beginning with the muon stop in the target and ending with the muon decay. An example of the event as seen by the detectors is presented in Fig. 7. The most important criteria for the event to be accepted were presence of a reliable signal for the μ-decay electron.

(2) Creation of the charge and time spectra for neutrons from the $d + t$ reaction and for electrons from the μ-decay.

(3) Fit of these spectra to determine the "effective" MCF parameters λ_c, ω, and Y_n.

(4) Analysis of these parameters as functions of the tritium concentration to obtain the $dt\mu$ formation rates $\lambda_{dt\mu-d}$ and $\lambda_{dt\mu-t}$ and the muon-to-helium sticking probability ω_s.

5.1. Analysis Methods

The most popular and practically the only method used by most groups involved in the study of the MCF $d + t$ process is the so-called standard method, where the yield and time distribution of all detected neutrons from reactions (5), (6) are recorded and analyzed. This distribution has the well-known one-exponent form (8). The number of μ-decay electrons N_e is used for normalization,

$$\frac{N_n}{N_e} = \frac{\epsilon_n \Lambda_c}{\lambda_0 + \omega\Lambda_c}. \tag{13}$$

The decay rate λ_n of the exponential in (8) and the normalized neutron yield Y_n are the measured parameters. The values of Λ_c, ω, and Y_n are extracted from (8), (9), and (13):

$$\epsilon_n \varphi \lambda_c = \frac{N_n}{N_e}\lambda_n, \quad \frac{\omega}{\epsilon_n} = \frac{\lambda_n - \lambda_0}{\epsilon_n \varphi \lambda_c}, \quad \epsilon_n Y_n = \frac{N_n}{N_e}. \tag{14}$$

In the Dubna experiments, we also used the standard method. To obtain spectrum (8), we created a time distribution of the neutron detector charge $Q(t)$. For this, we summed the amplitude spectra for each neutron detector *ND*1 and *ND*2. Then, the spectrum $Q(t)$ was transformed to the time distribution of the number of events

$$N_n(t) = Q(t)/\bar{q}$$

using the unit charge $\bar{q}$ [59]. The latter was measured under special conditions providing a low neutron multiplicity, where each charge pulse corresponded to one neutron. Charge distributions obtained in such exposures were compared with the calculated ones to obtain the experimental value of ϵ_n as a function of the threshold.

The number of electrons N_e was obtained from an analysis of the electron time spectra $N_e(t)$ using the distribution $B_{\mathrm{empty}}(t)$ measured with an empty target,

$$N_e^{\mathrm{total}}(t) = kB_{\mathrm{empty}}(t) + A_e \exp(-\lambda_e t) + F, \tag{15}$$

where λ_e is the muon disappearance rate and F is the accidental background. In this fit, k, A_e, λ_e, and F are parameters. The observed muon disappearance rates λ_e are close to the muon decay rate $\lambda_0 = 0.455\ \mu\mathrm{s}^{-1}$ and depend on the mixture purity. In exposures with liquid D/T, where the purity is maximum, λ_e is obtained equal to λ_0 within 1%.

A typical example of the fitted time distributions of decay electrons and fusion neutrons for the D/T filled target is shown in Fig. 8. The dashed line corresponds to the electrons from decays of muons stopped in the target walls (empty target).

The principal disadvantage of the standard method is that the main MCF parameters (cycling rate and effective muon losses) are not obtained directly; only their product is measured directly. In our measurements, we employed two novel independent methods proposed and developed in Dubna [60, 61]. These analysis methods make it possible to directly measure the values of λ_c and ω.

A proposal in [60] was to measure the distribution $N_{ne}(t)$ which was a function of the interval $t = t_e - t_n$ between the last detected neutron of the series and the μ-decay electron. This distribution has the form of a sum of two exponentials with significantly different slopes [60, 61],

$$\frac{dN_{ne}}{dt} = \frac{\lambda_0}{\lambda_n} \times [\omega\Lambda_c \exp(-\lambda_0 t) + \epsilon_n \Lambda_c (1-\omega)\exp(-(\lambda_0 + \lambda_n)t)], \tag{16}$$

where λ_n is expressed as

$$\lambda_n = (\epsilon_n + \omega - \epsilon_n \omega)\Lambda_c. \tag{17}$$

The first ("slow") exponential corresponds to the events with muon sticking and the second ("fast") one to the events without sticking. The cycling rate is determined from the fast component slope, and the muon loss is

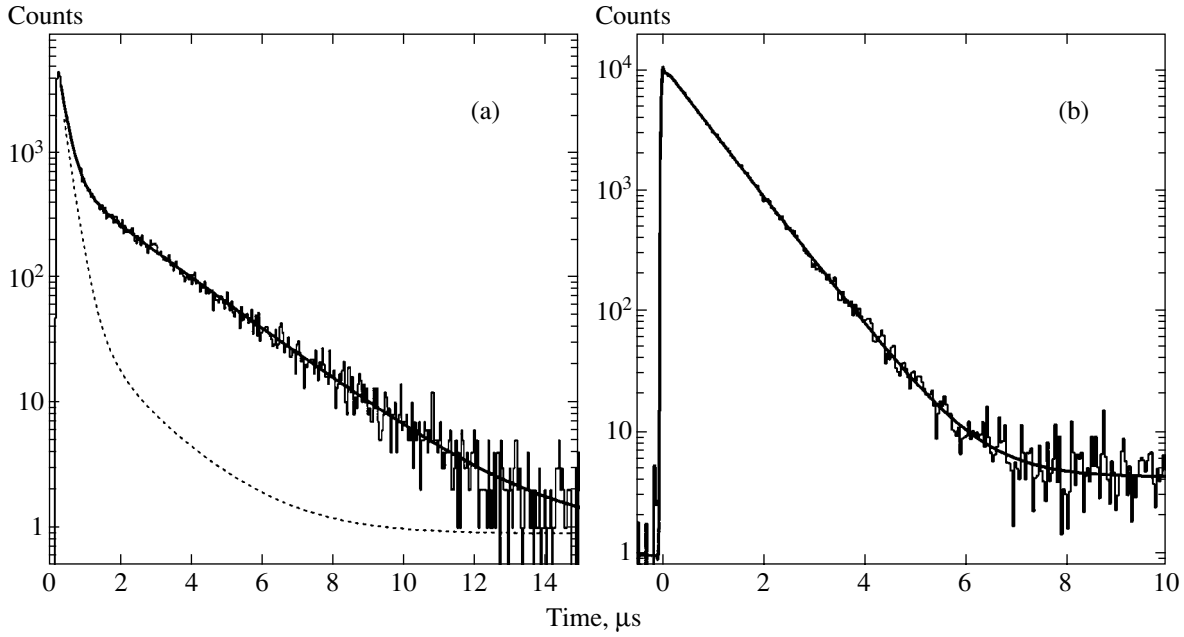


Fig. 8. Example of electron (a) and neutron (b) time distributions. Solid lines are the optimum fits with expressions (15) (a) and (8) (b); the dashed line corresponds to the electrons from empty target.

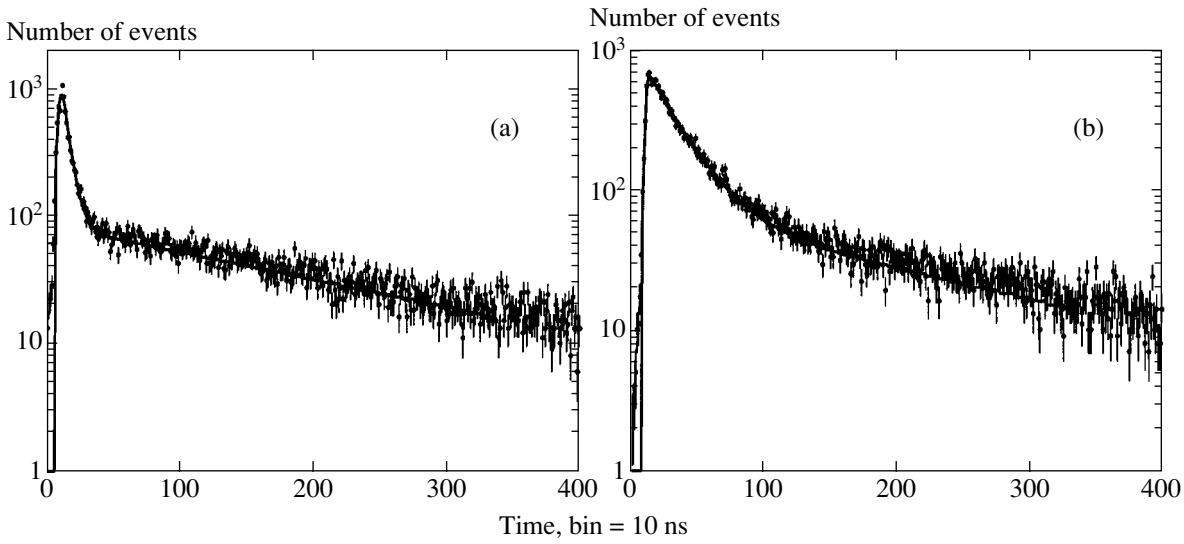


Fig. 9. Electron–last neutron timing spectra measured with a liquid D/T mixture. Spectrum (a) corresponds to the exposure with C_t = 35.2%, and variant (b) was selected for C_t = 85.5%. Lines represent the fits with expressions (16) and (17) and the optimum parameters $\epsilon_n\Lambda_c$ and ω/ϵ_n.

obtained from the ratio between the amplitudes of the slow and fast exponentials:

$$\frac{A_s}{A_f} = \frac{\omega}{\epsilon_n(1-\omega)}.$$

Examples of such distributions obtained in a liquid D/T mixture are presented in Fig. 9. As is seen from the figures, the events with and without sticking are clearly separated. Different slopes of the fast components of the spectra reflect the different values of the cycling rate realized for tritium concentrations of C_t = 35.2% and C_t = 85.5%. The advantage of the method is that charge calibration is not necessary in this case.

Another idea [61] was to measure the neutron multiplicity distribution (the number k of detected neutrons per muon) in some definite time interval T. If one selects the events for which the muon does not decay in this interval, then this distribution is a sum of two terms. One of them, which is Gaussian (Poisson) with a mean of $m = \epsilon_n\Lambda_c T$, corresponds to the events without

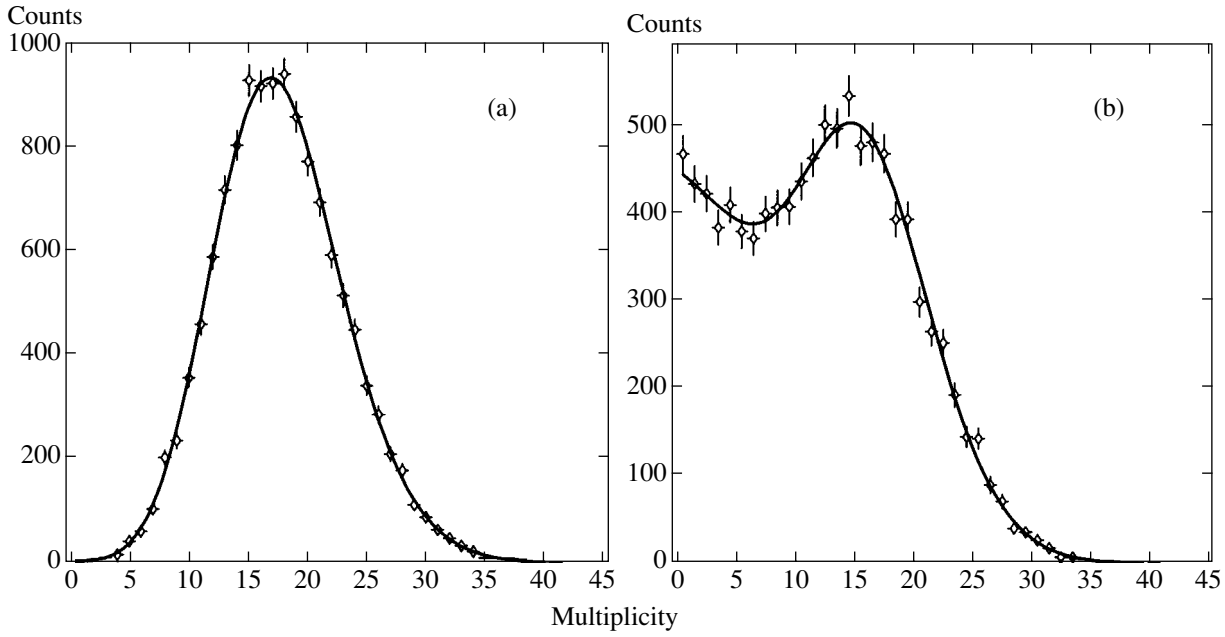


Fig. 10. Multiplicity distributions in time interval $T = 1$ μs: (a) distribution without sticking events, (b) with sticking events. Curves are the optimum fits.

sticking, and the other, depending on ω and falling with k, is the distribution of events with muon sticking.

The rigorous expression for the multiplicity distribution was obtained in [62]. It has the form

$$f(k) = \frac{[\epsilon_n(1-\omega)]^k}{(\epsilon_n + \omega - \epsilon_n\omega)^k}P(k) + \frac{[\epsilon_n(1-\omega)]^{k-1}\omega}{(\epsilon_n + \omega - \epsilon_n\omega)^k}F(k), \quad (18)$$

where $P(k)$ is the Poisson distribution with the mean $m = \lambda_n T$,

$$P(m) = \frac{(\lambda_n T)^k}{k!}e^{-\lambda_n T},$$

$$F(k) = 1 - e^{-\lambda_n T}\sum_{i=0}^{k-1}\frac{(\lambda_n T)^i}{i!},$$

and λ_n is given by formula (17).

Formula (18) corresponds to the "event mode" where the number of detected neutrons was considered. Actually measured in experiments were the distributions of the neutron detector charge; they were divided by the unit charge to obtain a multiplicity distribution. The real response function of the detector results in diffusion of the measured spectra as compared with the ones obtained in the event mode. It turns out that, in a good approximation (with an accuracy of 2–3% in the cycling rate), the real distribution can be obtained as a convolution of formula (18) with a Gaussian function. The Gaussian width is varied to obtain the best agreement between experiment and calculations.

We can also select only nonsticking events. For this, we must exclude the requirement of the electron signal in the trigger and select only those neutron series whose duration is larger than the chosen interval T. Plotting and analyzing the multiplicity distribution of such events, we can directly obtain the cycling rate. The advantage of this method is that we do not need the fact of the μ-decay electron existence. Of course, only λ_c can be determined in this case, because events without muon sticking to helium are accepted. The examples of fitted multiplicity distributions are presented in Fig. 10.

The comparison of all methods that we used in the analysis is given in Table 2. The statistical power is practically the same for all methods. Indeed, in the standard method, the main factor for the statistical accuracy is the limited number of electrons; the number of neutrons is much higher under real experimental conditions. In two other methods, the full statistics is the number of the first or last neutrons, which are also approximately equal to the electron number.

In our investigations, we use all the three methods mentioned. This allows us to reliably analyze the data, with minimum systematic uncertainties. Of course, the full analysis is rather complicated and includes many tests with different selection rules for events to be accepted.

5.2. Electron Identification

A serious problem in the MCF data analysis is how to distinguish the real electron from a false one. Under

Table 2. Comparison of different methods used in the MCF study

Method	λ_c, ω determination	Charge calibration	Electron detection
Standard	Indirect	Necessary	Necessary
t_e–t_n	Direct	Not necessary	Necessary
Multiplicity	Direct	Necessary	Not necessary

the conditions where one muon can cause up to 100 reactions (5), it is possible to detect a neutron by the electron detector and accept it as an electron. Contrary to the measurements of other groups, we detect electrons with a proportional wire counter, having a very low sensitivity to neutrons. However, even in this case, a fraction of false electrons caused by the neutron counts was noticeable.

Only the last (in time) electron signal is accepted as real. It would be enough to exclude the false electrons if the electron detection efficiency would be $\epsilon_e = 100\%$. However, for different reasons (see [63]), this efficiency is not equal to unity. Thus, the situation can occur where a real electron is not detected and a false one is interpreted as real. The presence of false electrons results in distortion of λ_c, ω, and Y_n. The cycling rate determined according to formula (14) is distorted due to the error in N_n and N_e, and λ_n extracted from expressions (8) and (15) senses the error in λ_c. On the other hand, confusion of the real and false electrons leads to distortion in the relation between "stick" series (interrupted due to muon sticking) and "no-stick" series (ending with μ-decay). The latter are accepted more effectively. Thus, the results for the muon losses are also distorted. Finally, the distortion of the slope of the electron time distribution does not make it possible to correct the estimate of the D/T mixture purity and thus to check the parameters of the purification system.

Fortunately, the cycling rate determined from the peak position in the multiplicity spectrum is free of false electrons. This is a very important circumstance allowing reliable data on the cycling rate to serve as a source of "elementary" process parameters such as the $dt\mu$-molecule formation rate. Of course, it is very desirable to obtain a correct value for λ_c by different independent methods. Moreover, obtaining correct data on muon losses is an important independent task.

5.2.1. Selection by the energy loss in the neutron detector. An effective way to reject false electrons was elaborated and used in our work [23]. For this, we required the following when selecting events.

(1) Electron signals from the *PC* and *ND*1 or *ND*2 should coincide.

(2) The energy that the electron releases in the neutron detector should be greater than the maximum possible energy released by a 14 MeV neutron in this detector. This allows reliable discrimination of false electrons. The use of these selection criteria allowed us to obtain the data on λ_c and ω coinciding for all three analysis methods within 5% [23].

The disadvantage of this selection is a decrease in the statistics because an essential part of the useful events are rejected. This decrease becomes much more important in experiments with a high-pressure gaseous target having rather thick walls, for which the "output" electron energy spectra are noticeably distorted and the transparency of the target walls for electrons is noticeably smaller than for the liquid target. Therefore, for a gaseous target, reliable neutron–electron separation was connected with larger statistical losses than for a liquid target.

5.2.2. Selection by the time position of the electron signal relative to neutron series. To avoid losing statistics, we developed a new method [63] for false electron discrimination, which is most effective for the large neutron detection efficiency realized in our experiments. We now impose the criterion "electron inside neutron series." For this, we consider the neutron detector charge Q (the sum of the amplitudes) on some time interval (ΔT) close to the electron signal and delayed relative to it by Δt. The events were accepted under the condition that the charge Q is smaller than the threshold: $Q < Q_{\text{th}}$ Our examination [63] shows that the proper values are $\Delta t = 60$ ns and $\Delta T = 500$ ns. The largest values of Q_{th} correspond to events without selection for the false electron. In this case, the distortion in the electron yield and time spectrum (15) is the largest. The opposite case (low Q_{th}) corresponds to the smallest distortions for electrons and to the minimum value of the electron time slope λ_e, which nearly coincides with the one determined using selection by electron energy in the neutron detector.

The opposite situation occurs for the slope of neutron time distribution (8). In the case where the real electron is not detected, a false one is accepted as electron. This means that the long neutron series are predominantly detected, because the appearance of a false electron is most probable just in those series. Indeed, our considerations [63] show that the minimum Q_{th} (maximum false electron rejection) leads to the maximum slope λ_n. Again, the "correct" value of λ_n is in agreement with the one obtained with selection by electron energy in the neutron detector.

The main MCF parameters obtained under two different selection options coincide within an accuracy of 3–4%. The reliability of the data is confirmed by the fact that the value of the cycling rate determined by the standard method is identical to the one yielded by the multiplicity method, where it is independent of the selection criteria. The method considered gives statistics 4–5 times larger than in the case with energy discrimination (Section 5.2.1). This indicates that we have found the way, described in detail in [63], for correctly

obtaining the MCF parameters without essential loss in statistics.

5.3. Neutron Detection Efficiency

All methods discussed above give values of $\lambda_c\epsilon_n$ and ω/ϵ_n. To obtain the MCF parameters λ_c and ω, we must know the neutron detection efficiency ϵ_n.

Determining ϵ_n for organic scintillation counters is a nontrivial task, because it is influenced by factors like the geometry of the surrounding material, generation of light by various reaction products, and because many energy-dependent cross sections are involved. Because of the lack of neutron calibration sources with well-known intensity and sufficiently large energy, the efficiency had to be calculated. The Monte Carlo technique was used.

Calculations of ϵ_n for neutrons detected by the *ND* in the Dubna experiments are described in [64]. The CERN package GEANT was used in [64] for the simulation calculations. Because it lacks the appropriate low- and fast-neutron interaction cross-sections, GEANT was linked with the MICAP package. MICAP uses experimental neutron cross-sections from the ENDF/B-VI database from 20 MeV down to thermal energies (10^{-5} eV). This includes partial cross-sections, angular distributions, and energy distributions of reaction products and deexcitation photons. The preprocessed ENDF/B-VI data represent the experimental data within 2%.

After the calculation of the energy deposited inside the scintillator, the electronic output signal was obtained by first converting the energy into scintillation light considering the particle type, and then converting the total light output into an electric signal by applying the detector response function [49]. This function takes several factors into account, such as nonuniform light collection depending on the position of light generation inside the scintillator and photon statistics. The results of calculations [64] for 14 MeV neutrons from reactions (5), (6) are presented in Fig. 11 together with the measured spectra.

One neutron detected in a scintillator may generate a response from one detector or, due to scattering or to generated gamma-rays, from both detectors. This leads to a single and coincident rate. The corresponding spectra are shown in Fig. 11 together with the measured ones. As can be seen, there is good agreement between the measurements and the calculations in both cases (single and coincident). The intensity and amplitude calibration of the calculated single spectrum was normalized to single data. The normalization thus obtained is then applied to the calculated coincident spectrum, which then neatly coincides with the corresponding data. This means that the single-to-coincident ratio is well predicted, which should be considered as a sensitive validation check for the calculations. The estimated relative uncertainty in ϵ_n is no worse than 5–7%.

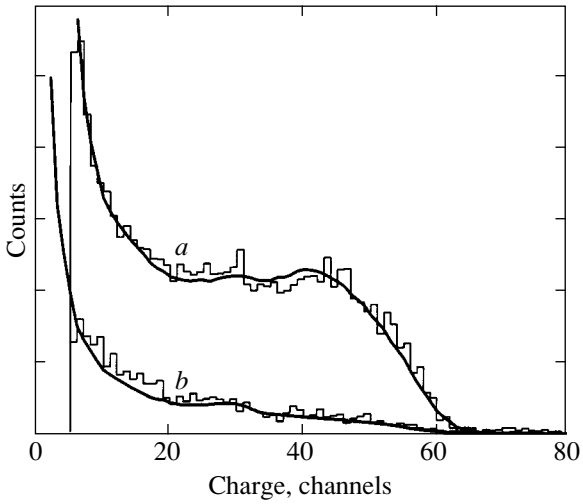


Fig. 11. Charge distribution for 14 MeV neutrons measured in [23] with the *ND* (histograms). Distributions are plotted for single (*a*) and coincident (*b*) events. Lines represent the Monte Carlo calculations [64].

The problem is how the neutron detection efficiency depends on the neutron multiplicity (cycling rate). The main idea of using the flash ADC is that the total charge per number of neutrons is conserved even when the ND signals mostly overlap. However, it is true only for zero charge threshold. In fact, the cluster charge should be limited to reduce the low-energy background. At a high neutron multiplicity, small-charge clusters can overlap with one or more other clusters and, hence, can be accepted (a noneffective threshold). Obviously, this results in an increase of the detection efficiency compared with the low neutron multiplicity. The actual increase depends on several factors, such as the shape of the *ND* signal, the form of the response function, the magnitude of the threshold, and the measured cycling rate. Because one would expect an essential correction to the value of ϵ_n, the problem required special consideration.

This was made in [65], where the fusion neutron registration was Monte Carlo simulated for a wide cycling rate range. All the three analysis methods were considered. It turned out that, in the standard and multiplicity methods, the corresponding corrections to the efficiency were not very large: even for the maximum possible measured cycling rate $\epsilon_n\Lambda_c = 40\ \mu s^{-1}$, they are only 12%.

6. RESULTS

6.1. The Effective MCF Parameters

The effective MCF parameters λ_c, ω, and Y_n were obtained from the fit of the distributions considered in the "standard", "multiplicity", and "$t_e - t_n$" analysis methods. Although the first two methods are more reliable, the results obtained by three different methods

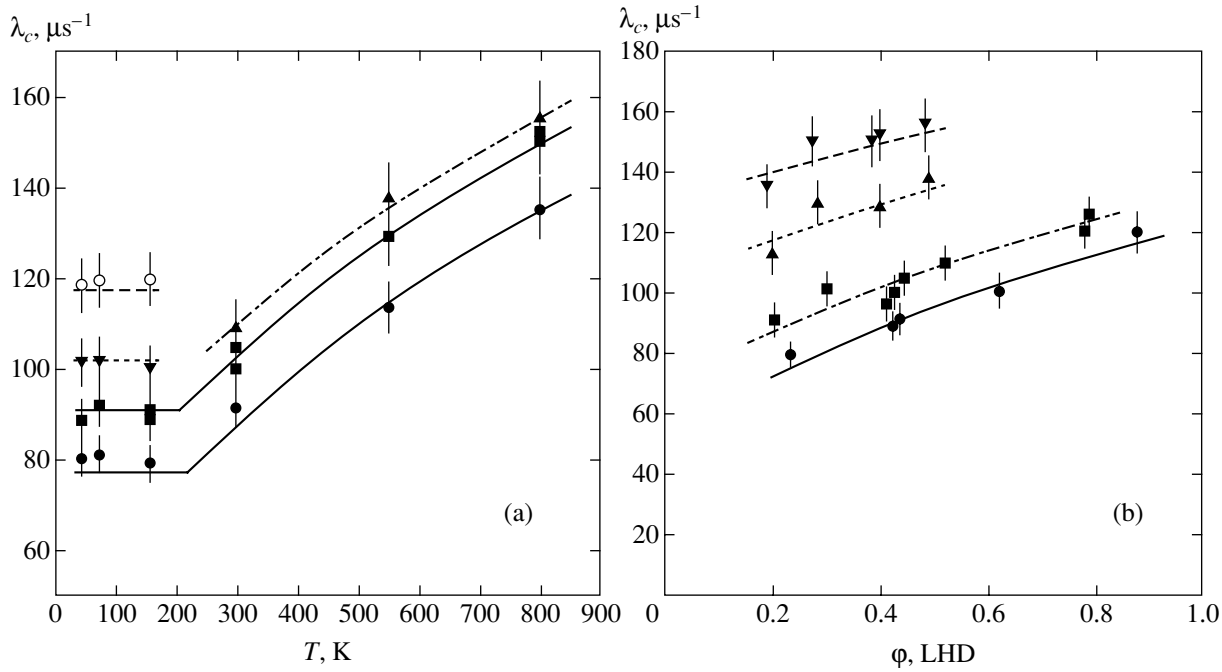


Fig. 12. (a) Normalized cycling rates as a function of temperature for the gaseous D/T mixture at $C_t \approx 33\%$ and different densities φ = 0.88–0.91 (○), 0.62–0.64 (▼), 0.49–0.52 (▲), 0.39–0.45 (■), 0.19–0.24 (●) LHD. (b) Normalized cycling rates as a function of density for the gaseous D/T mixture at $C_t \approx 33\%$ and different temperatures T = 800 K, C_t = 0.34–0.36 (▼); T = 550 K, C_t = 0.33–0.36 (▲); T = 300 K, C_t = 0.31–0.36 (■); T = 158 K, C_t = 0.31 (●). The curves are obtained with optimum parameters.

were in agreement with an accuracy 3–4%. They are presented in Table 1 and Fig. 12.

The statistical uncertainty in the results is determined by the number of events and the fit accuracy. In both main methods (standard and multiplicity), the statistics ensures that this error is not higher than 2%. We note that, in the multiplicity method, we do not need the muon number normalization.

The following factors contribute to systematic error.

(1) Uncertainty in the neutron detection efficiency makes the maximum contribution to systematic error. It was estimated from the calculation of ϵ_n and the accuracy in the energy threshold determination and is 6% in total.

(2) Uncertainty of the charge calibration procedure gives an error smaller than 3%.

(3) Uncertainty of the gas and liquid density (for normalized cycling rate) is about 3–4 and 2%, respectively.

(4) Uncertainty of the time zero position (only for the standard method) gives a systematic error smaller than 0.5%.

(5) Uncertainty due to the correct selection of muon decay electrons (only for the standard method; see Section 5.2) is 2%.

(6) Uncertainty caused by possible instability of detectors and electronics does not exceed 2%.

Therefore, the total uncertainty in ω, Y_n, and the absolute values of λ_c did not exceed 9%. Obviously, the relative dependences of the cycling rate on temperature and density are known with a better accuracy (4.5–5.5%).

6.2. The dtμ-molecule Formation Rate and Muon Sticking Probability

The usual way to determine the "physical" values $\lambda_{dt\mu}$ and ω_s is an analysis of the "effective" parameters λ_c and ω the use of expressions (10) and (11), representing their dependence on tritium concentration and density. For this purpose, it is, first of all, necessary to express q_{1S} as a function of C_t and φ. With the general expression (1) and the theoretical predictions in [32, 34, 66], as well as experimental results in [17, 35] taken into account, the parametrization of q_{1S} was chosen in the form

$$q_{1S}(C_t, \varphi) = \frac{1}{1 + (b + c\varphi)C_t}. \qquad (19)$$

6.2.1. Fit of the liquid D/T data. Muon sticking probability ω_s. As we have noted, the most expedient condition for the ω_s measurement is a liquid D/T mixture, where λ_c is high and the admixture content is negligible. Our first data for the liquid D/T mixture were given in [23]. In this paper, we correct the values of the

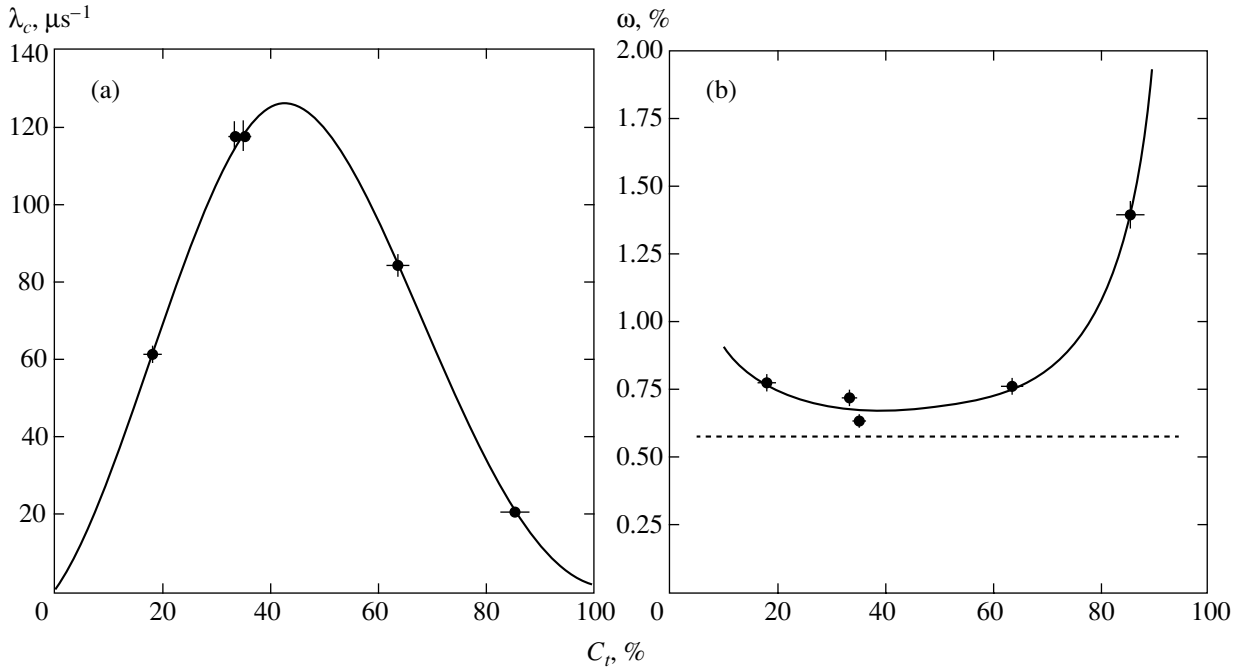


Fig. 13. Normalized cycling rates (a) and muon loss probability (b) as a function of the tritium concentration for the liquid D/T mixture (T = 22 K, $\varphi \approx 1.22$ LHD). Solid lines are optimum fits. The dashed line is the value (23).

molecular concentrations and, in addition, perform the common fit of λ_c and ω in liquid D/T using formulas (10) and (11). As for all liquid points, the density values were very close to each other, the parametrization for q_{1S} was chosen as

$$q_{1S}(C_t) = (1 + aC_t)^{-1}, \qquad (20)$$

where $a \equiv b + c\varphi$ from (19). We have performed a set of fits, varying different parameters according to the known theoretical and experimental data for λ_{1-0} [68, 69], $\lambda_{tt\mu}$ [70–72], and $\lambda_{dt\mu-t}$ [14, 17, 18, 70]. Such a variation does not lead to a significant change in the results. The systematic error $\Delta\epsilon_n$ (the same for all liquid data) was excluded from the data errors in these fits. In Table 3, the values used for the MCF cycle parameters and the results of one of the fits are shown.

The fit results are shown in Fig. 13 and Table 3. The main results for the liquid ($T \approx 22$ K, $\varphi \approx 1.22$ LHD) D/T mixture are

$$\lambda_{dt\mu-d} = (685 \pm 35^{\text{stat}} \pm 41^{\text{syst}})\ \mu s^{-1}, \qquad (21)$$

$$\lambda_{dt\mu-t} = (18 \pm 6^{\text{stat}} \pm 11^{\text{syst}})\ \mu s^{-1}, \qquad (22)$$

$$\omega_s = (0.573 \pm 0.021^{\text{stat}} \pm 0.032^{\text{syst}})\ \%. \qquad (23)$$

Our value of $\lambda_{dt\mu-d}$ in (21) is essentially higher than the PSI group data [14] but is in agreement with the LAMPF results [17] (see Fig. 15). The value of $\lambda_{dt\mu-t}$ in (22) is in satisfactory agreement with the values obtained in [17, 18], $\lambda_{dt\mu-t} = 20\ \mu s^{-1}$, and [70], $\lambda_{dt\mu-t} = 11^{+6}_{-11}\ \mu s^{-1}$. An unexpectedly high rate $\lambda_{dt\mu-t} = 160\ \mu s^{-1}$ was obtained by the RIKEN group [16, 73]. Fixing this value, we do not achieve any satisfactory agreement of fit to our data, and we therefore conclude that this value is uprated by about a factor of 5.

Table 3. Results of one from the set of common fits of the data for liquid D/T. Fixed parameters are given with references

Parameter	Value	Ref.
a	2.9 ± 0.4	free
$\lambda_Z C_Z$, μs^{-1}	0.08 ± 0.03	free
λ_{dt}, μs^{-1}	280	[18, 21, 35, 36]
λ_{1-0}, μs^{-1}	1200	[68, 69]
$\lambda^{3/2}_{dd\mu}$, μs^{-1}	3.5	[11]
$\lambda_{3/2-1/2}$, μs^{-1}	36	[11]
λ^{tt}_f, μs^{-1}	14	[72]
ω_{dd}, %	0.13	[11]
r	0.51	[11]
$\lambda_{tt\mu}\omega_{tt}$, μs^{-1}	0.28 ± 0.15	free
$\lambda_{dt\mu-d}$, μs^{-1}	650 ± 40	free
$\lambda_{dt\mu-t}$, μs^{-1}	21 ± 8	free
ω_s, %	0.574 ± 0.022	free

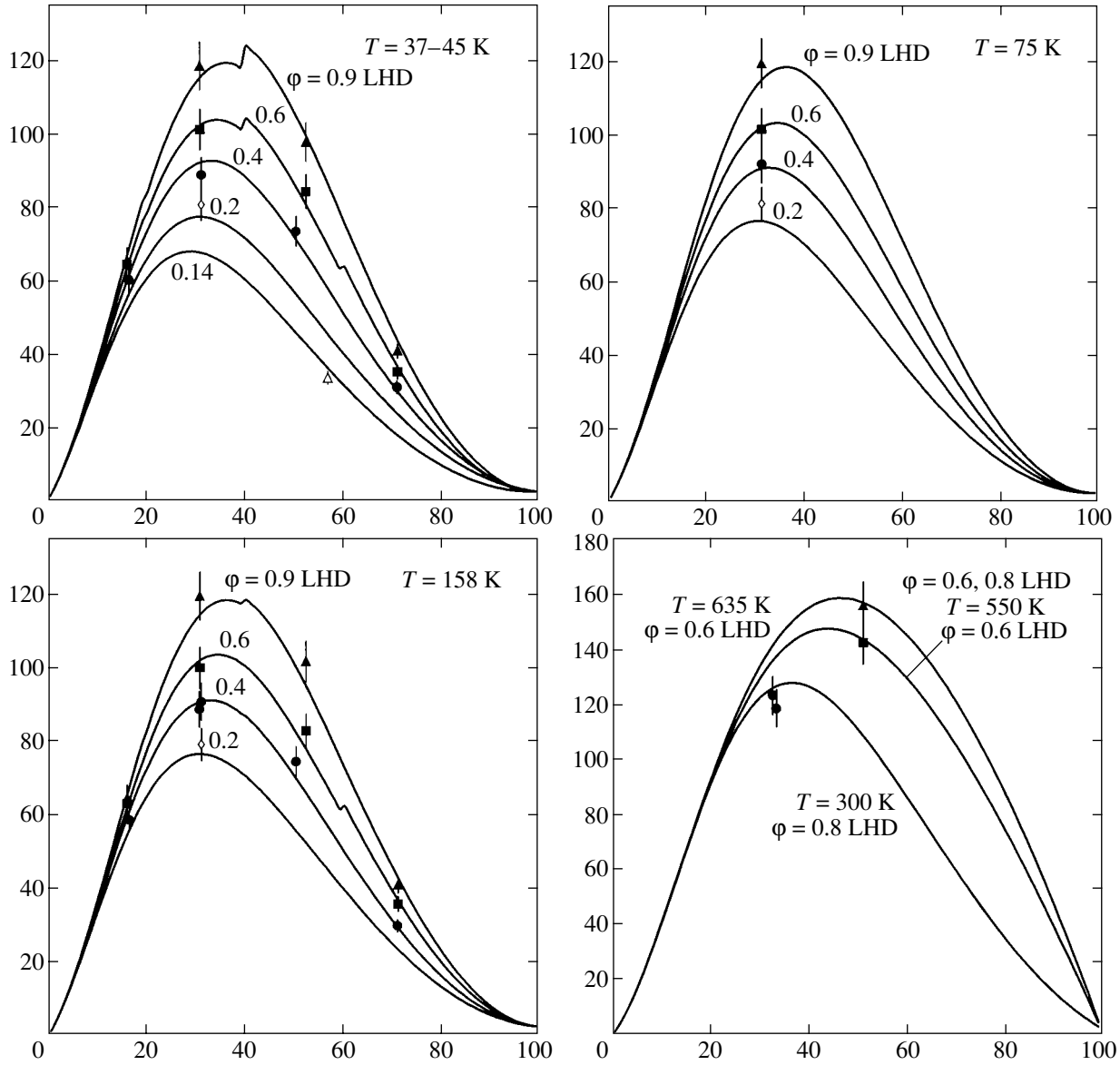


Fig. 14. Common fit of the normalized cycling rates as a function of the tritium concentration for all (76 points) data for the gaseous D/T mixture (T = 37–800 K, φ = 0.143–1.024 LHD). Lines are the optimum fit.

The probability ω_s of the effective muon-to-helium sticking in *dt*-fusion is one of the most important MCF characteristics, because it limits the number of fusions per muon. In theory, ω_s is considered as the product

$$\omega_s = \omega_s^0(1 - R),$$

where ω_s^0 is the "initial" sticking probability directly after fusion and R is the probability of the muon-from-helium stripping during the Heμ thermalization stage. R is density-dependent, and, hence, the theory predicts a slow, close to linear, decrease of ω_s with density. Comparison of different theoretical and experimental results on ω_s is presented in Table 4.

The mean value $\lambda_Z C_Z = 0.08 \pm 0.03\ \mu s^{-1}$ obtained in the fits is in agreement with the estimate $0.08 \pm 0.04\ \mu s^{-1}$ based on the analysis of the electron time spectra. The product $\lambda_{tt\mu}\omega_{tt}$, being free, was obtained as 0.28 ± 0.15, which agrees with [70–72].

6.2.2. Low-temperature gaseous D/T data. Here, we present new data related to the mixture temperature T = 45 and 158 K and different densities φ = 0.2–1 LHD. It is primarily interesting from the standpoint of the density dependence of the *dt*μ-molecule formation rate on D_2 molecules.

For all values of φ, approximation (20) for q_{1S} was used in the fit. The results are presented in Table 5. The data for T = 300 K recently presented in [26] is also

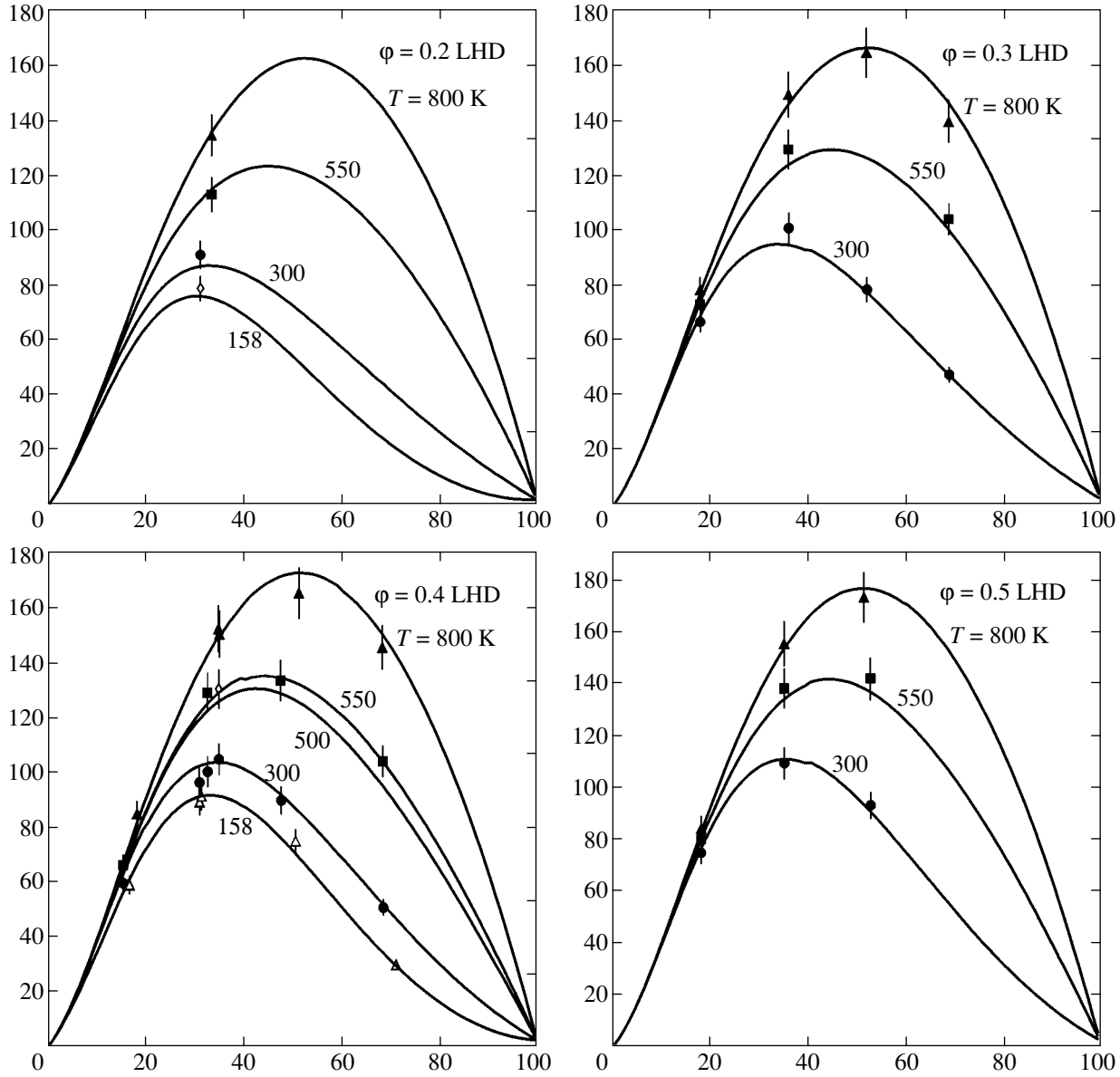


Fig. 14. (Contd.)

included. As can be seen, the data obtained for $\lambda_{dt\mu-d}$ demonstrate a strong density dependence and demonstrate practically no sensitivity to the mixture temperature.

6.2.3. High-temperature gaseous D/T data. Measurements for high temperatures ($T \geq 300$ K) are important for the determination of the $dt\mu$-molecule formation rate on DT molecules. According to the "standard" theory, the Maxwell distribution for the $t\mu$-energy ($E_{t\mu}$) overlaps the nearest resonance $\lambda_{dt\mu-t}(E_{t\mu})$ in this region. In addition, it is interesting to clarify for which temperatures $\lambda_{dt\mu-d}$ remains density-dependent and what is its temperature dependence.

The preliminary data for $T = 300$, 550, 800 K were presented in [26]. We now present the final data obtained from the common fit of the experimental dependences $\lambda_c(\varphi, T)$. The $dt\mu$-molecule formation rates on D_2 and DT molecules were assumed independent of density at temperatures $T > 300$ K. They are presented in Table 6.

The following conclusions can be made from their consideration.

(1) According to theory, $\lambda_{dt\mu-t}$ rises with temperature.

(2) Contrarily this, $\lambda_{dt\mu-d}$ does not reveal a temperature dependence.

6.2.4. Common fit of gaseous D/T data. To reliably extract $\lambda_{dt\mu}$ for given T and φ, each set of $\lambda_c(\varphi, T; C_t)$ should contain sufficient points corresponding to a

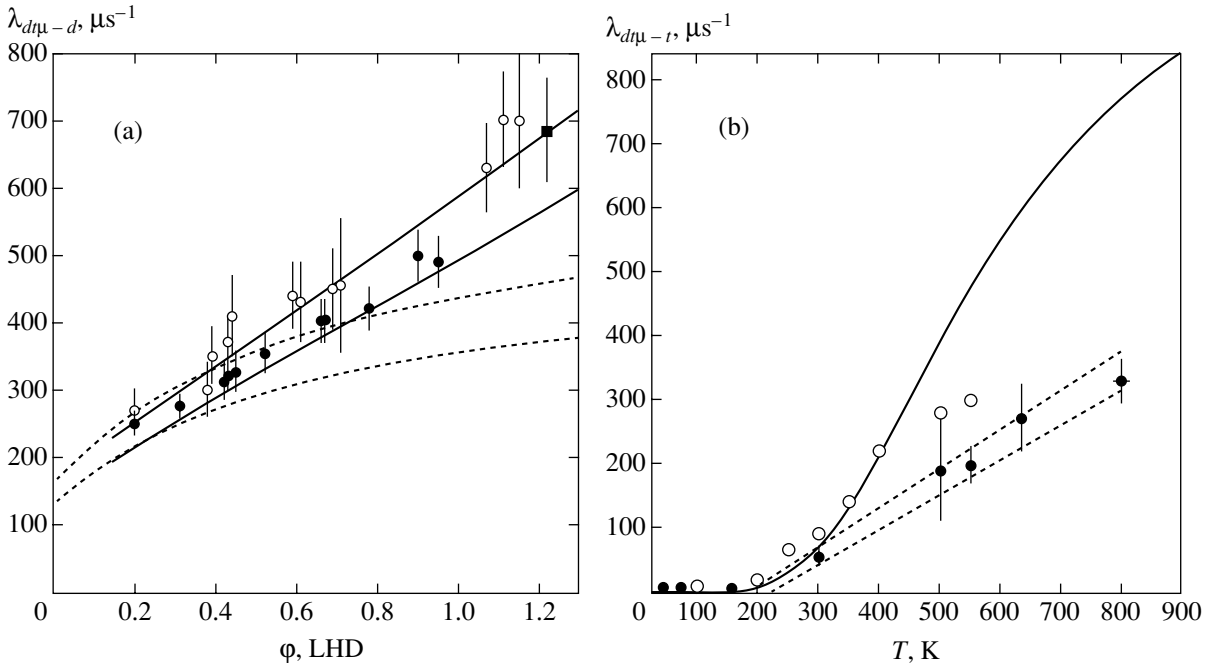


Fig. 15. (a) $\lambda_{dt\mu-d}$ as a function of density for $T \le 300$ K. Filled circles are our points for gas; empty circles are the results of LAMPF [17]; the square is the result of the present paper for liquid (21). Solid lines are the permissible values found from the fit. Dashed lines are limits for the $\lambda_{dt\mu-d}$ region obtained in [14]. (b) $\lambda_{dt\mu-t}$ as a function of temperature. Filled circles are our points; empty circles are the results of LAMPF [17]. The solid line is the theory result [13] for $\lambda^0_{dt\mu-t}$. Dashed lines are limits of parametrization (25).

wide range of C_t. Not all our data satisfied this condition. We could not conduct measurements with high C_t at "extreme" φ and *T*. Nevertheless, we could include all data in the analysis, making certain assumptions on the $\lambda_{dt\mu}$ density and temperature dependences.

Table 4. Comparison of the results on the ω_s obtained in different experimental and theoretical investigations

ω_s, %	Ref.	Comment
0.58	[74]	Theory for φ = 1.2 LHD
0.58	[75]	Theory for φ = 1.2 LHD
0.65	[76]	Theory for φ = 1.2 LHD
$0.43 \pm 0.05 \pm 0.06$	[77]	LAMPF experiment for φ = 1.2 LHD
$0.48 \pm 0.02 \pm 0.04$	[78]	PSI experiment for φ = 1.2 LHD
0.532 ± 0.030	[79]	RIKEN experiment for φ = 1.2 LHD
0.505 ± 0.029	[10]	PSI experiment for φ = 1.45 LHD
$0.573 \pm 0.021^{stat} \pm 0.032^{syst}$		This experiment for φ = 1.22 LHD

Based on our preliminary analysis [26, 67] and temperature and density dependences of λ_c obtained from the analysis of the total data, we can conclude that:

(i) $\lambda_{dt\mu-d}$ rises linearly with density in a wide temperature range up to $T = 300$ K;

(ii) at temperatures $T = 300$–800 K, $\lambda_{dt\mu-t}$ is very close to a linear function of temperature and does not depend on density.

We therefore chose the simplest linear parametrization for the temperature and density dependences of the formation rates:

$$\lambda_{dt\mu-d}(\varphi) = A_d + B_d\varphi \text{ at } T = 37\text{–}300 \text{ K}; \quad \lambda_{dt\mu-d} = C_d \text{ at } T > 300 \text{ K}, \tag{24}$$

$$\lambda_{dt\mu-t}(T) = A_t + B_t T \text{ at } T \ge 200 \text{ K}, \quad \varphi = 0.2\text{–}0.9. \tag{25}$$

The general expression (19) was used for q_{1S}, including the density-dependence term.

A total of 76 gaseous points of λ_c were under fitting by using formula (10). The systematic error due to the neutron detection efficiency $\Delta\epsilon_n$ (the same for all data) was excluded from the errors in the course of fitting. The results are presented in Table 6 and Figs. 14–16.

Figure 14 shows how the experimental values of λ_c are described by formula (10) with our parametrization of q_{1S} and $\lambda_{dt\mu}$. It follows from the fit that the experimental data are in satisfactory agreement with the used approximation: $\chi^2 = 84$ for 76 points and six variable parameters. The optimum values of our approximation are presented in Table 7. We note that q_{1S} turned out to be density-independent. The same was obtained in [14], but our values of q_{1S} are somewhat lower than in that study (see Fig. 16).

Figure 15 represents the fit results for the low-temperature ($T \le 300$ K) data for $\lambda_{dt\mu-d}(\varphi)$ (a) and $\lambda_{dt\mu-t}(T)$ (b). The area limited by the straight lines corresponds to permissible values found from the fit with taking the uncertainties in density and temperature into account. The obtained fit accuracy turned out to be

$$\delta(\lambda_{dt\mu-d}) = 8\text{–}9\%,$$
$$\delta(\lambda_{dt\mu-t}) = 20\%(T = 300\text{ K})\text{–}9\%(T = 800\text{ K}).$$

A systematic error of $\delta_{\text{syst}} = 7\%$ should be taken into account in comparing our data with theory and other measurements.

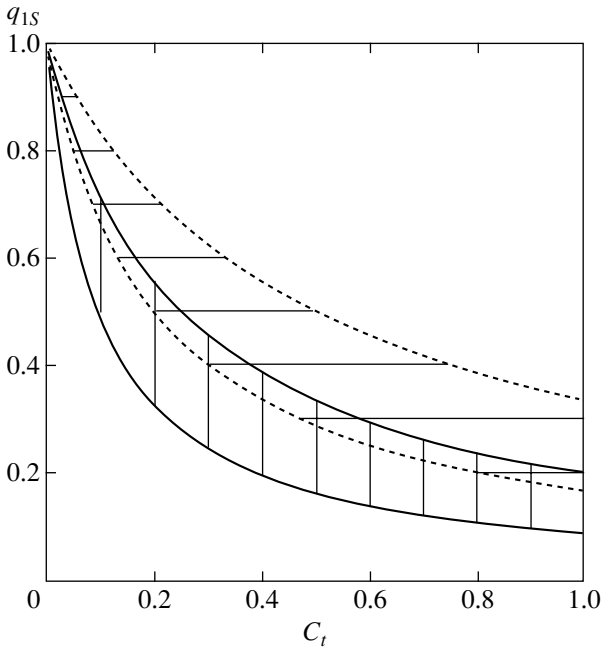


Fig. 16. q_{1S} as a function of the tritium concentration. The vertical shading is the parametrization obtained by the current fit. The horizontal shading is the PSI result [14] based on the measurements at low temperature ≤40 K.

7. DISCUSSION

7.1. Muon Sticking Probability

As can be seen from Table 4, the experimental values of ω_s obtained by different experimental groups are in satisfactory agreement with each other. The value obtained by the direct method remarkably coincides with the one determined from the analysis of the muon losses as a function of the tritium concentration. At the same time, some disagreement between experiment and theory remains.

7.2. q_{1S} and $\lambda_{dt\mu}$

Analysis of the experimental data confirms the theoretical conclusion about the significant role of the

Table 5. Results of the fit of the data for gaseous D/T at low temperatures (45–300 K) (*a* is the parameter of formula (20) for the q_{1S} approximation)

Parameter	Value	Conditions	
a	7.1 ± 2.5 free	*T*, K	φ, LHD
$\lambda_{dt\mu-d}$, μs^{-1}	326 ± 27 free	45	0.45
	403 ± 32 free	45	0.67
	490 ± 36 free	45	0.95
	320 ± 26 free	158	0.43
	402 ± 32 free	158	0.66
	499 ± 37 free	158	0.90
	292 ± 29 free	300	0.31
	313 ± 30 free	300	0.42
	380 ± 33 free	300	0.52
$\lambda_{dt\mu-t}$, μs^{-1}	7 ± 4 free	45–160	0.4–1.0
	52 ± 14 free	300	0.3–0.5

Table 6. Results of the fit of the data for gaseous D/T at high temperatures (300–800 K)

Parameter	Value	Conditions	
a	8.5 ± 2.8 free	*T*, K	φ, LHD
$\lambda_{dt\mu-t}$, μs^{-1}	56 ± 14 free	300	0.2–0.8
	190 ± 81 free	500	0.2–0.8
	198 ± 28 free	550	0.2–0.8
	270 ± 53 free	635	0.2–0.8
	328 ± 34 free	800	0.2–0.8
$\lambda_{dt\mu-d}$, μs^{-1}	251 ± 36 free	300	0.20
	277 ± 30 free	300	0.31
	293 ± 31 free	300	0.42
	354 ± 34 free	300	0.52
	420 ± 44 free	300	0.78
	319 ± 45 free	500–800	0.19–0.60

Table 7. Results of the common fit of all the data for gaseous D/T

Parameter	Fit result	
q_{1S}: b	7.2 ± 2.9	free
c	0 ± 1	free
λ_{dt}, μs^{-1}	280 [18, 21, 35, 36]	fixed
λ_{1-0}, μs^{-1}	1200 [68, 69]	fixed
$\lambda_{dt\mu-d}$: A_d ($T \le 300$ K), μs^{-1}	156 ± 14	free
B_d ($T \le 300$ K), LHD^{-1}	384 ± 21	free
C_d ($T > 300$ K), μs^{-1}	331 ± 32	free
$\lambda_{dt\mu-t}$ ($T \le 200$ K), μs^{-1}	6 ± 6	free
A_t ($T \ge 200$ K), μs^{-1}	-117 ± 9	free
B_t ($T \ge 200$ K), K^{-1}	0.577 ± 0.028	free

muon transfer from the excited $d\mu$-atom states. According to the theory, the intensity of this process turns out to depend on the tritium concentration. The probability q_{1S} of a muon reaching the $d\mu$-atomic $1S$ state is successfully described by rather simple expression (20) with the same parameter a for different C_t. At the same time, contrary to the theoretical predictions, q_{1S} does not show a noticeable density dependence. These conclusions coincide with those made in the PSI paper [14].

As follows from the theory, at low temperatures ($T <$ 300 K), the process of the $dt\mu$ formation on D_2 molecules dominates. A nontrivial density dependence $\lambda_{dt\mu-d}(\varphi)$ evidences in favor of the triple collision mechanism (7). Unfortunately, there is still no qualitative agreement between experiment and theory on the intensity of this process. It seems that the absence of a noticeable dependence of $\lambda_{dt\mu-d}$ on temperature is difficult to reconcile with the mechanism of $dt\mu$ formation on the "negative" resonance.

The experimental data on the dependence $\lambda_{dt\mu-d}(\varphi)$ obtained by different experimental groups are in satisfactory agreement. We note that the parametrization of this dependence suggested by the PSI group is not appropriate for the entire data set.

According to the theoretical predictions about the resonance positions, the process of the $dt\mu$-molecule formation on DT molecules manifests itself at high temperatures $T \approx 300$ K and rises with temperature. However, both the present results and the LAMPF data on $\lambda_{dt\mu-t}(T)$ turned out to be significantly lower than the calculated ones. This means that the intensity of the appropriate resonances is overestimated by the theory. The same conclusion follows from the analysis of the epithermal effects in the $dt\mu$ formation made by the PSI group [80]. At the same time, the TRIUMF group, making the TOF measurements of the MCF $d + t$ reaction yield as a function of the $t\mu$-atom energy [81], concluded that their data are in a satisfactory agreement with the theory. One should note, however, that the analysis in [81] is very complicated and can involve some systematic uncertainties. To clarify the situation, it is very important to make steady-state measurements at the highest temperatures T = 1000–2000 K, where the Maxwell distribution of the $t\mu$-atom energy overlaps the most intensive resonances.

7.3. Possible Influence of Epithermal Effects

The parameters obtained are related to the steady-state regime, when the $t\mu$-atoms formed with the initial energy $E \ge 1$ eV have already passed through the resonances and are thermalized. However, each time after muon regeneration in the fusion reaction, the μ-atoms go through the deceleration stage, again feeling the effect of the resonances. Obviously, this leads to an increase in the cycling rate compared to values related to the Maxwell-distributed μ-atoms. This is similar to the well-studied μ-catalysis in low-temperature deuterium related to the two $d\mu$-atom spin states [82]. As in that case, there are "upper" and "lower" states with sharply different cycling rates and quick degradation of the "upper" state. By analogy, one can express the steady-state cycling rate λ_c^{ss} as

$$\lambda_c^{ss} = \lambda_c(1 + \delta),$$

where λ_c is the "bare" value.

A relative increase in the cycling rate δ can be estimated as

$$\delta \approx \alpha_{ep}\lambda_{dt\mu}^{\text{res}}/\lambda_d, \tag{26}$$

where α_{ep} is the fraction of the $t\mu$-atoms passing through the resonances during thermalization, $\lambda_{dt\mu}^{\text{res}}$ is the effective $dt\mu$-molecule formation rate in the resonance region, and λ_d is the rate with which $t\mu$-atoms leave the resonance (thermalization and back decay after the $dt\mu$ formation).

Estimations made from the calculated values of $\lambda_{dt\mu}(E_{t\mu})$ [13] and the scattering cross sections $\sigma_{t\mu+d,t\mu+t}$ [83] are evidence that the corrections to the steady state can be as large as tenths of a percent. One can expect that the correction δ should be smallest for a low tritium concentration, because the deceleration rate in $t\mu + d$ collisions is significantly larger than in $t\mu + t$ collisions predominant at high C_t.

As a consequence, there arises a problem of correctly extracting the $dt\mu$-molecule formation rate from the dependence $\lambda_c(C_t)$ given by formula (10). The obvious conclusion is an overestimation of $\lambda_{dt\mu}$ as compared with the thermalized $t\mu$-atom situation. In addition, systematic errors in the parameters of (10) can occur. Fortunately, as is seen from Figs. 13 and 14, there is satisfactory agreement between the experimental values of λ_c and expression (10). Thus, it is believed that the cor-

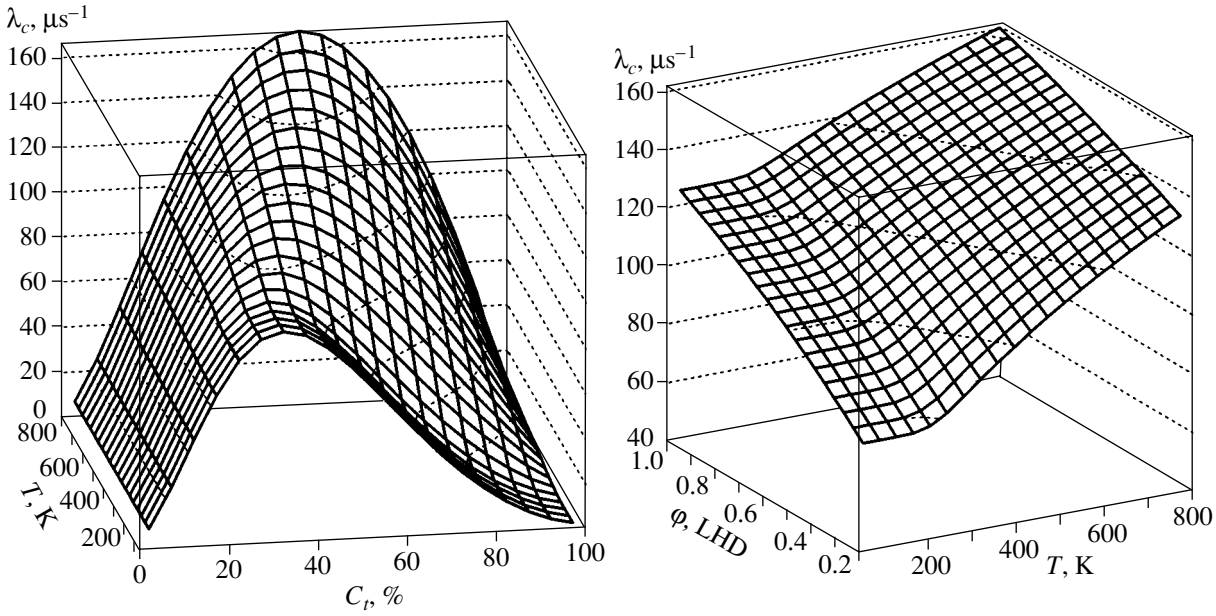


Fig. 17. Normalized cycling rate dependences on the D/T mixture conditions plotted with the use of optimum parametrizations obtained from the fit. (left) Normalized cycling rate as a function of tritium concentration and temperature for φ = 0.4 LHD. (right) Normalized cycling rate as a function of temperature and density for C_t = 0.35. The temperature region 160–300 K is smoothed.

responding distortions are not as large. According to our estimations, the appropriate corrections to $\lambda_{dt\mu}$ in the region C_t = 20–70% are $\delta \approx$ 10–20%.

8. CONCLUSIONS

The systematic experimental investigations of the MCF process in the D/T mixture have been conducted at the JINR phasotron by the novel method. Measurements were made in a wide range of the mixture parameters: density, temperature, and tritium concentration. The variety of the experimental conditions can be seen in Fig. 17, showing the cycling rate vs. mixture conditions.

Analysis of the data allows us to determine the basic MCF parameters. In general, they are in agreement with the ones obtained by other groups in the region where the experimental conditions were similar. Comparison of the experimental data with the theory confirms the efficiency of the main mechanisms considered in the MCF theory, but a full qualitative description of the process has not yet been achieved.

In our opinion, it will be very important to make measurements with a D/T mixture at the highest temperatures of T = 1000–2000 K, where the main resonances manifest themselves most effectively.

ACKNOWLEDGMENTS

The authors are grateful to Prof. L.I. Ponomarev for stimulating discussions and to D.V. Balin, G.G. Semenchuk, Yu.A. Smirenin, and N.I. Voropaev (PNPI, Gatchina), who took part in our first measurements. The work was supported by the ISTC (project no. 025-95), the RF Ministry of Atomic Energy (contract no. 6.25.19.19.99.969), the RF Ministry of Science and Technology (State Contract no. 103-7(00)-II), and the Russian Foundation for Basic Research (project nos. 97-02-16882, 98-02-16351, 00-02-17192, and 03-02-16876).

The authors dedicate this article to the memory of Prof. V.P. Dzhelepov, their coauthor and supervisor. Dzhelepov was the initiator of the long-term experimental research on muon-catalyzed fusion and until his final days headed this scientific branch at JINR.

REFERENCES

1. L. I. Ponomarev, Contemp. Phys. **31**, 219 (1990).
2. W. H. Breunlich, P. Kammel, J. S. Cohen, and M. Leon, Annu. Rev. Nucl. Part. Sci. **39**, 311 (1989).
3. S. S. Gerstein, Yu. V. Petrov, and L. I. Ponomarev, Usp. Fiz. Nauk **160**, 3 (1990) [Sov. Phys. Usp. **33**, 1 (1990)].
4. L. I. Ponomarev, Hyperfine Interact. **138**, 15 (2001).
5. L. I. Ponomarev, in *Proceedings of the International Workshop on Exotic Atoms, Vienna, 2002* (Austrian Academy of Science, Vienna, 2002).
6. V. V. Filchenkov and N. N. Grafov, JINR Commun. (Dubna), E15-2003-96 (2003).
7. L. N. Bogdanova and V. V. Filchenkov, Hyperfine Interact. **138**, 321 (2001).
8. C. Petitjean *et al.*, Fusion Technol. **25**, 437 (1994); Yu. V. Petrov and E. G. Sakhnovsky, Hyperfine Interact.

101/**102**, 647 (1996); V. Anisimov *et al.*, Fusion Technol. **39**, 198 (2001).

9. Yu. V. Petrov, Nature **285**, 466 (1980).
10. C. Petitjean, Hyperfine Interact. **138**, 191 (2001).
11. N. I. Voropaev *et al.*, Hyperfine Interact. **138**, 331 (2001).
12. M. P. Faifman, L. I. Menshikov, and T. A. Strizh, Muon Catal. Fusion **4**, 1 (1989).
13. M. P. Faifman *et al.*, Hyperfine Interact. **101**/**102**, 179 (1996).
14. P. Ackerbauer *et al.*, Nucl. Phys. A **652**, 311 (1999).
15. P. Ackerbauer *et al.*, Hyperfine Interact. **82**, 357 (1993).
16. N. Kawamura *et al.*, Hyperfine Interact. **138**, 235 (2001).
17. S. E. Jones *et al.*, Phys. Rev. Lett. **51**, 1757 (1983); **56**, 588 (1986).
18. S. E. Jones *et al.*, Muon Catal. Fusion **1**, 21 (1987); A. J. Caffrey *et al.*, Muon Catal. Fusion **1**, 53 (1987).
19. V. P. Dzhelepov *et al.*, Zh. Éksp. Teor. Fiz. **50**, 1235 (1966) [Sov. Phys. JETP **23**, 820 (1966)].
20. V. M. Bystritsky *et al.*, Zh. Éksp. Teor. Fiz. **76**, 460 (1979) [Sov. Phys. JETP **49**, 232 (1979)].
21. V. M. Bystritsky *et al.*, Phys. Lett. B **94B**, 476 (1980); Zh. Éksp. Teor. Fiz. **80**, 1700 (1981) [Sov. Phys. JETP **53**, 877 (1981)].
22. S. I. Vinitsky *et al.*, Zh. Éksp. Teor. Fiz. **74**, 849 (1978) [Sov. Phys. JETP **47**, 444 (1978)].
23. Yu. P. Averin *et al.*, Hyperfine Interact. **118**, 111 (1999).
24. V. R. Bom *et al.*, Hyperfine Interact. **118**, 103 (1999).
25. Yu. P. Averin *et al.*, Hyperfine Interact. **118**, 121 (1999).
26. V. R. Bom *et al.*, Hyperfine Interact. **138**, 213 (2001); D. L. Demin *et al.*, JINR Commun. (Dubna), E15-2000-157 (2000).
27. L. I. Menshikov and L. I. Ponomarev, Pis'ma Zh. Éksp. Teor. Fiz. **39**, 542 (1984) [JETP Lett. **39**, 663 (1984)].
28. L. I. Menshikov and L. I. Ponomarev, Pis'ma Zh. Éksp. Teor. Fiz. **42**, 12 (1985) [JETP Lett. **42**, 13 (1985)].
29. L. I. Menshikov and L. I. Ponomarev, Z. Phys. D **2**, 1 (1986).
30. V. E. Markushin, Phys. Rev. A **50**, 1137 (1994).
31. W. Czaplinski *et al.*, Phys. Rev. A **50**, 525 (1994); **50**, 518 (1994).
32. M. P. Faifman and L. I. Menshikov, Hyperfine Interact. **138**, 61 (2001).
33. V. E. Markushin and T. S. Jensen, Hyperfine Interact. **138**, 71 (2001).
34. A. V. Kravtsov *et al.*, Hyperfine Interact. **138**, 103 (2001).
35. W. H. Breunlich *et al.*, Phys. Rev. Lett. **58**, 329 (1987).
36. D. V. Balin *et al.*, Zh. Éksp. Teor. Fiz. **92**, 1543 (1987) [Sov. Phys. JETP **65**, 866 (1987)].
37. E. Vesman, Pis'ma Zh. Éksp. Teor. Fiz. **5**, 113 (1967) [JETP Lett. **5**, 91 (1967)].
38. L. N. Bogdanova *et al.*, Zh. Éksp. Teor. Fiz. **83**, 1615 (1982) [Sov. Phys. JETP **56**, 931 (1982)].
39. L. I. Menshikov and L. I. Ponomarev, Phys. Lett. B **167**, 141 (1986).
40. M. Leon, Muon Catal. Fusion **1**, 163 (1987); Phys. Rev. A **49**, 4438 (1994).
41. Yu. V. Petrov, V. Yu. Petrov, and H. H. Schmidt, Phys. Lett. B **331**, 266 (1994).
42. A. V. Demianov *et al.*, JINR Commun. (Dubna), P9-93-374 (1993).
43. D. L. Demin *et al.*, Hyperfine Interact. **119**, 349 (1999); Prib. Tekh. Éksp., No. 1, 21 (1999) [Instrum. Exp. Tech. **42**, 18 (1999)]; Preprint JINR P13-97-243 (Joint Inst. for Nuclear Research, Dubna, 1997).
44. V. V. Perevozchikov *et al.*, Hyperfine Interact. **119**, 353 (1999); Prib. Tekh. Éksp., No. 1, 28 (1999) [Instrum. Exp. Tech. **42**, 25 (1999)]; Preprint JINR D15-98-107 (Joint Inst. for Nuclear Research, Dubna, 1998).
45. V. V. Perevozchikov *et al.*, in *Abstracts of Presentations at the Second International Workshop on Interaction of Hydrogen Isotopes with Structural Materials, IHISM-04* (Sarov, 2004), p. 165; Fusion Sci. Technol. (in press).
46. A. A. Yukhimchuk *et al.*, Hyperfine Interact. **119**, 341 (1999).
47. A. D. Konin, JINR Commun. (Dubna), P13-82-634 (1982).
48. V. P. Dzhelepov *et al.*, Nucl. Instrum. Methods Phys. Res. A **269**, 634 (1988); V. V. Filchenkov, A. D. Konin, and V. G. Zinov, Nucl. Instrum. Methods Phys. Res. A **245**, 490 (1986); V. A. Baranov *et al.*, Nucl. Instrum. Methods Phys. Res. A **374**, 335 (1996); V. G. Zinov *et al.*, JINR Commun. (Dubna), P13-91-182 (1991).
49. V. V. Filchenkov, A. D. Konin, and A. I. Rudenko, Nucl. Instrum. Methods Phys. Res. A **294**, 504 (1990).
50. V. G. Zinov *et al.*, Prib. Tekh. Éksp., No. 3, 38 (1998) [Instrum. Exp. Tech. **41**, 327 (1998)].
51. M. P. Malkov *et al.*, *Guide to Physico-Technical Base of Cryogenics* (Énergoatomizdat, Moscow, 1973) [in Russian].
52. R. Prydz, K. D. Timmerhaus, and R. B. Stewart, Adv. Cryog. Eng. **13**, 384 (1967).
53. A. A. Yukhimchuk *et al.*, Preprint VNIIEF 83-2002 (All-Russian Research Inst. of Experimental Physics, Sarov, 2002).
54. D. L. Demin *et al.*, J. Low Temp. Phys. **120**, 45 (2000); Preprint JINR P8-99-179 (Joint Inst. for Nuclear Research, Dubna, 1999).
55. D. L. Demin and N. N. Grafov, Determination of Concentrations and Nuclear Density of Hydrogen Isotope Mixtures in MCF Experiments with Liquid Tritium Target: Inner Report (LNP JINR, 1997).
56. L. Schellenberg, Muon Catal. Fusion **5**/**6**, 73 (1990/1991); Hyperfine Interact. **82**, 513 (1993).
57. B. Gartner *et al.*, Hyperfine Interact. **119**, 103 (1999).
58. N. Kawamura *et al.*, Hyperfine Interact. **118**, 213 (1999); Phys. Lett. B **465**, 74 (1999).
59. V. V. Filchenkov, A. E. Drebushko, and A. I. Rudenko, Nucl. Instrum. Methods Phys. Res. A **395**, 237 (1997).
60. V. G. Zinov, Muon Catal. Fusion **7**, 419 (1992).
61. V. V. Filchenkov, Muon Catal. Fusion **7**, 409 (1992).

62. V. V. Filchenkov and S. M. Sadetsky, Nucl. Instrum. Methods Phys. Res. A **480**, 771 (2002); Preprint JINR E15-2000-223 (Joint Inst. for Nuclear Research, Dubna, 2000).
63. V. V. Filchenkov *et al.*, Preprint JINR E15-2002-285 (Joint Inst. for Nuclear Research, Dubna, 2002); Hyperfine Interact. **155**, 39 (2004).
64. V. R. Bom and V. V. Filchenkov, Hyperfine Interact. **119**, 365 (1999); Preprint JINR E15-98-338 (Joint Inst. for Nuclear Research, Dubna, 1998); V. V. Filchenkov and L. Marczis, JINR Commun. (Dubna), E13-88-566 (1988).
65. V. V. Filchenkov, JINR Commun. (Dubna), E15-2000-224 (2000).
66. T. S. Jensen and V. E. Markushin, Eur. Phys. J. D **21**, 271 (2002).
67. V. V. Filchenkov and N. N. Grafov, Hyperfine Interact. **138**, 241 (2001).
68. L. Bracci *et al.*, Phys. Lett. A **134**, 435 (1989); Muon Catal. Fusion **4**, 247 (1989).
69. D. I. Abramov, V. V. Gusev, and L. I. Ponomarev, Hyperfine Interact. **138**, 275 (2001).
70. W. H. Breunlich *et al.*, Phys. Rev. Lett. **53**, 1137 (1984); Muon Catal. Fusion **1**, 29, 67, 121 (1987).
71. T. Matsuzaki *et al.*, Hyperfine Interact. **118**, 229 (1999); Phys. Lett. B **557**, 176 (2003).
72. L. N. Bogdanova, Muon Catal. Fusion **3**, 359 (1988).
73. K. Ishida *et al.*, Phys. Rev. Lett. **90**, 043401 (2003); Hyperfine Interact. **118**, 203 (1999).
74. M. Struensee and J. S. Cohen, Phys. Rev. A **38**, 44 (1988).
75. V. E. Markushin, Muon Catal. Fusion **3**, 395 (1988).
76. M. Kamimura *et al.*, Hyperfine Interact. **118**, 217 (1999).
77. S. E. Jones, S. F. Taylor, and A. N. Anderson, Hyperfine Interact. **82**, 303 (1993).
78. C. Petitjean, Hyperfine Interact. **82**, 273 (1993).
79. K. Ishida *et al.*, Hyperfine Interact. **138**, 225 (2001).
80. M. Jeitler *et al.*, Phys. Rev. A **51**, 2881 (1995).
81. M. C. Fujiwara *et al.*, Phys. Rev. Lett. **85**, 1642 (2000).
82. V. V. Filchenkov, Hyperfine Interact. **101**/**102**, 37 (1996).
83. A. Adamczak *et al.*, At. Data Nucl. Data Tables **62**, 255 (1996); Preprint JINR E4-95-488 (Joint Inst. for Nuclear Research, Dubna, 1995).

Journal of Experimental and Theoretical Physics, Vol. 100, No. 4, 2005, pp. 688–696.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 4, 2005, pp. 780–789.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

High-Pressure Magnetic Properties and *P–T* Phase Diagram of Iron Borate¶

A. G. Gavriliuk[a,b], I. A. Trojan[a], I. S. Lyubutin[b], S. G. Ovchinnikov[c], and V. A. Sarkissian[b]

[a]*Institute for High-Pressure Physics, Troitsk, Moscow oblast, 142190 Russia*
[b]*Institute of Crystallography, Russian Academy of Sciences, Moscow, 119333 Russia*
[c]*Kirensky Institute of Physics, Siberian Division, Russian Academy of Sciences, Krasnoyarsk, 660036 Russia*
e-mail: lyubutin@ns.crys.ras.ru

Received July 6, 2004

Abstract—The high-pressure magnetic states of iron borate $^{57}FeBO_3$ single-crystal and powder samples have been investigated in diamond anvil cells by nuclear forward scattering (NFS) of synchrotron radiation at different temperatures. In the low-pressure ($0 < P < 46$ GPa) antiferromagnetic phase, an increase of the Neél temperature from 350 to 595 K induced by pressure was found. At pressures 46–49 GPa, a transition from the antiferromagnetic to a new magnetic state with a weak magnetic moment (magnetic collapse) was discovered. It is attributed to the electronic transition in Fe^{3+} ions from the high-spin $3d^5$ ($S = 5/2$, $^6A_{1g}$) to the low-spin ($S = 1/2$, $^2T_{2g}$) state (spin crossover) due to the insulator–semiconductor-type transition with extensive suppression of strong *d–d* electron correlations. At low temperatures, NFS spectra of the high-pressure phase indicate magnetic correlations in the low-spin system with a magnetic ordering temperature of about 50 K. A tentative magnetic *P–T* phase diagram of $FeBO_3$ is proposed. An important feature of this diagram is the presence of two triple points where magnetic and paramagnetic phases of the high-spin and low-spin states coexist. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Iron borate $FeBO_3$ is a rare magnetic material that is transparent in the visible range and possesses spontaneous magnetization at room temperature. Light modulation by magneto-optical effects is possible in this crystal. The crystal lattice of $FeBO_3$ has rhombohedral symmetry of the calcite type with the space group $R\bar{3}c$ (D_{3d}^6) and with the lattice parameters $a = 4.612$ Å and $c = 14.47$ Å [1, 2]. Iron ions Fe^{3+} are in oxygen octahedra, and the interionic distances are (Fe–O) = 2.028 Å and (Fe–Fe) = 3.601 Å, while the angles of the bonds (O–Fe–O) are 91.82° and 88.18° [2]. Thus, the oxygen surrounding of Fe is almost cubic. At ambient conditions, $FeBO_3$ is an easy-plane antiferromagnet with weak ferromagnetism and with a Neél temperature of about 348 K [3, 4]. Magnetic moments of two iron sublattices and the weak ferromagnetic moment lie in the basal (111) plane [5, 6].

At ambient pressure, iron borate is an insulator with an optical gap value of 2.9 eV [5]. Recently, a drop in the optical absorption edge approximately from 3 to 0.8 eV has been found in optical spectra at pressures near 46 GPa [7]. It was concluded from direct measurements of electroresistivity that a transition of the insulator–semiconductor type occurs at this pressure [7].

In the present paper, iron borate $^{57}FeBO_3$ single crystals and powder samples are studied under high pressures in a diamond-anvil cell by the technique of nuclear forward scattering (NFS) of synchrotron radiation (SR) in the temperature range 3.5–300 K. At pressures of $P = 46$–49 GPa, the sharp transition from the antiferromagnetic to a new magnetic state with a weak magnetic moment was discovered. The pressure dependence of the Neél temperature was calculated from the experimental data, and the magnetic *P–T* phase diagram was plotted and analyzed theoretically.

2. EXPERIMENTAL

The perfect quality light-green colored single crystals of $FeBO_3$ enriched with the ^{57}Fe isotope up to 96% were grown by the flux method. The crystals were plate-shaped, and the plane of the plate was the basal (111) plane. The thickness of the plates was about 10–40 μm with dimensions of about 8×8 mm^2. The NFS experiments were performed with both single crystals and powder samples obtained by grinding an $^{57}FeBO_3$ single crystal.

The experiments with nuclear forward scattering of SR were performed with the $^{57}FeBO_3$ samples at high pressures of up to 65 GPa created in diamond-anvil cells at temperatures in the range 3.5–300 K. The measurements were performed with ID18 nuclear reso-

¶ This article was submitted by the authors in English.

nance scattering equipment [8] at the European Synchrotron Radiation Facility (ESRF), Grenoble, France.

Two types of samples were used: a single crystal at room temperature and a powdered sample (a crushed single crystal) at low temperature. At room temperature, a $^{57}FeBO_3$ single crystal with dimensions of about $80 \times 40 \times 4$ μm^3 was placed into a high-pressure diamond-anvil cell. The diameter of the working surface of diamonds in the cell was about 300 μm, and the diameter of the hole in the rhenium gasket where the sample was placed was about 100 μm. In the low-temperature experiment, the gasket hole was filled with the sample powder to about one-third to ensure that all powder grains were surrounded by pressure liquid. To create quasi-hydrostatic pressure, the working volume of the cell was filled with PES-5 polyethylsilacsanic liquid. A standard technique of the shift of ruby fluorescence was used to measure the pressure value. For that, several crumbled ruby crystals with dimensions of about 5 μm were placed into the cell along with the sample. They were placed at different distances from the center of the working volume in order to evaluate the pressure gradient in the chamber. The accuracy in the pressure measurements was about 3–4 GPa.

In the NFS experiments, the pressure value was varied up to 65 GPa. The basal plane (111) of the $^{57}FeBO_3$ single crystal was oriented perpendicular to the synchrotron radiation beam, and the vector of polarization of gamma rays was in the sample plane. At every pressure value, the NFS spectra of the powdered sample were measured in the temperature range from 3.5 to 300 K. The Mössbauer time spectra of resonance forward scattering from ^{57}Fe nuclei were measured without an external magnetic field at the sample. The measurements were made in the 16-branch regime.

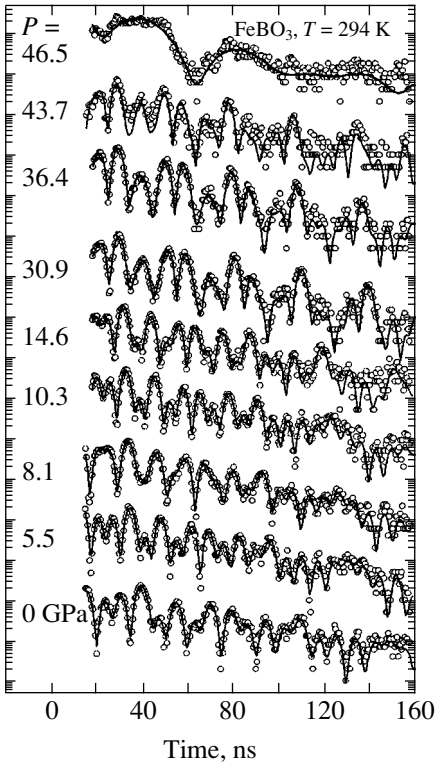


Fig. 1. Room-temperature NFS spectra of the $^{57}FeBO_3$ single crystal at different pressures. Symbols, experimental points; lines, the result of fitting to the MOTIF model.

3. RESULTS AND DISCUSSION

3.1. The Room-Temperature NFS Spectra

Time spectra of the NFS from ^{57}Fe nuclei in $^{57}FeBO_3$ have been recorded at different pressures in the temperature range 3.5–300 K. Figure 1 shows the room-temperature spectra. The spectra represent the intensity of scattered radiation depending on the time following the SR impulse. The damped decay of a nuclear excitation is modulated in time by quantum and dynamic beats. The quantum beats appear due to splitting of nuclear levels by a hyperfine interaction as a result of interference between scattered radiation components of sublevels with different frequencies. The period of quantum beats is inversely proportional to the value of hyperfine splitting energy, and in our case, to the magnetic field value at the iron nuclei. The dynamic beats are due to multiple processes of scattering in a "thick" sample (see details in [9]).

At pressures below 46 GPa, the main feature of the spectra is the evident quantum beats (Fig. 1). Because we used a thin sample, the dynamic beats are not present in the spectra. The NFS spectra were measured with different mutual orientations of the polarization vector of the SR beam and the crystal magnetization. The period of beats is about 8 ns in the case of random orientation of the (111) crystal plane with respect to the direction of the SR-beam polarization and about 15 ns when the crystal is rotated in the basal plane by 90° relative to the first ("random") orientation. The beats with a 15-ns period are 100% modulated, which means that the intensity of scattering in the beats' minimum tends to zero. This indicates that, at all pressures in the range $0 < P < 46$ GPa, the orientation of magnetic fields at the nuclei of iron ions remains in the basal (111) plane of the crystal normal to the radiation beam. At pressures of $P > 46$ GPa, the quantum beats disappear abruptly, showing a drop to zero of the hyperfine magnetic field at ^{57}Fe nuclei.

At ambient pressure, our NFS spectrum is similar to that obtained by Mitsui *et al.* [10] in iron borate. Some distinctions are due to a different thickness of the samples and the absence of an external magnetic field in our measurements.

At $P < 44$ GPa (in the low-pressure (LP) phase of $FeBO_3$), the spectra were processed by the MOTIF program developed by Shvyd'ko [11]. A large number of quantum beats in each spectrum (more than 15) pro-

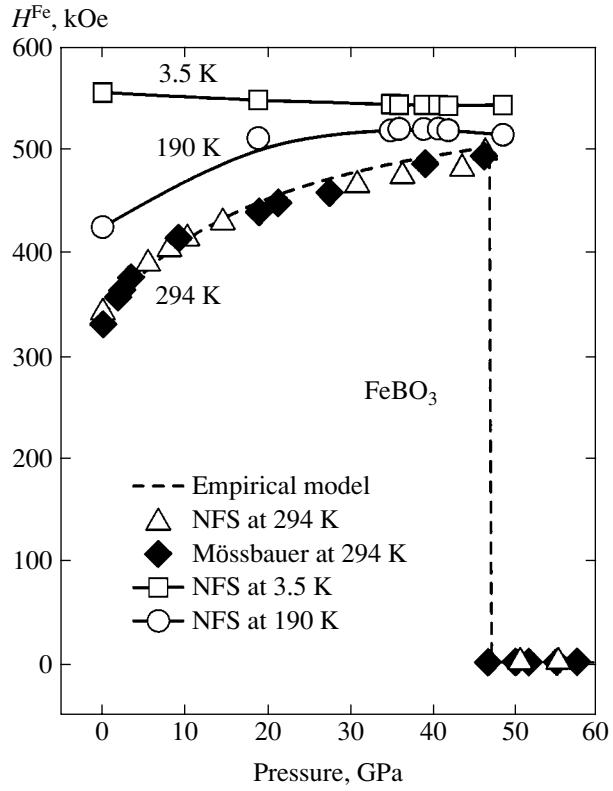


Fig. 2. Pressure dependences of the hyperfine magnetic field H_{hf} at ^{57}Fe nuclei in $FeBO_3$ at different temperatures.

vides high accuracy in determination of the hyperfine magnetic field H_{hf} at iron nuclei (with an error in the range 0.1 T). The H_{hf} values measured at the "random" orientation of the single crystal and after its 90° rotation are the same.

3.2. Magnetic Collapse

The pressure dependences of the hyperfine magnetic field H_{hf} at the iron nuclei are shown in Fig. 2 for different temperatures. At room temperature, the field H_{hf} increases nonlinearly from 34.1 T to its maximum value of 48.1 T as the pressure rises in the range $0 < P <$ 44 GPa. At P = 46–47 GPa, the field H_{hf} drastically falls down to zero, indicating a magnetic-to-nonmagnetic phase transition (magnetic collapse), obviously of the first-order type. At the transition, the parameter of the quadrupole interaction, which is near zero at $P <$ 44 GPa, increases significantly up to 2.1 mm/s.

From the Mössbauer absorption spectra of $^{57}FeBO_3$, we have found that the isomer shift *IS* and quadrupole splitting *QS* of the spectra drastically change at the critical pressure P_c along with the disappearance of the magnetic field H_{hf}(see details in [12]). At $P < 46$ GPa in the low-pressure (LP) phase, the parameters H_{hf}, *IS*, and *QS* are typical of the high-spin (S = 5/2) state of Fe^{3+} ions. At $P > 48$ GPa in the high-pressure (HP) phase of $FeBO_3$, the *IS* and *QS* values become typical of the low-spin state of the Fe^{3+} ions (S = 1/2). No indication of the appearance of Fe^{2+} ions was found in the Mössbauer absorption spectra [12]. Thus, the origin of the magnetic collapse at $P = P_c$ is the high-spin (HS) to low-spin (LS) transition of Fe^{3+} ions. A similar conclusion was obtained theoretically in the multielectron model [13], where it was shown that an increase of the crystal field with pressure results in the high-spin–low-spin crossover and an insulator–semiconductor transition.

The NFS spectra at temperatures 77 and 3.5 K are shown in Fig. 3 for different pressures. At $P > 48$ GPa, the effect of the disappearance of quantum beats is clearly seen in the low-temperature spectra. This indicates that the pressure-induced magnetic collapse is not an effect of the temperature but is due to changes in the electronic structure of iron ions. We also observed that, after this transition, the light-green color of the $^{57}FeBO_3$ crystal, typical of ambient pressure, disappeared and the crystal became opaque, which suggests an abrupt drop in the optical absorption gap. The drop of the optical absorption edge has been found recently in optical spectra at pressures just near 46 GPa [7].

At $P < 46$ GPa, the quantum beats in the NFS spectra of the powder sample cannot be fit perfectly to the calculated curves (Fig. 3) as for the single-crystal sample (Fig. 1). This is because the MOTIF program is not developed enough for powder samples when a distribution of magnetic moment and crystal field directions occurs in powder particles. Nevertheless, the frequencies of beats and, hence, the values of the hyperfine magnetic field H_{hf} at iron nuclei can be obtained with a rather high accuracy (with an error in the range of 0.4 T, which is within the limit of a symbol size in Fig. 2).

The pressure dependences of the field H_{hf} at low temperatures are shown in Fig. 2. Contrary to the room-temperature behavior, the field H_{hf} at T = 3.5 K in the LP phase is almost constant at a saturation value of about 55.5 T. In fact, the value of H_{hf} even decreases slightly as the pressure increases. This effect can be easily explained by an increase of the covalence contribution to H_{hf} due to decreasing interionic Fe–O distances. It was found that the critical pressure value P_c at which the magnetic transition occurs varies slightly with temperature and P_c becomes somewhat larger at helium temperature.

3.3. Pressure Dependence of the Neél Temperature in the Low-Pressure Phase

In the low-pressure phase of $FeBO_3$, the room-temperature NFS spectra show an increase of the field H_{hf} as the pressure increases. The magnetic field increase is naturally connected with an increase of the exchange interaction, which, in turn, must correlate with the increase in the Neél temperature T_N. In general, the

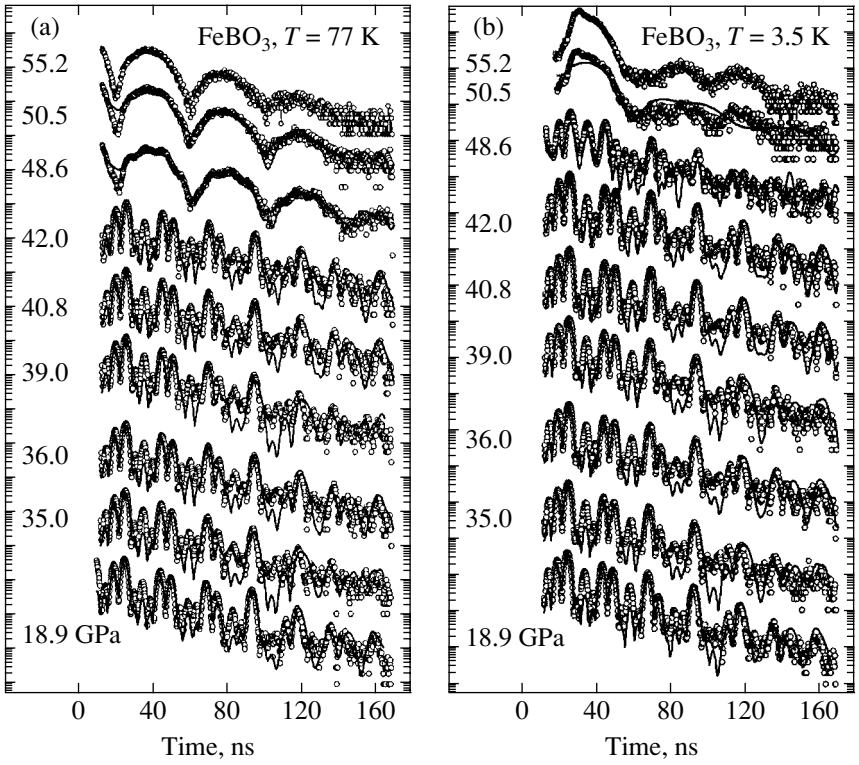


Fig. 3. NFS spectra of the $^{57}FeBO_3$ powder sample at different pressures: (a) $T = 77$ K, (b) $T = 3.5$ K. Symbols, experimental points; lines, the result of fitting to the MOTIF model.

pressure dependence of H_{hf} at room temperature is under the influence of two effects: the changes in T_N and a possible change of the saturation value of H_{hf} at 0 K [$H_{hf}(0)$]. The NFS spectra in the LP phase indicate that, at 3.5 K, the $H_{hf}(0)$ value only slightly depends on pressure. Then, starting with the room-temperature $H_{hf} = f(P)$ dependence and using the ambient-pressure $H_{hf} = F(T)$ dependence (which has been studied in detail by Eibschuts and Lines [6]), we can calculate the dependence of T_N on pressure. For that, we used an extrapolation procedure first suggested in [14, 15] and successfully applied to many experimental results.

We take $H_{hf}(P, T)$ as the empirical function

$$H_{hf}(P,T) = H_{hf}(P,0) \times \exp\left(-\alpha\frac{T}{T_N(P)}\right)\left[1-\frac{T}{T_N(P)}\right]^{\beta}. \quad (1)$$

The parameters α and β can be found from the fit of (1) to the experimental dependence $H_{hf}(T)$ at ambient pressure. Using the $H_{hf}(T)$ values for $FeBO_3$ in [6], we found $\alpha = -0.371$ and $\beta = 0.4308$. Then, we assume that α and β are independent of the pressure and take the experimental value $H_{hf}(P, 0) = 55.5$ T. For each experimental value of pressure and the corresponding values of $H_{hf}(P, T)$, Eq. (1) can be solved graphically for T_N. Figure 4 illustrates the calculation procedure, and the obtained pressure dependence of $T_N$is shown in Fig. 5.

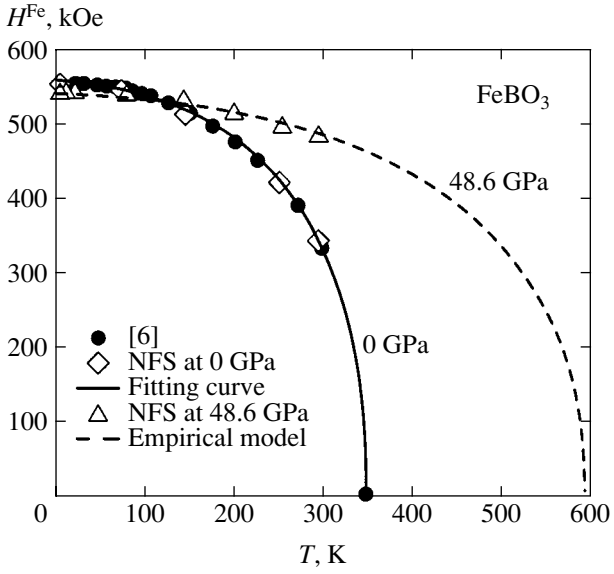


Fig. 4. Procedure for calculating the Neél temperature of $FeBO_3$ at different pressures by fitting the empirical function to the experimental temperature dependence of the hyperfine magnetic field at ^{57}Fe nuclei.

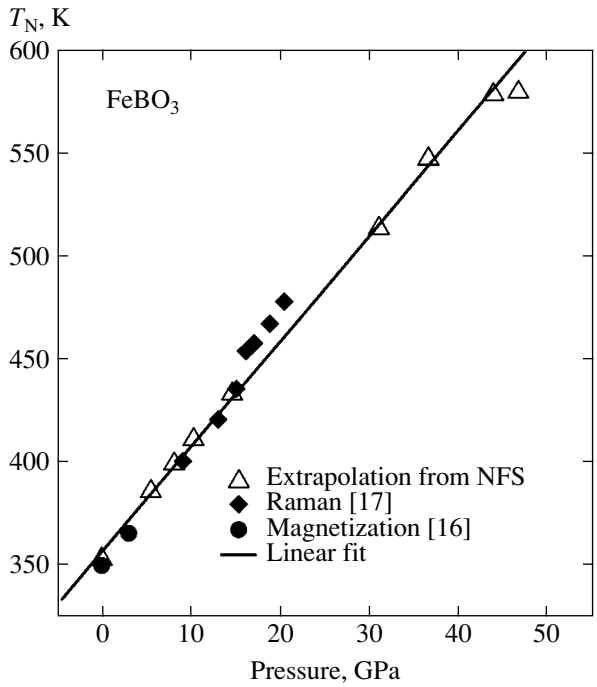


Fig. 5. Pressure dependence of the Neél temperature of $FeBO_3$ from NFS, Raman scattering, and magnetization experiments (symbols). The solid curve corresponds to the pressure slope dT_N/dP = 5.14 K/GPa.

In the pressure range $0 < P < 46$ GPa, the dependence $T_N(P)$ can be well fit to a linear function $T_N(P) = T_N(0) + P(dT_N/dP)$, with the parameters $T_N(0) = 355.0 \pm 1.5$ K and $dT_N/dP = 5.14 \pm 0.10$ K/GPa. The arrow value follows from the least-square fit. The maximum value of T_N, attained just before the magnetic collapse, is about 595 K.

From the magnetization measurements of $FeBO_3$ in the range of 0–3 kbar, Wilson and Broersma [16] have found that T_N grows linearly with the slope dT_N/dP = 5.3 K/GPa, which is close to our value. Massey *et al.* [17] measured the shift of the two-magnon Raman frequency Ω with the pressure increase in $FeBO_3$ at 99 K. In the range 0–13 GPa, the frequency shift can be approximated by a linear law $\Omega(P) = \Omega(0) + d\Omega/dP$, where $\Omega(0) = 530 \pm 20$ cm^{-1} and $d\Omega/dP = 8.15 \pm 0.7$ cm^{-1}/GPa.

It is interesting to compare the pressure behavior of Ω and T_N. We found that, below 13 GPa, the relative slopes of Ω and T_N are very close: $[1/\Omega(0)]d\Omega/dP = 0.0148$ GPa^{-1} and $[1/T_N(0)]dT_N/dP = 0.0150$ GPa^{-1}. This means that both these parameters are most probably proportional to the superexchange integral J. The $T_N(P)$ data obtained from the magnetization and Raman measurements are also shown in Fig. 5, and they are in good agreement with our studies.

3.4. Magnetic Properties of the High-Pressure Phase

After the magnetic transition, at pressures $P > P_c$, the low-spin state of Fe^{3+} (S = 1/2) is not diamagnetic, and one can expect some kind of magnetic correlations at low temperatures. For the HP phase of $FeBO_3$, the recent theoretical calculations of Parlinski [18] predicted a small magnetic moment at iron ions, which is about four times lower than that in the LP phase.

Figure 6 shows our NFS spectra of the powder sample of $^{57}FeBO_3$ taken at different temperatures with fixed pressures in the HP state (at $P > P_c$). The spectra in Fig. 6c above 50 K are typical of a pure quadrupole interaction without any trace of magnetic modulations. However, at low temperatures, an anomaly appears in the spectra, which cannot be fit to the quadrupole interaction. We tried to fit the NFS spectra at $T < 50$ K with different approximations and have found that the most appropriate is the model of magnetic correlations of Fe^{3+} ions with spin S = 1/2, and the magnetic state can be represented as a nonhomogeneous magnetic ordering with a distribution of the H_{hf} field values.

The theoretical calculations in [18] predicted a homogeneous antiferromagnetic ordering at low temperatures for the low-spin HP phase of $FeBO_3$. A nonhomogeneous magnetic state (the low-spin magnetic ordering, LS-MO) found in our experiment may be related to a powder state of the sample due to specific magnetic properties of small particles of $FeBO_3$ at high pressures.

To find the precise temperature of the magnetic ordering in the HP phase, the following procedure was suggested. We fit all spectra in Fig. 6 to the model of pure quadrupole interaction and plot the obtained quadrupole splitting parameter QS as a function of temperature (see Fig. 7). In the pure paramagnetic state, at $T >$ 50 K, the QS value is constant. When magnetic correlation appears, the "QS value" found this way starts to increase (Fig. 7), showing a deviation from the model. The point of deviation of QS from the constant value is then taken as the magnetic ordering temperature T_m of the low-spin HP phase.

3.5. Magnetic P–T Phase Diagram of $FeBO_3$

On the basis of all the data obtained, we can plot a tentative magnetic P–T phase diagram (Fig. 8), which shows various magnetic states of $FeBO_3$ at different pressures and temperatures. The almost vertical line at $P = P_c$ separates the left-hand HS insulating low-pressure phase and the right-hand LS semiconducting high-pressure phase. At $P < 46$ GPa, in the high-spin low-pressure phase, the T_N line separates the $T < T_N$ antiferromagnetic (AF) state and the $T > T_N$ paramagnetic (PM) state. At $P > 49$ GPa, in the high-pressure phase,

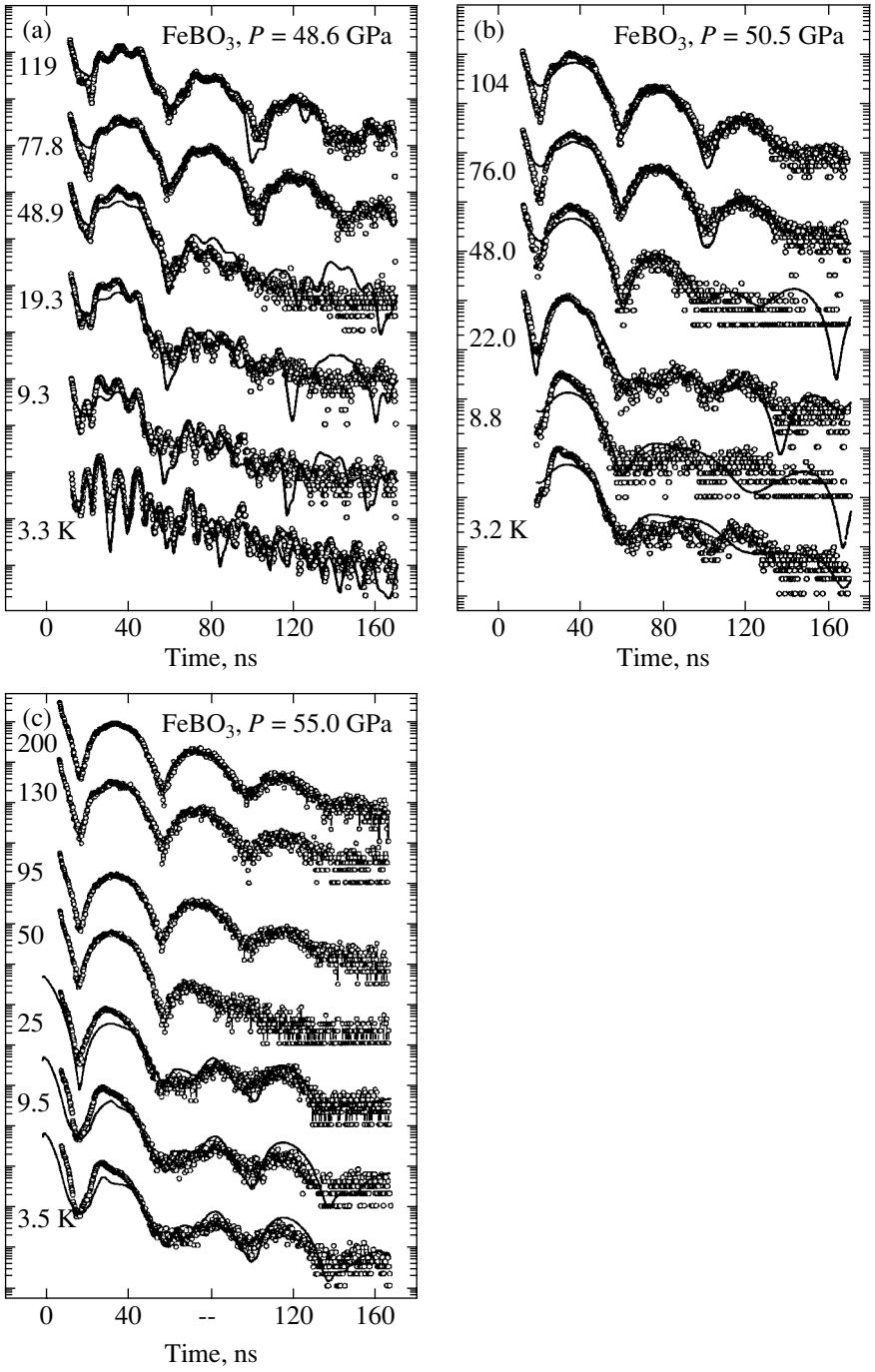


Fig. 6. NFS spectra of a powder $^{57}FeBO_3$ sample for different temperatures with the pressure values fixed at 48.6, 50.5, and 55 GPa. In the HP phase at 55 GPa, the spectra at 3.5, 9.5, and 25 K were fitted to the model of nonhomogeneous magnetic ordering of Fe^{3+} ions with spin $S = 1/2$, and with a distribution of the H_{hf} field values.

the T_m line separates the $T < T_m$ low-spin magnetically ordered state and the $T > T_m$ paramagnetic low-spin state.

An important conclusion follows from the diagram: one can expect two triple points with the coordinates ($P = 46$ GPa, $T = 600$ K) and ($P = 49$ GPa, $T = 50$ K), where three phases coexist. At the first point, the high-spin antiferromagnetic (HS-AF) and high-spin paramagnetic (HS-PM) phases coexist with the low-spin paramagnetic (LS-PM) phase. At the second point, the low-spin magnetically-ordered (LS-MO) and low-spin paramagnetic (LS-PM) phases coexist with the high-spin antiferromagnetic (HS-AF) phase.

4. THEORETICAL APPROACH

Because the crystal is a semiconductor in the HP phase of $FeBO_3$, its magnetic properties can be described in an approximation of the Heisenberg model both below and above the critical pressure P_c In the mean-field approximation,

$$T_N = JzS(S+1)/3, \tag{2}$$

where $z = 6$ is the number of nearest-neighbors of Fe^{3+},

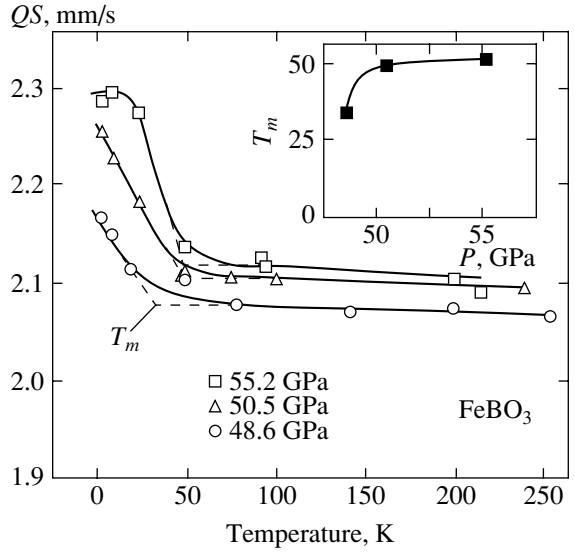


Fig. 7. Temperature dependences of the quadrupole splitting parameter at different pressures in the high-pressure phase of $FeBO_3$ from the fit of the NFS data to the pure quadrupole-interaction model. The point T_m of the deviation from the straight line corresponds to the onset of magnetic ordering of the low-spin HP phase.

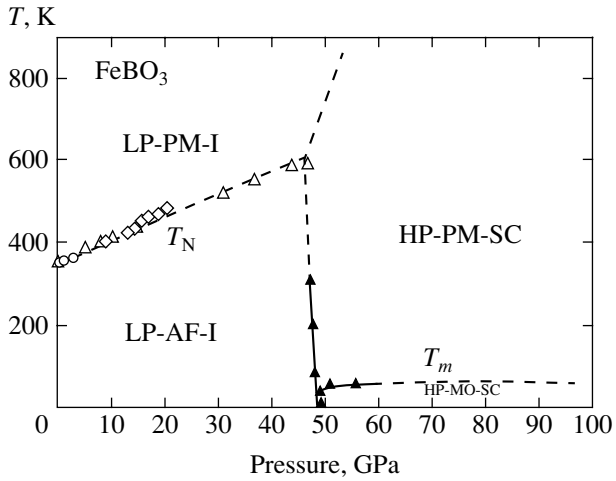


Fig. 8. Tentative magnetic P–T phase diagram of $FeBO_3$. LP-AF-I is the low-pressure antiferromagnetic insulating phase, LP-PM-I is the low-pressure paramagnetic insulating phase, HP-MO-SC is the high-pressure magnetically ordered low-spin semiconducting phase, HP-PM-SC is the high-pressure paramagnetic semiconducting phase. We note the existence of two triple points where three different phases coexist.

the spin $S = 5/2$ characterizes the LP phase, and $S = 1/2$ is the iron spin in the HP phase. The pressure-dependent exchange integral is $J = 2t^2/U_{\text{eff}}$, where t is the parameter of the electron jump between nearest Fe^{3+} ions governing the half-width of the d-band, $W_d = zt$, and $U_{\text{eff}} = \Omega_c - \Omega_v$ is the effective Hubbard parameter representing the gap between the upper Ω_c (conductivity) and lower Ω_v (valence) Hubbard bands. Here, the upper Hubbard band is the extra electron band due to $d^5 \longrightarrow d^6$ excitations and the lower Hubbard band is the electron removal band due to $d^5 \longrightarrow d^4$ excitations [19]. In the LP phase, both t and U_{eff} parameters depend on pressure as

$$\begin{aligned} t(P) &= t_0 + \alpha_t P, \\ U_{\text{eff}}(P) &= U_0 - \alpha_\Delta P, \end{aligned} \tag{3}$$

where $t_0 = 0.076$ eV and $U_0 = 4.2$ eV are the ambient pressure parameters [19]. The value of the pressure derivative of the crystal field Δ, $\alpha_\Delta = d\Delta/dP = 0.018$ eV/GPa, is found from the condition of crossover of the high-spin $^6A_1$ and low-spin $^2T_2$ terms at $P = P_c$, and the $\alpha_t = dt/dP = 0.00046$ eV/GPa value is found from the rise of T_N from 350 up to 600 K in the LP phase. These values of the derivatives ensure the increase of T_N in the LP phase and the collapse of the Fe^{3+} magnetic moment at P_c.

We now consider the change in magnetic properties of $FeBO_3$ under transition into the HP phase. Near P_c, a structural transition occurs with a jump of unit-cell parameters [20], and, therefore, a jump in the t and U_{eff} values can be expected.

We use "+" to denote the values of parameters on the right-hand side of P_c and "–" to denote those on the left-hand side. Then,

$$\begin{aligned} t_c^{(+)} &= t_0 + \alpha_t P_c + \delta t, \\ U_{\text{eff}}^{(+)} &= U_0 - \alpha_\Delta P_c - \delta U. \end{aligned} \tag{4}$$

Because the a- and c-unit-cell parameters decrease at the transition [20], the δt and δU values must be positive. Assuming $\delta t/t_0 \ll 1$ and $\delta U/U_0 \ll 1$, we write the exchange integral just after the transition as

$$J_c^{(+)} = J_0\left[1 + \left(\frac{2\alpha_t}{t_0} + \frac{\alpha_\Delta}{U_0}\right)P_c + 2\frac{\delta t}{t_0} + \frac{\delta U}{U_0}\right], \tag{5}$$

where

$$J_0 = \frac{2t_0^2}{U_0}.$$

The ratio of T_N above and below the transition is then given by

$$\frac{T_N^{(+)}}{T_N^{(-)}} = \frac{J^{(+)} \cdot 1/2 \cdot 3/2}{J^{(-)} \cdot 5/2 \cdot 7/2} = \frac{3}{35}\left(1 + \frac{2\delta t}{t_0} + \frac{\delta U}{U_0}\right), \quad (6)$$

where

$$T_N^{(-)} = 600 \text{ K}.$$

If the jumps in δt and δU are negligible and the change in T_N is only connected with the spin jump $5/2 \longrightarrow 1/2$, we can evaluate the magnetic ordering temperature of the HP phase as

$$T_N^{(+)} = 3T_N^{(-)}/35 = 51 \text{ K}. \quad (7)$$

Taking δt and δU into account could only increase the $T_N^{(+)}$ value. Thus, (7) is an estimate from below, that is, $T_N^{(+)} \geq 51$ K. It turns out that the experimental value of $T_N^{(+)}$ evaluated in Section 3.4 is about 50 K. This suggests that the δt and δU values are negligibly small.

In the HP phase, U_{eff} depends only on the electron transfer and does not depend on the crystal field and pressure [21], and therefore the pressure dependence of T_N is different from that in the LP phase,

$$T_N(P)/T_N^{(+)} = 1 + 2\alpha_t(P - P_c)/t_0. \quad (8)$$

The slope of $T_N(P)$ in the LP phase

$$\frac{dT_N(P)/T_N(0)}{dP} = \frac{2\alpha_t}{t_0} + \frac{\alpha_\Delta}{U_0} = 0.016 \frac{1}{\text{GPa}} \quad (9)$$

is different from that in the HP phase

$$\frac{dT_N(P)/T_N^{(+)}(P_c)}{dP} = \frac{2\alpha_t}{t_0} = 0.012 \frac{1}{\text{GPa}}. \quad (10)$$

Thus, the slope ratio is 4/3.

Now, the question is: How far is expression (8) valid as the pressure increases further, and what happens above T_N? In the HP phase at $P > P_c$ and $T > T_N$, the $FeBO_3$ crystal is a paramagnetic semiconductor with iron ions Fe^{3+} in the low-spin state ($S = 1/2$). However, with a further pressure increase, the semiconducting gap decreases, and it tends to zero at $P = P_M$, where P_M is the point of the transition into the metallic state. The experimental value of P_M evaluated from the thermoactivation gap is approximately 210 GPa [7].

The same value of P_M was found theoretically [21] in extrapolating the level Ω_c down to the crossing with the top of the valence band ε_v. The corresponding diagram of the electron structure below and above P_M is shown in Fig. 9. Here, the Ω_c level is related to the transitions $[Fe^{(3+)}(d^5), S = 1/2] \longleftrightarrow [Fe^{(2+)}(d6), S = 0]$, and it is smeared into a narrow band due to electron hopping (Fig. 9a). The spin–polaron effect in the antiferromagnetic phase gives rise to a sharp suppression of the *d*-band width [22].

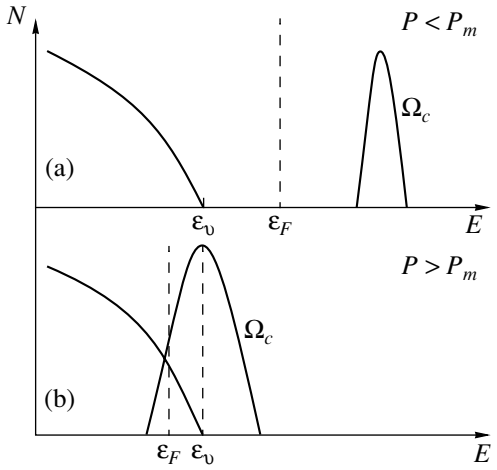


Fig. 9. Diagram of the density of states of $FeBO_3$ at high pressures below (a) and above (b) the semiconductor–metal transition.

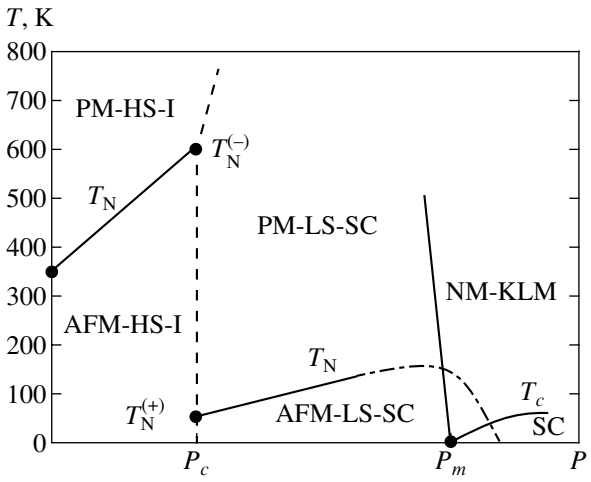


Fig. 10. Supposed phase diagram of $FeBO_3$ at high pressures. AFM, antiferromagnet; PM, paramagnet; HS-I, Fe^{3+} high-spin insulator; LS-SC, low-spin semiconductor; NM-KLM, nonmagnetic Kondo-lattice metal; SC, superconductor.

Above P_M, there are two types of carriers: oxygen holes at the top of the valence band and heavy electrons at the bottom of the *d*-band. The iron ion is in an intermediate valence state as a mixture of the p^6d^5 and p^5d^6 configurations (Fig. 9b). Because each hole at oxygen gives rise to the $S = 0$ state of the d^6 configuration in iron, one may consider the situation as a peculiar

Kondo effect, when carriers screen the iron spin. Because spins are located regularly, the system can be attributed to a Kondo lattice. At higher temperatures, one can expect a nonmagnetic Kondo metal state; at low temperatures, a competition between antiferromagnetism and superconductivity induced by spin fluctuations can be dominant [23, 24]. The corresponding phase diagram is shown in Fig. 10. Only the magnetic and electron properties of $FeBO_3$ are shown in the diagram, and the structural transitions are not discussed here. We note that, from the standpoint of modern terminology, the P_M point in the diagram of Fig. 10 is a typical quantum critical point.

5. CONCLUSIONS

Both experimentally and theoretically, we have shown that the magnetic collapse in $FeBO_3$ at high pressure does not transform the material into a nonmagnetic state with the disappearance of magnetic properties. At the transition, the low-pressure phase with a strong magnetic interaction transforms into the high-pressure phase with a weak magnetic interaction, and this transformation is accompanied by an insulator–semiconductor transition. The forthcoming metallization and unusual properties of the Kondo lattice metal state are subjects for future experimental study.

ACKNOWLEDGMENTS

We are grateful to R. Rüffer (ESRF, Grenoble, France) for providing the facility for the NFS studies and to O. Leupold, E. Schreier, and A. Chumakov for help in the NFS experiment and very fruitful discussions. The studies were made under the NATO grant no. PST.CLG.976560.

We thank E.A. Popov for processing the low-temperature NFS spectra by his program. This work was supported by the Russian Foundation for Basic Research (project nos. 02-02-17364a, 03-02-16286a, 04-02-16945a, and 05-02-16142a) and by a program of the Russian Academy of Sciences under the project "Strong Correlating Electron Systems."

REFERENCES

1. Database PDF-II, record 21-0423; I. Bernal, C. W. Struck, and J. G. White, Acta Crystallogr. **16**, 849 (1963).
2. R. Diehl, Solid State Commun. **17**, 743 (1975).
3. R. Wolfe, A. J. Kurtzig, and R. C. LeCraw, J. Appl. Phys. **41**, 1218 (1970).
4. M. Eibschuts, L. Pfeiffer, and J. W. Nielsen, J. Appl. Phys. **41**, 1276 (1970).
5. M. P. Petrov, G. A. Smolenskiĭ, A. P. Pagurt, *et al.*, Fiz. Tverd. Tela (Leningrad) **14**, 109 (1972) [Sov. Phys. Solid State **14**, 87 (1972)].
6. M. Eibschuts and M. E. Lines, Phys. Rev. B **7**, 4907 (1973).
7. I. A. Trojan, M. I. Eremets, A. G. Gavriliuk, *et al.*, JETP Lett. **78**, 13 (2003).
8. R. Ruffer and A. I. Chumakov, Hyperfine Interact. **97**/**98**, 589 (1996).
9. G. V. Smirnov, Hyperfine Interact. **123**/**124**, 31 (1999).
10. T. Mitsui, S. Kitao, M. Seto, *et al.*, J. Phys. Soc. Jpn. **68**, 1049 (1999).
11. Yu. V. Shvyd'ko, Phys. Rev. B **59**, 9132 (1999).
12. V. A. Sarkissyan, I. A. Trojan, I. S. Lyubutin, *et al.*, JETP Lett. **76**, 664 (2002).
13. S. G. Ovchinnikov, JETP Lett. **77**, 676 (2003).
14. A. G. Gavriliuk, G. N. Stepanov, V. A. Sidorov, and S. M. Irkaev, J. Appl. Phys. **79**, 2609 (1996).
15. A. G. Gavriliuk, G. N. Stepanov, I. S. Lyubutin, *et al.*, JETP **90**, 330 (2000).
16. D. M. Wilson and S. Broersma, Phys. Rev. B **14**, 1977 (1976).
17. M. J. Massey, R. Merlin, and S. M. Girvin, Phys. Rev. Lett. **69**, 2299 (1992).
18. K. Parlinski, Eur. Phys. J. B **27**, 283 (2002).
19. S. G. Ovchinnikov and V. N. Zabluda, JETP **98**, 135 (2004).
20. A. G. Gavriliuk, I. A. Trojan, R. Boehler, *et al.*, JETP Lett. **75**, 23 (2002).
21. A. G. Gavriliuk, I. A. Trojan, S. G. Ovchinnikov, *et al.*, JETP **99**, 566 (2004).
22. É. A. Nagaev, *Physics of Magnetic Semiconductors* (Nauka, Moscow, 1979) [in Russian].
23. P. Schlottmann, Phys. Rep. **181**, 1 (1989).
24. P. Fulde, *Electron Correlations in Molecules and Solids* (Springer, Berlin, 1991), Springer Ser. Solid-State Sci., Vol. 100.

Journal of Experimental and Theoretical Physics, Vol. 100, No. 4, 2005, pp. 697–702.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 4, 2005, pp. 790–795.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Simple Theory of Extremely Overdoped HTS¶

A. F. Andreev

Kapitza Institute for Physical Problems, Russian Academy of Sciences, Moscow, 119334 Russia
Low Temperature Laboratory, Helsinki University of Technology FIN-02015 HUT, Finland
e-mail: andreev@kapitza.ras.ru

Received November 15, 2004

Abstract—We demonstrate the existence of a simple physical picture of superconductivity for extremely overdoped CuO_2 planes. It has all the characteristic features of HTS, such as a high superconducting transition temperature, the $d_{x^2-y^2}$ symmetry of the order parameter, and the coexistence of a single-electron Fermi surface and a pseudogap in the normal state. The values of the pseudogap are calculated for different doping levels. Orbital paramagnetism of preformed pairs is predicted. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

In this work (also see earlier Letter [1]), we demonstrate that in the phase diagram of cuprate high-temperature superconductors (HTS), a small region exists where the characteristic features of HTS can be easily understood on the base of a simple theory. These characteristic features include a high superconducting transition temperature, the $d_{x^2-y^2}$ symmetry of the order parameter (see [2]), and the coexistence of a single-electron Fermi surface and a pseudogap in the normal state [3]. The last phenomenon is usually attributed to the presence of preformed (i.e., normal-state) electron pairs (in particular, bipolarons [4–8]).

The aforementioned small region in the phase diagram is situated in the vicinity of the maximal hole-doping level $x = x_c$ compatible with superconductivity. The superconducting transition temperature T_c is zero for $x \geq x_c$, and hence it is low in our region near $x = x_c$. However, T_c increases with decreasing x for $x < x_c$ such that it is quite high at the boundary of the region (i.e., for $x_c - x \sim 1$).

Two features of our small region are important to make a simple physical picture possible. These are relatively low T_c and the clear nature of the normal state as a mostly conventional Fermi liquid.

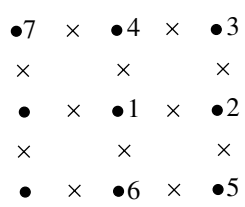
We calculate the pseudogap. With increasing x, the pseudogap decreases for $x < x_c$. As well as T_c, the pseudogap disappears at $x = x_c$. However, it reappears for larger doping levels $x > x_c$.

As a new prediction, we show the existence of an unusual orbital paramagnetism of the preformed (singlet) pairs, which can probably be experimentally separated from the Pauli spin paramagnetism of single electrons and the Landau diamagnetism of single electrons and pairs.

2. PAIR QUASIPARTICLES

The key point is the existence of very mobile pair quasiparticles in crystals under the tight-binding conditions, i.e., if the energy of the electron–electron interaction at a distance on the order of the atomic spacing considerably exceeds the electron tunneling amplitude to neighboring lattice cites. Quasiparticles of this type were studied earlier [9] in helium quantum crystals and more recently by Alexandrov and Kornilovitch [7] as a model of bipolarons in HTS (also see [10]).

We consider two electrons localized at neighboring (1 and 2 in the figure) copper atoms (to be more precise, in unit cells containing these atoms) forming a square lattice in the CuO_2 plane. The electron tunneling from 2 to 4 or 6 does not change the energy of the system in view of the crystal lattice symmetry. The same is true for the electron tunneling from 1 to 3 or 5. Owing to this type of transitions, an electron pair can move as a whole over the entire plane, because the $2 \longrightarrow 4$ transition can be followed by the transition $1 \longrightarrow 7$ or $1 \longrightarrow 3$, and so on. Because the transitions do not change the energy of the system, the motion is fully coherent. An electron pair behaves as a delocalized Bose quasiparticle.



CuO_2 plane: ●, Cu atoms and ×, O atoms.

¶ This article was submitted by the author in English.

To calculate the quasiparticle spectrum, we consider the localized states of a pair,

$$|\mathbf{r}, \mathbf{r}', \alpha\beta\rangle = c^+_{\mathbf{r}\alpha}c^+_{\mathbf{r}'\beta}|0\rangle, \quad (1)$$

where $c^+_{\mathbf{r}\alpha}$ are the electron creation operators with a spin projection $\alpha = \uparrow, \downarrow$ at a point $\mathbf{r}$ and $|0\rangle$ is the electron vacuum.

The effective tunneling Hamiltonian H_{eff} is defined by the matrix elements of the operator

$$H = t\sum_{\mathbf{r}\mathbf{r}'\alpha} c^+_{\mathbf{r}'\alpha}c_{\mathbf{r}\alpha}, \quad (2)$$

which correspond to the transitions of one of the electrons to copper atoms that are next-to-nearest neighbors of the initial atom, such that the energy of the system of two electrons remains unchanged. Here, t is the tunneling amplitude, which is known to be positive (see [2, p. 1004]).

Let $\mathbf{a}_n$ ($n = x, y$) be the square-lattice periods directed from point 1 to point 2 and from point 1 to point 4, respectively. We have

$$\begin{aligned} H_{\text{eff}}|\mathbf{r}, \mathbf{r}+\mathbf{a}_x, \alpha\beta\rangle &= t(|\mathbf{r}+\mathbf{a}_x+\mathbf{a}_y, \mathbf{r}+\mathbf{a}_x, \alpha\beta\rangle \\ &+ |\mathbf{r}+\mathbf{a}_x-\mathbf{a}_y, \mathbf{r}+\mathbf{a}_x, \alpha\beta\rangle + |\mathbf{r}, \mathbf{r}+\mathbf{a}_y, \alpha\beta\rangle \\ &+ |\mathbf{r}, \mathbf{r}-\mathbf{a}_y, \alpha\beta\rangle) = t(-|\mathbf{r}+\mathbf{a}_x, \mathbf{r}+\mathbf{a}_x+\mathbf{a}_y, \beta\alpha\rangle \\ &+ |\mathbf{r}+\mathbf{a}_x-\mathbf{a}_y, \mathbf{r}+\mathbf{a}_x, \alpha\beta\rangle \\ &+ |\mathbf{r}, \mathbf{r}+\mathbf{a}_y, \alpha\beta\rangle - |\mathbf{r}-\mathbf{a}_y, \mathbf{r}, \beta\alpha\rangle), \end{aligned} \quad (3)$$

where we used the antisymmetry of quantities (1) with respect to the arguments ($\mathbf{r}$, α) and ($\mathbf{r}'$, β). Analogously,

$$\begin{aligned} H_{\text{eff}}|\mathbf{r}, \mathbf{r}+\mathbf{a}_y, \alpha\beta\rangle &= t(-|\mathbf{r}+\mathbf{a}_y, \mathbf{r}+\mathbf{a}_x+\mathbf{a}_y, \beta\alpha\rangle \\ &+ |\mathbf{r}-\mathbf{a}_x+\mathbf{a}_y, \mathbf{r}+\mathbf{a}_y, \alpha\beta\rangle \\ &+ |\mathbf{r}, \mathbf{r}+\mathbf{a}_x, \alpha\beta\rangle - |\mathbf{r}-\mathbf{a}_x, \mathbf{r}, \beta\alpha\rangle). \end{aligned} \quad (4)$$

The complete set of localized states of an electron pair is determined by the state vectors

$$|\mathbf{r}, n, \alpha\beta\rangle \equiv |\mathbf{r}, \mathbf{r}+\mathbf{a}_n, \alpha\beta\rangle, \quad (5)$$

where $\mathbf{r}$ labels unit cells of the square lattice.

The problem obviously splits into two independent problems for singlet and triplet pairs that are characterized by quantities (5), which are respectively antisymmetric and symmetric in the spin indices α, β. Assuming that the required stationary states of a pair are superpositions of localized states,

$$\sum_{\mathbf{r}, n} \psi^{(n)}_{\alpha\beta} e^{i\mathbf{k}\mathbf{r}}|\mathbf{r}, n, \alpha\beta\rangle \quad (6)$$

with coefficients $\psi^{(n)}_{\alpha\beta}$ independent of $\mathbf{r}$ (this corresponds to a definite quasimomentum $\mathbf{k}$), we obtain

$$\begin{aligned} (E(\mathbf{k})-\epsilon_0)\psi^{(x)} &= t\psi^{(y)}(1 \pm e^{-ik_x})(1 \pm e^{i\kappa_y}), \\ (E(\mathbf{k})-\epsilon_0)\psi^{(y)} &= t\psi^{(x)}(1 \pm e^{ik_x})(1 \pm e^{-i\kappa_y}), \end{aligned} \quad (7)$$

where the upper or lower sign corresponds to a singlet or triplet state, respectively. The conditions for the existence of a nontrivial solution $\psi^{(x)}$, $\psi^{(y)}$ of system (7) determine the energy $E(\mathbf{k})$ of a pair quasiparticle. Here, ϵ_0 is the energy of the initial localized state; $\kappa_x = \mathbf{k}\cdot\mathbf{a}_x$ and $\kappa_y = \mathbf{k}\cdot\mathbf{a}_y$. Everywhere in formulas (7), we omit identical spin indices $\alpha\beta$.

The minimal energy $\epsilon_m = \min E(\mathbf{k}) = \epsilon_0 - 4t$ of a singlet pair is attained at $\kappa_x = \kappa_y = 0$. The same minimal energy of a triplet pair is attained at the nonzero quasimomentum $\kappa_x = \kappa_y = \pi$. This degeneracy is removed by taking the electron exchange in the initial localized pair into account. It is well known that this exchange is of an antiferromagnetic nature, and hence singlet pairs have the minimal energy.

Thus, solitary Bose quasiparticles can exist in the CuO_2 plane; these particles are characterized by a doubled electric charge and by zero momentum and spin in the ground state. It readily can be seen from Eqs. (7) that the effective mass of quasiparticles is $m = \hbar^2/ta^2$, where $a = |\mathbf{a}_x| = |\mathbf{a}_y|$. In addition, quasiparticles have a specific quantum number $n = x, y$, which determines the orientation of a two-electron "dumb-bell." Substituting $E(\mathbf{k}) = \epsilon_m$ and $\mathbf{k} = 0$ in Eqs. (7), we obtain $\psi^{(x)} = -\psi^{(y)}$ in the ground state. Because the orientations $n = x$ and $n = y$ are transformed into each other under lattice rotation through the angle $\pi/2$ and under reflection in the diagonal plane passing through points 1 and 3 in the figure, the ground-state wavefunction $\psi \equiv \psi^{(x)} = -\psi^{(y)}$ of quasiparticles transforms in accordance with the nontrivial $1D$ representation (usually denoted by $d_{x^2-y^2}$) of the symmetry group of the CuO_2 plane (see [2]).

3. SUPERCONDUCTIVITY

We further assume that all other two-electron, three-electron, etc., configurations localized at distances on the order of the atomic spacing are energetically disadvantageous compared to the pair configuration considered above. In addition, we assume that electrons are repulsed at large distances such that the electron–electron interaction energy is on the order of the one-electron tunneling amplitude. Under these conditions, only

single-electron Fermi particles and the pair Bose particles considered above play a significant role.

Finally, we assume that the minimal energy ϵ_m of pair quasiparticles is such that $\epsilon_m/2$ is within the single-electron energy band. We note the following. Under tight-binding conditions, there are two different situations in which $\epsilon_m/2$ can be within the single-electron energy band. First, if single electrons and electrons in pairs correspond to the same energy band, the single-electron tunneling amplitude must be on the order of the electron–electron interaction energy in the pairs, while the one-electron tunneling amplitude t in pairs introduced in Section 2 must be much smaller than the interaction. The latter condition, which is the condition of the applicability of the procedure used in Section 2, can be a result of the large polaron effect in pairs. Second, if single electrons and electrons in pairs correspond to different bands, both one-electron amplitudes can be of the same order. The analysis carried out by Alexandrov and Kornilovitch in [7] shows that the conditions formulated above are likely to be realistic.

We now trace the change of the state of the system at $T = 0$ as the number of electrons increases (the hole-doping level decreases). Until $\epsilon_m/2 > \epsilon_F$, only single-electron quasiparticles are present and the system behaves as an ordinary Fermi liquid. The condition $\epsilon_m/2 = \epsilon_F$ determines the minimal hole-doping level compatible with the state of a normal Fermi liquid. Let n_c denote the corresponding electron density n. Upon a further decrease in the hole-doping level, all additional $n - n_c$ electrons pass into a Bose–Einstein (BE) condensate of pair quasiparticles (we everywhere consider the case of small $n - n_c$ values, for which the concentration of pairs is low and their interaction can be disregarded). The system becomes a superconductor. The superconducting order parameter is given by the boson ground state wavefunction $\psi \equiv \psi^{(x)}$ normalized by the condition $|\psi|^2 = (n - n_c)/2$; the wavefunction transforms in accordance with the $d_{x^2-y^2}$ representation of the symmetry group of the CuO_2 plane.

It is important to note the following. In the system ground state (i.e., for complete filling of all fermion states with energies smaller than ϵ_F), the uncertainty in the energy of a boson quasiparticle with a low excitation energy $\epsilon = k^2/2m$ arising due to its collisions with single-electron Landau quasiparticles is proportional to ϵ^2. As in the conventional theory of a Fermi liquid, this is, first, because of a low density of fermions in an order-ϵ neighborhood of ϵ_F, with which the given boson can collide due to energy conservation. Second, the statistical weight of the final states to which fermionic transitions are possible is small. The probability of boson decay into two fermions per unit of time is also small: as suggested at the beginning of this section, the boson must overcome a significant energy barrier. Thus, the proposed picture of superconductivity in the vicinity of the maximal doping level remains valid even in the region of appreciable densities of fermions, where the interaction between bosons and fermions is significant. The critical electron density n_c is determined from the condition that the electron chemical potential is equal to half the minimal boson energy. In the general case, this energy is a functional of the distribution function for single-electron Landau quasiparticles.

In calculating the superconducting transition temperature, the fermion distribution function may be considered as corresponding to $T = 0$, because the temperature corrections (proportional to T^2) to the thermodynamic functions of the Fermi liquid are considerably smaller than the corrections included below.

The density of uncondensed bosons at a finite temperature $T < T_c$ is

$$N' = \int \frac{2\pi k dk}{(2\pi\hbar)^2} \frac{1}{e^{\epsilon/T} - 1} = \frac{mT}{2\pi\hbar^2} \log \frac{T}{\tau}. \quad (8)$$

The integral in Eq. (8) diverges at small ϵ and is therefore cut off at $\epsilon \sim \tau$, where τ is a small tunneling amplitude of electrons in the direction perpendicular to the CuO_2 plane.

The excess number $n - n_c$ of electrons in the system is equal to the doubled sum of N' and the number N_0 of bosons in the condensate. This leads to the dependence of the superconducting transition temperature on the doping level for small values of $n - n_c$:

$$n - n_c = \frac{mT_c}{\pi\hbar^2} \log \frac{T_c}{\tau}. \quad (9)$$

The number of pairs in the condensate,

$$N_0 = \frac{n - n_c}{2} \left(1 - \frac{T}{T_c} \frac{\log T/\tau}{\log T_c/\tau}\right) \quad (10)$$

determines the modulus of the order parameter $|\psi|^2 = N_0$ at finite temperatures. The superconducting transition temperature defined by Eq. (9) is quite high. To within the logarithmic term, this temperature is on the order of the one-electron tunneling amplitude t at the boundary of the applicability region (i.e., for $n - n_c \sim a^{-2}$). The possibility that the superconducting transition temperature may have such an order of magnitude was pointed out in the aforementioned paper by Alexandrov and Kornilovitch [7].

The interaction of fermions with the BE condensate (effective electron–electron interaction), which is described by the order parameter ψ, creates an effective potential $\Delta_{\mathbf{k}}$ acting on fermions as in conventional superconductors:

$$H_{\text{int}} = \sum_{\mathbf{k}} (\Delta_{\mathbf{k}} c^+_{\mathbf{k}\uparrow} c^+_{-\mathbf{k}\downarrow} + \text{H.c.}) \quad (11)$$

In view of the symmetry of ψ, we have

$$\Delta_{\mathbf{k}} = V(\hat{k}_x^2 - \hat{k}_y^2)\psi, \quad (12)$$

where $\hat{\mathbf{k}} = \mathbf{k}/|\mathbf{k}|$ and V is invariant under the symmetry group.

Owing to this interaction, fermions in the superconducting state considered acquire features typical of an ordinary superconductor with the $d_{x^2-y^2}$ symmetry.

4. NORMAL STATE THERMODYNAMICS. THE PSEUDOGAP

The total number of pairs for $T < T_c$ is independent of the temperature and is equal to $(n - n_c)/2$. The electron chemical potential for $T < T_c$ is also temperature-independent and equal to $\mu = \mu(n_c) = \epsilon_m(n_c)/2$, where $\epsilon_m = \epsilon_m(n)$ is the pair minimal energy, which depends on the fermion density, as shown above.

For $T > T_c$, the fermion distribution function, as above, corresponds to $T = 0$, but with the temperature-dependent chemical potential. The pair energy spectrum is $E = \epsilon_m(\mu) + \epsilon$, where $\epsilon = k^2/2m$. The pair density above T_c is given by

$$N = \int_0^\infty \frac{2\pi k dk}{(2\pi\hbar)^2} \frac{1}{e^{(\epsilon+\zeta)/T} - 1} = \frac{mT}{2\pi\hbar^2} \log \frac{1}{1 - e^{-\zeta/T}}. \quad (13)$$

The parameter ζ ($\zeta \gg \tau$) is defined by

$$\zeta = \frac{\partial \epsilon_m}{\partial \mu} \delta\mu - 2\delta\mu, \quad (14)$$

where $\delta\mu = \mu - \mu(n_c)$. With changing temperature, the total electron number conservation gives

$$n - n_c = 2N + \frac{\partial n}{\partial \mu} \delta\mu. \quad (15)$$

From the last equation, we find $\zeta = \zeta(T)$ and then all the other quantities.

For $n > n_c$ and not too high temperature $T \ll T_c \log(T_c/\tau)$, the pair density is determined by

$$\frac{N(T) - N(T_c)}{N(T_c)} = \frac{\partial n/\partial \mu}{2(2 - \partial \epsilon_m/\partial \mu)} T e^{-\Delta_p/T}, \quad (16)$$

where $N(T_c) = (n - n_c)/2$ and

$$\Delta_p = T_c \log \frac{T_c}{\tau} = \frac{\pi\hbar^2}{m}(n - n_c) \quad (17)$$

is the pseudogap for $n > n_c$. As well as T_c, it is zero at the critical value of the doping level $n = n_c$. For a higher doping level $n < n_c$ ($T_c = 0$), we have

$$N(T) = \frac{mT}{2\pi\hbar^2} e^{-\Delta_p'/T}, \quad (18)$$

where

$$\Delta_p' = \left(2\frac{\partial \mu}{\partial n} - \frac{\partial \epsilon_m}{\partial n}\right)(n_c - n) \quad (19)$$

is the pseudogap for $n < n_c$. Equation (18) holds in the low-temperature region $T \ll \Delta_p'$. For $n < n_c$, the pseudogap Δ_p' is the gap in the energy spectrum of the pair quasiparticles. For high temperatures $T \gg \Delta_p$, Δ_p' (but $T \ll t$), the pair density is a linear function of temperature,

$$N(T) = \frac{z\partial n/\partial \mu}{2(2 - \partial \epsilon_m/\partial \mu)} T \quad (20)$$

where z is the solution of the equation $\lambda z = e^{-z}$ with

$$\lambda = \frac{\pi\hbar^2}{m} \frac{\partial n/\partial \mu}{2 - \partial \epsilon_m/\partial \mu}. \quad (21)$$

The entropy of pairs is determined by the equation

$$S(T) = \frac{m}{2\pi\hbar^2} \int_0^\infty d\epsilon \{(1 + f)\log(1 + f) - f \log f\}, \quad (22)$$

where $f = \{e^{(\epsilon+\zeta)/T} - 1)\}^{-1}$. For $n > n_c$ in the low-temperature region $T \ll \Delta_p$, we have

$$\frac{S(T)}{T} - \left(\frac{S}{T}\right)_{T=T_c} = -\frac{m}{2\pi\hbar^2} \frac{\Delta_p}{T} e^{-\Delta_p/T}, \quad (23)$$

where

$$\left(\frac{S}{T}\right)_{T=T_c} = \frac{\pi m}{12\hbar^2}. \quad (24)$$

The function $S(T)$ is almost linear in T, with exponentially small deviations. For $n < n_c$, the pair entropy is exponentially small at low temperatures $T \ll \Delta_p'$:

$$S(T) = \frac{mT}{2\pi\hbar^2} e^{-\Delta_p'/T}. \quad (25)$$

At high temperatures $T \gg \Delta_p$, Δ_p', the entropy is

$$S(T) = \frac{m\sigma}{2\pi\hbar^2} T. \quad (26)$$

The temperature-independent factor σ is determined by

$$\sigma = \int_z^\infty \frac{x dx}{e^x - 1} - \lambda z^2. \quad (27)$$

The entropy is again a linear function of temperature.

5. ORBITAL PARAMAGNETISM OF PAIRS

In this section, we show that the orbital motion of electrons inside the pairs cause a peculiar paramagnetism. Let a pair be at rest as a, whole. For singlet pairs at $\mathbf{k} = 0$, the Hamiltonian in Eqs. (3) and (4) can be written as the 2×2 matrix

$$H = 4t \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \equiv 4t\sigma_1, \quad (28)$$

acting on a state vector

$$\psi = \psi^{(x)} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \psi^{(y)} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad (29)$$

where $\psi^{(n)}$, $n = x, y$ are quantum amplitudes of two orientations of the two-electron dumb-bell and σ_1 is a Pauli matrix.

In the x state, coordinates of two electrons (with respect to the center of gravity of the pair) are $x_1 = -a/2$, $y_1 = 0$ and $x_2 = a/2$, $y_2 = 0$, respectively. In the y state, we have $x_1 = 0$, $y_1 = -a/2$ and $x_2 = 0$, $y_2 = a/2$. From this, we find the coordinate operators for both electrons:

$$x_1 = -x_2 = -\frac{a}{2} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad (30)$$

$$y_1 = -y_2 = -\frac{a}{2} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

The velocity operators are determined by the commutators

$$\dot{\mathbf{r}}_{1,2} = \frac{i}{\hbar}[H, \mathbf{r}_{1,2}]. \quad (31)$$

Simple calculation gives

$$\dot{x}_1 = -\dot{x}_2 = -\dot{y}_1 = \dot{y}_2 = -\frac{2at}{\hbar} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \equiv -\frac{2at}{\hbar}\sigma_2. \quad (32)$$

The operator of the pair magnetic moment, which is directed along the z axis, is

$$\mu \equiv \mu_z = \frac{e}{2c}\sum_{1,2}(x\dot{y} - y\dot{x}) = -\frac{eta^2}{\hbar c}\sigma_2, \quad (33)$$

where e is the electron charge and c is the velocity of

In the presence of an external magnetic field $B \equiv B_z$, the Hamiltonian of the pair is

$$H = 4t\sigma_1 - \mu B. \quad (34)$$

The energy eigenvalues are

$$E = \epsilon_0 \mp 4t\left[1 + \left(\frac{ea^2}{4\hbar c}B\right)^2\right]^{1/2}. \quad (35)$$

In weak fields, the minimal energy is

$$E_{\min} = \epsilon_0 - 4t - t\frac{e^2a^4}{8\hbar^2c^2}B^2. \quad (36)$$

The average magnetic moment of the pair is

$$\langle \mu \rangle = -\frac{\partial E_{\min}}{\partial B} = \alpha B, \quad (37)$$

where

$$\alpha = \frac{e^2a^4}{4\hbar^2c^2}t = \frac{e^2a^2}{4mc^2} \quad (38)$$

is the pair paramagnetic polarizability.

We note that pairs with $\mathbf{k} = 0$ in the upper energy band (the lower sign in (35)) are diamagnetic.

The pair contribution to the paramagnetic susceptibility of a 3D sample is

$$\chi = \frac{e^2a^2}{4mc^2}N^{(3)}, \quad (39)$$

where $N^{(3)} = N(T)/L$ is the 3D density of pairs and $N(T)$ is the 2D density determined by formulas (16), (18), and (20). Here, L is the distance between neighboring CuO_2 planes.

Generally, we have three competing contributions to the magnetic susceptibility: the orbital paramagnetism of pairs considered above, the Pauli spin susceptibility of single electrons (pairs are singlet), and the Landau diamagnetism of single electrons and pairs. Spin susceptibility is isotropic. Orbital paramagnetism and Landau diamagnetism are both strongly anisotropic (the magnetic moment is directed along the z axis independently of the direction of the magnetic field) because of a 2D character of single electrons and pairs. However, Landau diamagnetism, especially in the 2D case, is very sensitive to inhomogeneities. For example, it is easily suppressed by localization of charge carriers. Orbital paramagnetism is finite at zero velocity of a pair as a whole. Therefore, it has to be much more stable against inhomogeneities. We hope that orbital paramagnetism can be experimentally separated from the other two contributions to susceptibility.

ACKNOWLEDGMENTS

This study was supported by INTAS (grant no. 01686), the CRDF (grant no. RP1-2411-MO-02), Leverhulme Trust (grant no. S-00261-H), the Russian Foundation for Basic Research (project no. 03-02-16401), and the President Program Supporting Leading Scientific Schools.

REFERENCES

1. A. F. Andreev, JETP Lett. **79**, 88 (2004).
2. C. C. Tsuei and J. R. Kirtley, Rev. Mod. Phys. **72**, 969 (2000).
3. T. Timusk and B. Statt, Rep. Prog. Phys. **62**, 61 (1999).
4. A. S. Alexandrov, Phys. Rev. B **48**, 10571 (1993).
5. V. Emery and S. Kivelson, Phys. Rev. Lett. **74**, 3253 (1995); Nature **374**, 434 (1995).
6. V. B. Geshkenbein, L. B. Ioffe, and A. I. Larkin, Phys. Rev. B **55**, 3173 (1997).
7. A. S. Alexandrov and P. E. Kornilovitch, J. Supercond. **15**, 403 (2002).
8. T. Domanski and J. Ranninger, Physica C (Amsterdam) **387**, 77 (2003).
9. A. F. Andreev, in *Progress in Low Temperature Physics*, Ed. by D. F. Brewer (North-Holland, Amsterdam, 1982), Vol. 8, Sect. 4.4.
10. K. P. Sinha, Indian J. Phys. **35**, 434 (1961).

Journal of Experimental and Theoretical Physics, Vol. 100, No. 4, 2005, pp. 703–714.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 4, 2005, pp. 796–808.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Superconductivity in a System of *p–d* Electrons

R. O. Zaitsev

Russian Research Center Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia
e-mail: zaitsev@mbslab.kiae.ru

Received November 15, 2004

Abstract—The electronic structure of compounds is studied taking into account the Hubbard energy as the largest energy parameter. The conditions for the emergence of Cooper instability are obtained. The phase diagram for the superconducting state is calculated for various degrees of filling of the *d* and *p* shells of transition and nontransition elements. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The classical Hubbard model, in which direct hopping between the nearest cations of the *t*- and *e*-shells of a transient element is taken into account, makes it possible to correctly describe the electron properties of superconductors of the $Ln_{2-x}^{3+}M_x^{4+}CuO_4$ type [1, 2]. With increasing concentration x of the tetravalent cation M^{4+}, the upper half of the Hubbard subband, which corresponds to collectivization of the $(x^2 - y^2)$ states of copper, is populated. In this case, the $2p^6$ shell of oxygen remains completely filled, which corresponds to the assumption on the high energy of hole p excitations as compared to the energy of the hole states in the $3d^{10}$ shell of Cu^+.

For other compounds such as $Ln_{2-x}^{3+}M_x^{2+}CuO_4$, $YB_2Cu_3O_{7-\delta}$, and Sr_2RuO_4, the assumption concerning complete filling of the p shell of O^{2-} anions should be rejected, and hopping between the oxygen anions and copper cations should be taken into account explicitly [3, 4]. The energy shift of the hole states of oxygen with energy ϵ_p relative to cation states with energy ϵ_d is assumed to be preset ($r = \epsilon_p - \epsilon_d$) so that $r \gg U$ corresponds to the classical Hubbard model. The real situation corresponds to the opposite limit, when Hubbard energy U is larger than or on the order of the energy shift r. In this study, the X operator method is used in the so-called *p–d* model. The main assumption made in choosing the Hamiltonian is that the Hubbard energy as a function of the principal quantum number n can be estimated by the formula $U(n) \approx U(1)/n$, where $U(1)$ is the Hubbard energy of s electrons, which is equal to 17 eV. Consequently, the corresponding values for the $2p$ electrons of oxygen and for $3d$ electrons of copper are 8.5 and 5.67 eV, respectively.

In accordance with these estimates, we can assume that the Hubbard energy for the d as well as p electrons is higher than the hopping energy t_{pd}. For this reason, a correct choice of the zeroth approximation and a transition to the X operators make it possible to obtain the scattering amplitude on the order of $t_{pd}(\mathbf{p})$ both in the case of (*d–d*) and (*p–p*) scattering. The same method allows us to write the equation of state, i.e., to establish the relation between the chemical potential, parameter $\Delta = \epsilon_p - \epsilon_d$, and the number of holes in the $2p^6$ and $3d^{10}$ shells of oxygen and copper.

As regards the value of U_{pd} (i.e., the Coulomb energy of the interaction between electrons belonging to neighboring copper cations and oxygen anions), all these matrix electrons are assumed to be small and will not be considered here together with the energy of interactions between electrons (V_{dd} and V_{pp}) belonging to adjacent atoms.

Here, the Emery parameter $r = \epsilon_p - \epsilon_d$ is arbitrary. As a matter of fact, it contains the contribution to the crystal field due to the difference in the potentials at the sites occupied by copper cations and oxygen anions; this contribution has not been reasonably estimated so far due to screening effects.

In such a formulation, our task is to construct the phase diagram depending on two parameters, viz., $r = \epsilon_p - \epsilon_d$ and chemical potential $\mu = -(\epsilon_p - \epsilon_d)/2$. The equations of state make it possible to transform the resultant phase diagram to variables h_p and h_d.

Experiments on studying the resistance in compounds Sr_2RuO_4 and $Nd_{2-x}^{3+}Ce_x^{4+}CuO_2O_2^{2-}$ reveal a quadratic temperature dependence [5], indicating the weakness of the electron–phonon interaction. The characteristic value of energy, which determines the superconducting transition temperature by the BCS formula for the $Nd_{2-x}Ce_xCuO_4$ compound, has the same order of magnitude as the well-known compound $La_{2-x}^{3+}Me_x^{2+}CuO_4$ doped with bivalent Ba, Sr, or Hg cations. These facts indicate that the key role is played by the electron–electron interaction, whose intensity considerably exceeds the width of the e_g shell being populated.

In the $SrRuO_4$ compound, the Hubbard energy of 4*d* electrons of ruthenium is found to be lower than the Hubbard energy of 3*d* electrons of copper; however, it is higher, as before, than the width of underpopulated t_{xy} shell.

The number of holes (h_p) in the $2p^6$ shell of oxygen in the plane of the CuO_2 layer is connected with the number of holes in the $3d^{10}$ shell of Cu cations via the electroneutrality condition

$$2h_p + h_d = 1 - x \text{ for } Nd_{2-x}Ce_xCuO_4, \quad (1a)$$

$$2h_p + h_d = 1 + x \text{ for } La_{2-x}Me_x^{2+}CuO_4. \quad (1b)$$

The energy of the 4*s* states is higher than the energy of the 3*d* states. Accordingly, $0 < h_d < 2$, and we must trace the extent of underpopulation of the $p_{x,y}$ shells of oxygen in the CuO_2 plane, taking into account their twofold degeneracy and noting that the Hubbard energy is substantially higher than the Fermi energy.

An analogous situation takes place in the $Sr_2^{2+}RuO_4$ compound, in which the degenerate $4t_{zx}$ and $4t_{yz}$ shells are filled, while the population of the $4t_{xy}$ shell depends on the number of holes in the $2p^6$ shells of oxygen anions lying in the RuO_2 plane. Taking into account the electroneutrality condition, we can write the relation between the number of holes h_t in the t_{xy} shell of ruthenium cations and the number of holes h_p in oxygen anions:

$$2h_p + h_t = 2 \text{ for } Sr_2RuO_4. \quad (1c)$$

The crystal structure of $Sr_2^{2+}RuO_4$ is the same as the structure of $La_2^{3+}CuO_4$; consequently, the phase diagrams of these compounds are identical. The main difference is determined by different arrangements of electroneutrality lines (1b) and (1c).

2. HAMILTONIAN AND FORMULATION OF THE PROBLEM

For simplicity, we assume that this energy is infinitely large as compared to hopping integrals $t^{ik}(r)$, $t^i(r)$, and $\tau(r)$ appearing in the definition of the Hamiltonian:

$$\begin{aligned}\hat{H} = &\sum_{i,j,\mathbf{r},\mathbf{r}',\sigma;\,\mathbf{r}\neq\mathbf{r}'} \hat{p}^+_{i\mathbf{r}\sigma}\hat{p}_{j\mathbf{r}'\sigma}t^{(ij)}(\mathbf{r}-\mathbf{r}') \\ &+ \sum_{i,\mathbf{r},\mathbf{r}',\sigma} [\hat{a}^+_{\mathbf{r}\sigma}\hat{p}_{i\mathbf{r}'\sigma}t^{(i)}(\mathbf{r}-\mathbf{r}') + \text{h.c.}] \\ &+ \sum_{\mathbf{r},\mathbf{r}',\sigma;\,\mathbf{r}\neq\mathbf{r}'} \hat{a}^+_{\mathbf{r}\sigma}\hat{a}_{\mathbf{r}'\sigma}\tau(\mathbf{r}-\mathbf{r}') \\ &+ \epsilon_p\sum_{i,\mathbf{r},\sigma}\hat{p}^+_{i\mathbf{r}\sigma}\hat{p}_{i\mathbf{r}\sigma} + \epsilon_d\sum_{i,\mathbf{r},\sigma}\hat{a}^+_{\mathbf{r}\sigma}\hat{a}_{\mathbf{r}\sigma}.\end{aligned} \quad (2)$$

Here, *i*, *j* are the vector indices corresponding to the t_{1u} states and ϵ_p and ϵ_d are the energies of one-hole *p* and *d* states.

If the mean occupation numbers h_p and h_d are smaller than unity, it is sufficient to consider transitions between completely filled and four possible one-hole states. Accordingly, we express the creation and annihilation operators in terms of the *X* operators corresponding to transitions between the completely filled and one-hole states:

$$\begin{aligned}\hat{a}^+_{\mathbf{r}\sigma} &= \hat{X}^{\sigma,0}, \quad \hat{a}_{\mathbf{r}\sigma} = \hat{X}^{0,\sigma}, \\ \hat{p}^+_{k\mathbf{r}\sigma} &= \hat{X}^{k\sigma,0}_{\mathbf{r}}, \quad \hat{p}_{k\mathbf{r}\sigma} = \hat{X}^{0,k\sigma}_{\mathbf{r}}.\end{aligned} \quad (3)$$

Here, $k = x, y$ and indices $(k\sigma, 0)$ and $(0, k\sigma)$ denote transitions between the completely filled $2p^6$ shell and one-hole $|p_{x,y;\sigma}\rangle$ states (and back).

For region $1 < h_d < 2$, we must consider transitions between two one-hole $(x^2 - y^2)$ states and a two-hole state with zero total spin.

Emery [3] was the first to note that the matrix elements *w* of hopping between oxygen and copper play the main role in the formation of the elementary excitation spectrum of the CuO_2 complex. The matrix elements proportional to w_p and w_d are smaller and have a complex angular dependence on quasi-momentum. For the latter reason, to simplify our analysis, we will assume that the diagonal matrix elements *p–p* and *p–d* are proportional to the same dimensionless function $t_{\mathbf{p}}$.

On account of these simplifications, we can write the reciprocal one-particle Green function, which has the following form in the zero-loop approximation of the self-consistent field (Hubbard I) for U_p, $U_d = +\infty$:

$$G^{-1}_{\omega}(\mathbf{p}) = \begin{matrix}(0,\sigma)\\(ax)\\(by)\\(ay)\\(bx)\end{matrix} \times \begin{pmatrix} \Omega_d(\mathbf{p}) & -f_d w\tau_x(\mathbf{p}) & -f_d w\tau_y(\mathbf{p}) & 0 & 0 \\ -f_p w\tau^*_x(\mathbf{p}) & \Omega_p(\mathbf{p}) & 0 & 0 & 0 \\ -f_p w\tau^*_y(\mathbf{p}) & 0 & \Omega_p(\mathbf{p}) & 0 & 0 \\ 0 & 0 & 0 & \Omega_p & 0 \\ 0 & 0 & 0 & 0 & \Omega_p \end{pmatrix}. \quad (4)$$

Here, the following convenient notation has been

introduced:

$$\omega_n = (2n+1)\pi T,$$
$$\Omega_p = i\omega - \epsilon_p, \quad \Omega_d(\mathbf{p}) = i\omega - w_d f_d t_\mathbf{p} - \epsilon_d, \quad (5)$$
$$\Omega_p(\mathbf{p}) = i\omega - w_p f_p t_\mathbf{p} - \epsilon_p.$$

The sum $\epsilon_p + \tau_p f_p t_\mathbf{p}$ consists of two anisotropic branches $p_{x,y}$ of degenerate states. The other two branches are not collectivized and have energy ϵ_p.

For an infinitely large Hubbard energy, the end factors f_p and f_d are linear functions of $h_{p,d}$ [6]. At the ends of the integral interval $k < h_{p,d}^{(k)} < k+1$, these factors are equal to the reciprocal degeneracy of the lower k-hole state or the $(k+1)$ hole state:

$$f_d^{(0)} = 1 - \frac{h_d}{2}, \quad 0 < n_d < 1;$$
$$f_p = 1 - \frac{3}{4}h_p, \quad 0 < n_p < 1;$$
$$f_d^{(1)} = \frac{h_d}{2}, \quad 1 < h_d < 2; \quad (6)$$
$$f_p^{(1)} = \frac{h_p + 2}{12}, \quad 1 < h_p < 2.$$

In the subsequent analysis, we will confine ourselves to the region $0 < h_d < 2$, $0 < h_p < 1$.

3. EQUATIONS OF STATE

Apart from two noncollectivized $p_{x,y}$ branches, which have the same energy $E^{(1,2)} = \epsilon_p$, the poles of one-particle Green function (4) define three more branches:

$$E_\mathbf{p}^{(3)} = \xi_\mathbf{p}^{(p)} = w_p f_p t_\mathbf{p} + \epsilon_p, \quad \xi_\mathbf{p}^{(\pm)} = \frac{\xi_\mathbf{p}^{(p)} + \xi_\mathbf{p}^{(d)}}{2}$$
$$\pm \frac{1}{2}\sqrt{(\xi_\mathbf{p}^{(p)} - \xi_\mathbf{p}^{(d)})^2 + 4w^2 f_p f_d |\tau_\mathbf{p}|^2}, \quad (7)$$

where

$$|\tau_\mathbf{p}|^2 = |\tau_x(\mathbf{p})|^2 + |\tau_y(\mathbf{p})|^2.$$

Using the one-loop approximation, we can find the relation between the chemical potential $\mu = -(\epsilon_d + \epsilon_p)/2$, the temperature, and the mean numbers of particles per unit cell. Taking into account the degeneracy in spin and crystallographic indices, we obtain the following expression for the sums of the diagonal matrix elements of the one-particle Green function:

$$h_d = 2f_d T \sum_{\omega,\mathbf{p}} G_\omega^{d,d}(\mathbf{p}) \exp(i\delta\omega),$$
$$h_p = f_p T \sum_{\omega,\mathbf{p}} \left\{ G_{xx} + G_{yy} + \frac{2}{i\omega - \epsilon_p} \right\} \exp(i\delta\omega). \quad (8)$$

The reason for the disappearance of factor 2 in front of the sum over p electronic states of the diatomic CuO_2 molecule is that h_p is the mean number of p electrons corresponding to an oxygen anion and not to the entire unit cell.

Using the explicit expression for reciprocal Green function (4), we obtain

$$G_{dd} = \sum_{\lambda=\pm} A_\mathbf{p}^{(-\lambda)} \frac{1}{i\omega_n - \xi_\mathbf{p}^{(\lambda)}},$$
$$G_{xx} = G_{yy} = \frac{1}{2(i\omega_n - \xi_\mathbf{p}^{(p)})} + \frac{1}{2}\sum_{\lambda=\pm} A_\mathbf{p}^{(\lambda)} \frac{1}{i\omega_n - \xi_\mathbf{p}^{(\lambda)}}, \quad (9)$$
$$A_\mathbf{p}^{\pm} = \frac{1}{2}\left[1 \pm \frac{\xi_\mathbf{p}^{(p)} - \xi_\mathbf{p}^{(d)}}{\sqrt{(\xi_\mathbf{p}^{(p)} - \xi_\mathbf{p}^{(d)})^2 + 4w^2 f_p f_d |\tau_\mathbf{p}|^2}} \right].$$

As a result of substitution and summation over $i\omega_n$, we obtain the equations of state expressed in terms of sums of the Fermi function $n_F(\epsilon)$.

For the d-hole states, it is necessary to consider the following two regions:

$$h_d = 2f_d \sum_{\mathbf{p},\lambda=\pm} A_\mathbf{p}^{(-\lambda)} n_F(\xi_\mathbf{p}^{(\lambda)}),$$
$$\text{for } 0 < h_d < 1, \quad f_d = 1 - \frac{h_d}{2}, \quad (10a)$$
$$h_d = 1 + f_d \sum_{\mathbf{p},\lambda=\pm} A_\mathbf{p}^{(-\lambda)} n_F(\xi_\mathbf{p}^{(\lambda)}),$$
$$\text{for } 1 < h_d < 2, \quad f_d = \frac{h_d}{2}.$$

For the p-hole states, it is sufficient to consider the region $0 < h_p < 1$, for which we have $f_p = 1 - 3h_p/4$,

$$h_p = f_p \left\{ 2n_F(\epsilon_p) + \sum_\mathbf{p} \left[n_F(\xi_\mathbf{p}^{(p)}) + \sum_{\lambda=\pm} A_\mathbf{p}^{(\lambda)} n_F(\xi_\mathbf{p}^{(\lambda)}) \right] \right\}. \quad (10b)$$

On account of the square symmetry, we can express the hopping integrals in terms of the single function $\epsilon(\mathbf{p})$ of quasi-momenta:

$$\xi_\mathbf{p}^{(p)} = w_p \epsilon(\mathbf{p}), \quad \xi_\mathbf{p}^{(d)} = w_d \epsilon(\mathbf{p}),$$
$$w^2 |\tau_\mathbf{p}|^2 = w^2(1 - \epsilon(\mathbf{p})).$$

For this function, we can define the initial functions of the density of states,

$$\rho_0(\epsilon) = \sum_\mathbf{p} \delta(\epsilon - \epsilon(\mathbf{p})),$$

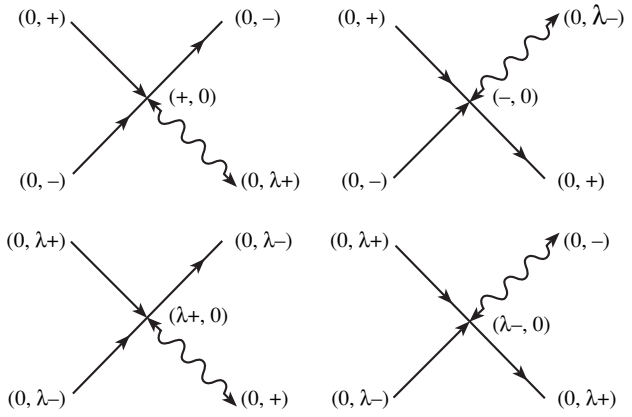


Fig. 1. Born amplitudes of the kinematic interaction for $n_d < 1$; for the upper Hubbard subband ($n_d > 1$), the substitution $(0, \sigma) \longrightarrow (-\sigma, 2)$ should be carried out.

after which the equations of state can be represented in the form of single integrals.

The classical Emery model corresponds to the approximation $w \gg w_p \sim w_d$; however, a more realistic situation corresponds to the inequalities $w \sim w_p \gg w_d$.

4. SUPERCONDUCTIVITY CRITERION

For an infinitely large Hubbard energy, the interaction between the *s* and *t* excitations is manifested in scattering and a strong dependence of the scattering amplitude on the position of the Fermi level. We can find the Cooper instability from the condition for the emergence of singularity in the two-particle Green function [7]. In the ladder approximation, the problem is reduced to determining the conditions for nonvanishing solutions in the corresponding homogeneous system of equations. For zero values of the total momentum, total projection of spin, and energy, this system in our problem has the form

$$\Gamma_{\alpha\bar{\alpha};\,\nu\bar{\nu}}(\mathbf{p}) = g_{\alpha\bar{\alpha};\,\nu\bar{\nu}}(\mathbf{p}) - \sum_{\mathbf{p}',\beta,\bar{\beta}} g_{\alpha\bar{\alpha};\,\beta\bar{\beta}}(\mathbf{p}')K^{\gamma,\bar{\gamma}}_{\beta,\bar{\beta}}(\mathbf{p}')\Gamma_{\gamma\bar{\gamma};\,\nu,\bar{\nu}}(\mathbf{p}'), \tag{11}$$

where

$$K^{\gamma,\bar{\gamma}}_{\beta,\bar{\beta}}(\mathbf{p}') = T\sum_{\omega} G^{\beta\gamma}_{-\omega}(-\mathbf{p}')G^{\bar{\beta}\bar{\gamma}}_{\omega}(\mathbf{p}').$$

Indices α and $\bar{\alpha}$ correspond to transitions with opposite signs of the variation of the spin projection. The components of Green function $G^{\alpha\beta}_{\omega}(\mathbf{p})$ can be defined in terms of reciprocal matrix (4). According to Dyson [8], the scattering amplitude $g_{\alpha\bar{\alpha};\,\beta\bar{\beta}}$ is defined by the double commutator $\{\hat{X}^{\alpha}_{\mathbf{r}}[\hat{X}^{\bar{\alpha}}_{\mathbf{r}}\hat{H}]\}$. Ultimately, this relation can be expressed in terms of structural constants $N^{\pm}_{\alpha,\beta}$ of the corresponding superalgebra [1, 2]:

$$[\hat{X}^{\alpha}_{\mathbf{r}}, \hat{X}^{\beta}_{\mathbf{r}'}]_{\pm} = N^{\pm}_{\alpha,\beta}\hat{X}^{\alpha+\beta}_{\mathbf{r}}\delta_{\mathbf{r},\mathbf{r}'}. \tag{12}$$

In our case, Hamiltonian (2) can be represented as the sum of the products of Fermi operators corresponding to different unit cells:

$$\hat{H} = \sum_{\mathbf{r},\mathbf{r}';\,\mathbf{r}\neq\mathbf{r}'} (t(\mathbf{r},\mathbf{r}')^{\nu,\beta}\hat{X}^{\nu}_{\mathbf{r}}\hat{X}^{\beta}_{\mathbf{r}'} + t(\mathbf{r},\mathbf{r}')^{\bar{\nu},\bar{\beta}}\hat{X}^{\bar{\nu}}_{\mathbf{r}}\hat{X}^{\bar{\beta}}_{\mathbf{r}'}). \tag{13}$$

Here, the subscripts with bars (e.g., $\bar{\alpha}$) indicate transitions with the reverse sign of the spin projection (relative to α).

Direct calculation leads to the expressions

$$[X^{\bar{\alpha}}_{\mathbf{r}}, \hat{H}] = \sum_{\mathbf{r}',\nu,\beta;\,\mathbf{r}'\neq\mathbf{r}} N^{+}_{\bar{\alpha},\nu}t(\mathbf{r},\mathbf{r}')^{\nu,\beta}\hat{X}^{\bar{\alpha}+\nu}_{\mathbf{r}}\hat{X}^{\beta}_{\mathbf{r}'} + \sum_{\mathbf{r}',\bar{\nu},\bar{\beta};\,\mathbf{r}'\neq\mathbf{r}} N^{+}_{\bar{\alpha},\bar{\nu}}t(\mathbf{r},\mathbf{r}')^{\bar{\nu},\bar{\beta}}\hat{X}^{\bar{\alpha}+\bar{\nu}}_{\mathbf{r}}\hat{X}^{\bar{\beta}}_{\mathbf{r}'}. \tag{14}$$

$$\{\hat{X}^{\alpha}_{\mathbf{r}}[\hat{X}^{\bar{\alpha}}_{\mathbf{r}}, \hat{H}]\} = \sum_{\mathbf{r}',\nu,\beta;\,\mathbf{r}'\neq\mathbf{r}} N^{-}_{\alpha,\bar{\alpha}+\nu}N^{+}_{\bar{\alpha},\nu}t(\mathbf{r},\mathbf{r}')^{\nu,\beta}\hat{X}^{\alpha+\bar{\alpha}+\nu}_{\mathbf{r}}\hat{X}^{\beta}_{\mathbf{r}'} = \sum_{\mathbf{r}',\bar{\nu},\bar{\beta};\,\mathbf{r}'\neq\mathbf{r}} N^{-}_{\alpha,\bar{\alpha}+\bar{\nu}}N^{+}_{\bar{\alpha},\bar{\nu}}t(\mathbf{r},\mathbf{r}')^{\bar{\nu},\bar{\beta}}\hat{X}^{\alpha+\bar{\alpha}+\bar{\nu}}_{\mathbf{r}}\hat{X}^{\bar{\beta}}_{\mathbf{r}'}. \tag{15}$$

In our case, for transitions between states with completely filled shells and one-hole states, we can easily calculate the double commutators. Figure 1 shows the graphic representation of the Born amplitudes.

In the case of the (*d*–*d*) scattering, using Hamilton (2), we obtain

$$[\hat{X}^{(0,-\sigma)}_{\mathbf{r}}\{\hat{X}^{(0,\sigma)}_{\mathbf{r}}, \hat{H}\}] = -\sum_{i,\mathbf{r}'}\hat{X}^{(0,\sigma)}_{\mathbf{r}}\hat{p}_{i\mathbf{r}'-\sigma}t^{(i)}(\mathbf{r}-\mathbf{r}') - \sum_{i,\mathbf{r}'}\hat{p}_{i\mathbf{r}'\sigma}\hat{X}^{(0,-\sigma)}_{\mathbf{r}}t^{(i)}(\mathbf{r}-\mathbf{r}') - \sum_{\mathbf{r}';\,\mathbf{r}'\neq\mathbf{r}}\hat{X}^{(0,\sigma)}_{\mathbf{r}}\hat{a}_{\mathbf{r}'-\sigma}\tau(\mathbf{r}-\mathbf{r}') - \sum_{\mathbf{r}';\,\mathbf{r}\neq\mathbf{r}'}\hat{a}_{\mathbf{r}',\sigma}\hat{X}^{(0,-\sigma)}_{\mathbf{r}}\tau(\mathbf{r}-\mathbf{r}'). \tag{16}$$

In the case of the (*p–p*) scattering, we have

$$
\begin{aligned}
&[\hat{X}_{\mathbf{r}}^{(0;\,i,-\sigma)}\{\hat{X}_{\mathbf{r}}^{(0;\,i,\sigma)},\hat{H}\}]\\
&= -\sum_{\mathbf{r}'}\hat{X}_{\mathbf{r}}^{(0;\,i,\sigma)}\hat{p}_{i\mathbf{k}'-\sigma}t^{(i)}(\mathbf{r}-\mathbf{r}')\\
&-\sum_{\mathbf{r}'}\hat{p}_{i\mathbf{r}'\sigma}\hat{X}_{\mathbf{r}}^{(0;\,i-\sigma)}t^{(i)}(\mathbf{r}-\mathbf{r}')\\
&-\sum_{j,\,\mathbf{r}',\,\sigma;\,\mathbf{r}'\neq\mathbf{r}}\hat{X}_{\mathbf{r}}^{(0;\,i,\sigma)}\hat{p}_{j\mathbf{r}',-\sigma}t^{(ij)}(\mathbf{r}-\mathbf{r}')\\
&-\sum_{j,\,\mathbf{r}',\,\sigma;\,\mathbf{r}'\neq\mathbf{r}}\hat{p}_{j\mathbf{r}'\sigma}\hat{X}^{(0;\,i,-\sigma,\,\mathbf{r})}t^{(ij)}(\mathbf{r}-\mathbf{r}').
\end{aligned}
\tag{17}
$$

After the transition to Fourier components and to the center-of-mass system, it turns out that the scattering amplitudes depend only on the momentum of scattered particles and differ from hopping integrals $t^{\alpha,\beta}(\mathbf{p})$ only by factors:

$$
\begin{aligned}
g_{\alpha,\alpha;\,\gamma\nu}(\mathbf{p};\mathbf{p}') &= -\delta_{\alpha,\gamma}\{t^{\alpha,\nu}(\mathbf{p}')+t^{\alpha,\nu}(-\mathbf{p}')\}\\
&-\delta_{\alpha,\nu}\{t^{\alpha,\gamma}(\mathbf{p}')+t^{\alpha,\gamma}(-\mathbf{p}')\}.
\end{aligned}
\tag{18}
$$

Here, indices α, γ, and ν are the numbers of one-particle transitions (*d*, *x*, *y*).

Subsequent simplification of the kernel of integral equation (11) are associated with the possibility of carrying out summation over internal indices β and $\bar{\beta}$. For this purpose, we write the Dyson equation that was used for determining reciprocal Green function (10). In the one-loop approximation (Hubbard I), we have

$$
\sum_{\beta}f_{\nu}t^{\nu,\beta}(\mathbf{p})G_{\omega}^{\beta,\gamma}(\mathbf{p}) = \delta_{\nu,\gamma}-[(\hat{G}_{\omega}^{(0)})^{-1}\hat{G}_{\omega}(\mathbf{p})]_{\gamma}^{\nu};
\tag{19a}
$$

$$
\begin{aligned}
&\sum_{\bar{\nu}}f_{\bar{\lambda}}t^{\bar{\lambda},\bar{\nu}}(-\mathbf{p})G_{-\omega}^{\bar{\nu},\bar{\gamma}}(-\mathbf{p})\\
&= \delta_{\bar{\lambda},\bar{\gamma}}-[(\hat{G}_{-\omega}^{(0)})^{-1}\hat{G}_{-\omega}(-\mathbf{p})]_{\bar{\gamma}}^{\bar{\lambda}}.
\end{aligned}
\tag{19b}
$$

Multiplying relation (16) by $[G_{-\omega}(-\mathbf{p})]_{\bar{\gamma}}^{\bar{\beta}}$ and relation (17) by $[G_{\omega}(\mathbf{p})]_{\gamma}^{\beta}$, we will be able to exclude the sums over the number of transitions containing hopping integrals $t^{\nu,\beta}(\mathbf{p})$ and $t^{-\bar{\lambda},\bar{\nu}}(-\mathbf{p})$.

Excluding these sums and ignoring nonlogarithmic terms containing the first powers of Green functions, we arrive at the following system of equations:

$$
\begin{aligned}
\Gamma_{\alpha\bar{\alpha};\,\nu\bar{\nu}}(\mathbf{p}) &= g_{\alpha\bar{\alpha};\,\nu\bar{\nu}}(\mathbf{p})\\
&+T\sum_{\omega,\,\mathbf{p}',\,\beta,\,\bar{\beta}}\left\{\frac{1}{f_{\alpha}}[(\hat{G}_{\omega}^{(0)})^{-1}\hat{G}_{\omega}(\mathbf{p})]_{\gamma}^{\alpha}[\hat{G}_{-\omega}(-\mathbf{p})]_{\bar{\gamma}}^{\bar{\alpha}}\right.\\
&\left.+\frac{1}{f_{\bar{\alpha}}}[(\hat{G}_{-\omega}^{(0)})^{-1}\hat{G}_{-\omega}(-\mathbf{p})]_{\bar{\gamma}}^{\bar{\alpha}}[\hat{G}_{\omega}(\mathbf{p})]_{\gamma}^{\alpha}\right\}\Gamma_{\gamma\bar{\gamma};\,\nu,\,\bar{\nu}}(\mathbf{p}').
\end{aligned}
\tag{20}
$$

To determine the conditions for the emergence of Cooper instability, it is sufficient to consider the system of homogeneous equations corresponding to inhomogeneous system (20):

$$
\begin{aligned}
\Gamma_{\alpha\bar{\alpha}} &= T\sum_{\omega,\,\mathbf{p}',\,\gamma,\,\bar{\gamma}}\left\{\frac{1}{f_{\alpha}}[(\hat{G}_{\omega}^{(0)})^{-1}\hat{G}_{\omega}(\mathbf{p})]_{\gamma}^{\alpha}[\hat{G}_{-\omega}(-\mathbf{p})]_{\bar{\gamma}}^{\bar{\alpha}}\right.\\
&\left.+\frac{1}{f_{\bar{\alpha}}}[(\hat{G}_{-\omega}^{(0)})^{-1}\hat{G}_{-\omega}(-\mathbf{p})]_{\bar{\gamma}}^{\bar{\alpha}}[\hat{G}_{\omega}(\mathbf{p})]_{\gamma}^{\alpha}\right\}\Gamma_{\gamma\bar{\gamma}}.
\end{aligned}
\tag{21}
$$

In zero magnetic field, all matrix elements are independent of the value of the spin projection. The reciprocal zeroth Green function has only diagonal matrix elements $[(\hat{G}_{-\omega}^{(0)})^{-1}]_{\alpha}^{\alpha} = i\omega - \epsilon_{\alpha}$; consequently, in the framework of the logarithmic approximation used here, the problem of determining the superconducting transition temperature is reduced to writing the solvability condition of the homogeneous system of algebraic equations

$$
\Gamma_{\alpha} = -2\frac{\epsilon_{\alpha}}{f_{\alpha}}T\sum_{\omega,\,\mathbf{p}',\,\gamma}[\hat{G}_{\omega}(\mathbf{p})]_{\gamma}^{\alpha}[\hat{G}_{-\omega}(-\mathbf{p})]_{\gamma}^{\alpha}\Gamma_{\gamma}.
\tag{22}
$$

Thus, the problem of determining the superconducting transition temperature is reduced to determining the solvability condition for system of algebraic equations (22).

The kernel of this equation has strong degeneracy in indices α and γ. Direct calculations show that the final solvability condition does not contain the products of nondiagonal minors of the matrix of the operator on the right-hand side of Eq. (22). As regards the diagonal components, the solvability condition can be expressed in terms of the simplest sum of diagonal components:

$$
1 = -2\sum_{\alpha}\frac{\epsilon_{\alpha}}{f_{\alpha}}T\sum_{\omega,\,\mathbf{p}}[\hat{G}_{\omega}(\mathbf{p})]_{\alpha}^{\alpha}[\hat{G}_{-\omega}(-\mathbf{p})]_{\alpha}^{\alpha}.
\tag{23}
$$

It is convenient to express the diagonal components of the one-particle Green functions in terms of normal

coordinates and eigenvalues of the corresponding reciprocal matrix:

$$[\hat{G}_\omega(\mathbf{p})]_k^k = \sum_\lambda \frac{A_k^{(\lambda)}(\mathbf{p})}{i\omega_n - \xi_\mathbf{p}^{(\lambda)}}, \quad \xi_\mathbf{p}^{(\lambda)} = \epsilon_\mathbf{p}^\lambda - \mu, \qquad (24)$$

where $\epsilon_\mathbf{p}^\lambda$ are the eigenvalues of reciprocal matrix (4).

In substituting this expansion into Eq. (23), it should be noted that, to within the logarithmic accuracy adopted here, we must omit the terms with different number of pole addends:

$$\begin{aligned} 1 &= -2\sum_{k\lambda}\frac{\epsilon_k}{f_k}T\sum_{\omega,\mathbf{p}}\frac{A_k^{(\lambda)}(\mathbf{p})A_k^{(\lambda)}(-\mathbf{p})}{\omega_n^2 + (\xi_\mathbf{p}^{(\lambda)})^2} \\ &= -\sum_{k\lambda}\frac{\epsilon_k}{f_k}\sum_\mathbf{p}\frac{A_k^{(\lambda)}(\mathbf{p})A_k^{(\lambda)}(-\mathbf{p})}{\xi_\mathbf{p}^{(\lambda)}}\tanh\left(\frac{\xi_\mathbf{p}^{(\lambda)}}{2T_c}\right). \end{aligned} \qquad (25)$$

Thus, for a given number of the energy band, we have a set of singular terms determining the superconducting transition temperature. In other words, in the metal conduction range, for a given value of the chemical potential, a fixed number λ exists, for which the condition $\xi_\mathbf{p}^{(\lambda)} = 0$ defines the Fermi surface.

Separating such terms, we can rewrite the Cooper instability condition for the given number of energy band λ:

$$1 = -\sum_k\frac{\epsilon_k}{f_k}\sum_\mathbf{p}\frac{A_k^{(\lambda)}(\mathbf{p})A_k^{(\lambda)}(-\mathbf{p})}{\xi_\mathbf{p}^{(\lambda)}}\tanh\left(\frac{\xi_\mathbf{p}^{(\lambda)}}{2T_c}\right). \qquad (26)$$

This relation can be written in the classical BCS form: $T_c = \bar{\epsilon}\exp(-1/\Lambda)$, where

$$\Lambda = -2\sum_k\frac{\epsilon_k}{f_k}\sum_\mathbf{p}A_k^{(\lambda)}(\mathbf{p})A_k^{(\lambda)}(-\mathbf{p})\delta(\xi_\mathbf{p}^\lambda). \qquad (27)$$

In our simplest case, when the lower Hubbard subband is populated ($h_d < 1$), the normal coordinates and the eigenfunctions can be determined using relations (9). As a result, for a given λ, we can determine the BCS constant and the condition for the existence of superconductivity at $T = 0$: $\Lambda_{(\lambda)} > 0$, where

$$\begin{aligned} \Lambda_{(\lambda)} &= -2\frac{\epsilon_d}{f_d}\sum_\mathbf{p}(A_k^{(-\lambda)}(\mathbf{p}))^2\delta(\xi_\mathbf{p}^{(\lambda)}) \\ &\quad -\frac{\epsilon_p}{f_p}\sum_\mathbf{p}(A_k^{(\lambda)}(\mathbf{p}))^2\delta(\xi_\mathbf{p}^{(\lambda)}). \end{aligned} \qquad (28)$$

It can easily be traced that, as we pass to filling of the upper Hubbard subband, the superconductivity condition has the same form (28), but with the reverse sign of ϵ_d:

$$\begin{aligned} \Lambda_{(\lambda)} &= 2\frac{\epsilon_d}{f_d}\sum_\mathbf{p}(A_k^{(-\lambda)}(\mathbf{p}))^2\delta(\xi_\mathbf{p}^{(\lambda)}) \\ &\quad -\frac{\epsilon_p}{f_p}\sum_\mathbf{p}(A_k^{(\lambda)}(\mathbf{p}))^2\delta(\xi_\mathbf{p}^{(\lambda)}). \end{aligned} \qquad (29)$$

In both cases, $\lambda = -1$ for filling of the lower hybridization subband and $\lambda = +1$ for filling of the upper hybridization subband.

The condition for the existence of superconductivity for independent filling of the nonhybridized *p* band has the form $\epsilon_p < 0$.

5. EMERY–HIRSH MODEL

Let us consider the most simplified model, in which we can disregard direct jumps w_p and w_d altogether as compared to the jumps between 3d electrons of copper and $p_{x,y}$ electrons of oxygen.

Such a formulation of the problem was proposed by Emery [3], as well as Hirsh [4], who studied the system with preset one-particle energies ϵ_p and ϵ_d, which corresponds to the Hamiltonian

$$\begin{aligned} \hat{H} &= -t\sum_{\mathbf{r},\mathbf{r}',\lambda,\sigma}[\hat{p}^+_{\mathbf{r},\sigma,\lambda}\hat{a}_{\mathbf{r}',\sigma} + \text{h.c.}] \\ &\quad + \epsilon_p\sum_{\mathbf{r},\lambda,\sigma}\hat{p}^+_{\mathbf{r},\sigma,\lambda}\hat{p}_{\mathbf{r},\sigma,\lambda} + \epsilon_d\sum_{\mathbf{r},\sigma}\hat{a}^+_{\mathbf{r},\sigma}\hat{a}_{\mathbf{r},\sigma}. \end{aligned} \qquad (30)$$

Here, $\hat{p}^+_{\mathbf{r},\sigma,\lambda}$ and $\hat{a}^+_{\mathbf{r},\sigma}$ are the creation operators for hole *p*- and *d*-excitations in the $\mathbf{r}$ cell with spin projection σ and with orbital constants $\lambda = x, y$.

We assume that the energy of electrostatic repulsion of electrons belonging to the same atom is high as compared to the energy $\epsilon_{p,d}$ of one-electrons states and as compared to hopping integral *t*.

The equation for determining the one-particle excitation spectrum can be derived from the poles of four-component Green functions, which will be defined in terms of the reciprocal matrix

$$G_\omega^{-1}\mathbf{p} = \begin{matrix}(0;\sigma)\\(\sigma;2)\\(0;j,\sigma)\end{matrix}\begin{pmatrix}\Omega_0 & 0 & -\tau_\mathbf{p}^{(0,k)}\\ 0 & \Omega_2 & -\tau_\mathbf{p}^{(2,k)}\\ -\tau_\mathbf{p}^{(j,0)} & -\tau_\mathbf{p}^{(j,2)} & \delta_{(j,k)}\Omega_p\end{pmatrix}. \qquad (31)$$

For the CuO_2 layer, we have

$$\tau_\mathbf{p}^{(\alpha,k)} = tf_\alpha(1-\exp(ip_k)),$$

$$\tau_\mathbf{p}^{(j\alpha)} = tf_j(1-\exp(-ip_j)).$$

Crystallographic indices k and j independently run through two values (x and y). For the particular case of a square unit cell, the end factors for p electrons are independent of the crystallographic index $f_j = f_p = 1 - 3h_p/4$. As regards the $(x^2 - y^2)$) hole states, these states differ substantially for the lower and upper Hubbard subbands; consequently, we have

$$f_0 = 1 - h_d/2, \quad f_2 = h_d/2.$$

In the zero-loop Hubbard I approximation, the average occupation numbers are defined self-consistently in terms of the products of the virtual Green function at coinciding points and the corresponding end factors. For example, for the average number of $(x^2 - y^2)$ holes, we have

$$h_d = 2T\sum_{\omega, \mathbf{p}} \exp(i\omega\delta)\Omega_p(f_0\Omega_2 + f_2\Omega_0)\Phi^{-1}. \quad (32)$$

For the average number of p holes, we can write

$$h_p = f_p\left\{3n_F(\epsilon_p) + T\sum_{\omega, \mathbf{p}} \exp(i\omega\delta)\Omega_0\Omega_2\Phi^{-1}\right\}. \quad (33)$$

The determinant of matrix (31) gives the equation for the elementary excitation spectrum $\Phi_{-i\omega}(\mathbf{p}) = 0$, where

$$\Phi = \Omega_0\Omega_2\Omega_p - f_p t_\mathbf{p}^2(f_0\Omega_2 + f_2\Omega_0),$$
$$t_\mathbf{p}^2 = 2t^2\left(2 - \sum_k \cos p_k\right). \quad (34)$$

Introducing chemical potential μ, we obtain

$$\Omega_0 = i\omega + \frac{U}{2} + \mu, \quad \Omega_2 = i\omega - \frac{U}{2} + \mu,$$
$$\Omega_p = i\omega - \epsilon_p. \quad (35)$$

In the case of extremely high energy ϵ_p, when $\Omega_p \longrightarrow -\infty$, we obtain the equations for the excitation spectrum, which can be reduced to the classical Hubbard model with the effective hopping integral $-\tau_\mathbf{p} = t_\mathbf{p}^2/\epsilon_p$.

In the limiting case of the extremely high Hubbard energy of d electrons ($\Omega_2 \longrightarrow -\infty$), we arrive at the version with simultaneous population of the p and d lower Hubbard subband with the spectrum

$$\xi_\mathbf{p}^\pm = \pm\sqrt{\frac{r_0^2}{4} + f_0 f_p t_\mathbf{p}^2} - \mu_0, \quad h_d < 1, \quad (36)$$

where

$$r_0 = \epsilon_p + \frac{U}{2}, \quad \mu_0 = \mu - \frac{\epsilon_p}{2}.$$

In this limit, the equations of state can be written in the form

$$h_p = f_p\left\{3n_F(\epsilon_p) + \sum_{\mathbf{p};\,\lambda=\pm}[A^{(\lambda)}(\mathbf{p})n_F(\xi^{(\lambda)}(\mathbf{p}))]\right\}, \quad (37a)$$

$$h_d = 2f_d\sum_{\mathbf{p};\,\lambda=\pm} A^{(-\lambda)}(\mathbf{p})n_F(\xi^{(\lambda)}(\mathbf{p})). \quad (37b)$$

If the lower Hubbard band is filled and the upper is populated simultaneously with the p states, we have $\Omega_0 \longrightarrow +\infty$, and we again have two branches,

$$\xi_\mathbf{p}^\pm = \pm\sqrt{\frac{r_2^2}{4} + f_2 f_p t_\mathbf{p}^2} - \mu_2, \quad 1 < h_d < 2, \quad (38)$$

where

$$r_2 = \epsilon_p - \frac{U}{2}, \quad \mu_2 = \mu - \frac{\epsilon_p}{2}.$$

In this limit, the equations of state have the form

$$h_p = f_p\left\{3n_F(\epsilon_p) + \sum_{\mathbf{p};\,\lambda=\pm}[A^{(\lambda)}(\mathbf{p})n_F(\xi^{(\lambda)}(\mathbf{p}))]\right\}, \quad (39a)$$

$$h_d = 1 + f_d\sum_{\mathbf{p};\,\lambda=\pm} A^{(-\lambda)}(\mathbf{p})n_F(\xi^{(\lambda)}(\mathbf{p})). \quad (39b)$$

In both limiting cases, the normal coordinates are independent of the chemical potential and are defined as

$$A_\mathbf{p}^{(\pm)} = \frac{1}{2}\left[1 \pm \frac{r}{\sqrt{r^2 + 4f_p f_d t_\mathbf{p}^2}}\right]. \quad (40)$$

The condition for the emergence of the superconducting state is determined from the requirement of the emergence of singularity in the two-particle vertex part $\Gamma_{\alpha,\beta}$ for zero total frequency, momentum, and spin.

Using the ladder approximation, we obtain a linear inhomogeneous integral equation for $\Gamma_{\alpha,\beta}$. The point of emergence of a singularity corresponds to the possibility of solving the system of homogeneous equations, which corresponds to the system of integral equations (11).

For any two Hubbard operators $\hat{X}_\mathbf{r}^\alpha$, we have the identity

$$[\hat{H}, \hat{X}_\mathbf{r}^\alpha\hat{X}_\mathbf{r'}^\beta] = -\hat{X}_\mathbf{r'}^\beta[\hat{H}, \hat{X}_\mathbf{r}^\alpha]$$
$$+ \hat{X}_\mathbf{r}^\alpha[\hat{H}, \hat{X}_\mathbf{r'}^\beta] + \{[\hat{H}, \hat{X}_\mathbf{r}^\alpha]\hat{X}_\mathbf{r'}^\beta\}. \quad (41)$$

If we take into consideration that $\hat{X}_\mathbf{r}^\alpha$ are the creation operators for one-hole Fermi excitations and act on the wave function of the ground state by the left-hand side

of Eq. (41), we obtain the energy of two-hole excitations plus the additional energy stemming from the third term on the right-hand side of Eq. (41). This term precisely determines the Born amplitudes of the kinematic interaction.

Let us express Hamiltonian (30) in terms of the Hubbard operators,

$$\hat{p}^{+}_{\mathbf{r},\sigma,\lambda} = \hat{X}^{(\lambda,0)}_{\mathbf{r},\sigma}, \quad \hat{\alpha}^{+}_{\mathbf{r},\sigma} = \hat{X}^{(\sigma,0)}_{\mathbf{r}} + \sigma\hat{X}^{(2,-\sigma)}_{\mathbf{r}}, \qquad (42)$$

and then substitute it into the last term of relation (41). As a result, we obtain the following nonzero amplitudes for scattering with opposite values of spin projection σ:

$$\begin{aligned}
\{[\hat{H}, \hat{X}^{(\sigma,0)}_{\mathbf{r}}]\hat{X}^{(-\sigma,0)}_{\mathbf{r}'}\} &= -\delta_{\mathbf{r},\mathbf{r}'} \\
\times\sum_{r_2,\lambda} t_\lambda(\mathbf{r}_2,\mathbf{r})(\hat{X}^{(\lambda,0)}_{\mathbf{r}_2,\sigma}\hat{X}^{(-\sigma,0)}_{\mathbf{r}} - \hat{X}^{(\tilde{\lambda},0)}_{\mathbf{r}_2,-\sigma}\hat{X}^{(\sigma,0)}_{\mathbf{r}}); \\
\{[\hat{H}, \hat{X}^{(\lambda,0)}_{\mathbf{r},\sigma}]\hat{X}^{(\tilde{\lambda},0)}_{\mathbf{r}',-\sigma}\} &= -\delta_{\mathbf{r},\mathbf{r}'}\delta_{\lambda,\lambda r'} \\
\times\sum_{r_2} t_\lambda(\mathbf{r}_2,\mathbf{r})(\hat{X}^{(\sigma,0)}_{\mathbf{r}_2}\hat{X}^{(\lambda,0)}_{\mathbf{r},-\sigma} - \hat{X}^{(-\sigma,0)}_{\mathbf{r}_2}\hat{X}^{(\lambda,0)}_{\mathbf{r},\sigma}); \\
\{[\hat{H}, \hat{X}^{(2,-\sigma)}_{\mathbf{r}}]\hat{X}^{(2,\sigma)}_{\mathbf{r}'}\} &= \delta_{\mathbf{r},\mathbf{r}'}\sum_{r_2,\lambda} t_\lambda(\mathbf{r}_2,\mathbf{r}) \\
\times(\hat{X}^{(\lambda,0)}_{\mathbf{r}_2,\sigma}\hat{X}^{(2,\sigma)}_{\mathbf{r}} - \hat{X}^{(\tilde{\lambda},0)}_{\mathbf{r}_2,-\sigma}\hat{X}^{(2,-\sigma)}_{\mathbf{r}}).
\end{aligned} \qquad (43)$$

Passing to the momentum representation $\hat{X}^{\alpha}_{\mathbf{r}} = \sum_{\mathbf{p}}\hat{X}^{(\alpha)}_{\mathbf{p}}\exp(i\mathbf{p}\cdot\mathbf{r})$, we find that each term on the right-hand side of Eq. (43) gives the Born amplitude of the kinematic interaction $g_{\alpha,\beta,\lambda,\nu}(\mathbf{p})$. Pairs (α, β) correspond to the initial sates on the left-hand side of Eq. (43):

$$[(-\sigma,0),(\sigma,0)]; \quad [(\lambda,0),(-\lambda,0)]; \quad [(2,-\sigma),(2,\sigma)].$$

Indices λ and ν correspond to all possible two-particle states on the right-hand side of Eq. (43).

Combining all possible pairs of transitions, we obtain six algebraic equations in all. This simplification stems from the fact that the amplitudes of kinematic interaction (41) differ from zero only for transitions with identical crystallographic indices λ and for coinciding cells (see Fig. 1).

For unknown vertex parts Γ_{00}, Γ_{22}, Γ_{20}, Γ_{02}, and $\Gamma_{\lambda\lambda}$, we have three relations,

$$\Gamma_{\lambda\lambda} = \Gamma_p, \quad 2\Gamma_{02} = \Gamma_{00} - \Gamma_{22} = -2\Gamma_{20}. \qquad (44)$$

The solvability condition for the resulting equations can be written so that the following relation holds between vertices Γ_{00} and Γ_{22}: $\Gamma_{00}/\Gamma_{22} = -f_0\xi_2^2/f_2\xi_0^2$.

The final system of equations for Γ_p and $\Gamma = \Gamma_{00}\xi_2^2/f_0 = -\Gamma_{22}\xi_2^2/f_0$ has the form

$$\begin{aligned}
\Gamma &= 2T\sum_{\omega,\mathbf{p}} t_{\mathbf{p}}^2\Big[t_{\mathbf{p}}^2 f_p^2(\Omega_2-\Omega_0)\Big(\frac{f_0\Omega_2}{\Omega_0}+\frac{f_2\Omega_0}{\Omega_2}\Big)\Gamma \\
&\quad + \Omega_p\Omega_0^2\Omega_2^2\Gamma_p\Big]|\Phi_\omega(\mathbf{p})|; \\
\Gamma_p &= 2f_pT\sum_{\omega,\mathbf{p}}\Big[t_{\mathbf{p}}^2 f_p^2(\Omega_2-\Omega_0)\Big(\frac{f_0\Omega_2}{\Omega_0}+\frac{f_2\Omega_0}{\Omega_2}\Big)\Gamma \\
&\quad + \frac{\Omega_p\Omega_0^2\Omega_2^2}{f_p^2}\Gamma_p\Big]|\Phi_\omega(\mathbf{p})|,
\end{aligned} \qquad (45)$$

where

$$\Phi_\omega(\mathbf{p}) = [\Omega_0\Omega_2\Omega_p - (\Omega_0 f_2 + \Omega_2 f_0) f_p t_{\mathbf{p}}^2]^{-2}.$$

The solvability conditions for the system of equations derived above has the form

$$\begin{aligned}
&2f_pT\sum_{\omega,\mathbf{p}}\Big[t_{\mathbf{p}}^4 f_p(\Omega_2-\Omega_0)\Big(\frac{f_0\Omega_2}{\Omega_0}+\frac{f_2\Omega_0}{\Omega_2}\Big) \\
&\quad + \Omega_p\frac{\Omega_0^2\Omega_2^2}{2f_p^2}\Big]|\Phi_\omega(\mathbf{p})| = 1.
\end{aligned} \qquad (46)$$

Taking into account condition $\Omega_0 - \Omega_2 = U$ and assuming that

$$t_{\mathbf{p}}^2 f_0(f_0\Omega_2 + f_2\Omega_0) \approx \Omega_0\Omega_2\Omega_p,$$
$$\Omega_\nu \approx -\xi_\nu \quad (\nu = 0, 2, p),$$

we obtain the conditions for the existence of the superconducting state:

$$2Uf_p\frac{\xi_p^2(f_0\xi_2^2 + f_2\xi_0^2)}{\xi_0\xi_2} + \xi_p(f_2\xi_0 + f_0\xi_2)^2 < 0. \qquad (47)$$

In the limiting case $\xi_p \gg U$, we derive from relation (47) the BCS formula $T_c \propto \exp(-1/\Lambda_0)$, which contains constant Λ_0, coinciding with the general expression derived in our recent publication [2]:

$$\Lambda_0 = 2\rho_0\frac{U(U^2-4\mu^2)}{[U(1-n)-2\mu]^2}, \qquad (48)$$

where $1 - n = x$ is the extent of underpopulation of the lower Hubbard subband and ρ_0 is the initial density of states with the effective hopping integral $-t_p^2/\epsilon_p$.

The case when the excitation energy lies within the correlation gap is of special importance. In the limiting

case $U \longrightarrow \infty$, when the lower Hubbard subbands are populated simultaneously both for d and for p excitations, we have spectrum (36), $n_{p,d} < 1$, and superconductivity condition (47) can be transformed to [9]

$$\frac{\epsilon_p}{f_0} + \frac{\epsilon_d}{2f_p} < 0, \quad r = \epsilon_p - \epsilon_d, \quad \mu = -\frac{\epsilon_p + \epsilon_d}{2}. \quad (49)$$

It should be noted that this condition is satisfied only if the upper $\xi_{\mathbf{p}}^{(+)}$ subband (36) is filled. Using the equations of state in form (37), we observe that superconductivity can exist under these conditions only for values of h_p sufficiently close to unity:

$$4(8 - 5h_d)/(32 - 19h_d) < h_p < 1, \quad 0 < h_d < 1. \quad (50)$$

If the lower d Hubbard subband is filled and the p states are populated simultaneously with the upper subband $1 < h_d < 2$, we have $\xi_0 \approx -U \longrightarrow -\infty$, while superconductivity condition (47) is transformed to a condition analogous to (49):

$$-\frac{\epsilon_p}{f_2} + \frac{\epsilon_d}{2f_p} < 0, \quad r = \epsilon_p - \epsilon_d, \quad \mu = -\frac{\epsilon_p + \epsilon_d}{2}. \quad (51)$$

In this region, superconductivity can set in both during filling of $\xi_{\mathbf{p}}^{(-)}$ and during filling of $\xi_{\mathbf{p}}^{(+)}$ (see Fig. 2).

If, however, the energy of p states exceeds the energy of one-particle states in the upper d subband ($\epsilon_p > \epsilon_d$), superconductivity exists for $h_d > 4/3$ in the entire $\xi_{\mathbf{p}}^{(-)}$ subband:

$$0 < h_p < 4(2 - h_d)/(6 + h_d), \quad 4/3 < h_d < 2. \quad (52)$$

In the case when the energy of p states lies within the correlation gap ($1 < h_d < 4/3$), $\epsilon_p < \epsilon_d$, the boundary in h_p is determined by relation (52) and is observed only for the lower part of the $\xi_{\mathbf{p}}^{(-)}$ subband.

The corresponding curve $h_p = h_p(h_d)$, which was obtained in the plane band model, begins from point $h_d = 1$, $h_p = 0$, attains its maximum value $h_p^{\max} = 0.39$ for $h_d \approx 1.23$, and is transformed into curve (52) at point $h_d = 4/3$, $h_p = 4/11$ (Fig. 2).

In this region, the number of p excitations is quite small, their scattering amplitude is positive, and superconductivity is due to the kinematic interaction of d excitations. As the number of p excitations increases; their "repulsive" role increases, setting a limit in h_p on the region of existence of superconducting solutions.

It should also be noted that superconductivity condition (49) is also observed for $h_d > 4/3$, $h_p \approx 1$, i.e., for filling of the upper subband (38). However, in this concentration range, the charge $Q = 2h_p + h_d - 3$ of the CuO_2 complex is positive everywhere, which makes this region uninteresting for physical analysis.

An important property of the phase diagram obtained here is the absence of symmetry relative to the particle–hole transformation $h_d \longrightarrow 2 - h_d$ for a preset number h_p of p holes. Superconductivity turns out to be most effective in the region $h_d > 1$, while in the region $h_d < 1$, $h_p < 0.9$ it is absent altogether.

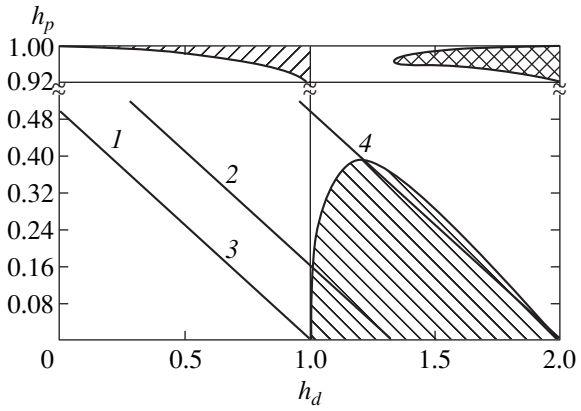


Fig. 2. Phase diagram for the CuO_2 layer at $T = 0$, calculated in the limit of an infinitely large Hubbard energy and with a rectangular density of states. The following electroneutrality lines are shown: $2h_p + h_d = 1$ (Ln_2CuO_4) (*1*), $2h_p + h_d = 4/3$ ($YBa_2Cu_3O_7$) (*2*); $2h_p + h_d = 1$ ($YBa_2Cu_3O_{6.5}$ (*3*), and $2h_p + h_d = 2$ (Sr_2RuO_4) (*4*).

This situation qualitatively corresponds to the experiment on $Ln_{2-x}^{3+}Me_xCuO_4$. When this compound is doped with bivalent Me^{2+} cations at $T = 0$, superconductivity appears and exists in a wide concentration range of $0.04 < x < 2/3$.

In the case of yttrium-based $Y^{3+}Ba_2Cu_3O_{7-\delta}$ compounds, the value of h_d is greater than or equal to unity for all values of $\delta \le 1/2$ attainable in experiments; this corresponds to hole-type superconductivity. The stoichiometric compound $YBa_2Cu_3O_7$ has the highest possible T_c, which is in qualitative agreement with our results.

For the ruthenium-based compound Sr_2RuO_4, the electroneutrality line passes near the right boundary of the superconducting phase; this explains the fairly low superconducting transition temperature ($T_c \approx 1$ K).

In the case of doping with tetravalent Me^{4+} cations, superconductivity is observed in a narrow concentration range of $0.14 < x < 0.18$. The existence of this effect (so-called electron-type superconductivity), which is not manifested in the Emery model, should be attributed to direct hopping between copper cations.

6. PHASE DIAGRAM BASED ON GENERALIZED *p–d* MODEL

To obtain a phase diagram of the general form, which would be applicable for comparison with experiments associated with doping with trivalent and tetravalent cations, we must take into account not only

direct oxygen–oxygen jumps w_p, but also jumps between copper cations w_d.

To construct the phase diagram, it is sufficient to note that the condition $\Lambda = 0$ establishes the relation between the energy difference $r = \epsilon_p - \epsilon_d$ and the chemical potential $-(\epsilon_p + \epsilon_d)/2$. Substitution of this relation into equation of state (10) leads to the emergence of a dependence between h_p and h_d, which corresponds to the superconducting phase diagram.

For subsequent calculations, it is convenient to introduce instead of two end factors f_p and f_d the dimensionless variables u and v:

$$u = \frac{w_p f_p - w_d f_d}{|r|}, \quad v = \frac{4w^2 f_p f_d}{r^2}. \tag{53a}$$

From these equations, we obtain the quantities $w_p f_p$ and $w_d f_d$:

$$\begin{aligned} w_p f_p &= |r|\left\{\frac{u}{2} + \frac{1}{2}\sqrt{u^2 + av}\right\}, \\ w_d f_d &= |r|\left\{-\frac{u}{2} + \frac{1}{2}\sqrt{u^2 + av}\right\}, \end{aligned} \tag{53b}$$

where $a = w_p w_d / w^2$.

Instead of the chemical potential at $T = 0$ we will use a dimensionless variable y, which can be determined from the condition $\xi^{(\pm)}(y) = 0$.

This relation, which is written in new variables using the explicit form of excitation spectrum (7), has the form

$$\begin{aligned} \mu &= -\frac{\epsilon_p + \epsilon_d}{2} = -y\frac{w_p f_p + w_d f_d}{2} \\ &\pm \frac{|r|}{2}\sqrt{(1 + uy\,\mathrm{sgn}\, r)^2 + v(1-y)}. \end{aligned} \tag{54}$$

As a result, the chemical potential, as well as ϵ_p and ϵ_d, are functions of variables u, v, r, y.

After elementary calculations using formulas (53) and (54), we obtain $\epsilon_p = -\mu + r/2$ and $\epsilon_d = -\mu - r/2$, or

$$\begin{aligned} \epsilon_p &= -\frac{y|r|}{2}\sqrt{u^2 + av} \\ &\mp \frac{|r|}{2}\sqrt{(1 + uy\,\mathrm{sgn}\, r)^2 + v(1-y)} + \frac{r}{2}, \\ \epsilon_d &= -\frac{y|r|}{2}\sqrt{u^2 + av} \\ &\mp \frac{|r|}{2}\sqrt{(1 + uy\,\mathrm{sgn}\, r)^2 + v(1-y)} - \frac{r}{2}, \end{aligned} \tag{55}$$

where $a = w_p w_d / w^2$ and the upper and lower sign correspond to the upper and lower hybridization subbands, respectively.

The end factors and average occupation numbers can also be written in terms of variables u, v, y. For example, for filling of the lower hybridization subband and at $T = 0$, we have

$$\begin{aligned} h_p &= \frac{4K_{(-)}}{4 + 3K_{(-)}}, \quad f_p = \frac{4}{4 + 3K_{(-)}} \\ &(0 < h_p < 1), \\ h_d &= \frac{2K_{(+)}}{1 + K_{(+)}}, \quad f_d = \frac{1}{1 + K_{(-)}} \\ &(0 < h_d < 1), \\ h_d &= \frac{2}{2 - K_{(+)}}, \quad f_d = \frac{1}{2 + K_{(+)}} \\ &(1 < h_d < 2). \end{aligned} \tag{56}$$

Quantities $K_\pm$ introduced here are generally functions of four parameters (u, v, y, r). However, in the limiting case when $T = 0$, quantities $K_\pm$ depend only on three variables (u, v, y).

In the limit $T = 0$ and in the case of filling of the lower hybridization subband, we have the following definitions:

$$K_{(\pm)}(u, v, y) = \int_{-1}^{y} A^{(\pm)}(u, v, x)\rho_0(x)dx. \tag{57}$$

Here, we have introduced the normalized initial density of states $\rho_0(\epsilon)$; the normal coordinates are found to be independent of the energy difference $r = \epsilon_p - \epsilon_d$:

$$\begin{aligned} A^{(\pm)}(u, v, x) &= \frac{1}{2} \\ &\times \left[1 \pm \mathrm{sgn}\, r\frac{1 + ux\,\mathrm{sgn}\, r}{\sqrt{(1 + ux\,\mathrm{sgn}\, r)^2 + v(1-x)}}\right]. \end{aligned} \tag{58}$$

In the simplest case of a plane band, when $\rho_0(\epsilon) = \theta(1 - \epsilon^2)/2$, at $T = 0$, we find the following expressions for functions $K_\pm(u, v, y)$:

$$\begin{aligned} K_\pm(u, v, y) &= \frac{1}{4}\left\{\pm\frac{v}{2}\right. \\ &\times \ln\left[\frac{2u(1+uy) - v + 2|u|\sqrt{(1+uy)^2 + v(1-y)}}{2u(1-u) - v + 2|u|\sqrt{(1-u)^2 + 2v}}\right] \\ &+ (1+y) \pm \frac{1}{u}\left[\sqrt{(1+uy)^2 + v(1-y)}\right. \\ &\left.\left. - \sqrt{(1-u)^2 + 2v}\right]\right\}. \end{aligned} \tag{59}$$

Thus, for preset values of three dimensionless energies $w_p/|r|$, $w_d/|r|$, and $a = w_p w_d / w^2$, Eqs. (54) together

with relations (56)–(59) make it possible to determine variables u, v, as well as the mean occupation numbers h_d, h_p, as functions of only one parameter y.

To construct the superconducting phase diagram, we first eliminate variable $|r|$ from Eqs. (53):

$$w_d f_d[u + \sqrt{u^2 + av}] = w_p f_p[-u + \sqrt{u^2 + av}]. \quad (60)$$

Further, we use condition (29) for the emergence of superconductivity for the lower hybridization subband: $\Lambda_{(-)} = 0$. When written in variables ϵ_p/r and ϵ_d/r, this condition will depend only on u, v, y:

$$\mp 2\frac{\epsilon_d}{|r| f_d}(A^{(+)}(u, v, y))^2 - \frac{\epsilon_p}{|r| f_p}(A^{(-)}(u, v, y))^2 = 0. \quad (61)$$

Here, the upper and lower signs correspond to regions $0 < h_d < 1$ and $1 < h_d < 2$, respectively.

The explicit dependence of quantities ϵ_p/r and ϵ_d/r on u, v, y, which is obtained from the definition of chemical potential, has the form

$$\frac{\epsilon_p}{|r|} = -\frac{y}{2}\sqrt{u^2 + av} \mp \frac{1}{2}\sqrt{(1 + uy\,\mathrm{sgn}\,r)^2 + v(1-y)} + \frac{\mathrm{sgn}\,r}{2},$$
$$\frac{\epsilon_d}{|r|} = -\frac{y}{2}\sqrt{u^2 + av} \mp \frac{1}{2}\sqrt{(1 + uy\,\mathrm{sgn}\,r)^2 + v(1-y)} - \frac{\mathrm{sgn}\,r}{2}. \quad (62)$$

Thus, the system of equations (60) and (61) makes it possible to determine two quantities as functions of the third quantity (e.g., $u = u(y)$ and $v = v(y)$).

Substituting these functions into equations of state (56)–(58) makes it possible to determine the form of the phase diagram in variables h_p, h_d (Figs. 3 and 4).

The general relation is valid provided that $h_p < 1$. Superconductivity exists when $\Lambda_k > 0$.

In the limiting case $v = 0$, the subbands are populated independently when the chemical potentials corresponding to the p and d hole states are identical. In this case, the symmetry of the phase diagram relative to the particle–hole transformation $h_d \longrightarrow 2 - h_d$ is preserved.

If we assume that direct jumps are small as compared to *p–d* jumps between oxygen anions and copper cations, we must set $u = 0$. In this limiting case, we obtain the results pertaining to the classical Emery model. The phase diagram exhibits the maximal asymmetry in this case. Superconductivity is absent in the range of small values of h_p and for $h_d < 1$. However, for a small number h_p and for $h_d > 1$, superconductivity exists in a wide range of variation of variables $1 < h_d < 2$ (see Fig. 2).

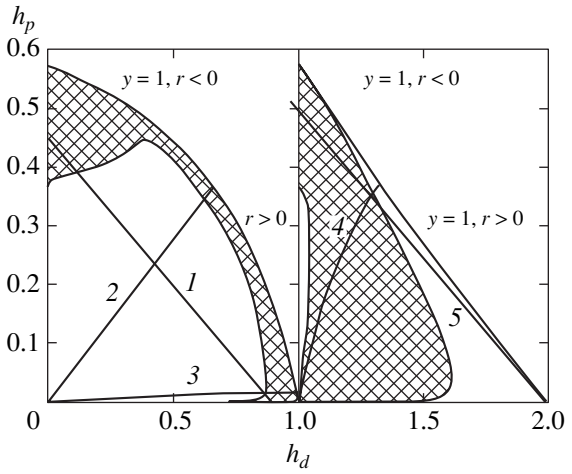


Fig. 3. Phase diagram in the generalized Emery model for hopping integrals $w = w_p = w_d$ at $T = 0$: $2h_p + h_d = 0.9$ (*1*), $r = 0$ (*2*), $r = 2.5w_p$ (*3*), $\tilde{r} = 0$ (*4*), and $2h_p + h_d = 2$ (*5*).

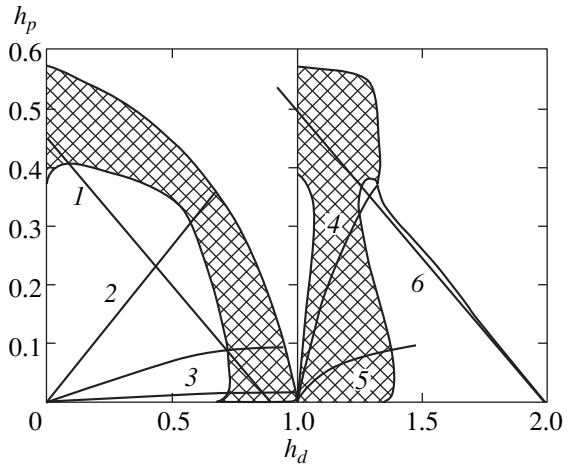


Fig. 4. Phase diagram in the generalized Emery model for hopping integrals $w = w_p = w_d$ at $T = 0$: $2h_p + h_d = 0.9$ (*1*), $r = 0$ (*2*), $r = w_p$ (*3*), $r = 0$ (*4*), $r = w_p$ (*5*), and $2h_p + h_d = 2$ (*6*).

7. RESULTS

In the previous section, we have considered the most realistic case when direct jumps between copper cations are small as compared to direct jumps between oxygen anions. Calculations are made for the plane band model.

It can be seen from Figs. 3 and 4 that, for $h_p < 1$ and $h_d < 1$, superconductivity exists only in a narrow region adjoining the band $2/3 < h_d < 1$, $h_p \ll 1$. As regards the

region $1 < h_d < 2$, superconductivity is absent only for small electron concentrations $2 - h_d \ll 1$, $h_p < 1$.

It should be noted in conclusion that the substantial dependence of T_c on the concentration of doping impurity, which is observed for the compounds being studied as well as for all high-temperature superconductor, is completely due to the strong dependence of the electron–electron scattering amplitude on the position of the Fermi level relative to the middle of the conduction band. The characteristic energy determining the value of T_c is on the order of the hopping integral equal to 10^3 K. Dimensionless constant Λ_k appearing in the BCS formula and defined by formula (29) does not exceed 1/4. It strongly depends on the number of *p* and *d* electrons scattered at energies close to the Fermi energy. All these facts are in qualitative agreement with experiments on the $Ln^{3+}_{2-x}Me_xCuO_4$, $YBa_2Cu_3O_{7-\delta}$, and Sr_2RuO_4 compounds.

8. CONCLUSIONS

This study is in fact a continuation and generalization of the previous publication [2] of the author, where the limiting case $U_{dd} \gg t_{dd}$ is considered and the scattering amplitude for two electrons with opposite spins is calculated. It was found that this amplitude is positive for small energies (in the gaseous phase); however, as the Fermi energy increases, it changes sign, which corresponds to the Ramsauer effect in the lattice. It should be noted that this result was obtained by the author in 1976 [6] using the Hubbard model for the gas phase. In the same article, the equivalence of the standard scattering theory and the method of the *X* operators corresponding to an infinitely high Hubbard energy was proved. Similar results were also obtained in the so-called *t–J* model [10], in which the scattering amplitude was calculated with the help of the *T* matrix and then a transition to the limit $U_{dd} \longrightarrow \infty$ was carried out.

The main result of the present study is the construction of the phase diagram both for the upper and the lower bands of the $(x^2 - y^2)$ electrons for a weak underpopulation of the $2p^6$ shell of oxygen anions. In this case, the main difference between the phase diagram of the Emery model and the classical Hubbard model is observed; this difference lies in partial breaking of the particle–hole symmetry of the *d* electron subsystem for a preset number of *p* holes. This result corresponds to the asymmetry observed upon a transition from lanthanum–strontium to neodymium–cerium cuprates. All these results were obtained from the Emery–Hirsh Hamiltonian [3, 4] for zero values of intercellular Coulomb matrix elements: $U_{pd} = 0$, $V_{dd} = 0$, and $V_{pp} = 0$. It is sufficient to assume that the Hubbard energy is the largest energy parameter.

It should be noted that the calculation of the superconducting transition temperature in the *t–J* model [11] is not consistent because it takes into account the second-order corrections in the hopping integral (since $J \sim t^2/U$), while the expansion of the scattering amplitude for $U \gg |t|$ begins with first-order terms in t_{pd}. The same drawback is inherent in the Zhang–Rice theory [12] in which analysis is carried out under the following two assumptions: (i) $|t_{pd}| \ll r = \epsilon_p - \epsilon_d$ and (ii) $U_{dd} \gg r$. In the present study, assumption (ii) is also used, but the expansion of the scattering amplitude is found to be linear in parameter t_{pd}.

The application of the ladder approximation for determining the conditions for the emergence of Cooper instability can be substantiated rigorously. In the Emery model, the corrections of the parquet type do not contain logarithmic divergences since the necessary nesting condition is not fulfilled in the given problem for any location of the Fermi level. The non-Born corrections to the kernel of the integral Gor'kov equations lead to the emergence of a finite relaxation time with spin flip. The corresponding reciprocal relaxation time is proportional to the second power of temperature. Hence, it can be concluded that the above-mentioned corrections are insignificant at $T = 0$. As regards the paramagnetic corrections of the Ruderman–Kittel type, their effect is substantial precisely in the region when Cooper instability is absent. The proof of this statement forms the subject of a special study.

REFERENCES

1. R. O. Zaitsev, Phys. Lett. A **134**, 199 (1988).
2. R. O. Zaitsev, Zh. Éksp. Teor. Fiz. **125**, 891 (2004) [JETP **98**, 780 (2004)].
3. V. J. Emery, Phys. Rev. Lett. **58**, 2794 (1987).
4. J. E. Hirsch, Phys. Rev. Lett. **59**, 228 (1987).
5. Y. Maeno, H. Hashimoto, K. Yoshida, *et al.*, Nature **372**, 532 (1994).
6. R. O. Zaitsev, Zh. Éksp. Teor. Fiz. **70**, 1100 (1976) [Sov. Phys. JETP **43**, 574 (1976)].
7. L. P. Gor'kov, Zh. Éksp. Teor. Fiz. **34**, 735 (1958) [Sov. Phys. JETP **7**, 505 (1958)].
8. F. Dyson, Phys. Rev. **102**, 1217 (1956); **102**, 1230 (1956).
9. R. O. Zaitsev, Solid State Commun. **76**, 795 (1990).
10. M. Yu. Kagan and T. M. Rice, J. Phys.: Condens. Matter **6**, 3771 (1994).
11. H. Fukuyama and K. Yosida, Jpn. J. Appl. Phys. Lett. **26**, 371 (1987).
12. F. C. Zhang and T. M. Rice, Phys. Rev. B **37**, 3759 (1988).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 100, No. 4, 2005, pp. 715–721.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 4, 2005, pp. 809–816.

ELECTRONIC PROPERTIES OF SOLIDS

Electric Conductivity of Inhomogeneous Two-Component Media in Two Dimensions¶

V. V. Kabanov, K. Zagar, and D. Mihailovic
Institut Jozef Stefan 1000, Ljubljana, Slovenia
e-mail: victor.kabanov@ijs.si
Received August 9, 2004

Abstract—The electric conductivity is calculated for regular inhomogeneous two-component isotropic medium in which droplets of one phase with conductivity σ_2 are embedded in another, with conductivity σ_1. An expression is formulated that can be used in many different situations and is of particular relevance in the case where the relative proportion of the components is temperature-dependent and varies over a wide range. Behavior of the effective conductivity depends on the spatial arrangements and the shape of the inclusions.

1. INTRODUCTION

Determination of the effective conductivity σ_{eff} of spatially inhomogeneous heterophase systems is an old, but increasingly important problem of theoretical physics. With the advent of new nanoscale probes of condensed matter systems, it has become apparent that many very diverse systems that were previously thought to be homogeneous are in fact either statically or dynamically inhomogeneous. The effective conductivity in such cases cannot be dealt with in terms of homogeneous medium theory, is not trivial, and solutions are presently known only in some rather special cases. Different aspects of the theory and different limiting cases are extensively discussed in [1].

In this paper, we focus on the problem of calculating the effective conductivity of an inhomogeneous two-dimensional (2*D*) plane. The classical problem can be formulated as follows. We assume that a 2*D* system contains a mixture of N ($N \geq 2$) different phases or materials with different conductivities σ_i, $i = 1, 2, \ldots, N$. The arrangements of different phases can be random or regular. The question that we wish to address is how the effective conductivity of the plane depends on the conductivities of the phases, their concentration, and the spatial arrangements.

In the past, a number of different approaches have been used to tackle this problem. The exact result for the effective conductivity of a two-component system with a symmetric and isotropic distribution of components was obtained by Dykhne [2]. He found that the effective conductivity of the system is determined by the simple relation

$$\sigma_{\text{eff}} = \sqrt{\sigma_1\sigma_2}.$$

A symmetric distribution used in this problem amounts to the case where the two components can be interchanged without changing the end result. Obviously, one requirement for a symmetric distribution is that the two components have equal proportions, but it also means that more general cases cannot be considered with this model.

Further investigations have shown that a more general duality relation is valid for 2*D* heterogeneous conductors than that initially considered by Keller and Dykhne [1]. More recently, it was shown that a more general relation for the effective conductivity tensor exists that is valid for multicomponent and anisotropic systems [3, 4]. The effective conductivity of several examples of ordered two-component systems was also calculated exactly [5–7]. It was shown in [5, 6] that for a chessboard plane and for a plane constructed of triangles, the relation derived by Dykhne is also valid.

A similar relation to the Dykhne formula for the effective conductivity of a system consisting of randomly distributed metallic and dielectric regions near a metal-to-insulator transition was derived by Efros and Shklovskii [8]. They generalized the expression of Dykhne on the basis of scaling arguments to the case of arbitrary concentrations of the two phases near a percolation threshold, such that the effective conductivity becomes

$$\sigma_{\text{eff}} = \sigma_1(\sigma_2/\sigma_1)^s, \quad (1)$$

where s is a universal scaling exponent. Critical exponents are also relatively well-known for this type of systems [9]. This relation is not applicable when the system is driven away from the percolation threshold and the general solution of the effective conductivity of an inhomogeneous medium thus remains an open problem.

¶ This article was submitted by the authors in English.

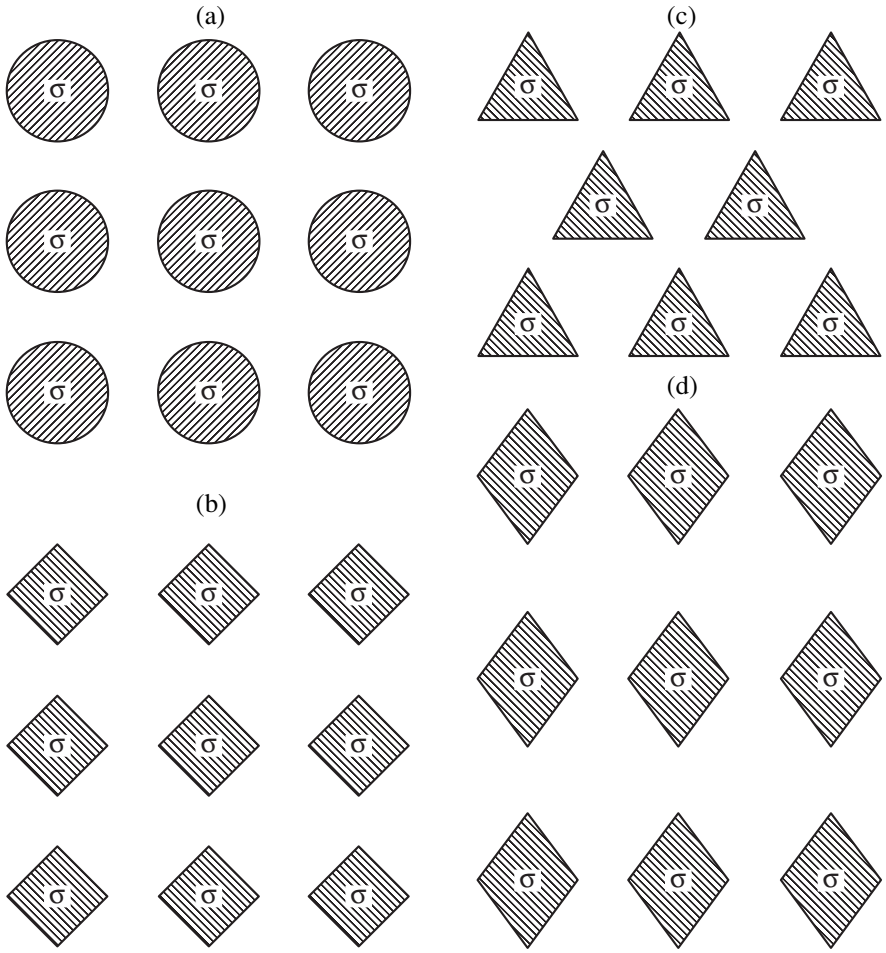


Fig. 1. Spatial arrangements of phases with conductivities $\sigma_1 = 1$ and $\sigma_2 = \sigma$ for four considered cases.

Sen and Torquato [10] derived an expression that allows an explicit calculation of the effective conductivity tensor from the *n*-point probability functions $S_n(\mathbf{r}_1, \ldots, \mathbf{r}_n)$. These functions give the probability of the points at $\mathbf{r}_1, \ldots, \mathbf{r}_n$ of belonging to the same phase, and are therefore uniquely determined by the spatial distribution of the phases. Unfortunately, the application of this method is limited because the computations with $n > 5$ are fairly time-consuming.

Different expansions of the effective conductivity in terms of a small parameter have been used in the past [1, 11, 12]. In most cases, the low-order terms depend weakly on the microgeometry. A diagrammatic expansion for the effective conductivity developed by Khalatnikov and Kamenshchik [13] promises to give more generally applicable results. The perturbative approach seems to be quite effective because it allows analyzing random and nonsymmetric distributions with different conductivities.

The problem was also discussed in the case where $N = 2$ and $N = 3$ on the basis of numerical calculations [3, 14]. It was shown that the effective conductivity for $N = 3$ is not universal and depends on the spatial arrangements of the phases. We have employed boundary element method for efficient numerical treatment of two-dimensional multiphase systems with an arbitrary arrangement of phases. More details on the method and its results can be found in [15].

In this paper, we consider the conductivity of a two-phase system in two dimensions for a wide range of concentrations and conductivities. One phase is assumed to be composed of droplets (of different shapes) with conductivity σ_2 embedded within a medium of conductivity σ_1 (see Fig. 1). We begin with calculating the effective conductivity σ_{eff} using a perturbation theory approach with the two phases having the respective volume fractions $(1 - \nu)$ and ν. Because the problem is linear, we can introduce a dimensionless conductivity σ, measured in units of $\sigma_1 = 1$, and the effective conductivity σ_{eff} is a function of $\sigma = \sigma_2/\sigma_1 = \sigma_2$ and ν. The volume-averaged conductivity

$$\bar{\sigma} = \frac{1}{V}\int \sigma dV$$

is given by

$$\bar{\sigma} = (1-\nu) + \nu\sigma. \quad (2)$$

If the conductivities of the two phases are not vastly different,

$$|\sigma - 1| \ll 1,$$

the effective conductivity can be calculated by perturbation theory [13]. To apply perturbation theory, we rewrite the spatial dependence of the conductivity as

$$\sigma(\mathbf{r}) = \bar{\sigma}(1-\alpha(\mathbf{r})), \quad (3)$$

where

$$\alpha(\mathbf{r}) = \frac{\sigma(\mathbf{r}) - \bar{\sigma}}{\bar{\sigma}}.$$

Assuming that the spatial distribution of conductivity is uncorrelated, we then obtain

$$\int d\mathbf{r}\alpha(\mathbf{r})\alpha(\mathbf{r}+\mathbf{r}') = \frac{(\sigma-1)^2\nu(1-\nu)}{\bar{\sigma}^2}\delta(\mathbf{r}'). \quad (4)$$

A straightforward calculation shows that up to the second order in a, the conductivity is given by

$$\sigma_{\text{eff}} = \bar{\sigma} - \frac{(\sigma-1)^2\nu(1-\nu)}{2\bar{\sigma}}. \quad (5)$$

This result has been known for many years and was derived for the dielectric function of dielectric mixtures [16]. In [11, 12], it was also derived using a systematic perturbative expansion, which showed it to be exact to the second order in α. The second term in Eq. (5) represents the first nonvanishing contribution due to the inhomogeneity of the distribution of the phases. In the case where $\nu = 0.5$, the result coincides with the expansion of the exact expression for the conductivity up to the second order in $(\sigma_2 - \sigma_1)$ [2]:

$$\sigma_{\text{eff}} = \sqrt{\sigma_1\sigma_2}. \quad (6)$$

2. CONDUCTIVITY OF A REGULAR ISOTROPIC TWO-COMPONENT SYSTEM IN TWO DIMENSIONS

Next, we exactly calculate the effective conductivity of the plane with different regular isotropic distributions. As before, we consider a 2*D* plane constructed from two different phases with different conductivities $\sigma_1 = 1$ and $\sigma_2 = \sigma$. The regions with the conductivity σ_2 have a circular shape with radius *R* and form a regular square lattice with the period *a* as shown in Fig. 1a. Changing the radius *R* from 0 to *a*/2, we can change the volume fraction of the second phase from $\nu = 0$ to the critical concentration $\nu_c = 0.785$, whereafter the regions with the conductivity σ_2 start to overlap and a percolation threshold is reached. In the case of metallic droplets, the total charge density must be zero, while a finite charge density can accumulate on the surface between different phases. This allows us to formulate the integral equation for the surface charge density [5, 6]. We define the surface charge density by the relation

$$\rho(\theta)Rd\theta = d\rho(\theta),$$

where $d\rho(\theta)$ is the charge on a small part of the surface between the two components with the length

$$dl = Rd\theta.$$

We recall that the scalar potential at the point $\mathbf{r}$ is determined by the relation

$$\phi = E_0 x - 2\int d^2r' \ln|\mathbf{r} - \mathbf{r}'|\rho(\mathbf{r}'), \quad (7)$$

where $\ln|\mathbf{r} - \mathbf{r}'|/2\pi$ is the 2*D* Green function. The boundary conditions on the surface between two phases are [16]

$$E_n^1 - E_n^2 = 4\pi\rho(\theta), \quad (8)$$

$$\sigma_1 E_n^1 = \sigma_2 E_n^2. \quad (9)$$

Substituting

$$\mathbf{r}' = \mathbf{i}(ma + R\cos\theta') + \mathbf{j}(na + R\sin\theta'),$$

$$\mathbf{r} = \mathbf{i}R\cos\theta + \mathbf{j}R\sin\theta$$

in Eqs. (7)–(9), we obtain an integral equation for the surface charge density in the form

$$\rho(\theta) = \frac{\kappa}{2\pi}\left[E_0\cos\theta + 2r\sum_{n,m=-\infty}^{\infty}\int_{-\pi}^{\pi} d\theta'\rho(\theta') \times \mathrm{Re}\left\{\frac{\exp(i\theta)}{m + r(\cos\theta' - \cos\theta) + i(n + r(\sin\theta' - \sin\theta))}\right\}\right], \quad (10)$$

where

$$r = \frac{R}{a}, \quad \kappa = \frac{1-\sigma}{1+\sigma}.$$

As shown in the Appendix, the sum over *m* can be calculated exactly and the integral equation for the surface charge density is reduced to the form

$$\rho(\theta) = \frac{\kappa}{2\pi}\left[E_0\cos\theta + 2r\sum_{n=-\infty}^{\infty}\int_{-\pi}^{\pi} d\theta' K(n,\theta,\theta')\rho(\theta')\right], \quad (11)$$

where

$$K(n,\theta,\theta') = \pi\frac{\cos\theta\sin(2\pi r(\cos\theta' - \cos\theta)) + \sin\theta\sinh[2\pi(n + r(\sin\theta' - \sin\theta))]}{\cosh[2\pi(n + r(\sin\theta' - \sin\theta))] - \cos(2\pi r(\cos\theta' - \cos\theta))}. \quad (12)$$

Expanding the surface density $\rho(\theta)$ in terms of Legendre polynomials $P_l(\cos\theta)$ and taking into account that

$$\rho(-\theta) = \rho(\theta), \quad \rho(\pi - \theta) = -\rho(\theta),$$

where

$$\rho(\theta) = \sum_{i=1}^{\infty} c_{2l-1} P_{2l-1}(\cos\theta), \tag{13}$$

we obtain the linear set of algebraic equations for the coefficients c_{2l-1}

$$\frac{2c_{2l-1}}{4l-1} = \frac{\kappa}{2\pi}\left[\frac{2}{3}E_0\delta_{l,1} + 2r\sum_{k=1}^{\infty} c_{2k-1}K_{l,k}\right], \tag{14}$$

where

$$K_{l,k} = \sum_{n=-\infty}^{\infty}\int_0^{\pi} d\theta' \int_0^{\pi} d\theta K(n, \theta, \theta') \times \sin\theta P_{2l-1}(\cos\theta) P_{2k-1}(\cos\theta'). \tag{15}$$

Solving Eq. (14) with a finite number of Legendre polynomials taken into account, we obtain the surface charge density in Eq. (13). As a result, the effective conductivity is evaluated by calculating the total current

$$j = \sigma_1 E_n = E_n$$

through the semicircular surface with a radius $R' = a/2$ (see Fig. 1a). Calculations similar to that of Eq. (12) lead to the expression for the effective conductivity

$$\sigma_{\text{eff}} = \frac{\kappa}{4\pi}\int_{-\pi/2}^{\pi/2} d\theta \times\left[\cos\theta + \frac{2r}{E_0}\sum_{n=-\infty}^{\infty}\int_{-\pi}^{\pi} d\theta' K'(n, \theta, \theta')\rho(\theta')\right], \tag{16}$$

where

$$K'(n, \theta, \theta') = \pi\frac{\cos\theta\sin(2\pi(r\cos\theta' - (1/2)\cos\theta)) + \sin\theta\sinh(2\pi(n + r\sin\theta' - (1/2)\sin\theta))}{\cosh(2\pi(n + r(\sin\theta' - (1/2)\sin\theta)) - \cos(2\pi(r\cos\theta' - (1/2)\cos\theta))}. \tag{17}$$

The result above applies to the case of a uniform distribution of circular droplets within the plane. To see how the effective conductivity depends on the shape of the regions with conductivity σ_2, we performed calculations in the case where circular droplets were replaced with squares, triangles, and rhombuses with the ratio of diagonals $\tan\alpha = a/b$, where a and b are translation vectors along the x and y axes, respectively (see Figs. 1b–1d). In all these cases, Eqs. (10)–(17) are slightly modified because in a polar coordinate system, $r(\theta)$ is a function of the angle. Unlike in the case of circles, the percolation threshold for cases *b*, *c*, and *d* is $\nu_c = 0.5$. We note that in the case of rhombuses, the lattice is anisotropic and $\sigma_{\text{eff}}^{11} \neq \sigma_{\text{eff}}^{22}$.

3. DISCUSSION

The results of the calculations of the effective conductivity are presented in Fig. 2 as a function of σ^ν for different values of the volume fraction ν. It is easy to check that the results satisfy the generalized duality relation [3, 4]

$$\sigma_{\text{eff}}^{11}(\sigma_1, \sigma_2)\sigma_{\text{eff}}^{2}(1/\sigma_1, 1/\sigma_2) = 1. \tag{18}$$

For circles, squares, and triangles,

$$\sigma^{22} = \sigma^{11}.$$

For rhombuses,

$$\sigma^{22}(\alpha) = \sigma^{11}(\pi/2 - \alpha).$$

Figures 2a–2d show that for small κ, perturbation theory [11–13] (Eq. (5)) gives the correct result independent of the geometry.

3.1. Approximate Expression for Effective Conductivity

Although the predictions in Fig. 2 represent the results of precise numerical calculation, they are not very tractable when it comes to comparing with experimental data, being the result of numerical calculations. It is therefore helpful to try to obtain a functional form for describing the behavior predicted in Fig. 2, which also includes all the relevant parameters, such as the volume fraction ν and the two conductivities σ_1 and σ_2. Such an expression can then be used in a wide range of problems, provided the validity range is taken into account. We describe the properties of such a heuristically determined function and determine its validity range in terms of the parameters ν, σ_1, and σ_2.

As can be seen from Fig. 2, the dependence of the effective conductivity on σ shows similar behavior independently of the particular geometry of the phases. First, we observe that when κ is small, all the curves are linear in σ^ν with the same slope. In the relatively wide

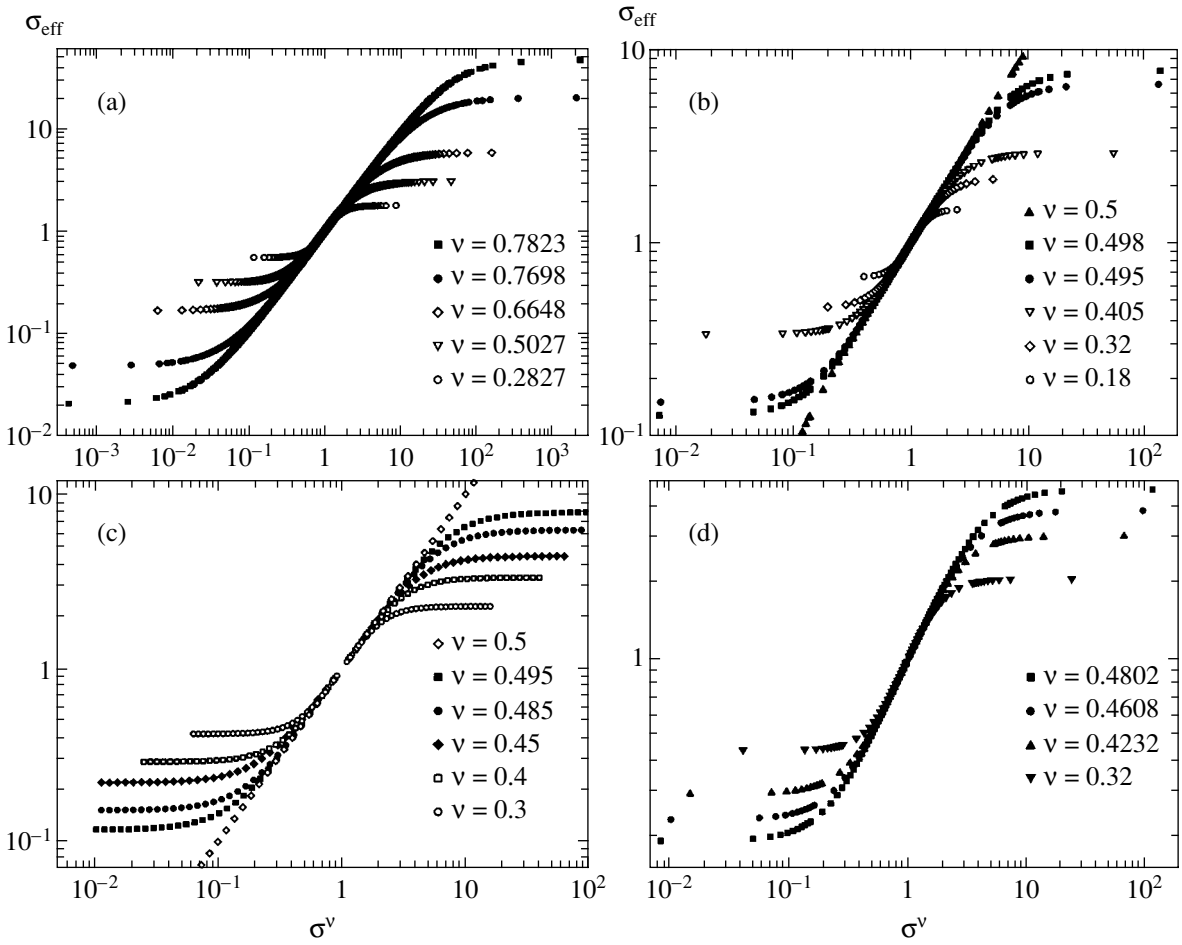


Fig. 2. Effective conductivity of the plane as a function of σ^ν for different volume fractions and four considered geometries.

interval $0.1 < \sigma < 10$, the effective conductivity is determined by the equation

$$\sigma_{\text{eff}}(\sigma) = \sigma_1^{(1-\nu)}\sigma_2^{\nu}. \tag{19}$$

The range of applicability of this formula becomes wider as we approach the percolation threshold ν_c. When $\sigma = \sigma_2/\sigma_1 \gg 1$, the effective conductivity saturates at σ_{sat}. The value of σ_{sat} is not universal and depends on the geometry. It was pointed out recently that in the case of circles with $\nu < 0.5$ in the entire range of σ, the effective conductivity may be approximated by the formula [18]

$$\sigma_{\text{eff}}(\kappa) = \frac{1-\nu\kappa}{1+\nu\kappa}. \tag{20}$$

To derive an approximate expression for the effective conductivity, we assume that Eq. (20) remains correct if we replace ν with the effective volume fraction $\nu_{\text{eff}}(\kappa, \nu)$. We require that

$$\nu_{\text{eff}}(\kappa, \nu) \approx \nu \quad \text{as} \quad \kappa \longrightarrow 0$$

or

$$\nu \longrightarrow 0,$$

and

$$\nu_{\text{eff}}(\kappa, \nu) \approx \frac{1}{\kappa}\frac{1-\sigma^{\nu_c}}{1+\sigma^{\nu_c}} \quad \text{as} \quad \nu \longrightarrow \nu_c$$

to satisfy Eq. (19), which is valid at $\nu = \nu_c$. It is easy to see that the function

$$\nu_{\text{eff}}(\kappa, \nu) = \nu + \frac{1}{\kappa} \times \frac{1-\left(\dfrac{1-(1-p(\nu))\kappa}{1+(1-p(\nu))\kappa}\right)^{\nu_c}}{1+\left(\dfrac{1-(1-p(\nu))\kappa}{1+(1-p(\nu))\kappa}\right)^{\nu_c}} - (1-p(\nu))\nu_c, \tag{21}$$

where $p(\nu) \longrightarrow 0$ as $\nu \longrightarrow \nu_c$ and $p(\nu) \longrightarrow 1$ as $\nu \longrightarrow 0$, satisfies all the above requirements. The function $p(\nu)$ is not universal and depends on the geometric shape of the region with conductivity σ and on the particular arrangement of these inclusions in the $2D$ plane.

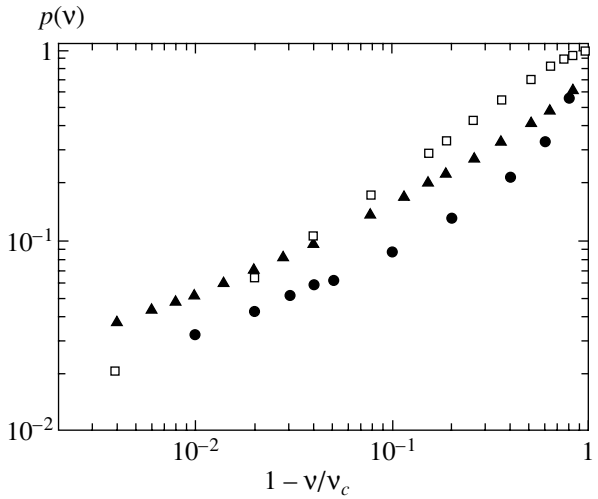


Fig. 3. Dependence of the function $p(\nu)$ on $1 - \nu/\nu_c$ for cases *a* (circles), *b* (squares), and *c* (triangles).

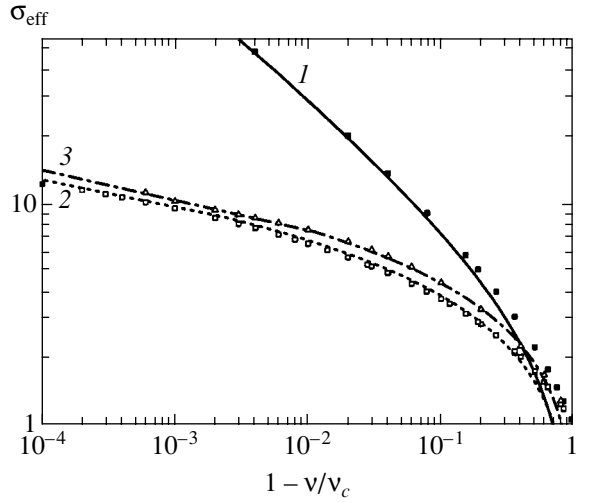


Fig. 4. Saturated effective conductivity as $\sigma \longrightarrow \infty$ for cases *a*, *b*, and *c*. Full, dotted, and dashed lines show different analytic asymptotic behavior for these cases: *1*—$(\pi/2 + \arcsin((\nu/\nu_c)^{1/2}))/(1 - \nu/\nu_c)^{1/2} - \pi/2$ (case *a*), *2*—$1.3\ln(1/(1 - (\nu/\nu_c)^{1/2}))$ (case *b*), *3*—$\sqrt{2}\ln(2/(\sqrt{3}(1 - (\nu/\nu_c)^{1/2})))$ (case *c*).

In Fig. 3, we plot $p(\nu)$ as a function of $1 - \nu/\nu_c$ in cases *a*, *b*, and *c*. Case *d* is different because the effective conductivity is anisotropic. As is clearly seen from Fig. 3, the behavior of the function $p(\nu)$ for circles (case *a*) is different from the cases of squares and triangles (cases *b* and *c*). On the other hand, in cases *b* and *c*, $p(\nu)$ shows a similar behavior.

3.2. Shape Dependence of Effective Conductivity

The function $p(\nu)$ is related to the value of

$$\sigma_{sat} = \frac{1 + \nu_{eff}(\kappa = 1, \nu)}{1 - \nu_{eff}(\kappa = 1, \nu)}.$$

Therefore, the behavior of $p(\nu)$ close to the percolation threshold should be different for different geometries. In Fig. 4, we plot the value of σ_{sat} as a function of $(1 - \nu/\nu_c)$ in the case of circles, squares, and triangles. There is an important difference between these two cases. In the case of circles, σ_{sat} has a power-like divergence $(1 - \nu/\nu_c)^{-k}$ ($k \approx 0.5$). For squares and triangles, this behavior is logarithmic. In both cases, close to percolation threshold, σ_{sat} is proportional to the average inverse distance between boundaries of the neighboring circles or squares,

$$\sigma_{sat} \propto \int \frac{dy}{1 - 2f(y)},$$

where

$$f(y) = \sqrt{r^2 - y^2}$$

for circles and

$$f(y) = r - |y|$$

for squares. We here assume that the period of the system is 1, and the dimensionless size of the circle and the square is *r*. Direct integration leads to the following results:

$$\sigma_{sat} \propto \frac{\pi/2 - \arccos((1 - \nu/\nu_c)^{1/2}}{(1 - \nu/\nu_c)^{1/2}} - \pi/2 \qquad (22)$$

for circles and

$$\sigma_{sat} \propto -\ln(1 - (\nu/\nu_c)^{1/2}) \qquad (23)$$

for squares (Fig. 4). For triangles (*c*), the asymptotic formula is similar to Eq. (23) with different numeric coefficients. Interestingly, this observation suggests that behavior of the function $p(\nu)$ is different depending on the curvature of the embedded regions.

4. CONCLUSIONS

From calculations of the effective conductivity of inhomogeneous two-phase systems in two dimensions, we find that the results of precise numerical calculations can be approximated by a universal function for σ_{eff}, Eqs. (20) and (21), where the function $p(\nu)$ depends on the spatial arrangements of the 2*D* plane and on the shape of the inclusions with conductivity σ. It is shown that in a large interval of the conductivity σ, the effective conductivity σ_{eff} is determined by the spatial average of the logarithm of individual conductivities. The closer the system is to the percolation threshold, the larger the validity range of this result. For large values of the conductivity σ, σ_{eff} is saturated at a value a σ_{sat}. The value of σ_{sat} near the percolation threshold is determined by the average inverse distance between

boundaries of neighboring regions with the conductivity σ in the direction of the field (Eqs. (22) and (23)).

The model that we have developed is quite generally applicable and can be applied in some interesting situations, such as cuprates and other two-dimensional complex transition metal oxides that exist near a phase-separation threshold. Importantly, there appears to be a significant amount of experimental evidence that many anomalous properties of oxides are associated with the coexistence of two or more phases. The application of the presented model may help understanding the transport properties of such systems.

APPENDIX

Here, we show how the sum over m in Eq. (10) can be calculated exactly. We represent the sum as

$$S + \sum_{m=-\infty}^{\infty} \mathrm{Re}\left\{\frac{\exp(i\theta)}{m+\beta+i\alpha}\right\},$$

where

$$\beta = r(\cos\theta' - \cos\theta), \quad \alpha = n + r(\sin\theta' - \sin\theta).$$

The sum over m is calculated using the definition of the digamma function. As a result, we express the sum as

$$\begin{aligned} S &= \mathrm{Re}\{\exp(i\theta)[\psi(-\beta - i\alpha) - \psi(1+\beta+i\alpha)]\} \\ &= \pi\mathrm{Re}\{\exp(i\theta)\cot(\pi(\beta+i\alpha))\}. \end{aligned}$$

Calculating the imaginary part of the previous equation, we obtain the result in Eq. (12),

$$S = \pi\frac{\cos\theta\sin(2\pi\beta) + \sin\theta\sinh(2\pi\alpha)}{\cosh(2\pi\alpha) - \cos(2\pi\beta)}.$$

REFERENCES

1. D. J. Bergman and D. Stroud, Solid State Phys. **46**, 147 (1992).
2. J. B. Keller, J. Math. Phys. **5**, 548 (1964); A. M. Dykhne, Pis'ma Zh. Éksp. Teor. Fiz. **32**, 68 (1970) [JETP Lett. **32**, 63 (1970)].
3. L. G. Fel, V. S. Machavariani, and D. J. Bergman, J. Phys. A **33**, 6669 (2000).
4. V. G. Marikhin, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 391 (2000) [JETP Lett. **71**, 271 (2000)].
5. Yu. N. Ovchinnikov and A. M. Dyugaev, Zh. Éksp. Teor. Fiz. **117**, 1013 (2000) [JETP **90**, 881 (2000)].
6. Yu. N. Ovchinnikov and I. A. Luk'yanchuk, Zh. Éksp. Teor. Fiz. **121**, 239 (2002) [JETP **94**, 203 (2002)].
7. Yu. N. Ovchinnikov, Zh. Éksp. Teor. Fiz. **125**, 181 (2004) [JETP **98**, 162 (2004)].
8. A. L. Efros and B. I. Shklovskii, Phys. Status Solidi B **76**, 475 (1976).
9. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **120**, 945 (2001) [JETP **93**, 824 (2001)].
10. A. K. Sen and S. Torquato, Phys. Rev. B **39**, 4504 (1989).
11. W. F. Brown, J. Chem. Phys. **23**, 1514 (1955).
12. D. J. Bergman, Phys. Rev. B **14**, 4304 (1976).
13. I. M. Khalatnikov and A. Yu. Kamenshchik, Zh. Éksp. Teor. Fiz. **118**, 1456 (2000) [JETP **91**, 1261 (2000)].
14. S. Torquato, C. Kim, and D. Cule, J. Appl. Phys. **85**, 1560 (1999).
15. K. Zagar, BSc Thesis (Slovene, 2003).
16. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 2nd ed. (Nauka, Moscow, 1982; Pergamon, Oxford, 1984).
17. S. A. Bulgadaev, cond-mat/0212104.
18. Yu. P. Emets, Zh. Éksp. Teor. Fiz. **121**, 1339 (2002) [JETP **94**, 1149 (2002)]; Zh. Éksp. Teor. Fiz. **118**, 1207 (2000) [JETP **91**, 1046 (2000)].

Journal of Experimental and Theoretical Physics, Vol. 100, No. 4, 2005, pp. 722–730.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 4, 2005, pp. 817–826.

ELECTRONIC PROPERTIES OF SOLIDS

Hopping Conductivity and Coulomb Correlations in 2D Arrays of Ge/Si Quantum Dots

A. I. Yakimov[a], A. V. Dvurechenskiĭ[a], G. M. Min'kov[b], A. A. Sherstobitov[b], A. I. Nikiforov[a], and A. A. Bloshkin[a]

[a]*Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia*

[b]*Ural State University, pr. Lenina 51, Yekaterinburg, 620083 Russia*

e-mail: yakimov@isp.nsc.ru

Received August 27, 2004

Abstract—The temperature and magnetic-field dependences of the conductivity associated with hopping transport of holes over a 2D array of Ge/Si(001) quantum dots with various filling factors are studied experimentally. A transition from the Éfros–Shklovskiĭ law for the temperature dependence of hopping conductivity to the Arrhenius law with an activation energy equal to 1.0–1.2 meV is observed upon a decrease in temperature. The activation energy for the low-temperature conductivity increases with the magnetic field and attains saturation in fields exceeding 4 T. It is found that the magnetoresistance in layers of quantum dots is essentially anisotropic: the conductivity decreases in an increasing magnetic field oriented perpendicularly to a quantum dot layer and increases in a magnetic field whose vector lies in the plane of the sample. The absolute values of magnetoresistance for transverse and longitudinal field orientations differ by two orders of magnitude. The experimental results are interpreted using the model of many-particle correlations of holes localized in quantum dots, which lead to the formation of electron polarons in a 2D disordered system. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

If the resistance of a 2D disordered system is much higher than the value determined by the resistance quantum h/e^2, where h is the Planck constant and e is the elementary charge, the system is on the insulator side of the metal–insulator transition and conductivity σ of the system tends to zero upon a decrease in temperature T. The effect of the electron–electron interaction on the conductivity of 2D systems, which remains an important problem, has become of special interest after a recent heated discussion concerning the existence of the metal–insulator transition in the 2D case. The role of correlation effects must be especially significant on the insulator side of the metal–insulator transition since localized electron states are ineffective in screening processes as compared to extended states.

Analysis of fundamental aspects of charge transport as well as Coulomb and spin correlations in semiconducting nanostructures form the basis for the development of nanoelectronics—one of the latest trends in the physics and technology of nanometer-size electronic devices. The layers of self-organizing quantum dots (QD) obtained as a result of heteroepitaxy of elastically stresses system are the most suitable objects for determining the main regularities in the processes of 2D charge transfer as well as the role of Coulomb correlations in a system of localized electrons. This is due to the fact that (i) QDs lie exactly in the same (growth) plane in view of peculiar epitaxy and the disorder factor associated with disorder in the vertical direction of the nanostructure (growth direction) is absent; (ii) introducing electrons or holes into QDs, it is possible to controllably obtain ensembles of localized charge carriers in a priori known quantum states with preset wavefunctions; (iii) in contrast to most of impurities in semiconductors, QD arrays may behave as a system of multiply charged localization centers, in which the role of Coulomb and spin correlations is most significant; and (iv) since, as a rule, the ratio of the height to the lateral size in self-organizing QDs is much smaller than unity, the wavefunctions of charge carriers in QDs are strongly anisotropic and two-dimensional in contrast to those for impurities.

The simplest manifestation of correlation effects is the formation of a Coulomb gap in the spectrum of localized states of a disordered system owing to the long-range part of the electron–electron interaction [1]. However, the Éfros–Shklovskiĭ one-electron model ignored possible many-particle correlations in electron hopping in the case when, for example, simultaneous or consecutive jumps of some electrons over small distances facilitate the motion of other electrons over long distances, lowering the corresponding energy barriers on the path of current by their Coulomb potential. One of the possible manifestations of many-particle excitations is the formation of electron polarons in the system, viz., a "coat" of polarized pairs of localized states, which is entrained by an electron moving over a percolation cluster [2, 3].

In this study, we analyze hopping transport of holes in 2D arrays of Ge/Si(001) QDs. Analysis of the temperature and magnetic-field dependences of low-temperature conductivity reveals the presence of many-particle Coulomb correlations of holes localized in QDs.

2. EXPERIMENTAL TECHNIQUE

Figure 1 shows a schematic diagram of the structures on which experiments were carried out. The samples were grown by molecular-beam epitaxy on a Katun'-type setup designed at the Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences. The substrates were Si plates with the (001) orientation and a resistivity of 20 Ω cm, which were doped with boron to a concentration of approximately to 10^{14} cm^{-3}. Before their loading into the epitaxial chamber, the substrates were degreased in acetone and coated with a layer of surface oxide having a thickness of about 10 Å by etching in an $H_2O_2 : HNO_3 : H_2O$ solution. After loading into the epitaxial chamber, the plates were heated to 800°C and held for 3 min in a silicon flow of approximately 10^{13} atm/(cm^2 s). The surface purity was monitored using diffraction of fast electrons. After the preepitaxial treatment, the substrates were reloaded to the growing vessel.

A Ge layer with a thickness of eight monolayers (~10 Å) was introduced into a Si layer grown on a 250-nm substrate at 300°C at a distance of 200 nm from its surface. Judging from the reflection high-energy electron diffraction pattern from the surface of the Ge island film, germanium nanoclusters that formed according to the Stranski–Krastanov growth mechanism had the shape of pyramids. Structural analysis using high-resolution electron microscopy revealed that the mean size of the base of Ge nanoclusters in the growth plane was 12 nm, the height was approximately 2 nm, and the nanocluster layer density was $N_{qd} \approx 3 \times 10^{11}$ cm^{-2}. Figure 2 shows the image of the cross section of a Ge/Si heterostructure, obtained with the help of high-resolution electron microscopy.

The controllable filling of Ge islands with holes was carried out by introducing into the sample a Si layer δ-doped with boron at a distance of 10 nm below the layer of Ge islands. Since the ionization energy E_B of boron impurities in silicon amounts to only 45 meV, and the energies E_i corresponding to the first ten layers of holes in germanium QDs of such a size, which are measured from the top of the valence band in Si, range between 200 and 320 meV [4], the holes leaving impurity sites below room temperature populate the levels in QDs. The boron concentration in the δ-layers of two prepared experimental samples with QDs was approximately 6.9×10^{11} and 8.4×10^{11} cm^{-2}, respectively. Calculations based on solving the system of the Poisson and electroneutrality equations as well as on statistical distribution of holes over the energy levels in the system proved that the average number of holes (filling factor ν) per QD was 2.3 and 2.8 for two different samples. The system formed by the charged QD layer and the charged δ-Si:B layer was simulated by a parallel-plate capacitor with distance d between the plates equal to the spacing between the QD and δ-Si:B layers. In calculating filling factor ν, we took into account the first five states of holes in QDs, which were determined earlier in [4, 5]. In addition, we assumed that the temperatures are so low ($k_BT \ll E_i$) that free holes are absent in the valence band. We solved the system of equations

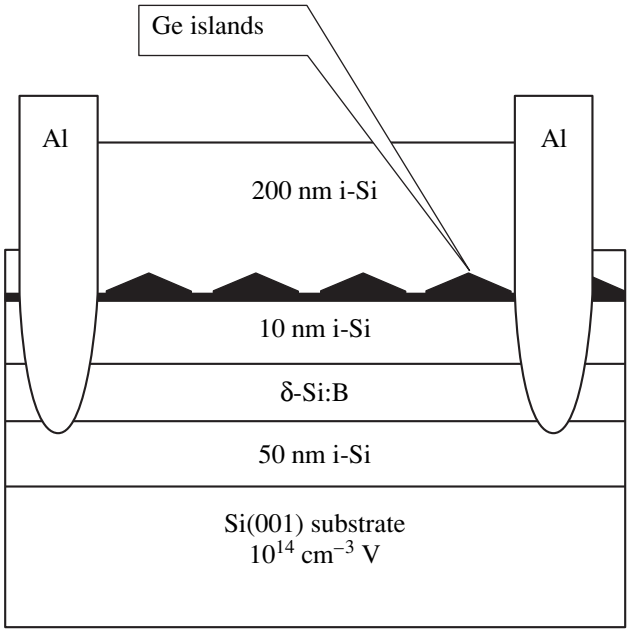


Fig. 1. Schematic diagram of the samples used for measurements of the conductivity of QD layers.

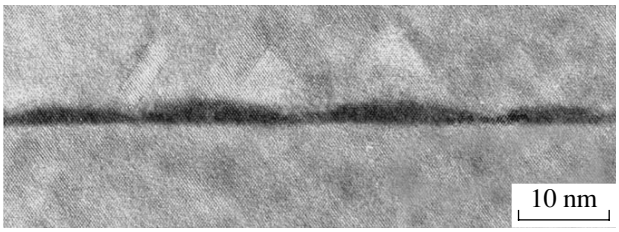


Fig. 2. Image of the structure cross section, obtained using high-resolution electron microscopy. Dark regions are Ge islands.

$$n_B = \frac{N_B}{1 + 0.5\exp[-(E_B - \varepsilon_F)/k_BT]},$$

$$n_{qd} = \sum_{1 \le i \le 5} \frac{2N_{qd}}{1 + \exp[(E_i - \varepsilon_F + u)/k_BT]},$$

$$u = \frac{n_{qd}e_d}{\kappa\varepsilon_0}, \quad n_{qd} + n_B = N_B,$$

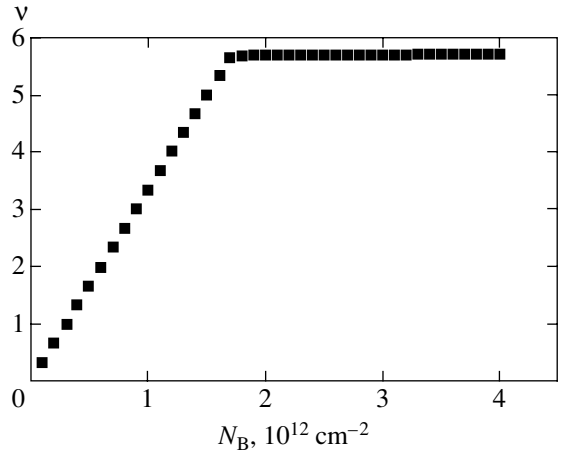


Fig. 3. Calculated dependence of the factor of filling of QDs with holes on the boron concentration in the δ layer. The distance between the Ge QD layer and the δ layer is 10 nm. The results are independent of temperature at $T < 20$ K.

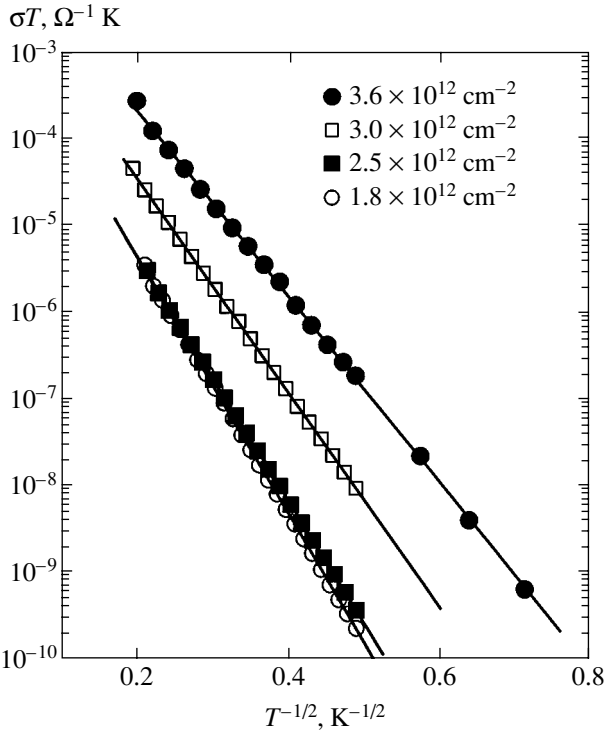


Fig. 4. Dependence of quantity σT, where σ is the conductivity and T is the temperature, on $T^{-1/2}$ for check δ-Si:B samples free of germanium QDs. The experimental data are plotted in semilogarithmic coordinates.

where k_B is the Boltzmann constant, n_B is the concentration of holes remaining at boron impurity atoms, n_{qd} is the hole concentration in the QD layer, N_B is the boron concentration in the δ layer, ε_F is the Fermi layer, u is the electrostatic potential of the QD layer, and $\varepsilon_0 = 8.87 \times 10^{-12}$ F/m is the absolute permittivity of vacuum. The results of calculations are depicted in Fig. 3. The factor of QD filling with holes for the prepared samples was calculated using precisely this figure.

The structures were prepared photolithographically in the form of Hall bridges. Contacts were formed by etching the Si layer to a depth of 0.15 μm. Then an aluminum layer was deposited in vacuum and annealing was carried out at a temperature of 600°C.

The monitoring of the composition of Ge islands with the help of Raman spectroscopy showed that the fraction of Ge in the islands was approximately 70%. The concentration of doping impurity was determined using secondary-ion mass spectrometry.

In addition to samples containing Ge QDs, a batch of check samples free of a Ge layer was grown. The boron concentration in the δ-doped Si layer in check samples varied from 1.8×10^{12} cm^{-2} to 4.2×10^{12} cm^{-2}.

Conductivity measurements were carried out in planar geometry (along the QD layer) using the dc four-point technique in a temperature range of 1.5–20 K. The magnetoresistance was measured in a constant magnetic field up to 5.5 T, which was perpendicular ($H_\perp$ orientation) or parallel ($H_\parallel$ orientation) to the plane of the structure. In the latter case, the magnetic field was directed parallel or perpendicular to the electric current.

3. EXPERIMENTAL RESULTS AND DISCUSSION

3.1. Temperature Dependence of Conductivity and Magnetic Transport in δ-Si:B Structures

Our measurements show that the dependence of conductivity σ on temperature T for all check samples free of Ge nanoclusters obeys the Éfros–Shklovskiĭ law

$$\sigma(T) = \frac{\gamma}{T}\exp\left[-\left(\frac{T_1}{T}\right)^{1/2}\right], \qquad (1)$$

where $\gamma = (0.3\text{–}3) \times 10^{-2}$ K/Ω and $T_1 = 600$–1000 K (Fig. 4). This law indicates the emergence of variable range hopping conduction mechanism provided that a "soft" Coulomb gap exists in the density of impurity states of a 2D disordered system [6]. The Si layer δ-doped with boron apparently plays the role of such a 2D system.

Figure 5 shows the typical magnetic-field dependences of the resistance for one of the test samples (with a boron concentration in the δ-doped layer of 3.6×10^{12} cm^{-2}). Although the magnetoresistance differs insignificantly for different field orientations, it is still not the same and is a combination of two contributions of opposite signs. The negative magnetoresistance is apparently associated with suppression of "destructive" interference of various tunneling "trajectories" in a magnetic field, which include subbarrier scattering events [7, 8]. This mechanism essentially resembles the negative magnetoresistance mechanism in the case of diffusion charge transfer under weak localization con-

ditions and is associated with a change in the phase of the wavefunction under the conditions when we can speak of coherence of tunneling with allowance for scattering from impurities. The positive contribution is due to "compression" of the wavefunctions of localized charge carriers in directions perpendicular to the magnetic field [9, 10]. As a result of such compression, the degree of overlapping of the wavefunctions of carriers at different sites separated by long distances becomes lower and the percolation threshold in the system is shifted towards higher values. Additional experiments revealed that magnetoresistance anisotropy increases upon cooling; however, since the relative magnetoresistance of the check samples is small (on the order of a few percent), the problem of temperature stabilization arises in obtaining correct experimental results at $T < 4$ K.

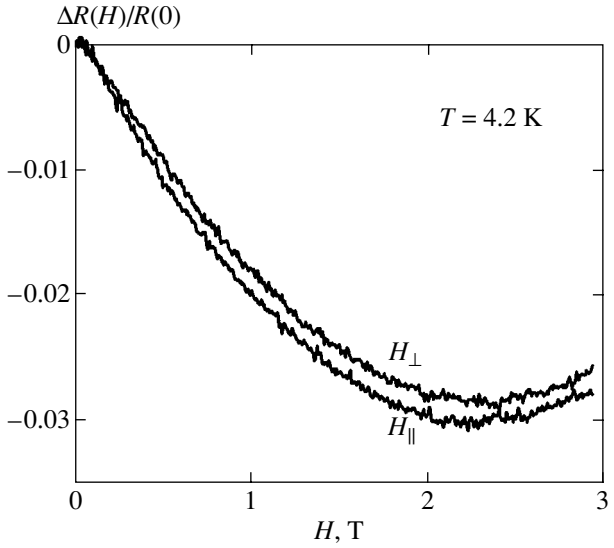


Fig. 5. Relative variation of the resistance in magnetic fields of different orientations for one of test samples free of Ge QDs. The measuring temperature is 4.2 K. The boron concentration in the δ-doped layer is 3.6×10^{12} cm^{-2}.

3.2. Temperature Dependence of Conductivity in Ge/δ-Si:B Samples with Quantum Dots

Our measurements of the temperature dependence of conductivity made for samples with Ge QDs proved that the conductivity of these samples above $T_c = 4$–5 K obeys the Éfros–Shklovskiĭ law

$$\sigma(T) = \sigma_0 \exp[-(T_1/T)^{1/2}]$$

with a temperature-independent preexponential factor σ_0 on the order of e^2/h. For example, $\sigma_0 = 2.5e^2/h$ and $T_1 = 220$ K for a sample with a filling factor of $\nu = 2.3$, while $\sigma_0 = e^2/h$ and $T_1 = 100$ K for a sample with $\nu = 2.8$. Figure 6 shows an example of such a behavior of the sample with $\nu = 2.8$. The universal value of the preexponential factor (on the order of e^2/h) was predicted earlier by Kozub *et al.* [11] and was one of indications of the existence of consecutive many-electron Coulomb correlations in the system.

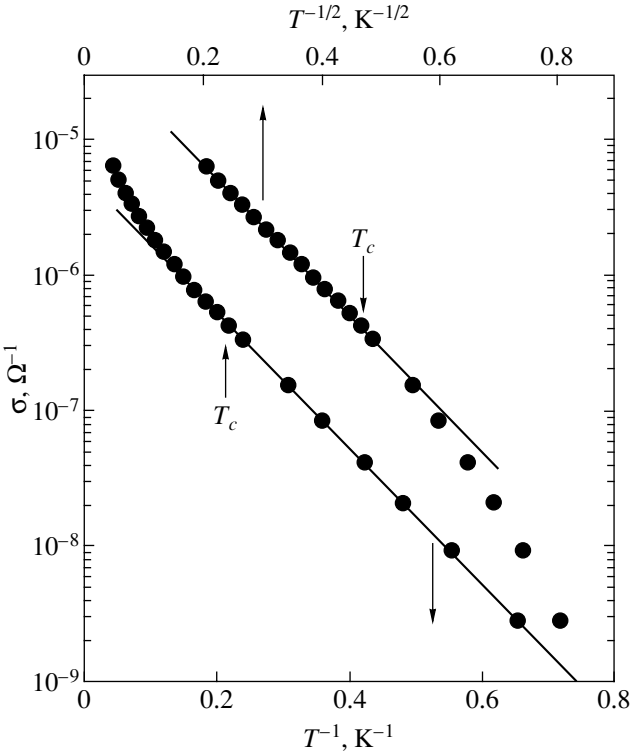


Fig. 6. Dependence of the logarithm of conductivity on $T^{-1/2}$ (upper abscissa axis) and T^{-1} (lower abscissa axis) for a sample with Ge QDs and with filling factor $\nu = 2.8$.

As the temperature decreases, the temperature dependence of conductivity becomes stronger; at $T < T_c$, a transition from the Éfros–Shklovskiĭ to the Arrhenius law

$$\sigma(T) = \sigma_0 \exp(-E_a/k_B T)$$

takes place. This phenomenon can be traced most clearly on the temperature dependence of the local conductivity activation energy (Fig. 7). In the general case, the temperature dependence of hopping conductivity is described by the law

$$\sigma(T) \propto \exp[-(T_1/T)^x].$$

The activation energy is defined as

$$W = \frac{\partial \ln \sigma}{\partial (1/k_B T)} = x k_B T^{1-x} T_1^x. \qquad (2)$$

With such an approach, we have

$$\log W(T) = A + (1-x)\log T \quad \text{and} \quad A = \log(x k_B T_1^x).$$

Plotting $\log W(T)$ as a function of $\log T$, we can determine the exponent x describing the temperature dependence of conductivity from the slope of the straight line. It can be seen from Fig. 7 that the activation energy

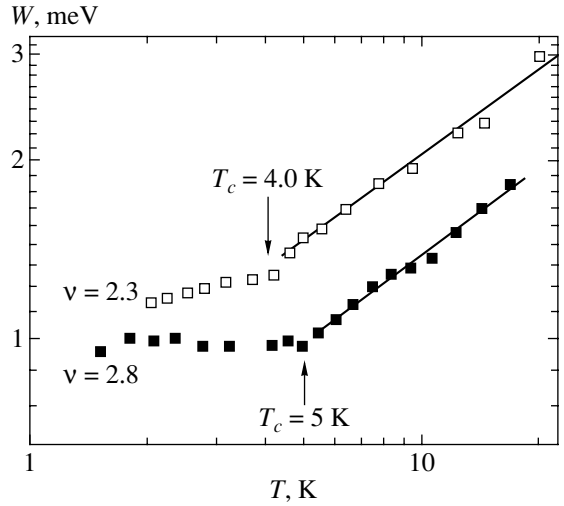


Fig. 7. Temperature dependence of the hopping conduction activation energy $W = \partial \ln\sigma/\partial(1/k_BT) = xk_BT^{1-x}T_1^x$ plotted in the log–log coordinates. Solid lines describe the approximation of experimental data on $W(T)$ by formula (2).

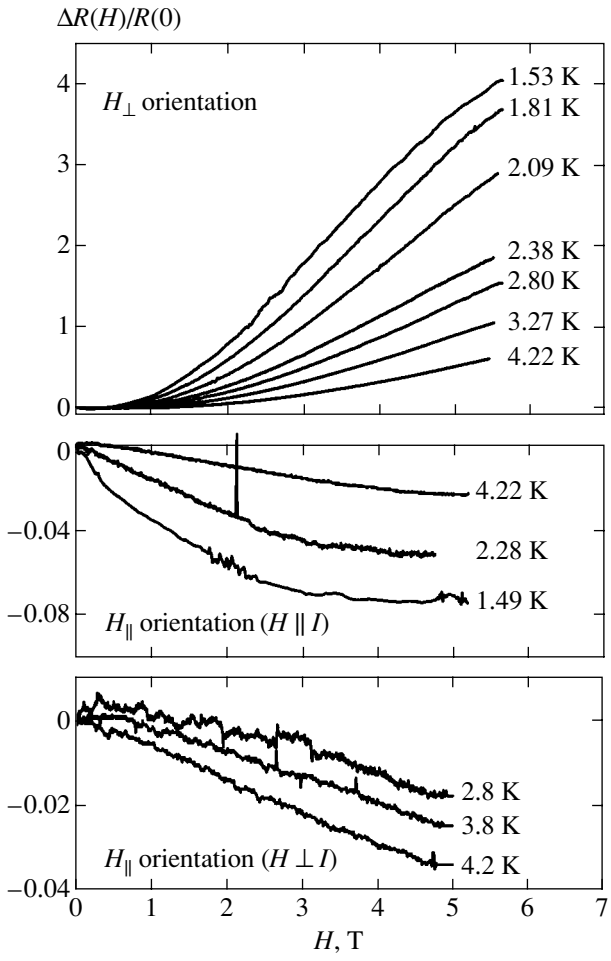


Fig. 8. Relative variation of the resistance in a magnetic field at various temperatures for a Ge/Si sample with Ge QDs and QD filling factor ν = 2.8.

at high temperatures decreases with T; i.e., the variable range hopping conduction mode is realized. At temperatures below T_c = 4 K for the sample with ν = 2.3 and T_c = 5 K and for the sample with ν = 2.8, the activation energy becomes virtually independent of temperature. This is manifested most clearly in the latter case. Approximation of experimental data on $W(T)$ by formula (2) at $T > T_c$ (solid lines in Fig. 7) gives exponents $x = 0.52 \pm 0.03$ and $x = 0.50 \pm 0.03$ for samples with ν = 2.3 and 2.8, respectively. This means that the value of x changes indeed from $x = 1/2$ to $x \approx 1$ upon system cooling.

3.3. Magnetoresistance in Ge/δ-Si:B Samples with Quantum Dots

The magnetoresistance for the Ge/δ-Si:B sample with ν = 2.8 is shown in Fig. 8. An analogous behavior is also observed for the structure with ν = 2.3. The following two circumstances are worth noting here. First, the magnetoresistance is positive in a transverse magnetic field and negative in a longitudinal magnetic field for both field orientations relative to the current in the latter case. The relative magnetoresistance for the transverse orientation of the magnetic field is almost two orders of magnitude higher than the value of $|\Delta R(H)/R(0)|$ for fields parallel to the plane of the structure. Such a giant anisotropy indicates an "ideal" 2D nature of hopping charge transfer in QD layers. Second, for the $H_\perp$ orientation, the magnetoresistance of the structure with QDs is much higher than the magnetoresistance of a check sample without QDs, which indicates a decisive role of the QD ensemble in the observed features of magnetic transport. In the subsequent analysis, we will confine ourselves to the discussion of experimental data for the transverse magnetic field orientation only.

A detailed analysis shows that the magnetic-field dependences $R(H)$ of the resistance for samples with QDs in the region of weak fields can be approximated by the expression

$$\ln(R(H)/R(0)) = H^2/B_0^2 - H/H^* \qquad (3)$$

(dashed lines in Fig. 9). The negative term linear in H takes into account the presence of a negative magnetoresistance in the vicinity of zero; as the magnetic field increases, the magnetoresistance rapidly changes its sign and becomes positive (see Fig. 9). The values of characteristic fields B_0 are shown in Fig. 10. It was found that parameter B_0 for both samples is a nearly linear function of temperature.

3.4. Activation Energy of Low-Temperature Conduction in a Magnetic Field

Figure 11 shows the temperature dependences of conductivity in the region where the Arrhenius law

holds (at $T < T_c$) in various magnetic fields. It can be seen that the activation nature of charge transport is also preserved in a magnetic field. Figure 12 shows the conduction activation energy E_a determined from the slope of the experimental curves. It turns out that the value of E_a increases in a magnetic field and attains saturation in fields $H > H_{sat} = 4$ T.

3.5. Discussion of Experimental Results

The activation dependence of the type $\sigma(T) = \sigma_0 \exp(-E_a/k_B T)$ observed in the region of hopping charge transfer is attributed as a rule to hopping conduction over nearest neighbors [1]. In our case, this conduction mechanism cannot explain the experimental results for the following reasons.

First, the mechanism of conduction over the nearest neighbors must change to the variable range hopping conduction upon cooling. Our experiments reveal the reverse.

Second, in the mechanism of hopping over the nearest neighbors, the characteristic magnetic field B_0 is independent of temperature [9]. However, Ge/Si samples with QDs exhibit a linear relation between B_0 and T (see Fig. 10).

Third, the value of activation energy ε_3 of hopping conduction over the nearest neighbors can be estimated as $\varepsilon_3 = e^2/\kappa l$, where l is the distance between the QD layer and the δ-doped Si layer [12]. For $l = 10$ nm and $\kappa = 12$, we obtain $\varepsilon_3 = 12$ meV, which is an order of magnitude higher than the experimental value of E_a.

The above arguments suggest that the energy states of holes are absent in the vicinity of the Fermi level in a band of width $2E_a$; in other words, the spectrum of states contains a "hard" energy gap. With decreasing temperature, the energy band that contains the energy levels ensuring optimal jumps of holes between QDs at a given temperature becomes narrower. Beginning with a certain temperature, at which the energy bandwidth becomes equal to the gap, the conduction activation energy and the hopping range become independent of temperature, which explains the transition from the Éfros–Shklovskiĭ law to the Arrhenius law. It should be noted that this gap cannot be due to formation of a magnetic polaron in the system since in this case it should vanish in a magnetic field due to screening of the exchange interaction [13–15].

We believe that the observed effects (such as the transition from the Éfros–Shklovskiĭ to the Arrhenius law upon system cooling and the increase on activation energy E_a in a magnetic field) are experimental evidence of the presence of a hard gap in the spectrum of hole states, which is associated with the formation of electron polarons [3]. To interpret this phenomenon, we must recollect that a disordered system contains, in addition to sites forming a through system of routes for the passage of current from site to site, compact pairs of

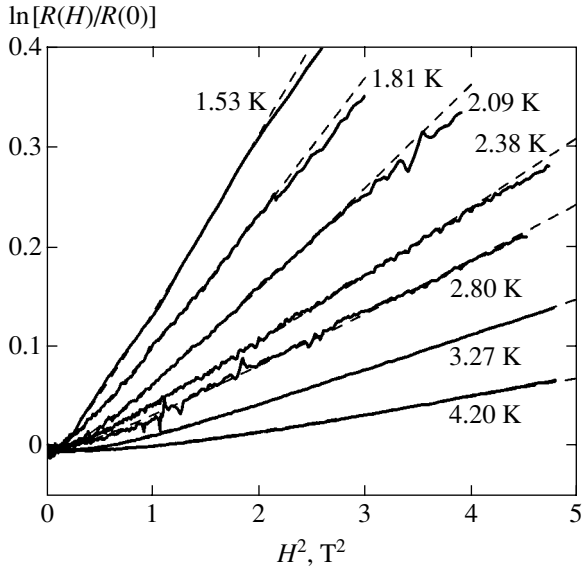


Fig. 9. Dependence of magnetoresistance for a Ge/Si sample with Ge QDs and QD filling factor $\nu = 2.8$ on H^2 in the range of weak magnetic fields.

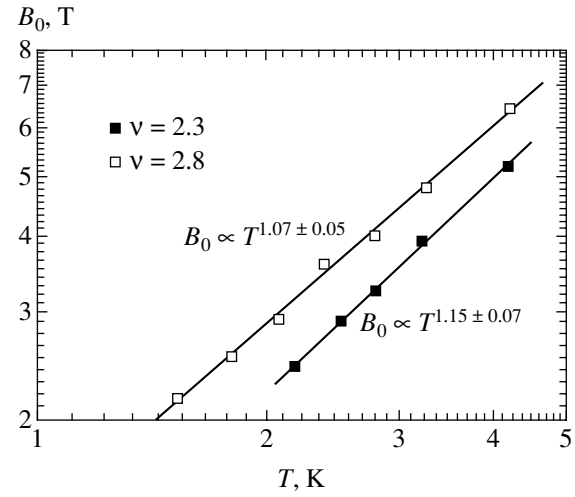


Fig. 10. Values of characteristic magnetic field B_0 obtained by approximating the experimental data by formula (3) for various temperatures. Solid lines show the results of approximation of experimental $B_0(T)$ dependences by a power function.

sites with a low excitation energy and a small spacing in a pair. Such compact pairs are separated by rather long distances. Transitions between such pairs do not occur; consequently, these pairs cannot participate in static conduction, but can affect the electrons moving over a percolation cluster through their dipole potential emerging during transitions in a pair. Let an electron hopping occur from a filled site to site i in a percolation cluster (Fig. 13a). Then it would be advantageous for many compact pairs in the vicinity of site i to transfer an electron from one site in a pair to the other site to

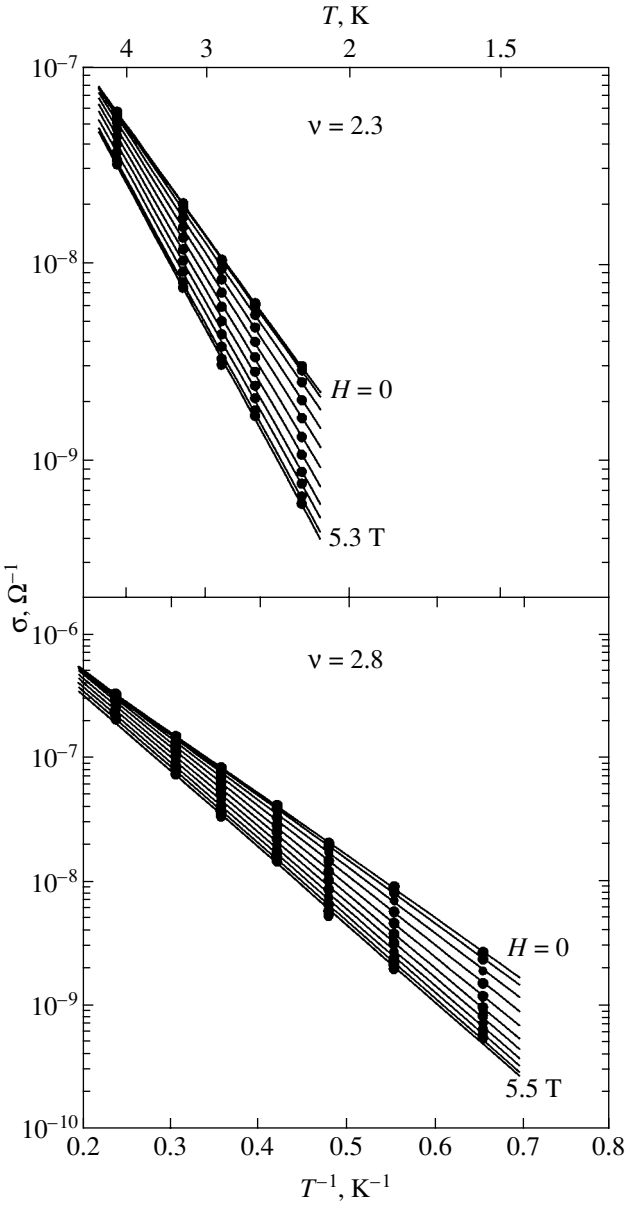


Fig. 11. Temperature dependences of conductivity at $T < 5$ K plotted in the Arrhenius coordinates in various magnetic fields. Symbols indicate the experimental data. The values of magnetic field H for the upper figure from top to bottom: 0, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, and 5.3 T; lower figure: H = 0, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, and 5.5 T.

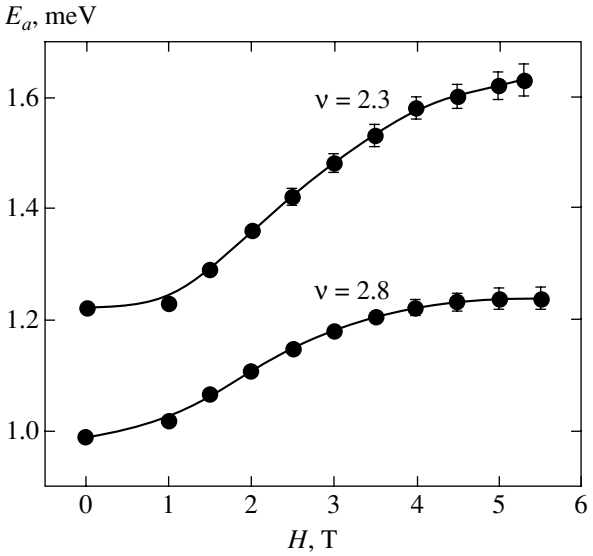


Fig. 12. Magnetic field dependence of the low-temperature conduction activation energy.

reduce the Coulomb repulsion of electrons and to lower the energy of an arriving electron. The "coat" of polarized pairs emerging in this case around the *i*th site and entrained by the electrons tunneling to the next site is known as an electron polaron. This phenomenon resembles in many respects the emergence of a polaron in ionic crystals, when a conduction electron moving over the lattice causes a redistribution of relative positions of positive and negative ions in the lattice. However, polarization in our case emerges not due to displacement of ions, but as a result of a redistribution of electrons in compact pairs.

Theoretical calculations show that, as a result of the polaron effect, filled states of electrons must be separated from empty states by a polaron gap of $2\Delta_p$ (Fig. 13b) [2, 3]. If the disorder in the system is not very strong and the energies of compact pairs in the vicinity of all sites differ insignificantly, the polaron gap is hard in the sense that it does not contain any states altogether. In systems with a strong disorder, the probability that a polaron shift in the vicinity of certain sites is small differs from zero; in this case, the density of states in the polaron gap turns zero in accordance with an exponential law [2, 3].

An increase in the activation energy of hopping conduction in a magnetic field can be interpreted on the basis of the following considerations. The magnetic field deforms the electron wavefunctions at a large distance so that the overlap integral for electrons at remote sites decreases. In turn, the decrease in the overlap integral for the electron wavefunctions in compact pairs with remote sites leads to a decrease in the energy of electron transitions in pairs [16], which increases the possible number of pairs participating in the formation of an electron polaron. The field hardly affects the overlap integrals inside the pairs. Formally speaking, density of states g_0 [17] and, hence, the Coulomb [1] and polaron gaps increase in the magnetic field (the polaron gap width is proportional to the Coulomb gap width [3], at least, for a 3D system).

Unfortunately, analytic models of the polaron effect, which would make it possible to calculate numerically the polaron gap expected for our samples, have not been developed as yet. However, to calculate Δ_p, we can use the results of simulation obtained in [3] and indicating that the polaron gap (at least, for a 3D disordered system) is smaller than the Coulomb gap Δ_C

approximately by a factor of 5. In turn, the value of Δ_C can be estimated from the relation $\Delta_C \sim k_B(T^*T_1)^{1/2}$, where T^* is the temperature above which the Éfros–Shklovskiĭ law is violated [18]. For our samples, T^* = 10–15 K [19]. Setting T^* = 15 K and T_1 = 200 K, we find that $\Delta_C \sim 5$ meV and $\Delta_p \sim 1$ meV, which is indeed close to the experimental value of E_a = 1.0–1.2 meV.

At first glance, the tendency in the conduction activation energy to saturation in strong fields appears as unexpected (Fig. 12). However, this phenomenon can also be explained on the basis of concepts of electron polarons. Following Raikh [16], let us consider the transition of an electron from site 1 to site 2 in a compact pair in the presence of filled center 3 near site 1. The energy required for this transition is defined by the formula

$$\Delta_{12} = \varepsilon_2 - \varepsilon_1 + \frac{|V_{13}|^2}{\varepsilon_3 - \varepsilon_1},$$

where energies ε_1, ε_2, and ε_3 of the sites are measured from the Fermi energy and their spread is determined by topological, composition, or Coulomb disorder in the system as well as by the overlap integrals for sites 1 and 2. The second term in this expression is just the correction associated with overlapping of the wavefunctions of holes at sites 1 and 3 and $|V_{13}|$ is the corresponding overlap integral. In a magnetic field, the overlap integral decreases and the energy of electron transitions in compact pairs tends to a value determined by the values of ε_1 and ε_2 alone. Consequently, the difference $E_a(H = H_{sat}) - E_a(H = 0)$ is an estimate of the typical overlap integral $|V|$ between neighboring QDs. It can be seen from Fig. 12 that $|V| \approx 0.2$–0.4 meV. On the other hand, the overlap integral can be written in the form

$$|V| = \frac{V_0 a}{r_0}\exp\left(-\frac{r_0}{a}\right), \quad (4)$$

where V_0 is the binding energy of a charge carrier at a site (in out case, it is the depth of a level in the quantum well) and r_0 is the mean distance between QDs. Setting $V_0 \sim 300$ meV [4], $|V| \sim 0.3$ meV, and r_0 = 20 nm, we obtain from relation (4) an estimate $a \sim 4$ nm for the radius of localization of a hole in an excited state, which is close to the value (approximately equal to 3 nm) obtained in [20] as a result of numerical simulation of the electron structure of hole states for analogous Ge/Si(001) QD layers.

After the manuscripts had been prepared, we learned that Dubrovskii *et al.*, who studied the tunneling between two disordered systems in a transverse magnetic field [21], observed the formation of a hard gap in the density of localized states in strong fields.

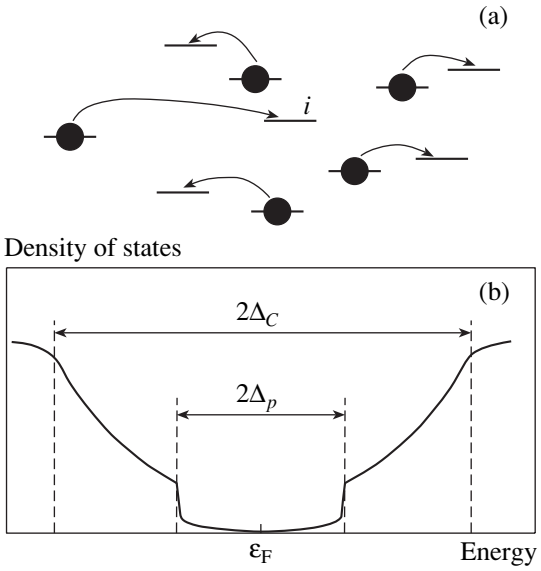


Fig. 13. (a) Formation of an electron polaron by placing an electron at the *i*th site. (b) Energy dependence of the density of states: ε_F is the Fermi level, Δ_C is the Coulomb gap, and Δ_p is the polaron gap in the spectrum of states.

Probably, the mechanisms of this phenomenon and the processes described above are of the same nature.

ACKNOWLEDGMENTS

The authors thank V.V. Kirienko and N.P. Stepina for their help in preparing the samples and V.A. Volodina, who measured the Raman spectra.

This study was financed by the Russian Foundation for Basic Research (project no. 03-02-16526), the program of the President of the Russian Federation in support of young doctors of science (grant no. MD-28.2003.02), and INTAS (grant nos. 2001-0615 and 03-51-5051). A.A. Sherstobitov is grateful to the Ministry of Education of the Russian Federation (grant no. A03-2.9-521).

REFERENCES

1. B. I. Shklovskiĭ and A. L. Éfros, *Electronic Properties of Doped Semiconductors* (Nauka, Moscow, 1979; Springer, New York, 1984).
2. S. D. Baranovskiĭ, B. I. Shklovskiĭ, and A. L. Éfros, Zh. Éksp. Teor. Fiz. **78**, 395 (1980) [Sov. Phys. JETP **51**, 199 (1980)].
3. R. Chicon, M. Ortuño, and M. Pollak, Phys. Rev. B **37**, 10520 (1988).
4. A. V. Dvurechenskii, A. V. Nenashev, and A. I. Yakimov, Nanotechnology **13**, 75 (2002).
5. A. V. Nenashev, A. V. Dvurechenskii, and A. F. Zinovieva, Phys. Rev. B **67**, 205301 (2003).
6. D. N. Tsigankov and A. L. Efros, Phys. Rev. Lett. **88**, 176602 (2002).

7. V. L. Nguen, B. Z. Spivak, and B. I. Shklovskiĭ, Zh. Éksp. Teor. Fiz. **89**, 1770 (1985) [Sov. Phys. JETP **62**, 1021 (1985)].
8. W. Schirmacher, Phys. Rev. B **41**, 2461 (1990).
9. V. L. Nguen, Fiz. Tekh. Poluprovodn. (Leningrad) **18**, 335 (1984) [Sov. Phys. Semicond. **18**, 207 (1984)].
10. B. I. Shklovskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **36**, 43 (1982) [JETP Lett. **36**, 51 (1982)].
11. V. I. Kozub, S. D. Baranovskii, and I. Shlimak, Solid State Commun. **113**, 587 (2000).
12. Qiu-yi, B. I. Shklovskii, A. Zrenner, *et al.*, Phys. Rev. B **41**, 8477 (1990).
13. A. N. Aleshin, A. N. Ionov, R. V. Parfen'ev, *et al.*, Fiz. Tverd. Tela (Leningrad) **30**, 696 (1988) [Sov. Phys. Solid State **30**, 398 (1988)].
14. P. Dai, Y. Zhang, and M. P. Sarachik, Phys. Rev. Lett. **69**, 1804 (1992).
15. A. I. Yakimov, T. Wright, C. J. Adkins, and A. V. Dvurechenskii, Phys. Rev. B **51**, 16549 (1995).
16. M. E. Raikh, Solid State Commun. **75**, 935 (1990).
17. M. E. Raikh, J. Czingon, Qiu-yi Ye, *et al.*, Phys. Rev. B **45**, 6015 (1992).
18. A. G. Zabrodskiĭ and K. N. Zinov'eva, Zh. Éksp. Teor. Fiz. **86**, 727 (1984) [Sov. Phys. JETP **59**, 425 (1984)].
19. A. I. Yakimov, A. V. Dvurechenskiĭ, A. I. Nikiforov, and A. A. Bloshkin, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 445 (2003) [JETP Lett. **77**, 376 (2003)].
20. A. I. Yakimov, A. V. Dvurechenskii, A. V. Nenashev, and A. I. Nikiforov, Phys. Rev. B **68**, 205310 (2003).
21. Yu. V. Dubrovskii, V. A. Volkov, L. Eaves, *et al.*, in *Proceedings of 12th International Symposium on Nanostructures: Physics and Technology* (St. Petersburg, 2004), p. 342.

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 100, No. 4, 2005, pp. 731–741.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 4, 2005, pp. 827–837.

ELECTRONIC PROPERTIES OF SOLIDS

Analytic Properties of the Effective Dielectric Constant of a Two-Dimensional Rayleigh Model

B. Ya. Balagurov and V. A. Kashin

Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Moscow, 119991 Russia
e-mail: balagurov@deom.chph.ras.ru

Received September 1, 2004

Abstract—Analytic properties of the dimensionless static effective dielectric constant $f(p, h)$ of a two-dimensional Rayleigh model (p is the concentration and h is the ratio of the dielectric constants of components) are considered as a function of the complex variable h. It is shown that the only singularities of the function $f(p, h)$ are first-order poles for real $h = h_n < 0$ ($n = 1, 2, \ldots$) with the condensation point $h = –1$, which form an infinite discrete (countable) set. The positions of the first ten poles of the function $f(p, h)$ and the residues at these points are calculated and represented graphically versus the concentration. Based on the results obtained, a pole-type approximate formula is proposed that describes the behavior of the function $f(p, h)$ over a wide range of p and complex h. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The experiments carried out in [1, 2] revealed an anomalous growth in the dielectric constant of samples when approaching a metal–insulator phase-transition point. This fact has stimulated interest in the more general problem, the study of the low-frequency electrophysical properties of such systems (in particular, of binary composites). Theoretical analysis of this problem was carried out within the similarity hypothesis [3, 4] and by the method of the effective-medium theory [5]. In [6], the low-frequency dispersion of the conductivity of a disordered two-component lattice model was analyzed by numerical methods. The studies of [4–6] revealed a number of essential features of this phenomenon and showed that investigating the dispersion of the dielectric constant (or the conductivity) of such systems is important, in particular, to gain a deeper insight into the metal–insulator phase transition.

In [7], a more general approach (compared with that of [4, 5]) has been applied to the problem of low-frequency dispersion of the dielectric constant of binary composites; the application of this approach is not restricted by the applicability domain of approximate methods, such as the similarity hypothesis or the effective-medium theory. This approach requires that one should know the properties of the dimensionless static effective dielectric constant $f(p, h)$ of a composite (p is the concentration and h is the ratio of the dielectric constants of the components) for complex values of the variable h. This is associated with the fact that, in a low-frequency (quasi-stationary [8]) electric field, the dielectric constants of individual components, as well as of the composite as a whole, are complex quantities.

The function $f(p, h)$, considered as a function of complex frequency ω, is analytic in the upper half-plane $\mathrm{Im}\omega > 0$, while its singularities lie in the half-plane $\mathrm{Im}\omega \le 0$. However, a natural independent variable for f is h rather than the frequency ω; we denote the complex value of h by ζ. It turns out that the static effective dielectric constant $f(p, \zeta)$, considered as a function of ζ, is analytic in the whole complex plane ζ except for the negative real half-line ($\mathrm{Re}\zeta \le 0$, $\mathrm{Im}\zeta = 0$) (see [7]). Thus, all singularities of the function $f(p, \zeta)$ lie on this half-line (and, possibly, at the infinite point $\zeta = \infty$) [7].

Knowledge of these singularities allows us to write out a dispersion relation. Using this relation, one can express the function $f(p, \zeta)$ with arbitrary complex ζ in terms of its imaginary part on the negative real half-line (more precisely, for $\mathrm{Re}\zeta < 0$ and $\mathrm{Im}\zeta = i0$). Note that the type of the singularities of $f(p, \zeta)$ is not universal but depends on the specific structure of a composite. For instance, according to [9, 10], the function $f(p, \zeta)$ has only a discrete set of first-order poles for finite samples. One may assume that infinite periodic systems also have a discrete set of such poles (which is confirmed, for example, by the results of the present study). For composites with a random distribution of components, one should expect that the poles merge into a cut $\mathrm{Re}\zeta \le 0$ and $\mathrm{Im}\zeta = 0$, so that, on the upper (or lower) bank of this cut, $\mathrm{Im}f(p, \zeta)$, considered as a function of $t = –\mathrm{Re}\zeta$, is generally different from zero in the whole interval $0 < t < \infty$ (see, for example, [11]).

The study of the analytic properties of $f(p, \zeta)$ for concrete models faces difficulties associated with the absence of exact solutions for binary systems with finite (nonzero) values of the dielectric constants of

both components. An exception is given by the following model, which was first proposed by Rayleigh [12]: a two-dimensional matrix with a doubly periodic distribution of circular inclusions. In [12], Rayleigh actually found only the first few terms of a virial decomposition (for a small concentration of inclusions) for the effective dielectric constant of such a system. A full solution to this problem is given, for example, in [13, 14].

In the present paper, we apply the results obtained in [13, 14], which can also be applied in a complex domain, to study the analytic properties of the effective static dielectric constant of the two-dimensional Rayleigh model. We show that, according to the general considerations given in [7], all singularities of the function $f(p, \zeta)$ are concentrated on the negative real halfline and constitute an infinite discrete (countable) set of simple poles. In this paper, we determine the positions of the first ten poles and the residues at these poles as a function of the concentration of inclusions. To this end, we solve the basic equations from [13, 14] by numerical methods. In specific calculations, we single out a subset of 40 equations from the infinite set of these equations.

In the case of the periodic model considered here, the function $f(p, \zeta)$ is represented as an infinite sum of polar terms. Taking into account a finite number of terms, we obtain a relatively simple approximate formula for $f(p, \zeta)$. In this case, the complex three-parameter function $f(p, \zeta)$ is reduced to a real function of a single variable (the concentration). The calculation of the position of the first few poles and the corresponding residues as a function of the concentration of inclusions allows us to describe $f(p, \zeta)$ over a wide range of the variables p and ζ.

The model under discussion is investigated within the dielectric-constant problem (with the use of appropriate terminology). Similar problems concerning the conductivity, thermal conductivity, steady-state diffusion, etc., differ from the dielectric-constant problem by an obvious change of notation, and the solutions to these problems (for a given model) are expressed in terms of the same function $f(p, \zeta)$.

2. PRELIMINARY REMARKS

Let us express the static effective dielectric constant ε_e of an isotropic binary composite as

$$\varepsilon_e = \varepsilon_e(p; \varepsilon_1, \varepsilon_2) = \varepsilon_1 f(p, h); \quad h = \varepsilon_2/\varepsilon_1. \tag{1}$$

Here, p is the concentration of the first component and ε_i ($i = 1, 2$) is the dielectric constant of the ith component; the type of the function $f(p, h)$—the dimensionless effective static dielectric constant—depends on the specific structure of the composite. Note that, in the two-dimensional case, the function $f(p, h)$ satisfies the reciprocity relation [15] (see also [7])

$$f(p, h) f(p, 1/h) = 1, \tag{2}$$

which is valid for isotropic two-dimensional, two-component systems of arbitrary structure.

In a low-frequency (quasi-stationary [8]) electric field, the expression for the effective dielectric constant remains the same as (1); however, the parameters ε_e and ε_i are now complex functions of frequency ω. In this case, the argument $h = h(\omega)$ of the function $f(p, h)$ is also complex. For instance, for "poor" conductors (say, for semiconductors), it is reasonable to introduce the conductivity σ and the dielectric constant ε simultaneously [8]. This is conveniently done by introducing the complex dielectric constant $\varepsilon(\omega)$ [8]

$$\varepsilon(\omega) = \varepsilon + i\frac{4\pi\sigma}{\omega}. \tag{3}$$

One of the conditions of quasi-stationarity is the independence of ε and σ of frequency [8], which we assume to be satisfied. In this case, by ε and σ are meant the static values of these quantities.

As pointed out above, the determination of the effective low-frequency electrophysical characteristics of a medium (a composite) in the quasi-stationary approximation differs from that in the static case only by the change of the dielectric constant ε in (1) by the complex dielectric constant $\varepsilon(\omega)$. In this case, we obtain the following expression for $\varepsilon(\omega)$ given by (3):

$$\varepsilon_e + i\frac{4\pi\sigma_e}{\omega} = \left(\varepsilon_1 + i\frac{4\pi\sigma_1}{\omega}\right) f(p, h(\omega)), \tag{4}$$

$$h(\omega) = \left(\varepsilon_2 + i\frac{4\pi\sigma_2}{\omega}\right) \Big/ \left(\varepsilon_1 + i\frac{4\pi\sigma_1}{\omega}\right). \tag{5}$$

Hence, we obtain

$$\varepsilon_e = \varepsilon_1 \mathrm{Re}\, f(p, h(\omega)) - \frac{4\pi\sigma_1}{\omega}\mathrm{Im}\, f(p, h(\omega)), \tag{6}$$

$$\sigma_e = \sigma_1 \mathrm{Re}\, f(p, h(\omega)) + \frac{\varepsilon_1\omega}{4\pi}\mathrm{Im}\, f(p, h(\omega)) \tag{7}$$

with $h(\omega)$ from (5). Note that, in contrast to ε_i and σ_i, the effective characteristics ε_e and σ_e depend on frequency ω.

Thus, knowledge of the properties of the static function $f(p, \zeta)$ in the plane of the complex variable ζ allows us to apply formula (1) even for $\omega \neq 0$ provided that we make the substitution

$$\zeta \longrightarrow h(\omega) = \varepsilon_2(\omega)/\varepsilon_1(\omega). \tag{8}$$

For systems with a metal–insulator phase transition, the behavior of the function $f(p, h)$ in the critical domain is usually described within the similarity hypothesis [4]. The application of appropriate expressions (more precisely, their analytic continuations) for $f(p, h)$ allows us to describe, in this approximation, the behavior of the effective parameters ε_e and σ_e as functions of frequency ω in the entire critical domain (see [4, 7]. On the other hand, the substitution of the expression for $f(p, h)$ calculated within the effective-medium theory [16] into formulas (4)–(7) reproduces the theoretical results of [5].

3. ANALYTIC PROPERTIES OF THE FUNCTION $f(p, \zeta)$

It is well known that the dielectric constant, considered as a function of complex frequency ω, is analytic in the upper half-plane $\mathrm{Im}\omega > 0$ (see [8, Section 82]). Moreover, according to [8], the dielectric constant has no zeros for $\mathrm{Im}\omega > 0$ and finite ω. Therefore, the function $f = \varepsilon_e(\omega)/\varepsilon_1(\omega)$ is also analytic in the upper half plane of ω. To determine the analytic properties of $f(p, \zeta)$ as a function of the complex variable ζ, one should know to what domain of the plane ζ the transform $\zeta = h(\omega)$ with $h(\omega)$ from (8) maps the upper half-plane $\mathrm{Im}\omega > 0$. A consideration of specific transforms shows [7] that the half-plane $\mathrm{Im}\zeta > 0$ is generally mapped to the whole plane ζ with the removed negative real half-line ($\mathrm{Re}\zeta < 0$, $\mathrm{Im}\zeta = 0$). Hence, the function $f(p, \zeta)$ is analytic in this domain of the plane ζ, while the singularities of this function lie on the above-mentioned half-line; in particular, the imaginary part of f has a discontinuity on this half-line. The infinite point $\zeta = \infty$ requires separate consideration. In the two-dimensional case considered here, the function $f(p, \zeta)$ with $\zeta = \infty$ either is finite, or has a first-order pole [7] if the concentration p is different from the critical concentration ($p \neq p_c$). Note also that, according to [7], the following equality holds:

$$f(p, \zeta^*) = f^*(p, \zeta), \tag{9}$$

where the star denotes complex conjugation.

Knowledge of the analytic properties of $f(p, \zeta)$ allows us to write out a dispersion relation. For definiteness, consider a two-dimensional case ($D = 2$) and the domain of concentrations $p > p_c$ when there is no percolation in the second component, so that the function $f(p, \zeta)$ takes a finite value $f(p, \infty)$ as $\zeta \longrightarrow \infty$. According to [7], under these conditions, the dispersion relation has the form

$$f(p, \zeta) = f(p, \infty) - \frac{1}{\pi}\int_0^\infty \frac{\mathrm{Im}\, f^{(+)}(p, -t)}{t + \zeta}\, dt, \tag{10}$$

where $f^{(+)}(p, -t) = f(p, -t + i0)$. Hence, to determine $f(p, \zeta)$ on the whole plane of ζ, it suffices to know the imaginary part of $f^{(+)}$ and the quantity $f(p, \infty)$.

Note that the real part of impedance in the so-called LC model is expressed in terms of the imaginary part of $f^{(+)}$; one component of this model has inductive reactance, and the other has purely capacitive reactance (see [7]). According to [8], the real part of the impedance is nonnegative, so that [7, 11]

$$\mathrm{Im}\, f^{(+)}(p, -t) \geq 0. \tag{11}$$

As pointed out in [15] (see also [7]), the fact that the real part of the effective impedance is different from zero (i.e., the presence of real absorption of energy) is associated with the existence of impurity levels (local oscillations) in the LC model whose resonant excitation is responsible for the energy dissipation. Thus, the mathematical singularities of the function $f(p, \zeta)$ in the complex plane ζ are directly related to a physical phenomenon—the presence of local oscillations in the LC model.

In periodic systems, the polar terms in f correspond to resonances at certain natural frequencies ω_n in the LC model, which are expected to form a discrete spectrum. The frequencies ω_n correspond to real values (eigenvalues) $h_n = h(\omega_n) < 0$ with $h(\omega) = -\omega^2/\Omega^2$ [7], where Ω is the Thomson frequency (the resonance frequency of the LC circuit). At $\zeta = h_n$, the function $f(p, \zeta)$ has poles. The eigenvalues h_n, which also form a discrete spectrum in the case of a periodic system, depend only on the structure (i.e., on the geometrical characteristics) of a composite and are independent of the specific character of the problem (cf. [17]). Therefore, the values h_n, which represent dimensionless numbers, are the same (for a given structure of the composite) both for a frequency-dependent LC model and for static problems (problems of dielectric constant, conductivity, thermal conductivity, etc.). In this case, the eigenvalues h_n are primary quantities that characterize the geometry of the composite, while, for example, the resonance frequencies ω_n of the LC model are secondary quantities, which are expressed in terms of h_n: $\omega_n^2 = -h_n\Omega^2 > 0$.

4. PERIODIC MODELS

As pointed out above, one should expect that, in periodic systems, the eigenvalues h_n (and, hence, the poles of the function f) form a discrete spectrum. Therefore, if we consider, as in the previous section, the two-dimensional case and the domain of concentrations $p > p_c$ (which corresponds to $R < a$ in the Rayleigh model), we obtain the following expression for the function $f(p, \zeta)$:

$$f(p, \zeta) = f(p, \infty) - \sum_{n=1}^{\infty} \frac{F_n}{\zeta - h_n}, \tag{12}$$

where $h_n < 0$. Hence, the imaginary part of $f(p, \zeta)$ for $\zeta = -t + i0$ (where t is real and positive) is represented by a sum of delta functions:

$$\mathrm{Im}\, f^{(+)}(p, -t) = \pi \sum_{n=1}^{\infty} F_n \delta(t + h_n). \quad (13)$$

Formula (13) and inequality (11) imply that

$$F_n \geq 0. \quad (14)$$

The function $f(p, \zeta)$ given by (12) has all the properties established in Section 3; in particular, it satisfies equality (9) because $f(p, \infty)$, F_n, and h_n are real. Note also that the substitution of (13) into (10) reproduces formula (12).

Since $f = 1$ for $\zeta = h = 1$ (a homogeneous medium), the following equality must hold:

$$f(p, \infty) = 1 + \sum_{n=1}^{\infty} \frac{F_n}{1 - h_n}. \quad (15)$$

For $\zeta = h \longrightarrow 1$, expansion (12) up to the term $(1 - h)^2$ inclusive, combined with (15), yields

$$f(p, h) = 1 - (1 - h) \sum_{n=1}^{\infty} \frac{F_n}{(1 - h_n)^2} - (1 - h)^2 \sum_{n=1}^{\infty} \frac{F_n}{(1 - h_n)^3} + \ldots. \quad (16)$$

Comparing (16) with the general expression for a weakly inhomogeneous isotropic two-dimensional binary system (see, for example, [11])

$$f(p, h) = 1 - c(1 - h) - \frac{1}{2} c(1 - c)(1 - h)^2 + \ldots, \quad (17)$$

we arrive at the two identities

$$\sum_{n=1}^{\infty} \frac{F_n}{(1 - h_n)^2} = c = 1 - p, \quad (18)$$

$$\sum_{n=1}^{\infty} \frac{F_n}{(1 - h_n)^3} = \frac{1}{2} c(1 - c) = \frac{1}{2} p(1 - p), \quad (19)$$

where c is the concentration (for $D = 2$, the area fraction) of the second component.

Setting $\zeta = 0$ in (12), we obtain the relation

$$f(p, 0) - f(p, \infty) = \sum_{n=1}^{\infty} \frac{F_n}{h_n}, \quad (20)$$

which implies the obvious inequality $f(p, 0) < f(p, \infty)$ because $F_n > 0$ and $h_n < 0$. Eliminating $f(p, \infty)$ from (12) with the use of (20), we obtain

$$f(p, \zeta) = f(p, 0) - \zeta \sum_{n=1}^{\infty} \frac{F_n}{h_n(\zeta - h_n)}. \quad (21)$$

Using (12) with $\zeta = h$, we expand $f(p, h)$ and $f(p, 1/h)$ in powers of h and substitute the result into reciprocity relation (2). Taking into account (20), we obtain the following relation in the linear approximation in h:

$$f(p, 0) \sum_{n=1}^{\infty} F_n = f(p, \infty) \sum_{n=1}^{\infty} \frac{F_n}{h_n^2}. \quad (22)$$

The functions $f(p, 0)$ and $f(p, \infty)$ are related by the formula

$$f(p, 0) f(p, \infty) = 1, \quad (23)$$

which also follows from (2) for $h \longrightarrow 0$. Eliminating $f(p, \infty)$ from (22) with the use of (23), we obtain

$$[f(p, 0)]^2 = \sum_{n=1}^{\infty} \frac{F_n}{h_n^2} \Big/ \sum_{n=1}^{\infty} F_n. \quad (24)$$

Differentiating (12) with respect to ζ, we find

$$\frac{\partial f(p, \zeta)}{\partial \zeta} = \sum_{n=1}^{\infty} \frac{F_n}{(\zeta - h_n)^2}. \quad (25)$$

Hence, for real $\zeta = h$, we have

$$\frac{\partial f(p, h)}{\partial h} > 0 \quad (26)$$

for positive h and

$$\frac{\partial \mathrm{Re}\, f(p, h)}{\partial h} > 0 \quad (27)$$

for negative h. For $h < 0$, $\mathrm{Re} f(p, h)$ monotonically increases (as h increases) in any interval bounded by two neighboring poles.

The substitution of (15) into (12) yields the following expression for $f(p, \zeta)$:

$$f(p, \zeta) = 1 + \sum_{n=1}^{\infty} F_n \frac{\zeta - 1}{(1 - h_n)(\zeta - h_n)}, \quad (28)$$

which contains only the quantities h_n and F_n. When $\zeta = 0$, expression (28) implies the following representation for $f(p, 0)$:

$$f(p, 0) = 1 + \sum_{n=1}^{\infty} \frac{F_n}{h_n(1 - h_n)}. \quad (29)$$

Note that formulas (15), (18)–(20), (22)–(24), and (29) can be applied to control the correct determination of h_n and F_n when studying specific two-dimensional periodic models.

5. THE RAYLEIGH MODEL

The model investigated in the present study represents a two-dimensional isotropic matrix with dielectric constant ε_1 that contains circular inclusions of radius R with dielectric constant ε_2. The inclusions form a regular structure: the centers of inclusions are situated at the nodes of a square lattice with period $2a$ (Fig. 1). The problem of calculating the effective dielectric constant ε_e of such a model was first considered by Rayleigh in [12], where he determined the first few terms of the corresponding virial decomposition. A full solution within the conductivity problem is given, for example, in [13, 14].

According to [13, 14], outside inclusions, the complex potential of the problem is given by the expression ($z = x + iy$)

$$\Phi(z) = \beta\left\{z + \sum_{k=0}^{\infty} B_{2k}\zeta^{(2k)}(z)\right\} \quad (30)$$

with real coefficients β and B_{2k}. The function $\zeta^{(2k)}(z)$ in (30) is the $2k$th-order derivative of the Weierstrass zeta function [18]

$$\zeta(z) = \frac{1}{z} + \sum_{l,m}{}' \left[\frac{1}{z - z_{lm}} + \frac{1}{z_{lm}} + \frac{z}{z_{lm}^2}\right], \quad (31)$$
$$z_{lm} = 2(l + im)a.$$

The summation in (31) is performed over all (both positive and negative) integers l and m, except for the $l = m = 0$.

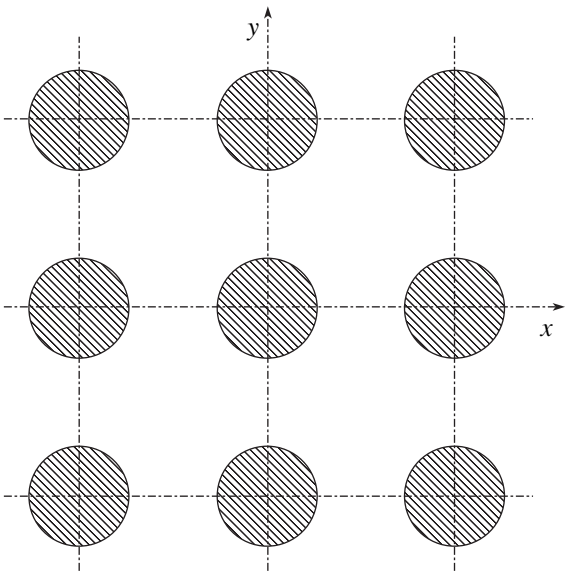


Fig. 1. Two-dimensional Rayleigh model.

For small z, the following expansion is valid for the function $\zeta(z)$ [18]:

$$\zeta(z) = \frac{1}{2} - \sum_{k=1}^{\infty} \frac{c_{2k}}{4k - 1} z^{4k-1}. \quad (32)$$

Here,

$$c_2 = g \equiv \frac{g_2}{20}, \quad c_4 = \frac{1}{3}g^2, \quad c_6 = \frac{2}{3 \cdot 13}g^3, \ldots, \quad (33)$$

where

$$g_2 = \frac{1}{a^4}\left[K\left(\frac{1}{\sqrt{2}}\right)\right]^4 \quad (34)$$

is an invariant of the Weierstrass zeta function [18]. In (34), $K(1/\sqrt{2}) = 1.85407\ldots$ is the complete elliptic integral of the first kind $K(k)$ with modulus $k = 1/\sqrt{2}$. The coefficients c_{2k} for $k \geq 2$ can be found from the recurrence relation [18]

$$c_{2k} = \frac{3}{(4k + 1)(2k - 3)} \sum_{m=1}^{k-1} c_{2m}c_{2k-2m} \quad (k \geq 2). \quad (35)$$

Note that expansion (32) does not contain coefficients with odd indices, so that, for the square lattice considered here (a lemniscatic case [18]), the coefficients c_{2k+1} vanish.

The coefficients B_{2k} ($k = 0, 1, 2, \ldots$) in (30) satisfy the infinite system of equations from [13, 14]. If,

instead of B_{2k}, we introduce "variables" ξ_k by the formulas

$$B_{2k} = \frac{R^{2k+2}\delta}{\sqrt{(2k)!(2k+1)!}}\xi_k, \quad \delta = \frac{1-h}{1+h}, \quad h = \frac{\varepsilon_2}{\varepsilon_1}, \tag{36}$$

then this system takes the form [14]

$$\xi_k + \sum_{l=0}^{\infty} S_{kl}\xi_l = \delta_{k0} \quad (k = 0, 1, 2, \ldots), \tag{37}$$

where δ_{k0} is the Kronecker delta and

$$S_{kl} = \frac{(2l+2k)!R^{2(k+l+1)}c_{k+l+1}}{\sqrt{(2k)!(2k+1)!(2l)!(2l+1)!}}\delta. \tag{38}$$

The coefficients c_{k+l+1} are defined in (32)–(35), and δ is defined in (36). The matrix S_{kl} is symmetric and is different from zero only when the indices k and l have different parities. In particular, $S_{00} = 0$ because $c_1 \equiv 0$.

According to [13], the dimensionless effective dielectric constant $f = \varepsilon_e/\varepsilon_1$ of the model under consideration is given by

$$f = \frac{1 - c\xi_0\delta}{1 + c\xi_0\delta}, \tag{39}$$

where

$$c = \pi R^2/4a^2 \tag{40}$$

is the concentration of inclusions. The roots of the equation

$$\xi_0^{-1} + c\delta = 0 \tag{41}$$

determine the positions of the poles h_n. As $h \longrightarrow h_n$, the function f takes the form

$$f \approx -\frac{F_n}{h - h_n}, \tag{42}$$

where

$$F_n = -\frac{2}{c}\left\{\left[\frac{d(\xi_0\delta)}{dh}\right]_{h=h_n}\right\}^{-1} \tag{43}$$

is the residue, taken with opposite sign, of the function f at the pole $h = h_n$.

Equations (37) remain valid for arbitrary complex h, including real $h < 0$. The sought-for quantities h_n and F_n are obtained by substituting $\xi_0 = \xi_0(h)$, which is determined from system (37) for negative h, into (41) and (43).

6. CALCULATION OF h_n AND F_n

For an isotropic two-dimensional system with circular inclusions, we have

$$f = 1 - 2c\frac{1-h}{1+h} \tag{44}$$

in the linear approximation in the concentration of inclusions (cf. a similar formula in the three-dimensional case [8]). Expression (44), which is valid for any (either random or periodic) distribution of disks, has a single pole at $h = -1$. This is associated with the fact that, in deriving (44), we actually considered an isolated inclusion. At the same time, $h = -1$ is an infinitefold degenerate eigenvalue for a disk, and only "interaction" with other disks removes the degeneracy. For instance, in the case of a pair of disks of radius R, there are two sets of eigenvalues [7]

$$\begin{aligned} h_{1n} &= -\tanh n\mu_0, \\ h_{2n} &= -\coth n\mu_0 \quad (n = 1, 2, \ldots), \\ \mu_0 &= \ln\frac{\varrho + \sqrt{\varrho^2 - 4R^2}}{2R}, \end{aligned} \tag{45}$$

where ϱ is the distance between the centers of the disks. In this case, $h = -1$ is a condensation point (as $n \longrightarrow \infty$) of h_{1n} and h_{2n}. Note that the eigenvalues of (45) correspond to the poles of the polarizability of a pair of disks (circular cylinders) [11].

A similar situation occurs in the Rayleigh model: a whole spectrum of poles arises for the function f in higher order approximations in the concentration. In this case, to evaluate h_n and F_n for $c \ll 1$, we will solve a finite subset of N equations for increasing N instead of (37). For further analysis, it is convenient to factor out the coefficient δ from the matrix S_{kl} defined in (38):

$$S_{kl} = A_{kl}\delta. \tag{46}$$

Taking into account (46), we obtain the following expressions from (38):

$$\begin{aligned} A_{01} &= \frac{1}{\sqrt{3}}c_2R^4, \quad A_{12} = \sqrt{15}c_4R^8, \\ A_{03} &= \frac{1}{\sqrt{7}}c_4R^8, \quad A_{23} = 6\sqrt{35}c_6R^{12}, \ldots, \end{aligned} \tag{47}$$

where c_{2k} is defined in (32)–(35).

If we restrict the analysis to a single equation (for $k = 0$) in (37), then $\xi_0 = 1$. In this case, (41) implies $\delta_1 = -1/c$, so that $h_1 = -(1 + 2c + 2c^2 + \ldots)$ and $F_1 =$

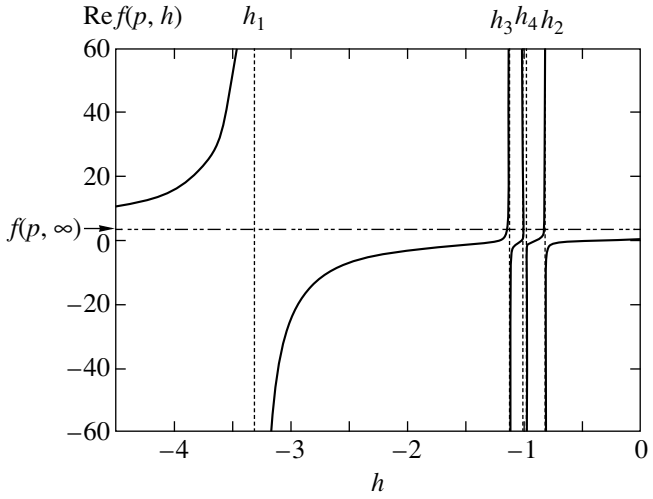


Fig. 2. Re$f(p, h)$ as a function of h on the negative real half-line for the concentration of inclusions $c = 1 - p = 0.5$. The figure represents only the first four poles; the remaining infinite family of poles is concentrated between h_3 and h_4.

$4c(1 + 2c + \ldots)$ for $c \ll 1$. In the next order of approximation ($k = 0, 1$ and $l = 0, 1$), we have

$$N = 2: \ \xi_0^{-1} = 1 - A_{01}^2 \delta^2. \tag{48}$$

The substitution of (48) into (41) yields a second-order algebraic equation in δ that has two real roots. Using these roots, we obtain the following expressions for $c \ll 1$:

$$h_1 = -\left(1 + 2c + 2c^2 + 2c^3 + \frac{2}{c}A_{01}^2 + \ldots\right), \tag{49}$$

$$F_1 = 4c\left(1 + 2c + 3c^2 - \frac{1}{c^2}A_{01}^2 + \ldots\right), \tag{50}$$

$$h_2 = -\left(1 - \frac{2}{c}A_{01}^2 + \ldots\right), \tag{51}$$

$$F_2 = \frac{4}{c}A_{01}^2 + \ldots. \tag{52}$$

For $c \longrightarrow 0$, we have $|h_1 + 1| \sim F_1 \sim c$, $|h_2 + 1| \sim F \sim c^3$, and $|\delta_2| \sim 1/c^3$.

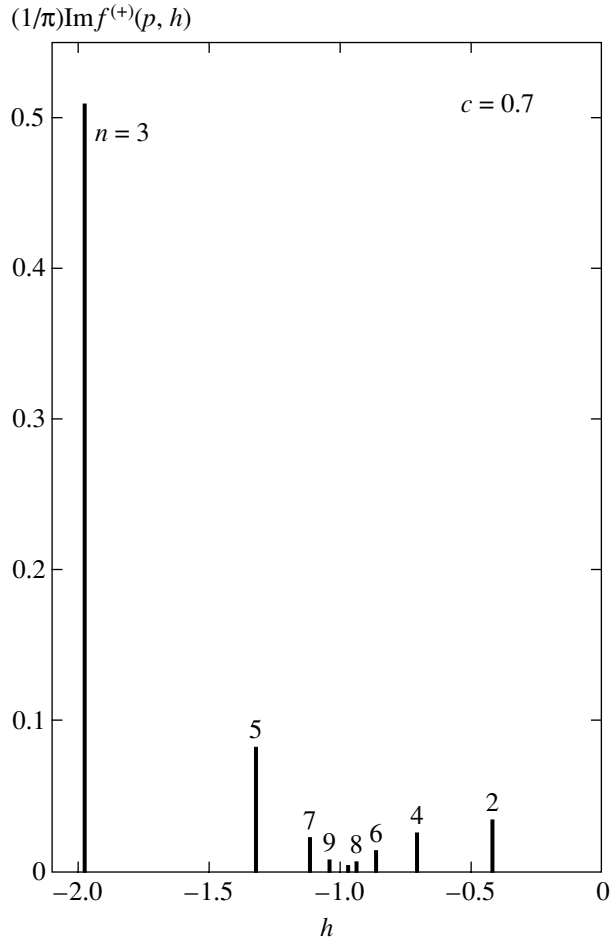


Fig. 3. Im$f^{(+)}(p, h)$ as a function of h for $c = 1 - p = 0.7$. The height of each peak is proportional to the "force" F_n of the corresponding delta function (see formula (13)).

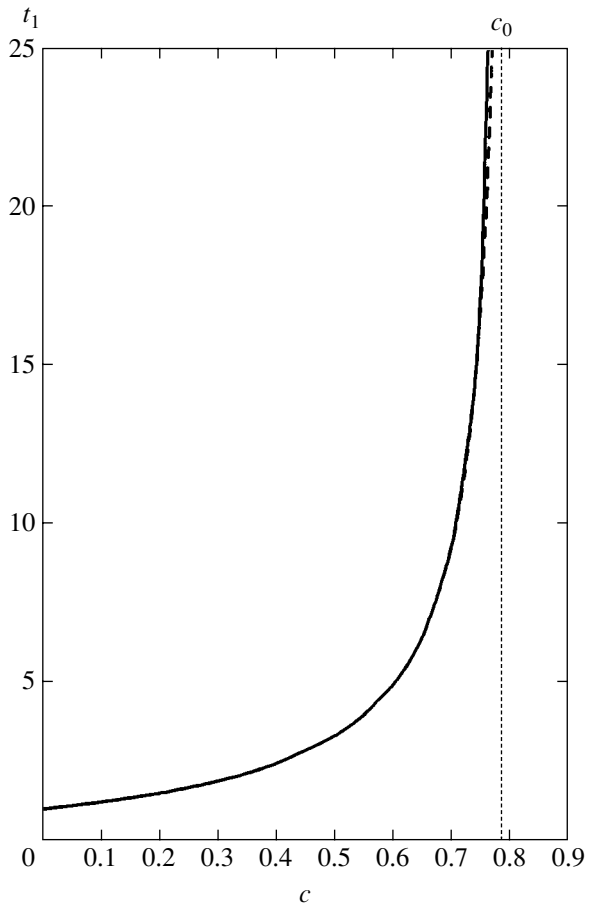


Fig. 4. $t_1 = -h_1$ as a function of concentration c. The dashed line represents t_1 as a function of c in the approximation of four equations. Here, $c_0 = \pi/4 \approx 0.785$ is the critical concentration (for $R = a$).

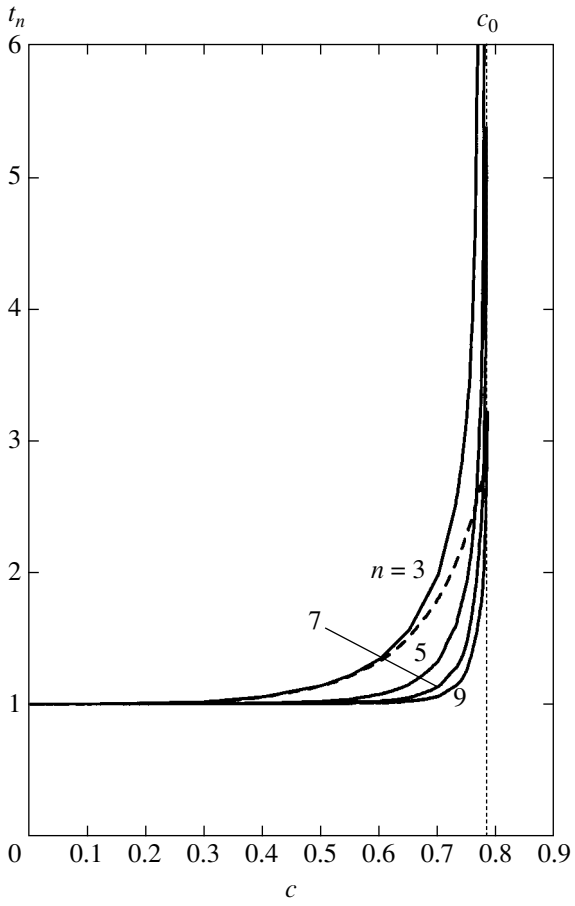


Fig. 5. $t_n = -h_n$ ($n = 3, 5, 7, 9$) as functions of concentration c. The dashed line represents t_3 as a function of c in the approximation of four equations.

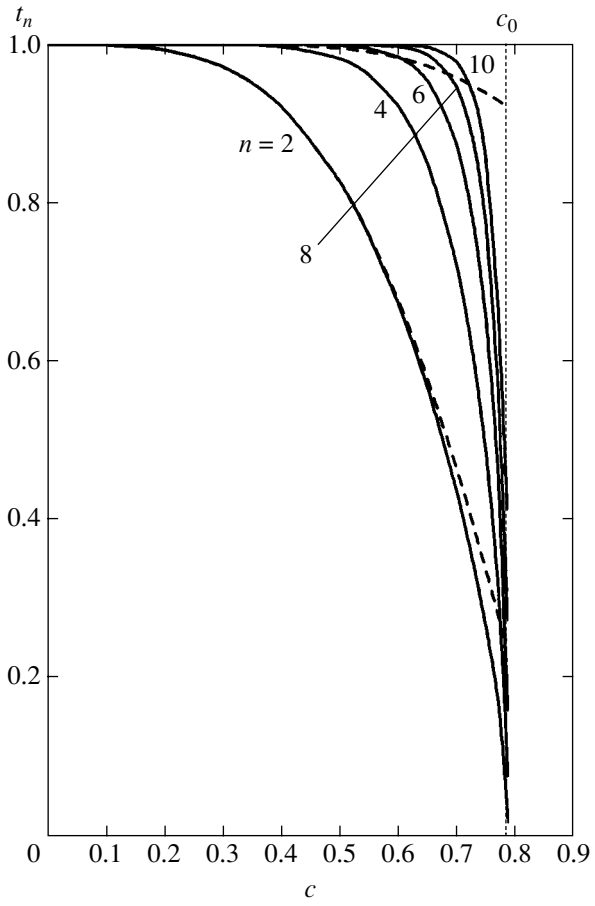


Fig. 6. $t_n = -h_n$ ($n = 2, 4, 6, 8, 10$) as functions of concentration c. The dashed lines represent t_2 and t_4 as functions of c in the approximation of four equations.

The set of three equations (in ξ_0, ξ_0, and ξ_2) yields

$$N = 3:$$
$$\xi_0^{-1} = [1 - (A_{01}^2 + A_{12}^2)\delta^2]/(1 - A_{12}^2\delta^2). \quad (53)$$

In this case, the substitution of (53) into (41) yields an equation of the third degree. In this way, we can obtain (for $c \ll 1$) corrections of orders c^4 and c^5 inclusive to quantities (49)–(52). For the third ($n = 3$) pole, we find

$$h_3 = -[1 + 2c(A_{12}/A_{01})^2 + \ldots], \quad (54)$$

$$F_3 = 4c(A_{12}/A_{01})^2 + \ldots, \quad (55)$$

so that $|h_3 + 1| \sim F_3 \sim c^5$ and $|\delta_3| \sim 1/c^5$ as $c \longrightarrow 0$.

For $n = 4$, we have

$$\xi_0^{-1} = \{1 - (A_{01}^2 + A_{12}^2 + A_{03}^2 + A_{23}^2)\delta^2 + (A_{01}^2A_{23}^2 + A_{12}^2A_{03}^2 - 2A_{01}A_{12}A_{03}A_{23})\delta^4\} \times [1 - (A_{12}^2 + A_{23}^2)\delta^2]^{-1}, \quad (56)$$

which corresponds to a fourth-degree equation in δ. For h_4 and F_4, we obtain rather cumbersome expressions even for $c \ll 1$; we do not present them here. Note, however, that $|h_4 + 1| \sim F_4 \sim c^7$ and $|\delta_4| \sim 1/c^7$ as $c \longrightarrow 0$. Higher order ($N \geq 5$) approximations can be considered analogously.

The analysis performed leads to the following conclusions. Choosing a subset of N equations, we obtain an algebraic equation of the Nth degree in δ that has N real roots δ_n ($n = 1, 2, \ldots, N$). This allows us to determine approximate positions of the first N poles (as well as the residues at these poles) of the function $f(p, \zeta)$; the accuracy of calculating h_n and F_n for fixed n increases with N. As N increases, the poles arise in a definite order: the poles with odd numbers lie to the left of the point $h = -1$, and the poles with even numbers, to the right of this point. For small c, we have the following estimates: $|h_n + 1| \sim F_n \sim c^{2n-1} \ll 1$ and $|\delta_n| \sim 1/c^{2n-1} \gg 1$ ($n = 1, 2, \ldots$). Thus, $h_n \longrightarrow -1$ as $n \longrightarrow \infty$, so that $h = -1$ is a condensation point of the poles. The quantities F_n rapidly decrease as n increases; this allows us to restrict the summation in expansions (12), (21), and

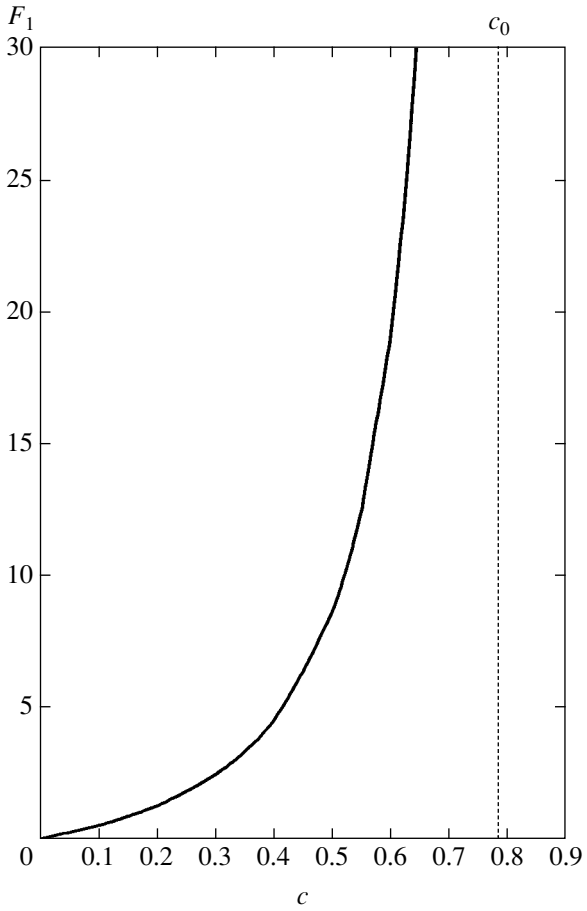


Fig. 7.

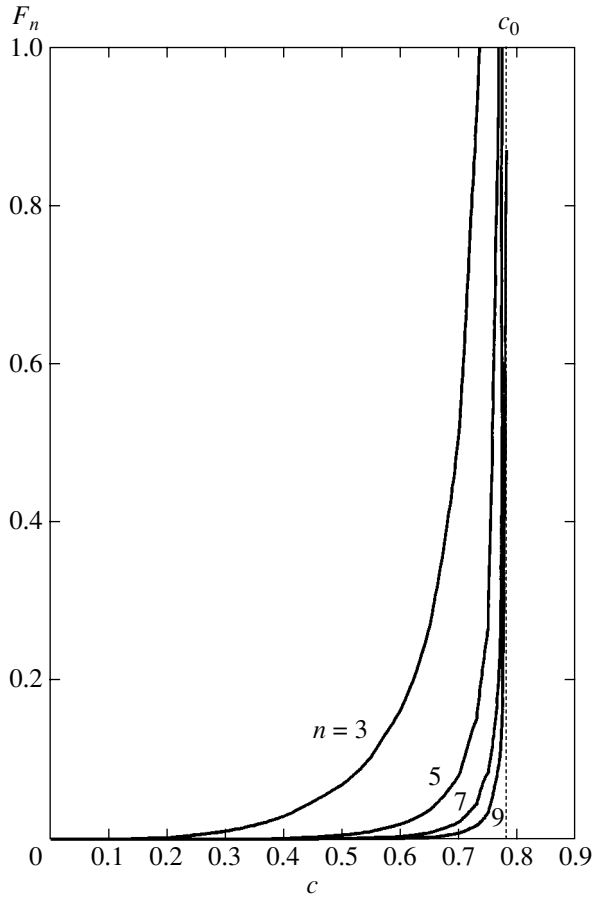


Fig. 8.

(28) for the function $f(p, \zeta)$ to a finite number of terms (outside a certain domain in the vicinity of the point $\zeta = -1$ in the complex plane ζ).

To determine h_n and F_n for large concentrations c, we solved system (37) by numerical methods. For convenience, we introduced, instead of ξ_n, the variables $x_n = \xi_n\delta$ that satisfy the following set of equations

$$x_k + \sum_{l=0}^{\infty} S_{kl}x_l = \delta\delta_{k0}. \tag{57}$$

In this case, the positions of the poles h_n are determined by the equation

$$x_0 = -1/c, \tag{58}$$

and F_n is determined by the following equality derived from (43):

$$F_n = \frac{1}{c}\left[(1+h)^2 \Big/ \frac{dx_0}{d\delta}\right]_{h=h_n}, \tag{59}$$

where we took into account that x_0 is a function of the variable δ.

Let us differentiate (57) with respect to δ:

$$\frac{dx_k}{d\delta} + \sum_{l=0}^{\infty} S_{kl}\frac{dx_l}{d\delta} + \frac{1}{\delta}\sum_{l=0}^{\infty} S_{kl}x_l = \delta_{k0}. \tag{60}$$

Hence, taking into account (57), we obtain

$$\frac{dx_k}{d\delta} + \sum_{l=0}^{\infty} S_{kl}\frac{dx_l}{d\delta} = \frac{x_k}{\delta}. \tag{61}$$

The set of equations (61) has the same structure as (57) and differs from the latter only by the form of the right-hand side. Substituting the values of x_k determined from (57) into the right-hand side of (61) and solving the set of equations obtained by the same method as that applied for solving (57), we determine the derivative $dx_0/d\delta$. The relevant calculations can be performed for a given $h = h_n$ (i.e., for $\delta = \delta_n$). This approach enables one to avoid numerical differentiation when determining F_n.

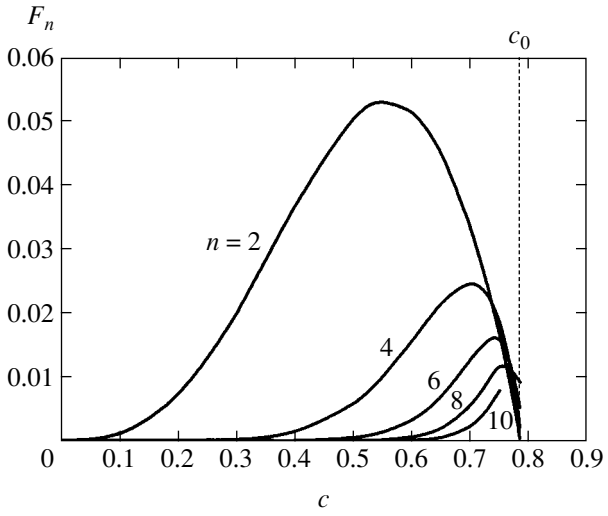


Fig. 9. F_n ($n = 2, 4, 6, 8, 10$) as a function of concentration c.

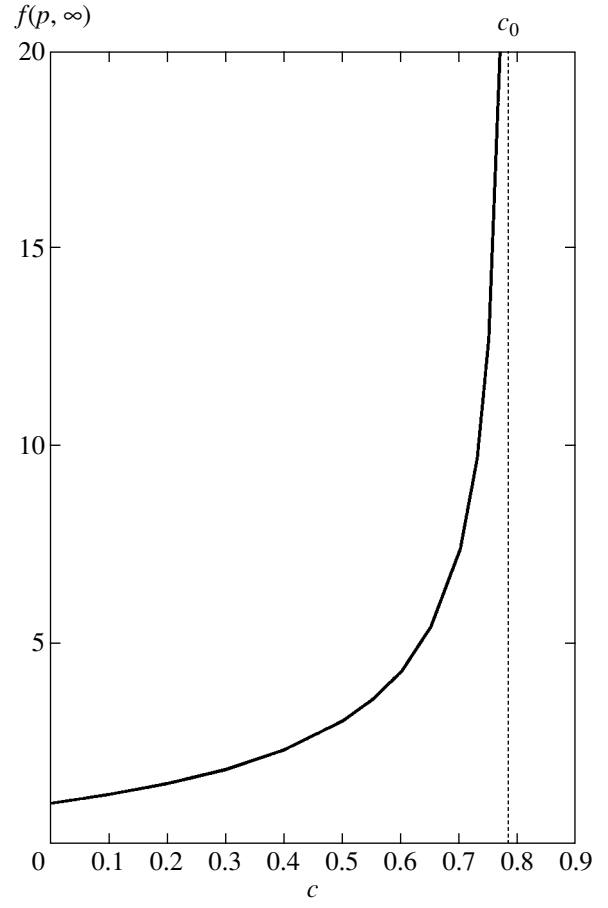


Fig. 10. $f(p, \infty)$ as a function of concentration $c = 1 - p$.

In specific calculations, we singled out a subset of 40 equations from (57) and (61). For $h < 0$, we determined all x_k ($k = 0, 1, 2, \ldots, 39$) as functions of h for a fixed concentration c (i.e., for a given ratio R/a). From Eq. (58), we determined the positions of the ten poles h_n ($n = 1, 2, \ldots, 10$). For example, for the first four h_n, we obtained the following values:

$$h_1 \approx -3.345, \quad h_2 \approx -0.823, \quad h_3 \approx -1.136, \quad h_4 \approx -0.981 \tag{62}$$

for $R/a = 0.8$ ($c \approx 0.503$) and

$$h_1 \approx -5.978, \quad h_2 \approx -0.594, \quad h_3 \approx -1.484, \quad h_4 \approx -0.873 \tag{63}$$

for $R/a = 0.9$ ($c \approx 0.636$).

Then, we determined the derivative $dx_0/d\delta$ for $h = h_n$ and, applying formula (59), calculated F_n ($n = 1, 2, \ldots, 10$). The application of a similar procedure for several values of concentration allowed us to determine h_n and F_n as a function of c. Using formula (39), we calculated $f(p, 0)$ and $f(p, \infty)$ as functions of $c = 1 - p$ and $\mathrm{Re} f(p, h)$ as a function of h for a fixed concentration. To control the correctness of calculations at all stages, we used formulas (15), (18)–(20), (22)–(24), and (29).

The results of the study of the analytic properties of the function $f(p, \zeta)$ carried out with the use of 40 equations form systems (57) and (61) are presented in Figs. 2–10 in graphical form. For comparison, we solved a fourth-degree algebraic equation in δ that was obtained by substituting ξ_0 from (56) into (41). The dependence of h_1, h_2, h_3, and h_4 on the concentration c in this approximation is shown by dashed lines in Figs. 4–6. Note that the main conclusions made for $c \ll 1$ (see above) remain valid for large concentrations c. The neighborhood of the metal–insulator phase-transition point $c_0 = 1 - p_c = \pi/4 \approx 0.785$ (for $R = a$) requires special consideration; here, one should take into account the larger number of equations from systems (59) and (61), the closer the concentration c to c_0.

Note that the comparison of truncated (for $n \le 10$) formulas (12), (21), and (28) with the results of the numerical analysis of $f(p, \zeta)$ with $\zeta = h > 0$ shows that the difference between them is no greater than about 1% for the concentration of inclusions of $c \le 0.73$ for any $0 \le h \le \infty$. This allows us to use such an approximation to calculate $f(p, h)$ in a wide range of variables p and h. Moreover, one may assume that, for given constraints on the concentration, this formula will satisfactorily describe the function $f(p, \zeta)$ also for complex ζ (except for a certain neighborhood of the point $\zeta = -1$).

REFERENCES

1. T. G. Castner, N. K. Lee, G. S. Cieloszyk, and G. L. Salinger, Phys. Rev. Lett. **34**, 1627 (1975).
2. A. Zilberstein, B. Pannetier, and P. Merenda, Phys. Lett. A **54**, 145 (1975).

3. B. I. Shklovskiĭ and A. L. Éfros, Usp. Fiz. Nauk **117**, 401 (1975) [Sov. Phys. Usp. **18**, 845 (1975)].
4. A. L. Efros and B. I. Shklovskii, Phys. Status Solidi B **76**, 475 (1976).
5. V. E. Dubrov, M. E. Levinshteĭn, and M. S. Shur, Zh. Éksp. Teor. Fiz. **70**, 2014 (1976) [Sov. Phys. JETP **43**, 1050 (1976)].
6. A. P. Vinogradov, A. M. Karimov, A. T. Kunavin, *et al.*, Dokl. Akad. Nauk SSSR **275**, 590 (1984) [Sov. Phys. Dokl. **29**, 214 (1984)].
7. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **88**, 1664 (1985) [Sov. Phys. JETP **61**, 991 (1985)].
8. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 3rd ed. (Nauka, Moscow, 1992; Pergamon, New York, 1984).
9. D. J. Bergman, Phys. Rep. **43**, 377 (1978).
10. D. J. Bergman, Ann. Phys. (N.Y.) **138**, 78 (1982).
11. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **89**, 1796 (1985) [Sov. Phys. JETP **62**, 1036 (1985)].
12. Lord Rayleigh, Philos. Mag. **34**, 481 (1892).
13. B. Ya. Balagurov and V. A. Kashin, Zh. Éksp. Teor. Fiz. **117**, 978 (2000) [JETP **90**, 850 (2000)].
14. B. Ya. Balagurov and V. A. Kashin, Zh. Tekh. Fiz. **71**, 106 (2001) [Tech. Phys. **46**, 101 (2001)].
15. A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 110 (1970) [Sov. Phys. JETP **32**, 63 (1971)].
16. S. Kirkpatrick, Rev. Mod. Phys. **45**, 574 (1973).
17. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **94** (7), 95 (1988) [Sov. Phys. JETP **67**, 1351 (1988)].
18. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun, 2nd ed. (Dover, New York, 1971; Nauka, Moscow, 1979).

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 100, No. 4, 2005, pp. 742–751.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 4, 2005, pp. 838–849.

ELECTRONIC PROPERTIES OF SOLIDS

The Special Features of the Hall Effect in GaMnSb Layers Deposited from a Laser Plasma

V. V. Rylkov[a], B. A. Aronzon[a], Yu. A. Danilov[b,c], Yu. N. Drozdov[c], V. P. Lesnikov[b], K. I. Maslakov[a], and V. V. Podol'skiĭ[b]

[a]*Russian Research Center Kurchatov Institute, pl. Akademika Kurchatova 1, Moscow, 123182 Russia*

[b]*Nizhni Novgorod Research Physicotechnical Institute, Nizhni Novgorod State University, pr. Gagarina 23/3, Nizhni Novgorod, 603950 Russia*

[c]*Institute for Physics of Microstructures, Russian Academy of Sciences, Nizhni Novgorod, 603950 Russia*

e-mail: rylkov@imp.kiae.ru; aronzon@imp.kiae.ru

Received September 23, 2004

Abstract—Epitaxial GaMnSb films with Mn contents up to about 10 at. % were obtained by deposition from a laser plasma in vacuum. The growth temperature T_s during deposition was varied from 440 to 200°C, which changed the concentration of holes from 3×10^{19} to 5×10^{20} cm^{-3}, respectively. Structure studies showed that, apart from Mn ions substituting Ga, the GaMnSb layers contained ferromagnetic clusters with Mn and shallow acceptor defects of the Ga_{Sb} type controlled by the T_s value. Unlike single-phase GaMnSb systems studied earlier with negative anomalous Hall effect values and Curie temperatures not exceeding 30 K, the films obtained in this work exhibited a positive anomalous Hall effect, whose hysteresis character manifested itself up to room temperature and was the more substantial the higher the concentration of holes. The unusual behavior of this effect was interpreted in terms of the interaction of charge carriers with ferromagnetic clusters, which was to a substantial extent determined by the presence of Schottky barriers at the boundary between the clusters and the semiconducting matrix; this interaction increased as the concentration of holes grew. The absence of this effect in semiconducting compounds based on III–V Group elements with MnSb or MnAs ferromagnetic clusters was discussed in the literature; we showed that this absence was most likely related to the low hole concentrations in these objects. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Diluted magnetic semiconductors containing magnetic impurities in high concentrations are disordered media, and their properties are therefore determined by disorder to a considerable extent (see review [1]). An enormous number of works have been concerned with disordered media in the presence of only Coulomb interaction, whereas such media under the conditions of joint magnetic and Coulomb interactions remain virtually unstudied, in spite of a fairly strong interest in them [1]. Solid solutions of manganese in semiconducting Group III and V element compounds (in particular, GaMnAs and GaMnSb) [2, 3] are among the diluted magnetic semiconductors that are most intensely developed and studied. The reason for this is comparatively high Curie temperatures of such semiconductors; they can be prepared as heteroepitaxial compositions on single crystalline substrates of the GaAs type, which offers prospects for their integration with instruments traditionally used in semiconducting micro- and optoelectronics [3, 4]. In these materials, Mn is an acceptor impurity; that is, the introduction of Mn into a semiconductor results in the appearance of both local magnetic moments and free holes, which can cause carrier-induced ferromagnetism [5].[1] It was, however, found [5] that the attainment of high Curie temperatures T_C in these materials ($T_C \geq 77$ K) required the introduction of Mn ions in a semiconducting matrix in concentrations of 10^{20}–10^{21} cm^{-3}, which was much higher than the limit of the equilibrium solubility of Mn. The successful preparation of supersaturated solid solutions of Mn in III–V compound semiconductors was performed using nonequilibrium methods for growing them. The most important among these is low-temperature molecular beam epitaxy at about 250°C [3, 4]. It was shown for the example of $Ga_{1-x}Mn_xAs$ layers [3] that there is an optimum manganese content x (0.05–0.06) at which single-phase monocrystalline films with a zinc blende structure were formed. In these films, Mn atoms substitute Ga at lattice sites and play the role of acceptors. The Curie temperature then increases to about 110 K at a concentration of holes $p = 3.5 \times 10^{20}$ cm^{-3} [3]. Recently, a special technique for decreasing the concentration of donors was used to reach $T_C = 159$ K [6],

[1] Direct exchange between Mn ions in these systems is antiferromagnetic in character. The introduction of compensating donors therefore results in the complete suppression of ferromagnetism in them [3].

which was close to the theoretically possible limit [7]. The Curie temperatures of similar layers of single-phase solid solutions of Mn in GaSb, where Mn atoms also predominantly occupy gallium sites and are acceptors, do not exceed 30 K [8]. The T_C value of so-called digital GaSb/Mn alloys (periodic structures) prepared likewise was 80 K [9]. High Curie temperatures can in principle be attained in digital alloys based on supersaturated solid solutions of Mn in III–V compounds (III–Mn–V) (e.g., see [9]). At the same time, at a high manganese content x and/or high growth temperatures, solid solution decay effects are observed and an additional magnetic phase is formed, whose nature is actively discussed at present [9, 10]. Most often, MnAs or MnSb ferromagnetic clusters with a NiAs (T_C = 318 and 587 K, respectively [11]) or zinc blende [9] structure play the role of the additional phase.

A key role in studies of the magnetic properties of diluted magnetic semiconductors based on III–Mn–V compounds is played by revealing and studying the special features of the behavior of the anomalous Hall effect, which is, as is well known, proportional to magnetization M for ferromagnetic metals and is related to the influence of spin–orbit coupling on the scattering of spin-polarized electrons [12]. The calculations performed recently [13] show that the anomalous Hall effect in III–Mn–V semiconductors can be caused by corrections to the velocity of carriers related to the so-called Berry phase. The anomalous Hall effect is then also determined by the exchange splitting of spin hole subbands, is proportional to magnetization, and, therefore, the Hall resistance R_H, as in ferromagnetic metals, obeys the equation [3]

$$R_H = \frac{R_0}{d}B + \frac{R_s}{d}M, \quad (1)$$

where d is the thickness of the diluted magnetic semiconductor layer; R_0 is the constant of the ordinary Hall effect caused by the Lorentz force, which is proportional to the magnetic induction B; and R_s is the anomalous Hall effect constant.

The anomalous Hall effect plays an important role in studies of ferromagnetism in diluted magnetic semiconductors because it is the most direct method for investigating the interaction of charge carriers with the magnetic subsystem. In addition, for thin films, when the influence of a diamagnetic substrate is strong, the anomalous Hall effect can more effectively be used to study magnetic ordering than magnetization measurements [3, 14, 15]. Another and more important reason for using the anomalous Hall effect is the complex character of the magnetic phase that may appear in III–Mn–V materials [9, 10]. For instance, $Ga_{1-x}Mn_xSb$ crystals grown by the Bridgman method (x = 0.03–0.14) were reported [16] to exhibit the Curie temperature $T_C \approx 540$ K, which was close to $T_C \approx 587$ K for MnSb clusters; this result was obtained by studying the temperature dependence of magnetization. It is pertinent to mention that the magnetic field dependences of the Hall effect (which were virtually linear) and magnetization presented in [16] were substantially different. In addition, there was no anomalous Hall effect manifestations of MnSb- or MnAs-type ferromagnetic clusters in GaMnSb [8] and GaMnAs [3], respectively, although the contribution of the clusters to magnetization was considerable and observable up to room temperatures. At the same time, the "Curie temperatures" obtained for these systems by measuring the anomalous Hall effect (T_C^*) were noticeably lower than room temperatures ($T_C^* < 10$ K for GaMnSb with MnSb clusters [8]). Accordingly, it is commonly supposed that the T_C^* parameter obtained from anomalous Hall effect measurements for multiphase solutions of Mn in III–V compounds characterizes magnetic ordering of only part of the magnetic subsystem, which nevertheless largely determines the spin polarization of carriers and is of the greatest importance for diluted magnetic semiconductor applications in spintronics. It is therefore no mere chance that, in recent works [6, 8, 14], preference is given to anomalous Hall effect measurements as a method for the observation of spin-polarized carriers and the determination of the magnetic state of III–Mn–V systems at various temperatures. For single-phase solutions, these measurements give the same T_C temperatures as those obtained in magnetization studies [3, 6].

In spite of the important role played by the anomalous Hall effect in III–Mn–V materials, the question of its nature remains open. It was shown recently [17] that the anomalous Hall effect value in GaMnAs is in close agreement with the calculation results [13], and its sign (positive) coincides with the sign of the ordinary Hall effect, in agreement with [13]. The GaMnSb system, however, exhibits a negative Hall effect, whose sign is opposite to that of the ordinary Hall effect [4, 8, 9]. The theory described in [13] does in principle predict a change in the sign of the anomalous Hall effect; this, however, requires that the Fermi energy be close to the top of the Γ_7 band split off because of spin-orbit coupling [18] (in GaSb, Δ_0 = 0.75 eV [19]), which is at variance with the experimental data [8, 9].[2] The question why MnSb- and MnAs-type clusters in the systems under consideration do not influence the anomalous Hall effect also remains open. Indeed, the anomalous Hall effect is observed quite distinctly in diluted ferromagnetic granulated alloys (nanoparticles of ferromagnetic metals in nonmagnetic metallic matrices) and substantially exceeds the ordinary Hall effect component [20]. The Curie temperatures of single-phase

[2] It is shown in [18] that the calculations made in [13] are equivalent to the consideration of the anomalous Hall effect in terms of the so-called side-jump model, in which the sign of the anomalous Hall effect should coincide with the sign of carriers. In the skew scattering model, the sign of the anomalous Hall effect can in principle be arbitrary [18].

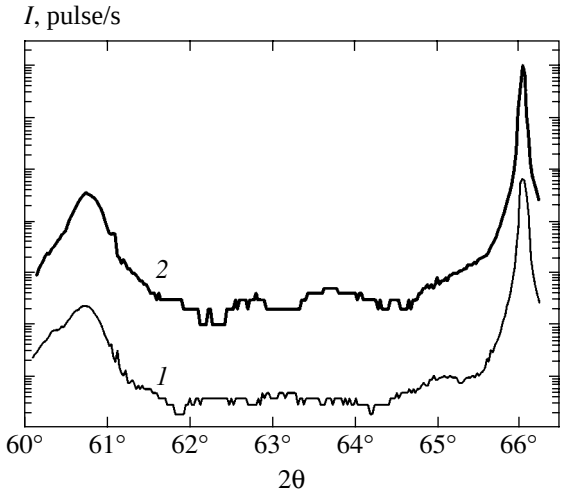


Fig. 1. X-ray diffraction spectra of GaSb/GaAs structures formed by depositing at T_s = 440°C: (*1*) undoped and (*2*) doped with manganese. For clarity, the spectra are spaced along the intensity axis: the initial spectrum of structure *2* was multiplied by 100.

diluted magnetic semiconductors are fairly low but can substantially exceed room temperatures in diluted ferromagnetic granulated alloys. It is therefore of interest to study supersaturated solid solutions of Mn in III–V semiconductors and, in particular, their galvanomagnetic properties.

The purpose of this work was to study the special features of the behavior of the Hall effect in Mn-supersaturated GaMnSb layers deposited by laser sputtering of undoped GaSb and metal Mn targets in a vacuum. It was found that acceptor-type defects were largely formed in the films prepared by laser plasma deposition. Moreover, the concentration of acceptor defects and, accordingly, the concentration of holes p could easily be controlled by changing the temperature of the substrate, which allowed p to be varied from 10^{19} to 5×10^{20} cm^{-3}. As distinct from the earlier results, we observed a positive high-temperature anomalous Hall effect that exhibited hysteresis up to room temperatures in the layers under study. Anomalous Hall effect data processing according to Belov and Arrott [21] allowed us to determine the value of T_C^* that characterized spontaneous manifestation of the (in the absence of a magnetic field) anomalous Hall effect. The T_C^* reached 330 K at $p = 5 \times 10^{20}$ cm^{-3} and decreased as the concentration of holes lowered. Physical reasons for the observed anomalous Hall effect characteristics and for its absence in similar structures that had been studied earlier will be considered.

2. EXPERIMENTAL PROCEDURE

GaMnSb films were deposited using a pulsed yttrium aluminum garnet laser operating in the Q-switching mode at 1.06 μm. We used a rotating combined target consisting of a plate of single crystalline undoped GaSb covered in part by a high-purity Mn metal. The trace of material vaporization with the laser was a circle, and the ratio between the lengths of the arcs of sputtered GaSb and Mn therefore determined the level of growing layer doping. The films were deposited on a plate of semi-insulating GaAs with the (100) orientation, and the substrate temperature T_s was varied from 200 to 440°C. The resulting films were 40–140 nm thick.

The structural characteristics and composition of the films were studied by X-ray diffraction on a DRON-4 instrument using the two-crystal scheme and Cu $K_{\alpha 1}$ radiation filtered with a Ge(400) monochromator, X-ray photoelectron spectroscopy on a MicroLab MK II unit (VG Scientific) using nonmonochromatized Al K_α radiation, and electron probe microanalysis on a GAMEBAX unit. The magnetization of the films was measured by a BHV-50 vibrating-coil magnetometer with a sensitivity no worse than 10^{-5} emu.

The samples for Hall effect measurements were prepared by photolithography, as mesostructures with the standard double cross form (the width and length of the conduction channel were W = 0.5 mm and L = 4.5 mm, respectively). Hall effect measurements in fields up to 1 T were performed using an automated unit by the method of digital filtration and signal accumulation. The voltage between the Hall (V_y) and potential (V_x) probes and current I_x that passed through the sample were synchronously recorded under constant voltage conditions at positive and negative magnetic field B values; the field was applied normally to the film surface (along the z axis). The measurement results were used to determine the resistance of the sample between the potential probes $R_{xx} = V_x/I_x$ and transverse resistance $R_{xy} = V_y/I_x$. Preliminary experiments showed that transverse resistance could exhibit hysteresis and the magnetic field dependence of longitudinal resistance was negligibly small (the magnetoresistance of the films under study did not exceed 0.1%). Considering possible hysteresis, the Hall resistance R_H was determined by subtracting the even signal component from R_{xy} (the even component appeared because of asymmetry in the arrangement of Hall probes); that is, $R_H = R_{xy} - (R_{xy}^+ + R_{xy}^-)/2$, where R_{xy}^+ and R_{xy}^- are the transverse resistance values corresponding to the positive and negative magnetic field directions obtained, for instance, in scanning over the field as it decreased in magnitude (from 1 to 0 T).

3. RESULTS AND DISCUSSION

The X-ray diffraction θ/2θ spectra of the GaSb/GaAs structures deposited at a 440°C substrate temperature are shown in Fig. 1. The spectrum of the structure with a GaSb layer undoped by Mn (curve *1*)

contains a substrate peak at $2\theta = 66.05°$ (the GaAs(400) reflection) and a layer peak at 60.74° (the GaSb(400) reflection). The Bragg angles of the 2θ peaks from the layer and substrate were refined in two stages. At the first stage, we refined the angle of crystal rotation with a broad slit in front of the detector (ω scan). At the second stage, the angle of detector rotation was determined with a narrow slit while the sample rotation angle was fixed at a value corresponding to maximum intensity (θ scan). The substrate was used as a reference to correct the 2θ angle of the layer.

Calculations of the lattice parameter of undoped GaSb from the 2θ angle gave $a_0 = 0.6096$ nm, which coincided with the value known from the literature [19]. The integral characteristic of the structural perfection of the layer was the rocking curve width (full width at half maximum, FWHM) measured from the ω-scan spectrum according to [22]. The FWHM value for the peak from the GaSb layer was $\Delta\omega \approx 0.4°$. It follows that GaSb is a mosaic single crystalline film, although the fairly large $\Delta\omega$ value may be evidence of nonideality of the crystal structure of the film likely caused by the difference in the lattice constants of GaSb and GaAs exceeding 7%.

The X-ray diffraction $\theta/2\theta$ spectrum of the structure with a GaMnSb layer also deposited at 440°C is shown by curve *2* in Fig. 1. This spectrum is similar to that of GaSb free of Mn (curve *1*). It follows that the introduction of Mn by laser plasma deposition does not cause noticeable structural imperfection of the deposited layers. With GaMnSb, the procedure for refining the Bragg angle of the 2θ peak GaSb(400) with the use of ω and θ scans, however, gives a value of $2\theta = 60.76°$, which is somewhat larger than that of the undoped layer. We used the equation $a(x) = a_0 - 0.00528x$, where a_0 is the lattice constant of undoped GaSb, for the lattice parameter of $Ga_{1-x}Mn_xSb$ with a zinc blende structure [23] and the $2\theta = 60.76°$ value obtained for GaMnSb layers grown at $T_s = 350$–440°C to estimate the content of Mn, $x \approx 0.04 \pm 0.01$.

The FWHM value monotonically increases from 0.4° at $T_s = 440$°C to 0.5° at $T_s = 300$°C as the temperature of GaMnSb layer growth lowers. The GaSb(400) peak, although low-intensity, is observed even at $T_s = 200$°C. This is evidence that the presence of manganese in a layer of gallium antimonide in the concentration specified above has no substantial influence on the character of its growth during laser plasma deposition over the temperature range 200–440°C.

The composition of the films was studied by electron-probe microanalysis with a spatial resolution of order 1 μm. The results showed that Mn was fairly uniformly distributed over the area of the samples (the spread of x values was about 1%). The thickness of the films was much less than the region of X-ray radiation excitation by accelerated electrons in the structure, which prevented exact calculations of manganese contents from the data obtained this way. Estimates, however, show that the content of Mn in GaSb films was $x \approx 0.10$. This is larger than the value obtained from X-ray diffraction. This discrepancy can arise because electron-probe microanalysis measurements determine the chemical composition of the films, whereas X-ray diffraction is sensitive to changes in the GaSb lattice parameter caused by the insertion of Mn ions into gallium sublattice sites. When a nonequilibrium method for depositing layers is used, manganese atoms can enter into other lattice sites (for instance, into interstices) and form clusters. It follows that the $x \approx 0.10$ value is more realistic.

The suggestion of the possible presence of clusters is substantiated by the X-ray photoelectron spectra of the samples. The X-ray photoelectron spectra of GaMnSb/GaAs structures are presented in Fig. 2. The Mn $2p$ line spectrum is shown in Fig. 2a. This line has a complex structure, in which at least two chemical states of manganese atoms are distinctly seen. The positions of the Mn $2p_{3/2}$ peaks for these two states are denoted by *A* and *B* in the figure. The actual positions of peaks in the spectra of compounds is determined by several factors, in particular, by changes in the energy levels caused by chemical interaction (the so-called chemical shift [24]) and exchange interaction [25] in magnetic materials. The experimental (Fig. 2a) E_B values for the Mn $2p_{3/2}$ line in states *A* and *B* are 638.9 and 640.8 eV, respectively. The first value coincides with E_B for Mn metal (638.9 eV [26]). A similar splitting of the Mn $2p_{3/2}$ peak was observed in [25] for ternary alloys containing Mn and Sb. The first peak in [25] coincided with the manganese metal peak. The second peak at larger E_B values was split off by approximately 2 eV in a situation close to that considered in this work. It appeared in compounds probably because of exchange interactions caused by the chemical state of manganese atoms in which they have a large local magnetic moment [25]. It follows that there are at least two states of Mn atoms in GaMnSb films, one characteristic of bonds between Mn and Sb atoms and the other similar to the state of Mn in Mn metal (Mn–Mn bonds). The conclusion can be drawn that the films under consideration are supersaturated solid solutions of Mn in GaSb that contain the GaSb matrix with 4% of Ga replaced by manganese and manganese-containing clusters, whose influence on the magnetic and galvanomagnetic properties of the layers is discussed below.

Another special feature of the GaMnSb films is their primordially *p*-type conduction, even in the absence of doping, in particular, with manganese. This is likely related to the formation of antisite Ga_{Sb} defects (Ga atoms in Sb sites) during film growth, which are shallow acceptors in GaSb [19]. Indeed, as follows from the X-ray photoelectron spectra of GaMnSb layers shown in Fig. 2b (Ga $3d$ and Sb $4d$ lines) and 2c (two Sb $3d$ lines) (spectra *2* and *3*) in comparison with the corresponding X-ray photoelectron spectra of single crystalline undoped GaSb (Fig. 2, spectrum *1*), the

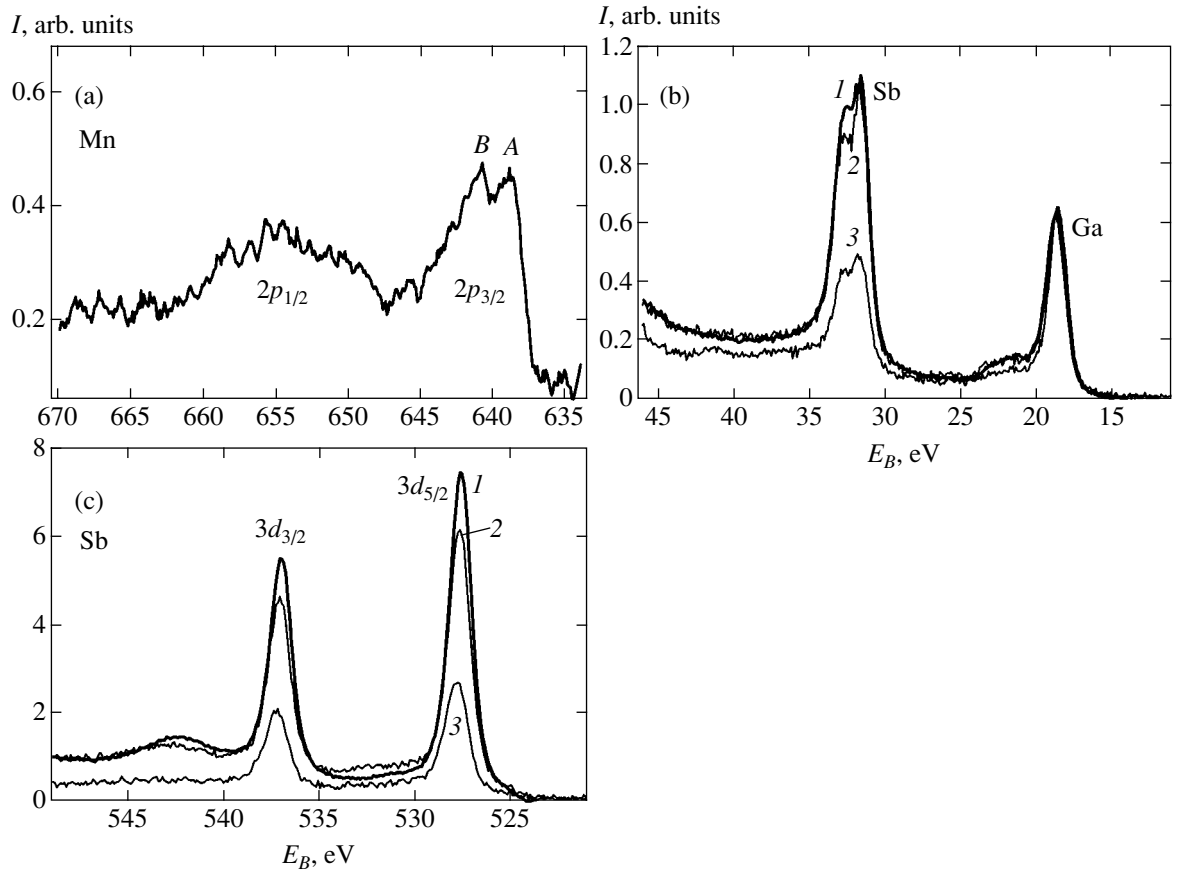


Fig. 2. X-ray photoelectron spectra: (a) Mn 2*p* line for a GaMnSb layer, T_s = 200°C; (b) Ga 3*d* and Sb 4*d* lines for (*1*) single crystalline GaSb, (*2*) a GaMnSb layer, T_s = 440°C, and (*3*) a GaMnSb layer, T_s = 200°C; and (c) Sb 3*d* lines for three types of structures, see (b).

content of Ga atoms in the films is higher than that of Sb atoms ($C_{Ga}/C_{Sb} > 1$). The nonstoichiometry of deposited layers (the C_{Ga}/C_{Sb} ratio) increases as the deposition temperature T_s decreases. According to the measurements performed at 300 K, the resistivity of the films ρ decreases from $\rho = 4 \times 10^{-2}$ Ω cm at T_s = 440°C to $\rho = 3 \times 10^{-3}$ Ω cm at T_s = 200°C, which is evidence of an increase in the concentration of holes.

Measurements of the magnetization of the GaMnSb films grown at various temperatures gave the results shown in Fig. 3. We see that, in spite of the substantial differences in the conductivities of the layers deposited at various T_s (differences in hole concentrations), the films exhibit ferromagnetic behavior, and their saturation magnetizations do not vary strongly, from M_s = 5.3 mT at T_s = 200°C to M_s = 3.6 mT at T_s = 440°C. We assume that the magnetic moment of the films is determined by Mn^{2+} ions [3] (*g*-factor = 2 and the total spin $S = 5/2$). On this assumption, calculations give Mn ion concentrations of $N_{Mn} = 1.1 \times 10^{21}$ cm^{-3} (T_s = 200°C) and $N_{Mn} = 7.8 \times 10^{20}$ cm^{-3} (T_s = 440°C), which is in agreement with maximum estimates of the concentration of Mn atoms as an impurity that replaces Ga, $N_{Mn} = 7.1 \times 10^{20}$–$1.8 \times 10^{21}$ cm^{-3} at x = 0.04–0.10. However note that if the magnetic properties of III–Mn–V semiconductors weakly depend on the concentration of holes, they are usually related to the presence of MnSb- or MnAs-type ferromagnetic clusters [3, 8].

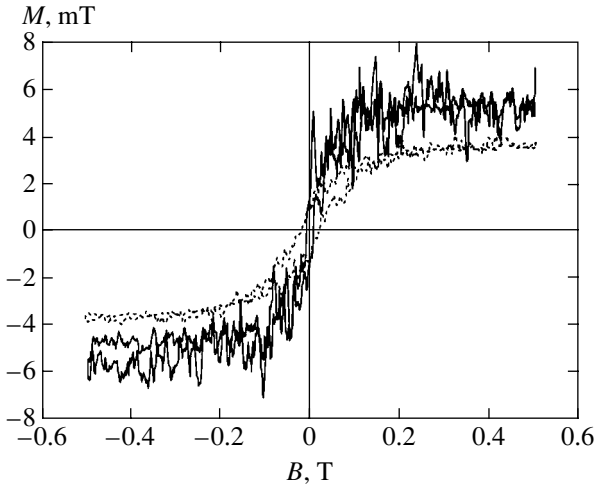


Fig. 3. Applied magnetic field dependences of magnetization for GaMnSb layers *d* = 40 nm thick, T_s = 200°C (solid line), and *d* = 140 nm thick, T_s = 440°C (dots).

As distinct from magnetization, the behavior of the Hall effect strongly depends on the concentration of holes (on the deposition temperature T_s). The magnetic field dependences of the Hall resistance $R_H(B)$ obtained at T = 77 and 293 K are shown in Figs. 4a and 4b, respectively, for samples 1–3 with hole concentrations $p = 5 \times 10^{20}$ cm^{-3} (curve *1*), 1.5×10^{20} cm^{-3} (curve *2*), and 3×10^{19} cm^{-3} (curve *3*). The concentration of carriers was determined from the slope of the $R_H(B)$ dependence in fields $B > 0.4$–0.5 T. The linear character of this dependence for sample 1 over the specified field range is illustrated by the upper inset to Fig. 4b. On the whole, it follows from Fig. 4 that the Hall effect in samples 1 and 2 is essentially anomalous in character over the temperature range T = 77–300 K, whereas the Hall effect in sample 3 with the lowest concentration of holes (T_s = 440°C) is ordinary. Indeed, the Hall resistance of this sample linearly depends on B over the field range 0–0.9 T, although its magnetization reaches saturation already at $B > 0.2$ T (Fig. 3). A comparison of the data on samples 1 and 2 presented in Fig. 4 shows that the hysteresis character of the behavior of the anomalous Hall effect also becomes suppressed as the concentration of holes decreases. For instance, for sample 1 ($p = 5 \times 10^{20}$ cm^{-3}), the coercive field reaches B_c = 0.29 T at T = 77 K and the anomalous Hall effect hysteresis manifests itself up to room temperature (B_c = 6.5 mT, see the lower inset to Fig. 4b). At the same time, for sample 2 ($p = 1.5 \times 10^{20}$ cm^{-3}), B_c = 0.058 T at T = 77 K and no anomalous Hall effect hysteresis is observed at T = 300 K.

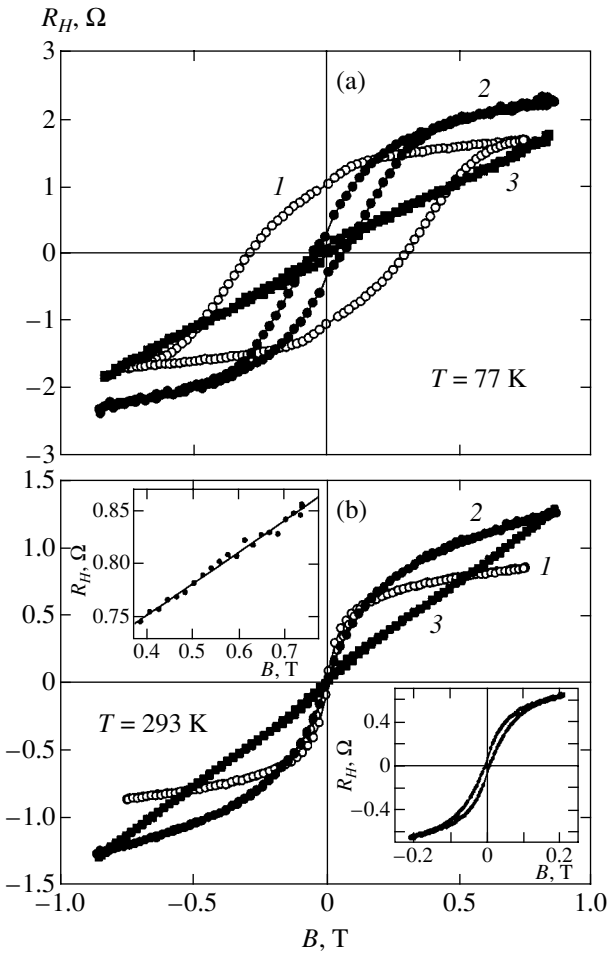


Fig. 4. Applied magnetic field dependences of Hall resistance for three GaMnSb/GaAs structure samples: (*1*) $p = 5 \times 10^{20}$ cm^{-3}, T_s = 200°C, d = 40 nm; (*2*) $p = 1.5 \times 10^{20}$ cm^{-3}, T_s = 200°C, d = 70 nm; and (*3*) $p = 3 \times 10^{19}$ cm^{-3}, T_s = 440°C, d = 140 nm. Curve numbers correspond to sample numbers. Measurement temperatures: (a) 77 K and (b) 293 K. The upper inset to Fig. 4b contains the $R_H(B)$ dependence for sample 1 at $B > 0.4$ T, and the lower inset, the $R_H(B)$ dependence for sample 1 at $-0.2 < B < 0.2$ T.

As the Hall resistivity is proportional to magnetization M when the anomalous Hall effect predominates (see (1)), it was suggested in [3, 15] that the procedure developed by Belov and Arrott [21] (the construction of the dependence of M^2 on B/M) can be used to determine the spontaneous Hall resistance R_H^s, which is proportional to the spontaneous magnetization M_s characteristic of ferromagnetic system ordering. According to [21], we must construct the dependence of R_H^2 on B/R_H and extrapolate its linear portion to the intersection with the axis of ordinates to determine R_H^s for our systems.

Examples of the dependences of R_H^2 on B/R_H for sample 1 at several measurement temperatures are shown in Fig. 5. At 267 and 293 K, the linear extrapolation of R_H^2 to B = 0 gives $(R_H^s)^2 > 0$, whereas linear extrapolation at T = 335 K gives a negative $(R_H^s)^2$ value, which means that there is no ferromagnetic ordering at this temperature. The temperature dependences of the spontaneous Hall resistance R_H^s obtained using the procedure suggested in [21] and the coercive field B_c are shown in Fig. 6 for sample 1. These results are evidence that the spontaneous Hall resistance in this sample persists up to the temperature $T_C^* \approx 330$ K.

The T_C^* parameter of single-phase III–Mn–V materials coincides with the Curie temperature T_C [3, 6]. This value can be determined by analyzing the behavior of the anomalous Hall effect in the paramagnetic temperature region [3, 14]. Indeed, the anomalous Hall effect constant is $R_s = cR_{xx}$ [12], where c is a temperature-independent coefficient, if the anomalous Hall effect is determined by the mechanism of skew scattering of carriers. For this reason, the ratio between the Hall R_H and longitudinal resistance R_{xx} is, according

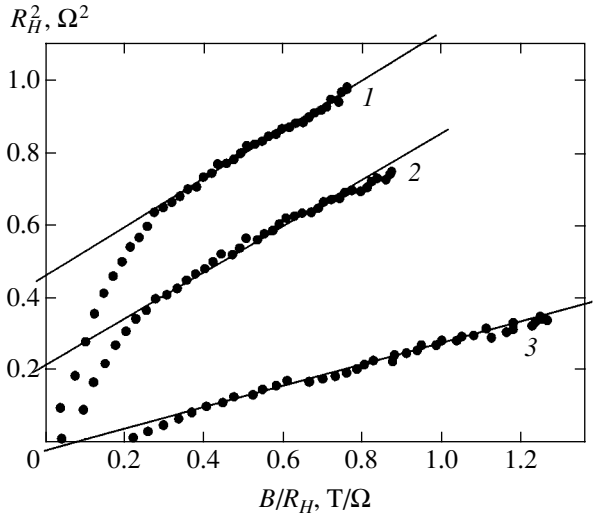


Fig. 5. Dependence of R_H^2 on B/R_H for sample 1 ($p = 5 \times 10^{20}$ cm^{-3}). Measurement temperatures (*1*) 267, (*2*) 293, and (*3*) 335 K.

to (1), $R_H/R_{xx} \approx cM/d$. It follows that, in the paramagnetic region, the magnetic susceptibility is

$$\chi \propto [d(R_H/R_{xx})/dB]_{B=0},$$

and the Curie–Weiss law $(1/\chi) \propto (T - T_C)$ can be used to determine T_C. It was shown in [3] for the example of GaMnAs that this approach gave the same T_C value as that obtained by analyzing the anomalous Hall effect in the ferromagnetic region.

The Arrott dependences for sample 2 ($p = 1.5 \times 10^{20}$ cm^{-3}) over the range of temperatures $T \geq 160$ K are shown in Fig. 7. The linear extrapolation of these dependences gives their intersection with the origin at $T \approx 180$ K; that is, we can expect, by analogy with single-phase diluted magnetic semiconductors, that the Curie temperature of this sample is $T_C \approx 180$ K, and, starting with this temperature, the sample turns paramagnetic. The dependence of the Hall resistance on the longitudinal resistance obtained at $B = 0.75$ T is shown in the inset to Fig. 7 in the double logarithmic coordinates. The R_{xx} resistance of this sample increases as the temperature lowers. It follows from the data given in the inset to Fig. 7 that the slope of the dependence of $\ln R_H$ on $\ln R_{xx}$ in the region of low temperatures, where magnetization should weakly depend on T, is close to one; that is, the suggestion of the predominant role played by the mechanism of skew scattering of carriers in the anomalous Hall effect is justified for our systems. Seemingly, the slope of the dependence of R_H/R_{xx} on B in low fields, which is proportional to χ, should then increase as the temperature lowers. However, in reality, this slope is virtually independent of the temperature (see Fig. 8), which distinguishes our systems from single-phase diluted magnetic semiconductors of the GaMnAs type [3] (see the data presented in Fig. 3 and borrowed from [3]).

More substantial differences become evident when the special features of the anomalous Hall effect described above are compared with an analogous effect in GaMnSb layers prepared by molecular beam epitaxy at various growth temperatures [8]. It was found in [8] that Mn is almost fully contained in ferromagnetic MnSb clusters in layers grown at high temperatures of $T_s = 560$°C. The concentration of holes in the GaSb matrix is then 2.4×10^{19} cm^{-3} (in the sample with the

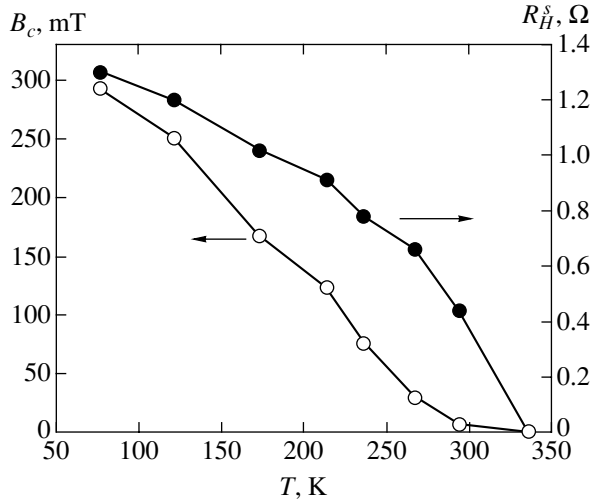


Fig. 6. Measurement temperature dependences of the coercive force of magnetization B_c (the left curve) and spontaneous Hall resistance component R_H^s (the right curve) for sample 1.

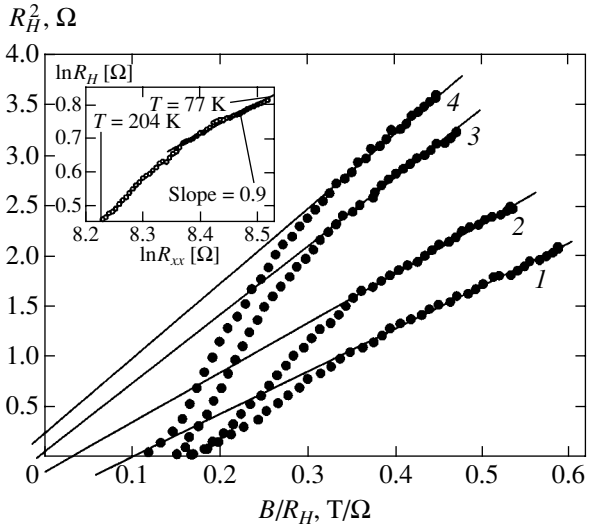


Fig. 7. Dependence of R_H^2 on B/R_H for sample 2 ($p = 1.5 \times 10^{20}$ cm^{-3}). Measurement temperatures (*1*) 254, (*2*) 226, (*3*) 180, and (*4*) 160 K; the relation between R_H and R_{xx} is shown in the inset to the double logarithmic coordinates.

total content of Mn $x = 0.013$), and the ordinary component predominates in the Hall effect to $T \approx 10$ K. In our systems, the Hall effect is only ordinary in sample 3 with a similar hole concentration ($p = 3 \times 10^{19}$ cm^{-3}), which is substantially lower than that in samples 1 and 2. The layers obtained in [8] at $T_s = 250$°C had the structure of zinc blende, in which Mn atoms largely replace Ga and play the role of acceptors. A well-defined negative anomalous Hall effect was then observed at low measurement temperatures (its sign was opposite to that of the ordinary Hall effect). The Belov–Arrott procedure was used in [8] to obtain $T_C^* \approx 25$ K at the hole concentration $p = 1.3 \times 10^{20}$ cm^{-3} (for the sample with the content of Mn $x = 0.016$). At the same time, sample 2 with approximately the same concentration of holes ($p = 1.5 \times 10^{20}$ cm^{-3}) had $T_C^* \approx 180$ K, and its anomalous Hall effect was positive.

Let us discuss the experimental data presented above. Note that, as in the samples studied in this work, the sign of the anomalous Hall effect in continuous MnSb films was positive [27]. It is reasonable to suggest that the sign of the anomalous Hall effect remains unchanged in passing from continuous to broken films. Indeed, we recently showed for the example of Fe nanoparticles in a SiO_2 matrix (the sign of the anomalous Hall effect in Fe was also positive) that the sign of the anomalous Hall effect did not change in the passage through the percolation threshold to tunnel conduction conditions [28]. The invariability of the sign of the anomalous Hall effect follows from the effective medium model [29]. Note also that, at the growth temperatures used ($T_s = 200$–440°C), the films contained a ferromagnetic phase in approximately equal concentrations (see the magnetization data given in Fig. 3).

The observations described above lead us to conclude that the anomalous Hall effect in the GaMnSb samples is related to the presence of MnSb-type clusters in them. The volume content of the ferromagnetic phase recalculated to MnSb ($M_s = 71$ mT [6]) is about 0.07, which is far below the critical value (0.6 [28]) corresponding to the metal–insulator percolation transition. The cardinal difference between our objects (the predominance of the anomalous Hall effect at fairly high temperatures) and samples with MnSb clusters [8] is the much higher concentration of holes in the GaSb matrix, which is related to the generation of acceptor-type defects (antisite Ga_{Sb} defects) during film growth by the laser plasma deposition method. This is in agreement with the observed strong dependence of the behavior of the anomalous Hall effect on the concentration of carriers at a constant ferromagnetic phase concentration (see Fig. 4). We can therefore naturally suggest that the interaction of carriers with ferromagnetic clusters in semiconductors with magnetic impurities is to a considerable extent determined by the presence of Schottky barriers at the boundary between the clusters and the semiconducting matrix (in our system, at the MnSb/GaSb boundary). Accordingly, this interaction depends not only on the content of the ferromagnetic phase but also on the concentration of holes, whose increase decreases the width of the Schottky barriers. Estimates show that the expected Schottky barrier width must be about 2 nm at an $N_A^- \approx p = 10^{20}$ cm^{-3} concentration of ionized acceptors. This estimate was obtained on the assumption that the height of the Schottky barriers φ was determined by the position of the maximum of the density of surface states in the forbidden band of GaSb. In the majority of covalent semiconductors, this maximum is shifted from the valence band edge by one-third of the forbidden bandwidth E_g [30] ($E_g = 0.7$ eV for GaSb [19]). At the same time, for these conditions, the effective depth l_ψ of a decrease in the wavefunction of heavy holes ($m_{hh} = 0.23m_0$) under the barrier can be estimated at 1.3–2.5 nm; that is, it can even be larger than the Schottky barrier width at $p = 10^{20}$ cm^{-3} and $\varphi = (1/3)E_g$. We can then naturally expect strong tunnel exchange between matrix carriers and ferromagnetic clusters.[3] (The estimates for l_ψ were obtained using the equation for the transparency of a triangular barrier [30]; the lower estimate corresponds to the mean electric field in the region of the spatial charge of the Schottky layer, and the upper, to the maximum field.)

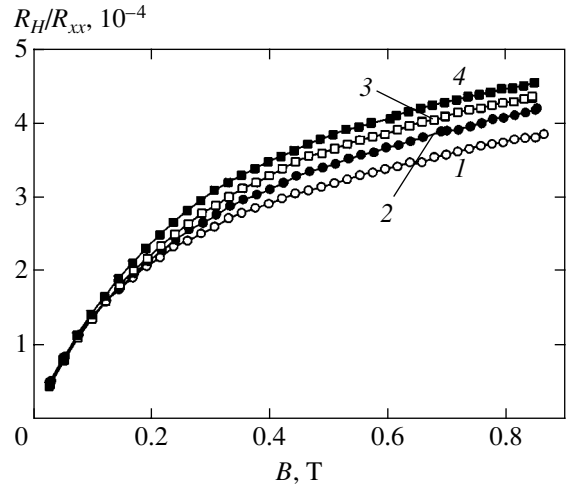


Fig. 8. Applied magnetic field dependences of the R_H/R_{xx} ratio for sample 2 at various measurement temperatures: (*1*) 293, (*2*) 254, (*3*) 226, and (*4*) 180 K.

The temperature at which there is no anomalous Hall effect hysteresis is interpreted as the blocking temperature of ferromagnetic granulated alloys [20]; at this temperature, the transition to the superparamagnetic

[3] Note that the mean distance between ionized acceptors at $p = 10^{20}$ cm^{-3} is also about 2 nm. This leads us to conclude that the above estimates are actually evidence of the absence of Schottky barriers. The origin of the anomalous Hall effect is then similar to that in magnetic granulated alloys [20].

limit occurs. Estimates show that MnSb clusters of size $a_c \approx 10$ nm give blocking temperatures of order 200–300 K observed as Curie temperatures in our experiments (the estimates were obtained only taking into account the magnetic anisotropy energy related to the shape of the clusters [31]). Also note that the spread of the clusters in shape and size substantially weakens the temperature dependence of magnetization [32], which probably explains the absence of temperature effects on the paramagnetic behavior of the anomalous Hall effect in low fields (the weak dependence $\chi(T) \propto [d(R_H/R_{xx})/dB]_{B=0}$) observed for sample 2 (Fig. 8).

At the same time, it should be noted that the interpretation of the data on the anomalous Hall effect in terms of isolated (noninteracting) MnSb clusters and blocking temperatures encounters obvious difficulties. Indeed, an increase in the growth temperature T_s accompanied by a decrease in the concentration of holes should cause the enlargement of clusters. The high coercive force values observed experimentally are evidence that the clusters are single-domain; the coercive field is then the higher the larger the size of the clusters [33]. For this reason, increasing T_s should make the hysteresis character of the behavior of the anomalous Hall effect more manifest, which has not been observed experimentally, although the contribution of the ordinary component to the Hall effect increases because of a decrease in the concentration of current carriers. This leads us to suggest that the size of MnSb clusters (and, accordingly, the distance between them) is noticeably smaller than 10 nm and that these clusters interact with each other. This interaction is mediated by carriers in the paramagnetic GaSb:Mn matrix. This results in an effective enlargement of the clusters and, simultaneously, increases hole spin polarization, which determines the anomalous Hall effect.

4. CONCLUSIONS

To summarize, we prepared epitaxial films of a supersaturated solid solution of Mn in GaSb by laser plasma deposition. The special feature of the layers grown was the presence of dissolved Mn atoms and ferromagnetic Mn-containing inclusions. The layers also contained acceptor-type defects controlled by the growth temperature. These defects to a substantial extent determined the concentration of holes in the GaSb matrix, which increased as the growth temperature lowered and reached 5×10^{20} cm^{-3} at $T_s = 200$°C. Unlike single-phase GaMnSb systems, the films exhibited a positive anomalous Hall effect. Its hysteresis character strongly depended on the concentration of holes and could be observed up to room temperatures.

We believe that the special features of the behavior of the anomalous Hall effect in our systems are related to the interaction of charge carriers with ferromagnetic clusters, determined to a significant extent by the presence of Schottky barriers at the boundaries between the clusters and the semiconducting matrix. The role played by these barriers becomes less important as the concentration of holes increases. It can be thought that the enigmatic absence of the anomalous Hall effect in diluted magnetic semiconductors with ferromagnetic inclusions (MnSb or MnAs clusters with high Curie temperatures) discussed in the literature is most likely related to the presence of Schottky barriers at the boundaries between the clusters and semiconducting matrices. At low carrier concentrations (10^{19} cm^{-3}), the Schottky barriers are fairly wide and prevent the interaction of carriers with the ferromagnetic clusters.

Further studies are, however, necessary to elucidate the nature of ferromagnetic inclusions in the synthesized layers and the mechanism of their interaction mediated by the semiconducting matrix, which contains free carriers and magnetic ions in considerable concentrations. Such studies would certainly be of fundamental interest, especially in light of the recently discovered long-range character of exchange interactions between thin ferromagnetic layers through a semiconducting spacer [34].

ACKNOWLEDGMENTS

The authors thank E.Z. Meilikhov and V.A. Ivanov for discussions and A.B. Granovskii for valuable comments.

This work was financially supported by the Russian Foundation for Basic Research (project nos. 04-02-16158, 03-02-17029, and 02-02-16974) and the program “Spin-Dependent Phenomena in Solids and Spintronics” of the Russian Academy of Sciences.

REFERENCES

1. C. Timm, J. Phys.: Condens. Matter **15**, R1865 (2003).
2. H. Ohno, Science **291**, 840 (2001).
3. H. Ohno and F. Matsukura, Solid State Commun. **117**, 179 (2001).
4. X. Chen, M. Na, M. Cheon, *et al.*, Appl. Phys. Lett. **81**, 511 (2002).
5. T. Dietl, H. Ohno, and F. Mutsukura, Phys. Rev. B **63**, 195205 (2001).
6. K. W. Edmonds, P. Boguslavski, K. Y. Wang, *et al.*, Phys. Rev. Lett. **92**, 037201 (2004).
7. P. M. Krstajie, F. M. Peeters, V. A. Ivanov, *et al.*, Phys. Rev. B **70**, 195215 (2004).
8. E. Abe, F. Matsukura, H. Yasuda, *et al.*, Physica E (Amsterdam) **7**, 981 (2000); F. Matsukura, E. Abe, Y. Ohno, and H. Ohno, Appl. Surf. Sci. **159–160**, 265 (2000).
9. H. Luo, G. B. Kim, M. Cheon, *et al.*, Physica E (Amsterdam) **20**, 338 (2004).
10. P. Mahadevan and A. Zunger, Phys. Rev. B **68**, 075202 (2003).

11. C. Kittel, *Introduction to Solid State Physics*, 5th ed. (Wiley, New York, 1976; Nauka, Moscow, 1978).
12. A. V. Vedyaev, A. B. Granovskiĭ, and O. A. Kotel'nikova, *Kinetic Phenomena in Disordered Ferromagnetic Alloys* (Mosk. Gos. Univ., Moscow, 1992) [in Russian].
13. T. Jungwirth, Q. Niu, and A. H. McDonald, Phys. Rev. Lett. **88**, 207208 (2002).
14. A. M. Nazmul, S. Sugahara, and M. Tanaka, Phys. Rev. B **67**, 241308R (2003).
15. F. Matsukura, D. Chiba, T. Omiya, *et al.*, Physica E (Amsterdam) **12**, 351 (2002).
16. T. Adhikari and S. Basu, J. Magn. Magn. Mater. **161**, 282 (1996).
17. K. W. Edmonds, R. P. Campion, K.-Y. Wang, *et al.*, J. Appl. Phys. **93**, 6787 (2003).
18. T. Dietl, F. Matsukura, H. Ohno, *et al.*, cond-mat/0306484.
19. P. S. Dutta, H. L. Bhat, and V. Kumar, J. Appl. Phys. **81**, 5821 (1997).
20. A. Gerber, A. Milner, M. Karpovsky, *et al.*, Phys. Rev. B **69**, 134422 (2004); J. Magn. Magn. Mater. **242–245**, 90 (2002).
21. K. P. Belov, *Magnetic Transitions* (Fizmatgiz, Moscow, 1959; Consultants Bureau, New York, 1961); A. Arrott, Phys. Rev. **108**, 1394 (1957).
22. D. K. Bowen and B. K. Tanner, *High Resolution X-Ray Diffractometry and Topography* (Taylor and Francis, London, 1998; Nauka, St. Petersburg, 2002).
23. S. Basu and T. Adhikari, J. Alloy Comp. **205**, 81 (1994).
24. D. Woodruff and T. Delchar, *Modern Techniques of Surface Science* (Cambridge Univ. Press, Cambridge, 1986; Mir, Moscow, 1989).
25. M. V. Yablonskikh, Yu. M. Yarmoshenko, E. G. Gerasimov, *et al.*, J. Magn. Magn. Mater. **256**, 396 (2003).
26. C. J. Powell, Appl. Surf. Sci. **89**, 141 (1995).
27. I. K. Kikoin, N. A. Babushkina, and T. N. Igosheva, Fiz. Met. Metalloved. **10**, 488 (1960).
28. B. A. Aronzon, D. Yu. Kovalev, A. N. Lagar'kov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 87 (1999) [JETP Lett. **70**, 90 (1999)].
29. A. B. Granovsky, A. V. Vedyaev, and F. Brouers, J. Magn. Magn. Mater. **136**, 229 (1994).
30. S. Sze, *Physics of Semiconductor Devices*, 2nd ed. (Wiley, New York, 1981; Mir, Moscow, 1984).
31. V. V. Ryl'kov, B. A. Aronzon, A. B. Davydov, *et al.*, Zh. Éksp. Teor. Fiz. **121**, 908 (2002) [JETP **94**, 779 (2002)].
32. E. Z. Meĭlikhov, Zh. Éksp. Teor. Fiz. **116**, 2182 (1999) [JETP **89**, 1184 (1999)].
33. Yu. I. Petrov, *Physics of Small Particles* (Nauka, Moscow, 1982) [in Russian].
34. P. Walser, M. Hunziker, T. Speck, *et al.*, Phys. Rev. B **60**, 4082 (1999); R. R. Gareev, D. E. Burgler, M. Buchmeier, *et al.*, Phys. Rev. Lett. **87**, 157202 (2001).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 4, 2005, pp. 752–759.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 4, 2005, pp. 850–859.

ELECTRONIC PROPERTIES OF SOLIDS

Electron and Lattice Stages in the Valence Transition in SmTe under a High Hydrostatic Pressure

O. B. Tsiok[a], L. G. Khvostantsev[a], I. A. Smirnov[b], and A. V. Golubkov[b]

[a] *Vereshchagin Institute of High-Pressure Physics, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia*

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

e-mail: tsiok@hppi.troitsk.ru

Received October 6, 2004

Abstract—The results of precision measurements of the resistivity, thermopower, volume, and thermal conductivity of the compound SmTe under truly hydrostatic pressure conditions at room temperature are reported. High quality stoichiometric and doped (*n*-type, $n \approx 8 \times 10^{18}$ cm^{-3}) single crystals are studied. It is found that the valence transition occurs as consecutive stages of rearrangement of the electron subsystem and the crystal lattice, which take place under different pressures. At the initial stage of the transition, metallization is observed, which is accompanied by anomalies in kinetic coefficients; the curve describing the pressure dependence of the volume deviates from the curve corresponding to the initial semiconductor phase only slightly. The next stage is accompanied by a substantial change in the sample volume (lattice collapse); in this pressure range, however, the resistivity and thermopower become independent of pressure. At the final stage of the transition, the sample compressibility decreases; the resistivity and thermopower become again functions of pressure; and a state emerging in the sample in this case corresponds to the "golden" phase of SmS in all the properties being measured. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The phenomenon of intermediate valence, which is inherent in rare-earth compounds and is manifested when the energy of localized *f* states becomes close to the energy of conducting *d* states, has been an object of intense study for more than three decades. This phenomenon was studied in a huge number of publications (see, for example, [1] for a review). The most striking manifestations of the intermediate-valence state are an intermediate value of the crystal lattice parameter and a specific peculiarity in the quasiparticle spectrum of carriers in the vicinity of the Fermi level in the form of peaks with a giant density of states, which are separated by a narrow (of a few millielectronvolts) gap. Samarium chalcogenides (in particular, SmTe) are classical objects for studying the intermediate-valence state. Under normal conditions, these compounds are semiconductors in which the gap is formed by $4f^6$ states of samarium ions with the conduction band formed by 5*d* states of samarium ions. The intermediate-valence state in samarium chalcogenides can be induced by a high pressure, leading to an increase in the 5*d* bandwidth and, ultimately, to overlapping of the energies of the $4f^6 5d^0$ and $4f^5 5d^1$ configurations. Another widely used method for obtaining the intermediate-valence state is the substitution of other rare-earth metals (such as lanthanum, yttrium, and cerium) for samarium. The substitution of a trivalent smaller-radius ion for the bivalent samarium ion in the compound gives rise to an additional "chemical" pressure. This method has made it possible to study intermediate-valence state using a wide set of experimental methods without specific limitations associated with high pressures. The given method for obtaining such states is not equivalent to the application of a high pressure (which was repeatedly noted in the literature; see, for example, [2]) since, in addition to the chemical pressure, electrons are supplied to the conduction band of the system. At early stages of investigations, this nonequivalence led to overestimation of the effects of the *f–d* hybridization in the theoretical description of intermediate-valence state. The concepts of the exciton origin of this state are also being developed [3–7].

At the same time, some aspects of the phenomenon have not been interpreted adequately as yet. We are talking about the interrelation between the rearrangement of the electron spectrum under pressure and volume anomalies in the region of a transition with a change in valence. These aspects are either ignored in publications altogether, or it is assumed that both processes occur synchronously since the effects associated with exciton condensation ($T \approx 70$ K) should not play a significant role at room temperature. Such a disregard of this problem is apparently due to the fact that the electron spectrum and the volume were studied on different samples using different high-pressure setups (the choice was dictated by the relevant experimental methods) and the problem of establishing the correlation in

the behavior of electron properties and the crystal lattice in the transition region was not specially posed. In our earlier publication [8], it was shown that transitions in the electron subsystems of SmSe and SmTe are not synchronous to the lattice collapse, but a satisfactory interpretation of this effect was not proposed.

Here, we attempt at a detailed study of the interrelation between the rearrangement in the electron spectrum and the volume anomaly during a valence transition in SmTe. The study is based on precision measurement of the volume and transport parameters (resistivity, thermopower, and thermal conductivity) of monocrystalline samples under the ideal hydrostatic conditions with a rigorously controlled pressure. The use of the doping (generally speaking, self-doping, i.e., a slight deviation from stoichiometry) as an additional parameter is aimed at studying the effect of conduction electrons on the process of transition in the absence of accompanying chemical pressure emerging in substituted solutions.

2. EXPERIMENT

High-pressure experiments were carried out on a Toroid-type device [9] with a central hole diameter of 25 mm (the useful volume was approximately 1 cm^3). The design of high-pressure apparatuses of this type ensured reliable operation of an ampoule with a liquid and 12–14 electric leads during an increase and decrease in pressure and, hence, makes it possible to implement various experimental techniques in a pressure range up to 9 GPa under ideal hydrostatic conditions. The design of the hydrostatic ampoule and its content are determined by the purpose of a specific experiment. By way of example, Fig. 1a shows an ampoule intended for high-temperature measurements. A partition separates two compartments of the ampoule filled with different liquids. The "hot" compartment contains a heater with an assembly for measuring resistance (Fig. 1b) or a sample with a strain gauge glued to it (Fig. 1c). The "cold" compartment contains a manganite pressure gauge. The experimental cell for room-temperature experiments has a simple design without the partition and the heater. The experimental techniques were described in detail in our earlier publications [10, 11]. Since we compare here the properties of the samples measured in different experiments, some remarks concerning the precision of pressure measurements will be appropriate. As a rule, an error of 0.2–0.3 GPa for a pressure range of 10 GPa is usually regarded as quite satisfactory. With such a precision of pressure measurements, a difference of a few kilobars in the positions of anomalies in various properties is attributed by some authors to shear stresses and to errors in pressure measurements. Comparison of the results obtained using high-pressure apparatuses of different types is even more difficult. In this study, all experimental results are obtained under hydrostatic conditions (i.e., in the absence of shear stresses in the sample). The reproducibility of the pressure scale in our experiments was at a level of 0.01 GPa (100 bar!). The absolute error of pressure measurements (about 1%) is determined by the accuracy of the reference scale (to calibrate the manganite pressure gauges, we used resistivity jumps during transitions in bismuth at 2.54 and 7.7 GPa).

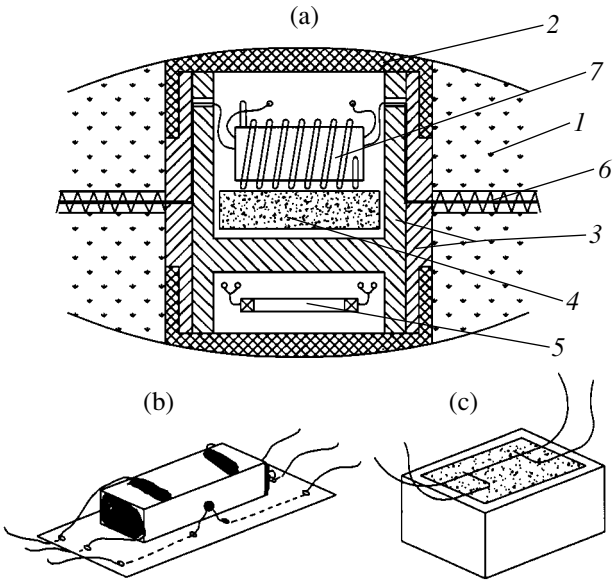


Fig. 1. Hydrostatic cell for electric measurements at high temperatures in Toroid-type apparatuses. (a) Schematic diagram: *1*—gasket (lithographic stone), *2*—ampoule lid (brass), *3*—ampoule shell (Teflon), *4*—thermal insulation (mineral wool), *5*—pressure gauge (manganin), *6*—leads, *7*—heater with a sample and a thermocouple. (b) Assembly for resistivity measurements. (c) Sample with a strain gauge.

We choose thermal conductivity as a property sensitive to variations in the phonon subsystem. Serious difficulties in studying this kinetic coefficient are associated with the fact that it is difficult to measure this parameter and even more difficult to interpret the results. The only advantage of thermal conductivity measurements in this study is the possibility of carrying out these measurements with our high-pressure setup on the same samples and with the same pressure scale. Since SmTe is a brittle material, the installment of a heater "inside" the sample (standard cylindrical geometry) appeared problematic. For our purposes, it was sufficient to measure the thermal conductivity on qualitative level (tendencies in variation, positions of anomalies, etc.); for this reason, we chose a relatively simple measuring technique.

Figure 2a shows a cell for thermal conductivity measurements; Fig. 2b illustrates the design of the assembly. To reduce the errors associated with heat fluxes through the lateral surface of the sample, we used samples of a nearly cubic shape. The typical size of the samples in a test series was within 1.8–2.2 mm. The end

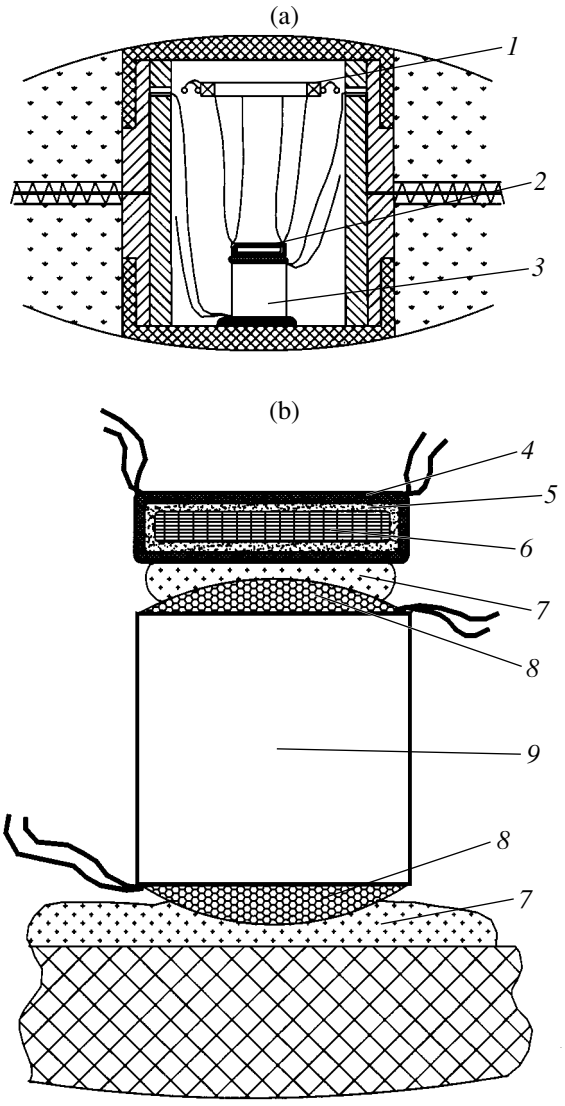


Fig. 2. Hydrostatic cell for thermal conductivity measurements. (a) Schematic diagram: *1*—pressure gauge, *2*—heater, *3*—sample. (b) Schematic of the assembly: *4*—copper thermal screen, *5*—insulation made of capacitor paper, *6*—heater, *7*—Wood alloy, *8*—indium, *9*—sample.

faces of sample *9* (Fig. 2b) were thoroughly tinned with indium *8*, into which a copper-constantan thermocouple made of a 50-μm wire were soldered. Heater *6* was in the form of a 50-μm constantan wire wound on a strip of paper. Current and potential copper leads (40-μm) were soldered to the heater ends. The heater, which has an insulation of capacitor paper *5*, was surrounded by copper screen *4* and the entire structure was impregnated with epoxy adhesive. The sample was soldered to the heater and the heat sink (ampoule lid) by Wood alloy *7*. To suppress convective flows in the liquid, the entire assembly was coated with mineral wool (not shown in the figure).

With such a design of the measuring cell, a considerable part of the heat flux passes through the liquid transmitting pressure to the sample. The pressure dependence of the heat flux through the liquid was measured in the same cell but without a sample. The heat flux through conducting leads (which was measured separately) amounted to a few percent of the total flux and was regarded as independent of pressure in data processing. At the beginning of the experiment (before the liquid was poured into the ampoule), we measured the initial thermal conductivity of the sample taking into account the heat flux in the leads. Then the ampoule was filled with the liquid and one more measurement was made under zero pressure. The initial heat flux through the liquid was determined as the difference between the results of these two measurements. To find the thermal conductivity of the sample under pressure, we must subtract the thermal conductivity of the leads and that of the liquid (calculated from the measured initial value and the known pressure dependence) from the measured overall value. The thermal conductivity coefficient was calculated with respect to the initial sample size (the sample compressibility was ignored). To reduce the influence of thermal effects in the liquid (which, however, were taken into account in data processing) during variation of pressure, measurements were made for a comparatively large (5–7 K) temperature difference at the end faces of the sample. Without going further into the analysis of experimental errors, it should be noted that, according to our estimates, a possible systematic error in thermal conductivity measurements amounts approximately to 30% for the maximal pressure. As regards the authenticity of the observed anomalies in the thermal conductivity of SmTe, it is beyond any doubt since the thermal conductivity of a liquid is a monotonic smooth curve without any peculiarities.

We used 4 : 1 methanol-ethanol and 1 : 1 pentane–petroleum ether mixtures as the pressure-transmitting medium. The second mixture has a hydrostatic limit approximately at 5 GPa, but is a good insulator. It was used only for measurements on a stoichiometric SmTe sample under pressures below 4 GPa. The mixture of alcohols exhibits a noticeable electrical conductivity, which affects the accuracy of resistivity and thermopower measurements on the stoichiometric sample under a pressure below 2.5 GPa because of shunting and electrochemical effects (this information has been removed from the graphs). In the intermediate pressure range (2.5–4 GPa), the results of measurements in various liquids coincide to a high degree of accuracy.

The samples of SmTe were synthesized and certified at the Ioffe Physicotechnical Institute, Russian Academy of Sciences. The initial samples were in the form of coarse (about 5 mm in size) single crystals, from which the samples of the required shape were cleaved for measurements. A single crystal of doped SmTe was obtained from the initial stoichiometric sample by annealing in samarium vapor. The charge carrier concentration in this samples was $n \approx 8 \times 10^{18}$ cm^{-3} according to the results of Hall measurements. The contacts

were soldered by rubbing molten indium into a fresh chip on the sample surface.

3. RESULTS

Figure 3 shows the results of measurement of the sample resistance. The resistivity was calculated using the initial size of the samples; i.e., the compressibility is disregarded (the correction for compressibility amounts to about 7% under maximal pressure). In addition, an error of about 5% may appear in the calculated values of ρ in determining the geometrical factor due to finite sizes of potential contacts to the sample. The inset shows the pressure dependence of activation energy E_a of conduction (thermal gap) for a stoichiometric sample. The curve was obtained from the slopes of the temperature dependences of resistance R in the temperature range 290–390 K. In the vicinity of 4 GPa, the temperature range narrows to 290–350 K due to a noticeable nonlinearity of the dependences. We assumed that the resistance is given by

$$R \propto \exp\left(\frac{eE_a}{kT}\right),$$

where e is the electron charge and activation energy E_a is expressed in electronvolts. This gives

$$E_a = \frac{k}{e}\frac{d(\ln R)}{d(1/T)}.$$

A conventional error in determining the thermal gap is disregard of the variation of pressure under a change in the sample temperature. The error becomes especially large when the sample resistance is independent of temperature ($E_a = 0$) and depends only on pressure, but a "fictitious gap" appears due to an increase in pressure in the cell during heating. In this study, the thermal gap is determined correctly from methodical point of view. The current value of pressure P_n was measured at each point n of the temperature dependence of the resistance, and the quantity

$$\ln R(P_n, T_0) = \ln[R(P_0, T_0)] + \frac{d\ln R(P, T_0)}{dP}(P_n - P_0)$$

was calculated. The $\ln R(P, T_0)$ dependence in the vicinity of P_0 at room temperature T_0 was determined in the same experiment. Then, for each point, we determined the value of

$$E_a(P_n) = \frac{k}{e}\frac{\ln R(P_n, T_0) - \ln R(P_n, T_n)}{1/T_0 - 1/T_n}$$

and averaged the set of $E_a(P_n)$ values with weight factor $(1/T_0 - 1/T_n)/(1/T_0 - 1/T_{\max})$, where $T_{\max}$ is the maximal temperature in a given measurement.

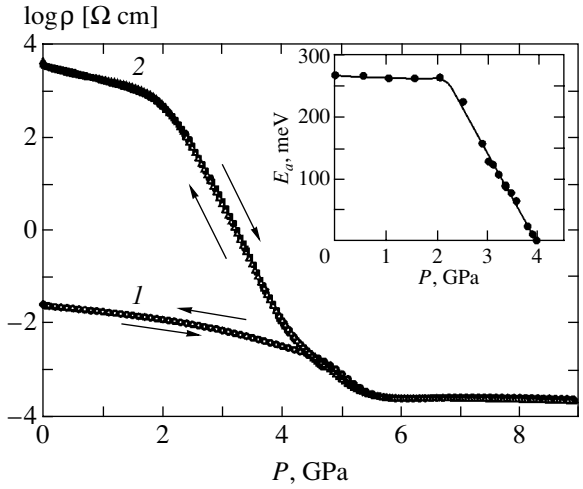


Fig. 3. Resistivity of SmTe samples at room temperature: *1*—doped sample upon an increase (●) and a decrease (○) in pressure; *2*—stoichiometric sample (results of three experiments with different liquid). The inset shows the activation energy for the stoichiometric sample.

It should be noted that the pressure dependences of resistivity obtained with increasing and decreasing pressure (see Fig. 3) coincided to a high degree of accuracy. No hysteresis loops or irreversible behavior were observed. In the low-pressure range ($P < 2$ GPa), the conductivity of the stoichiometric sample is ensured by an insignificant amount of *p*-type impurities and varies insignificantly under pressure. For a pressure of 2–4 GPa, an intrinsic conductivity region is observed, in which the logarithm of resistivity and the activation energy linearly decrease with increasing pressure (the conduction of the doped sample is of the impurity type in the entire range of pressure $P < 4$ GPa). It is important for subsequent analysis to note the following circumstance. Under a pressure of about 4 GPa, the value of E_a determined using the above technique vanishes. In the case of a classical semiconductor, the disappearance of the gap should lead to a strong anomaly on the $\log\rho(P)$ curve and to a pressure-independent resistance (on the given scale of the graph) in the "metallic" state. However, the experiment shows that the $\log\rho(P)$ curve exhibits only a small kink in the vicinity of 4 GPa; at a higher pressure (up to 5.5 GPa), the value of this quantity decreases exponentially with increasing pressure in spite of the metallic temperature behavior of resistivity. A similar anomaly is also observed for the doped sample. In the pressure range 4.5–5.5 GPa, the sample resistivity decreases exponentially with increasing pressure in spite of the metallic type of its temperature dependence. At a pressure of ~5.5 GPa, the resistivity of both samples attains a level of $2.45 \times 10^{-4}\ \Omega$ cm [$(2.35 \pm 0.1) \times 10^{-4}\ \Omega$ cm with allowance for the correction for compressibility] and varies only slightly upon a further increase in pressure.

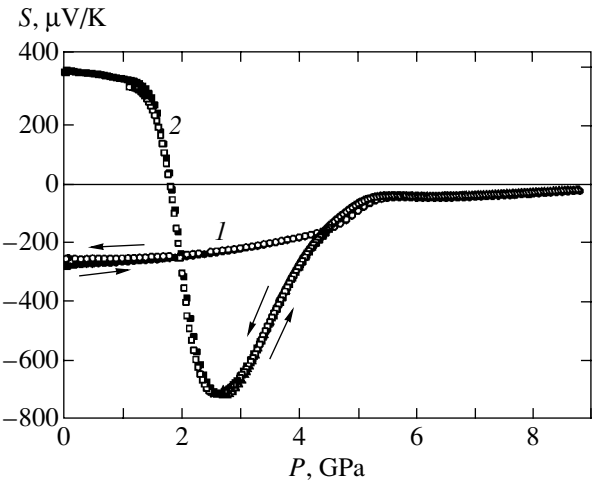


Fig. 4. Thermopower of a doped (*1*) and a stoichiometric (*2*) SmTe sample at room temperature upon an increase (dark symbols) and a decrease (light symbols) in pressure.

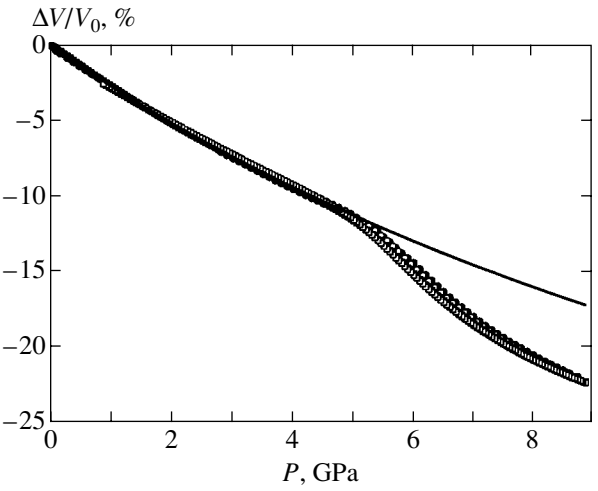


Fig. 5. Volume of a doped (●, ○) and a stoichiometric (■, □) SmTe sample upon an increase (●, ■) and a decrease (○, □) in pressure. The solid line, $V^{II}(P)/V_0$, describes the fitting of the data on $V(P)/V_0$ for $P < 4$ GPa to the Murnaghan equation with fitting parameters $K = 34.575 \pm 0.052$, $K' = 3.170 \pm 0.034$, $\chi^2 = 0.0004$.

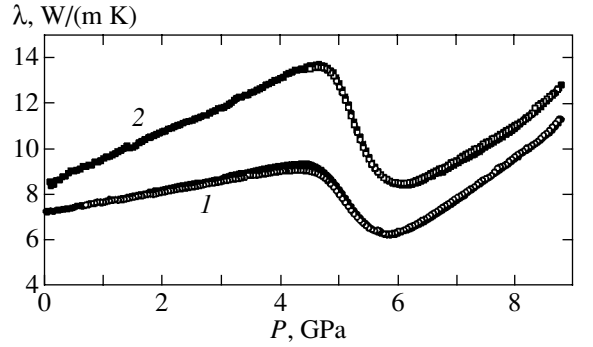


Fig. 6. Thermal conductivity of a stoichiometric (*1*) and a doped (*2*) SmTe sample at room temperature upon an increase (dark symbols) and a decrease (light symbols) in pressure.

Figure 4 shows the dependences of thermopower S on pressure P at room temperature. We also note a nonhysteretic behavior of the curves obtained upon an increase and a decrease in pressure (insignificant irreversibility at $P < 2$ GPa is apparently associated with the rearrangement of impurity states; however, these details are beyond the scope of this study). The variations of thermopower and resistivity with pressure in the semiconductor state ($P < 4$ GPa) occur in accordance with the same scenario. For low pressures, the values of ρ, S, and E_a are determined by the initial concentration of *p*-type impurities and depend on pressure only slightly. With increasing pressure, the width of the 5*d* band increases, leading to a decrease in the forbidden gap and to an increase in the intrinsic charge carrier concentration. On account of the difference in the mobilities of 4*f* holes and 5*d* electrons, transport phenomena for $P > 2$ GPa can be assumed to be determined by conduction electrons alone. Their concentration increases exponentially with pressure (E_a decreases linearly), which leads to a linear decrease in $\log\rho(P)$. The thermopower curve in the vicinity of 2 GPa exhibits "switching" from the impurity-type conduction to the intrinsic electron type. Behind the minimum in the vicinity of 2.7 GPa, the thermopower is determined only by intrinsic charge carriers and decreases with increasing concentration of these carriers (upon a decrease in E_a). In the vicinity of 4 GPa, the $S(P)$ curve has a small kink due to "metallization," after which the thermopower continues to increase up to a pressure of 5.15 GPa. The corresponding anomaly in thermopower is also observed for the doped sample in a pressure range of 4.50–5.35 GPa.

Figure 5 shows the pressure dependences of the volume obtained from strain measurements upon an increase and a decrease of pressure. Under a pressure below 4 GPa, the $V(P)$ dependence is of the semiconductor type. Anomalous behavior of the volume begins at $P > 4$ GPa in the form of a smoothly increasing deviation from the initial dependence. Figure 7c below shows the difference between the measured volume V of the samples and the extrapolated pressure dependence $V^{II}(P)$ of the volume for the semiconductor phase. The most substantial change in the volume, which is usually associated with the departure of an electron from the $4f^6$ level, is observed in the region of 5.5–7 GPa, although the compressibility remains anomalously high for higher pressures also.

It is appropriate to emphasize once again that the main variation of the volume begins only at $P = 5.5$ GPa, while the main variations of the electron properties have finished at this pressure.

Figure 6 shows the curves describing the pressure dependence of the thermal conductivity of SmTe samples. As in the case of resistivity, the initial size of the samples was used for calculating the specific value. Allowance for compressibility increases the size of the anomaly and slightly shifts the positions of the minima

towards higher pressures; however, this is immaterial for the subsequent discussion. In spite of their low accuracy (the systematic error is estimated at approximately 30% for 9 GPa), the thermal conductivity data nevertheless make it possible to judge the heat conduction mechanism and anomalies in the phonon spectrum. For $P < 4$ GPa, the heat conduction of the stoichiometric sample is of the phonon type; with increasing pressure, the thermal conductivity increases due to hardening of the phonon spectrum without a noticeable change in anharmonism. In the doped sample, part of the heat is transferred by conduction electrons. The faster increase in the thermal conductivity under pressure is due to the electron component and matches the decrease in the resistance of the doped sample in this pressure range (see Fig. 3). Metallization of the sample in the vicinity of 4 GPa is unnoticeable on the thermal conductivity curve. As in the case of the $V(P)$ dependence, the smooth deviation of the thermal conductivity from the initial dependence apparently increases, but cannot be noticed in view of insufficient accuracy of measurements (the opposite tendencies, viz., an increase in the electron component and an enhancement of scattering, probably compete in this case).

At this stage of the discussion, the following remarks can be made.

(i) In the pressure range where a decrease in the resistivity is observed, the thermal conductivity does not increase (owing to the electron component), but sharply decreases, the anomaly for the doped sample being noticeably larger. (ii) The minimum of the thermal conductivity corresponds to the maximum of compressibility. (iii) In the entire pressure range (including the region above 7 GPa), the heat conduction apparently remains mainly of the phonon type.

Generally speaking, it is impossible at present to unambiguously interpret the behavior of the thermal conductivity in such a complicated case. The increase in the thermal conductivity for $P > 6$ GPa can also be explained by an increase in the electron component; however, it is an order of magnitude smaller than the thermal conductivity of normal metals even in this case.

To give a visual idea on the mutual positions of anomalies, all measured dependences are presented in Fig. 7 on a magnified scale.

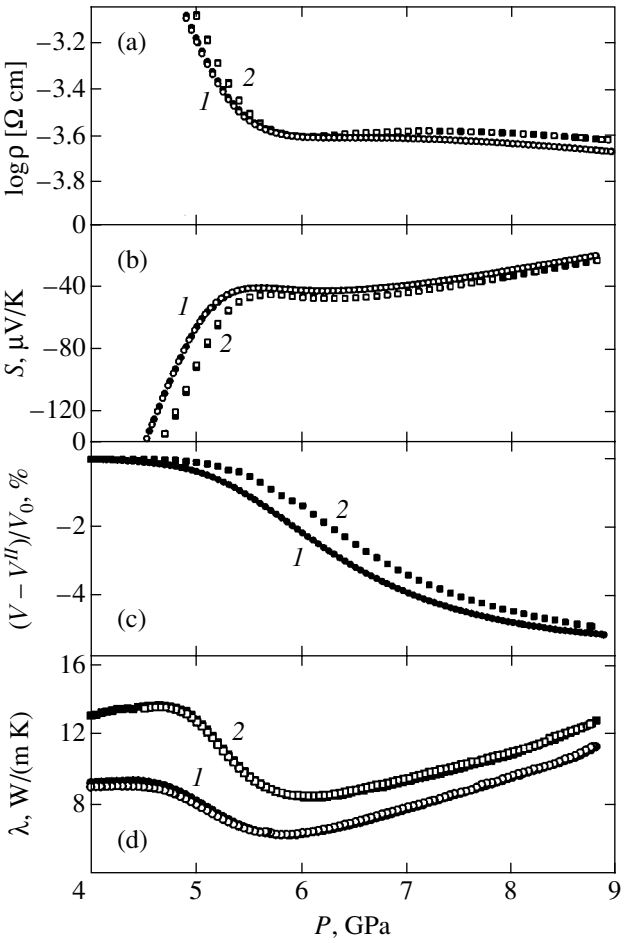


Fig. 7. Electron and lattice properties of a stoichiometric (*1*) and a doped (*2*) SmTe samples in the region of valence transition upon an increase (dark symbols) and a decrease (light symbols) in pressure.

4. DISCUSSION

An attempt at constructing a consistent explanation of experimental results (especially the above-mentioned separation of the "electron transition" and lattice collapse, the decrease in thermal conductivity upon an increase in electrical conductivity, and strange metallization followed by a decrease in $\rho(P)$ by almost two orders of magnitude) requires the acceptance of some model concepts on the mechanism of the valence transition. Unfortunately, it is impossible to directly compare the experiment with the results of calculations based on various theoretical models at the present stage. Even the calculation of resistivity (not to mention thermopower and thermal conductivity) requires taking into account various scattering mechanisms based on exact information on the phonon spectrum and the elementary excitation spectrum for charge carriers, which cannot be interpreted in the framework of the one-electron approximation. In this sense, experimental data on transport phenomena are "inconvenient" for estimating the adequacy of theoretical models. On the other hand, electric measurements can be carried out under ideal hydrostatic conditions with a high degree of accuracy and with a small pressure step, which provides detailed information on the interrelation between various stages of the transition. This is difficult and even impossible when "convenient" experimental methods are used (such as obtaining the phonon spectrum using the inelastic neutron scattering technique). With such an approach, we can try to estimate the adequacy of a theoretical model not by comparing the specific values of resistivity, thermopower, etc., with the results of calculations, but using a consistent qualitative description of

all the stages of the transition in the framework of this model.

In our opinion, it is difficult to explain the plateau on the thermopower and resistivity curves in the maximal compressibility region as well as other experimentally observed peculiarities in the framework of the model based on the concept of electron fluctuations between the $4f^6 5d^0$ and $4f^5 5d^1$ configurations (promotion model; see, for example, [1]). In this model, volume variations are usually attributed to the variation of the number of electrons in the conduction band. Upon a smooth variation of the volume in the pressure range 5.5–7 GPa, the corresponding variations of the electron kinetic coefficients must be observed. However, the $\rho(P)$ and $S(P)$ curves (see Figs. 7a and 7b) display a plateau-type anomaly in this pressure range.

Exciton models based of the presence of "shallow" (10–20 meV) exciton levels also fail to explain the experimental results since such models do not predict localized electron states in the high-temperature range (300 K ≈ 25 meV).

In the case of the intermediate-valence state, excitons with a structure strongly depending on the rearrangement of the phonon spectrum may exist. In the exciton models [4, 5, 7], a valence transition (a smooth transition as in SmSe, SmTe or an abrupt transition as in SmS) occurs as a result of softening of the fully symmetric exciton mode. This vibration ("breathing" mode) can be described as a radial displacement of anions closest to a given samarium cation. This process simultaneously involves the electron and phonon subsystems, which leads to a strong (resonant) electron–phonon interaction resulting in a radical rearrangement of the electron and phonon spectra for a certain relation between the energy parameters. In such a vibration, an electron excited from the $4f^6$ state does not reach the conduction band, but is localized on an orbital constructed from the 5*d* states of the nearest cations with the preserved symmetry of the ground state. The effective potential relief of atomic vibrations acquires several minima corresponding to several vibrational modes with different equilibrium positions and frequencies. Subsequent analysis of our experimental data requires allowance for an additional hard resonance mode in the high-temperature range, which was explained in [7]. In our opinion, our experimental results permit a consistent description based on the exciton model including two types of exciton excitations, viz., ordinary "shallow" excitons (playing a significant role at the beginning and at the end of the process) and "hard" excitons corresponding to the high-temperature resonance vibrational mode and determining the behavior of the system in the maximal anharmonism (maximal compressibility) region in a pressure range of 5.5–7 GPa.

Returning to the discussion of the results, we note that all experimental dependences obtained for the doped sample for $P > 4$ GPa virtually repeat the curves for the stoichiometric sample with a slight (about 0.2 GPa) shift towards high pressures. In formulating the problem, we assumed that the transition is "controlled" by the carrier concentration in the conduction band. The observed effect turned out to be opposite. It was found that excess electrons in the conduction band hamper the formation of exciton instability probably due to screening effects (although other explanations can also exist). An important result is that the transition is determined by the pressure dependence of the forbidden gap E_g, but it is exciton instability and not the transition of electrons to the conduction band that plays a decisive role.

Let us now consider all the "stages" of this transition, beginning with a strange "metallic" state formed under a pressure of 4 GPa. We can estimate the difference between the values of E_g (the gap between the 4*f* and 5*d* states obtained from optical measurements) and E_a (obtained from the temperature dependence of resistivity). According to Bucher *et al.* [12], $E_g = 0.63$ eV at zero pressure; with increasing pressure, the value of E_g decreases at a rate $dE_g/dP = -120$ meV/GPa (extrapolation to zero gives 5.2 GPa). According to our results, $dE_a/dP = -145$ meV/GPa, $E_a = 0$ for $P = 3.97$ GPa; extrapolation to zero pressure gives $E_a(0) = 575$ meV; i.e., $E_g - E_a \approx (55–150)$ meV. Thus, the metal-type conduction for $P > 4$ GPa is apparently associated with certain conducting states lying below the bottom of the 5*d* conduction band. If we assume that these states correspond to ionization of low-energy excitons, their concentration must be determined by the ratio E_g/Δ_{ex} (where Δ_{ex} is the exciton energy equal approximately to 30 meV (in accordance with the estimate $\hbar\omega/\Delta_{ex} \approx 0.15$ [5]) or 20 meV (the energy of the *f–f* transition [13])),

$$n \propto \exp(-E_g/\Delta_{ex}).$$

If we assume that $E_g = 0$ for $P = 5.5$ GPa (second kink on the $\rho(P)$ curve), a rough estimate gives

$$\ln\rho(4\text{ GPa}) - \ln\rho(5.5\text{ GPa}) = 2.3[\log\rho(4\text{ GPa}) - \log\rho(5.5\text{ GPa})] = 3.7.$$

Then $E_g(4\text{ GPa}) \approx 3.7\Delta_{ex} \approx (70–110)$ meV, which is in satisfactory agreement with the above estimate of the difference $E_g - E_a$. The $(V - V^{II})(P)$ dependence depicted in Fig. 7c displays a smoothly increasing defect of volume in this region, which is in qualitative agreement with the proposed interpretation. The anomaly of thermal conductivity λ (the peak near 4.5 GPa followed by a decrease in its value; see Fig. 7d) can also be explained by an increase in the number of fluctuating atoms. It was noted above that the heat conduction of the samples is of the phonon type with an insignificant (10–20%) electron contribution for the doped sample. Like the anomaly on the $V(P)$ curve, the thermal conductivity anomaly can be interpreted as a smoothly

increasing deviation from the linear dependence, beginning at $P = 4$ GPa; in this case, both the phonon–phonon and the electron–phonon scattering increases (the anomaly for the doped sample is noticeably larger). In this pressure range, the behavior of the kinetic coefficients can be explained equally well in the framework of hybridization (promotion) models since it is apparently impossible to separate shallow excitons from variations of the electron spectrum due to hybridization at $T \approx 300$ K. The above description of the 4–5.5 GPa region should be treated not as a proof of the aptitude of the exciton model, but as a possibility of describing the experimental data using this model also.

The next stage of the transition corresponds to the pressure range 5.5–7 GPa, in which the $V(P)$ dependence changes substantially (anomalous compressibility and lattice collapse are observed), while the $\rho(P)$ and $S(P)$ curves acquire a plateau (a plateau on the $S(P)$ curve in the anomalous compressibility region was detected earlier in [8]). In fact, in the pressure range in which the major part of electrons leave the $4f^6$ state, the density of states at the Fermi level and its derivative remain unchanged (the version of matched variation of the density of states and scattering from phonons is improbable since the thermal conductivity curves do not display the corresponding anomaly). In the framework of the exciton model [7], this effect can be explained by the fact that the maximal compressibility region corresponds to the maximal anharmonism in atomic vibrations. As a result of enhanced anharmonism, a resonant vibration is induced, which is simultaneously a lattice vibration and a charge density oscillation (hard exciton). Lattice collapse corresponds to a change in the position of the potential energy minimum of this vibrational mode, but the number of free charge carriers remains unchanged in this case (in other words, the electrons leaving the f level are involved in this vibrational process and do not make a substantial contribution to transport phenomena).

The final stage of the transition ($P > 7$ GPa) has no clearly defined boundaries. Since the term "valence transition" does not define the final state of the system, the question of "completion" of the transition should be treated as a matter of terminology. In the pressure range above 7 GPa, the resistivity begins decreasing, the thermopower increases, tending to the range of positive values, while the compressibility gradually decreases, remaining anomalously high. On quantitative and qualitative levels, these characteristics correspond to a stable intermediate-valence state analogous to the golden phase of SmS, which has been studied in detail.

It should be noted in conclusion that the proposed description of the transition in SmTe based on the exciton model of the intermediate-valence state should not be treated as the best possible and unique. Our main task was rather to formulate the problems concerning the interrelation of the stages in the valence transition and to obtain experimental data that can be useful for the development of theoretical concepts concerning the intermediate-valence state.

ACKNOWLEDGMENTS

The authors are grateful to A.F. Barabanov and K.A. Kikoin for discussion of the results concerning the theory of intermediate valence.

REFERENCES

1. P. Wachter, in *Handbook on the Physics and Chemistry of Rare Earth*, Ed. by K. A. Gschneidner, Jr., L. Eyring, G. N. Lander, and G. R. Choppin (Elsevier, Amsterdam, 1994), Vol. 19, p. 222.
2. P. Wachter, A. Jung, and P. Steiner, Phys. Rev. B **51**, 5542 (1995).
3. S. D. Mahanti, T. A. Kaplan, and M. Barma, J. Appl. Phys. **49**, 2084 (1978).
4. I. K. Kikoin, Zh. Éksp. Teor. Fiz. **85**, 1000 (1983) [Sov. Phys. JETP **58**, 582 (1983)].
5. K. A. Kikoin and A. S. Mishchenko, Zh. Éksp. Teor. Fiz. **94** (11), 237 (1988) [Sov. Phys. JETP **67**, 2309 (1988)].
6. J. Neuenschwander and P. Wachter, Phys. Rev. B **41**, 12693 (1990).
7. K. A. Kikoin and A. S. Mishchenko, Zh. Éksp. Teor. Fiz. **104**, 3810 (1993) [JETP **77**, 828 (1993)].
8. O. B. Tsiok, V. A. Sidorov, V. V. Bredikhin, *et al.*, Solid State Commun. **79**, 227 (1991).
9. L. G. Khvostantsev, L. F. Vereschagin, and A. P. Novikov, High Temp. High Press. **9**, 637 (1977).
10. L. G. Khvostantsev, V. A. Sidorov, and O. B. Tsiok, in *Properties of Earth and Planetary Materials at High Pressures and Temperatures*, Ed. by M. H. Manghnani and T. Yagi (Am. Geophys. Union, Washington, D.C., 1998), Geophys. Monogr. 101, p. 89.
11. O. B. Tsiok, V. V. Bredikhin, V. A. Sidorov, and L. G. Khvostantsev, High Press. Res. **10**, 523 (1992).
12. E. Bucher, V. Narayanamurti, and A. Jayaraman, J. Appl. Phys. **42**, 1741 (1971).
13. G. Trawaglini and P. Wachter, Phys. Rev. B **30**, 5877 (1984).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 100, No. 4, 2005, pp. 760–774.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 4, 2005, pp. 860–876.

ELECTRONIC PROPERTIES OF SOLIDS

High-Pressure Hydrogenated Fullerenes: Optical Spectra and Stability of $C_{60}H_{36}$ at High Pressure¶

K. P. Meletov[a] and G. A. Kourouklis[b]

[a]*Institute of Solid State Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*
e-mail: mele@issp.ac.ru

[b]*Physics Division, School of Technology, Aristotle University of Thessaloniki, Thessaloniki, Greece GR-540 06*

Received October 14, 2004

Abstract—The optical Raman and photoluminescence (PL) spectra of the high-pressure hydrogenated fullerene C_{60} are studied at normal conditions and at high pressure. The Raman spectrum of the most stable hydrofullerene $C_{60}H_{36}$ contains a large number of peaks related to various isomers of this molecule. Comparison of the experimental data with the results of calculations shows that the most abundant isomers have the symmetries S_6, T, and D_{3d}. The Raman spectrum of deuterofullerene $C_{60}D_{36}$ is similar to that of $C_{60}H_{36}$, but the frequencies of the C–H stretching and bending modes are shifted due to the isotopic effect. The PL spectrum of hydrofullerene $C_{60}H_{36}$ is shifted to higher energies by approximately 1 eV with respect to that of pristine C_{60}. The effect of hydrostatic pressure on the Raman and PL spectra of $C_{60}H_{36}$ has been investigated up to 12 GPa. The pressure dependence of the phonon frequencies exhibits peculiarities at approximately 0.6 and 6 GPa. The changes observed at approximately 0.6 GPa are probably related to a phase transition from the initial orientationally disordered body-centered cubic structure to an orientationally ordered structure. The peculiarity at approximately 6 GPa may be related to a pressure-driven enhancement of the C–H interaction between the hydrogen and carbon atoms belonging to neighboring molecular cages. The pressure-induced shift of the photoluminescence spectrum of $C_{60}H_{36}$ is very small up to 6 GPa, and a negative pressure shift was observed at higher pressure. All the observed pressure effects are reversible with pressure. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Hydrofullerenes have attracted considerable interest, particularly in relation to their potential use as hydrogen storage materials [1]. The existence of several hydrofullerenes has been predicted by theoretical calculations and some of them have been synthesized [2–5]. The most stable member, $C_{60}H_{36}$, is the hydrogenated derivative of the pristine C_{60} molecule, which can be prepared using either the high-pressure hydrogenation [6] or the hydrogen-atom transfer to C_{60} from other reagents in solution [7]. Hydrofullerene samples prepared by the use of both methods are usually rather nonuniform and may contain hydrides with various mass weights, as well as a small amount of reagents. Characterization of the hydrogenation reaction products is therefore of great importance and various methods such as electron, X-ray, and neutron diffraction; nuclear magnetic resonance; infrared spectroscopy; and laser desorption mass spectrometry have been used for this purpose [4–6, 8–10].

The $C_{60}H_{36}$ molecule can exist in a great number of isomeric forms, but only a small number of them are stable [11]. The isomeric form with the highest symmetry, T_h, has 12 double bonds, arranged as far apart as possible on its surface, while the form having the double bonds in four isolated aromatic six-member rings lacking hydrogen atoms and located at the corners of a tetrahedron has a T-symmetry structure. Between these two extremes are the isomers with the symmetry D_{3d} and S_6, which have two six-member rings at the threefold axis poles of the molecule, with the other six double bonds isolated in six pentagons. The presence of various isomers in the $C_{60}H_{36}$ specimens most likely depends on the preparation method and on the kinetic parameters controlling the hydrogen addition reaction. Thus, $C_{60}H_{36}$ prepared by transfer hydrogenation of C_{60} contains a mixture of the principal isomers D_{3d} and S_6 while $C_{60}H_{36}$ prepared by zinc reduction of C_{60} in aromatic solvents contains the S_6 isomer as the most abundant [12–14]. Concerning the solid-state phase of $C_{60}H_{36}$, Hall *et al.* [9] have suggested the body-centered cubic structure (BCC) with the cell parameter 11.785 ± 0.015 Å for the packing of the molecules in the crystalline state. Furthermore, they suppose, at least for the D_{3d} isomer, that the BCC crystal structure transforms into a body-centered tetragonal one at low temperatures.

¶ This article was submitted by the authors in English.

The optical characterization of hydrofullerenes is of great importance and Raman scattering was successfully used for studying the vibrational spectrum of $C_{60}H_{36}$ prepared by the transfer hydrogenation method [12]. The incorporation of 36 hydrogen atoms in the C_{60} cage lowers the molecular symmetry and activates Raman scattering from a variety of initially forbidden phonon modes. In addition, the appearance of the C–H stretching and bending modes and those related to various isomers of $C_{60}H_{36}$ results in a very rich Raman spectrum [12]. The comparison of the phonon frequencies for five principal isomers of $C_{60}H_{36}$ obtained by molecular dynamics calculations with the experimentally observed phonon frequencies has led to the conclusion that the material prepared by the transfer hydrogenation method mainly contains two isomers, those with the symmetries D_{3d} and S_6 [12].

In this paper, we study the optical Raman and PL spectra of the hydro- and deuterofullerene $C_{60}H_{36}$ and $C_{60}D_{36}$, respectively, prepared by high-pressure hydrogenation. The aim was to identify the phonon and electron energy spectra of the high-pressure hydrogenated fullerene, to clarify the isomer composition and homogeneity of samples, and to study the isotopic effects in the vibrational spectra. The Raman spectra of the high-pressure hydrogenated samples were compared with those obtained by transfer hydrogenation and with the molecular dynamics calculation data [12]. The Raman data show the presence of all principal isomers in the high-pressure hydrogenated fullerenes and a large isotopic shift for the C–H stretching mode, whereas the shift of the modes related to the fullerene molecular cage is negligible. We have also studied the pressure behavior of the Raman and PL spectra of $C_{60}H_{36}$ at pressure up to 12 GPa in order to obtain information about the structural and chemical stability of the material at high pressure. The incorporation of hydrogen in the fullerene molecular cage may play an important role in the stability of the material, in particular, it may prevent the pressure-induced polymerization that is typical of pristine C_{60} under high-pressure and high-temperature treatment [15]. The incorporation of hydrogen may also affect the pressure-induced phase transition of the rotational disorder–order nature analogously to the case of pristine C_{60}. We have studied the pressure behavior of phonon frequencies and the pressure-induced shift of electronic bands. The pressure coefficients of the phonon modes are positive and demonstrate singularities at approximately 0.6 and 6 GPa. The pressure shift of the luminescence spectrum is unusually small and increases somehow at $P \geq 6$ GPa. All the observed features are reversible with pressure, and $C_{60}H_{36}$ is stable in the pressure region investigated.

2. EXPERIMENTAL

The commercial material, C_{60} of 99.99% purity, was sublimed twice in vacuum better than 10^{-5} Torr at 800 K and was then compacted into pellets 12 mm in diameter and 1 mm in thickness. Each pellet was placed into a copper capsule, covered with a disc of 0.01-mm thick Pd foil, and then annealed in vacuum at 620 K for 2 h to expel desorbed gases. The remaining volume in the capsule was filled with AlH_3 or AlD_3 for hydrogenation or deuteration, respectively, and tightly plugged with a copper lid using gallium as a solder. This encapsulation procedure effectively prevents hydrogen or deuterium losses during subsequent treatment, because both Cu and Ga are largely impermeable to hydrogen.

The assembled capsules were pressurized to 3.0 GPa in a toroid-type high-pressure cell and maintained at 650 ± 10 K or 700 ± 10 K for a time of 24 or 48 h. AlH_3 decomposes above 400 K [16], producing hydrogen that reacts with the fullerite after permeating the Pd foil, which isolates the fullerite from chemically active Al. The amount of the hydrogen gas produced inside the capsule corresponds to a particle ratio of $H/C_{60} \approx 90$. Therefore, the available hydrogen quantity is always in excess of the number of C_{60} particles during the hydrogenation experiments. The hydrogenation procedure was repeated for a second run, with the product of the first run taken as the starting material for the second run. Preliminary mass-spectrometry data show that at least 95% of the material in the capsule is hydrofullerene $C_{60}H_{36}$, while the remaining 5% contains partially hydrogenated fullerenes. The X-ray analysis of the obtained material shows that it has the BCC structure, typical of $C_{60}H_{36}$ [9], with a lattice parameter of 11.83 Å.

For optical measurements, visually uniform, colorless, and transparent specimens were selected. Raman spectra were recorded using a DILOR XY-500 triple monochromator equipped with a CCD liquid-nitrogen-cooled detector system. The spectra were taken in the backscattering geometry by the use of the micro-Raman system comprising an OLYMPUS microscope equipped with an MSPlan100 objective with magnification 100 and spatial resolution approximately 1.7 μm. The spectral width of the system was approximately 2.0 cm^{-1}. The Raman frequencies were calibrated by the use of the low-pressure Ne lamp with the accuracy better than 0.2 cm^{-1}. To avoid interference from luminescence, the sample was excited by the 676.4 nm line of the Kr^+ laser, whose energy is below the fundamental absorption gap of the material. The laser power was varied from 5 to 10 mW, measured directly before the sample, to avoid the destruction of the samples by laser heating. The phonon frequencies were obtained by fitting Gaussian line shapes to the experimental Raman spectra. The PL spectra were recorded using a single JOBIN YVON THR-1000 monochromator equipped with a CCD liquid-nitrogen-cooled detector system. The 457.9 nm line of an Ar^+ laser was used for excitation of the luminescence spectra. The laser power was about 2 mW measured directly in front of the high-pressure cell; the spectral width of

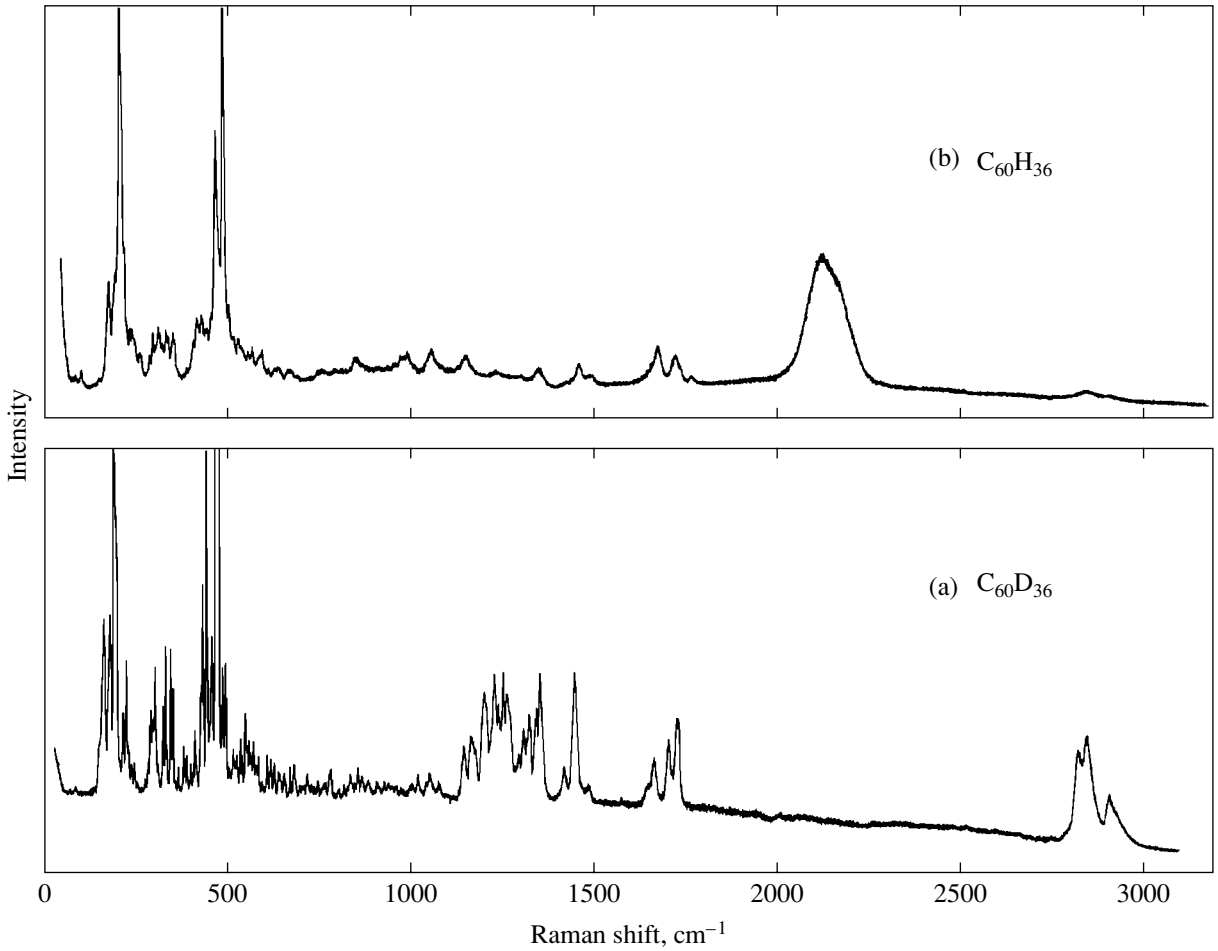


Fig. 1. Raman spectra of the $C_{60}H_{36}$ (a) and $C_{60}D_{36}$ (b) samples taken at ambient conditions in the energy range 50–3100 cm^{-1}. The samples were prepared by means of high-pressure hydrogenation at P = 3.0 GPa and T = 700 K over a reaction time of about 24 h.

the system was approximately 1.0 cm^{-1}. The measurements at high pressure were carried out using the diamond anvil cell (DAC) of the Mao–Bell type [17]. A 4 : 1 methanol–ethanol mixture was used as a pressure-transmitting medium, and the ruby fluorescence technique was used for pressure calibration [18]. The samples used for the high-pressure measurements had typical dimensions of approximately 100 μm.

3. RESULTS AND DISCUSSION

3.1. Isomeric Composition and Isotopic Effect in $C_{60}H_{36}$

The Raman spectra of $C_{60}H_{36}$ and $C_{60}D_{36}$ taken in the frequency region 50–3150 cm^{-1} at ambient conditions are shown in Fig. 1a and Fig. 1b, respectively. Both samples were synthesized at the pressure 3.0 GPa, temperature 700 K and reaction time approximately 24 hours. The spectrum in Fig. 1a consists of 126 sharp peaks with the lowest mode located at 86 cm^{-1} and the highest at 2912 cm^{-1}. For comparison, the Raman spectrum of pristine C_{60} contains only ten active modes, $H_g(1) - H_g(8)$ and $A_g(1) - A_g(2)$, with frequencies located in the region 273–1726 cm^{-1} [19]. The spectrum in Fig. 1a was taken at the best site of the best sample, selected from the content of the ampule by preliminary micro-Raman probing as having the lowest background and the clearest Raman signal. A number of rather good samples, taken from the same ampule, show less intense Raman peaks with respect to the relatively large background. The higher background is probably a result of a higher concentration of structural defects and impurities in the sample under study. These impurities may be microscopic amounts of partially hydrogenated fullerenes, which fluoresce under Kr-laser excitation in the spectral region under investigation. The majority of the selected samples from the content of the ampule show a Raman signal similar to that of the best one, but there are also many samples that give a large background obscuring the structure of the Raman spectrum. The samples obtained by high-pres-

sure hydrogenation are rather nonuniform, and great care should be exercised in sample selection. Using the micro-Raman probe, we checked a number of samples from different ampules in order to examine their quality in relation to the hydrogenation parameters, namely, temperature and reaction time. The results show that the sample quality does not depend significantly on the reaction time; on the contrary, temperature drastically affects their optical quality. Samples hydrogenated at 700 K show a better optical quality, and their Raman spectra show a rich structure, well-resolved intense peaks, and relatively small background.

The Raman spectrum of $C_{60}H_{36}$ differs drastically from that of pristine C_{60}. The most important differences in their spectra are as follows:

(i) The number of the Raman-active modes increases dramatically.

(ii) The low-energy radial modes (200–600 cm^{-1}) exhibit a considerable intensity enhancement with respect to the high-energy tangential modes (1400–1700 cm^{-1}). In addition, a number of new modes appear that are related to the C–H bending (1150–1350 cm^{-1}) and stretching (2800–3000 cm^{-1}) vibrations.

The vibrational data related to the Raman spectrum of $C_{60}H_{36}$ are summarized in Table 1. The first three columns of Table 1 contain the data related to the number, position, and intensity of the Raman peaks. The next two columns contain experimental results related to the positions and intensities of the Raman peaks of $C_{60}H_{36}$ reported in previous studies [12, 13]. The last three columns of Table 1 are related to the frequency, Raman cross section, and symmetry of the Raman-active modes of various isomers of the $C_{60}H_{36}$ molecule according to calculations [12] using the modified MNDO method. The comparison of the present experimental data with those in [12, 13] shows that the Raman spectrum of high-pressure hydrogenated $C_{60}H_{36}$ is more than five times richer than that of transfer hydrogenated $C_{60}H_{36}$. The majority of the experimentally observed Raman peaks (86 out of a total of 126 peaks) are very close, to an accuracy of approximately 5 cm^{-1}, to the calculated frequencies and cross sections of the Raman-active modes (their total number is approximately 400) [12]. The peaks that are close to the calculated frequencies are assigned to all principal isomers, but a majority of them belong to isomers with the symmetries S_6, T, and D_{3d}. We emphasize that the complexity of the calculated vibrational spectrum, the large number of isomers, and the accuracy of the molecular dynamics calculations might sometimes result in an accidental agreement (disagreement) of the experimental and calculated data. The peaks that are rather far from the calculated Raman frequencies belong mainly to the low-energy radial modes of the fullerene cage and are probably related to the presence of other isomers of $C_{60}H_{36}$ in the samples under study.

The Raman spectrum of deuterofullerene $C_{60}D_{36}$ taken at ambient conditions, is shown in Fig. 1b. A first glance at the Raman spectra of hydro- and deuterofullerene indicates that they have several similarities, but an important difference is also apparent. The spectrum is less rich in structure than that of $C_{60}H_{36}$. It contains about 80 peaks, probably due to a different isomer composition of $C_{60}D_{36}$ samples. The positions of the C–D stretching modes are shifted towards lower energies with respect to the C–H stretching modes. The related Raman peaks are located in the frequency regions 2113–2209 cm^{-1} and 2827–2912 cm^{-1} for $C_{60}D_{36}$ and $C_{60}H_{36}$, respectively. A similar shift is also observed for the C–D bending modes, which are located in the frequency regions 800–1200 cm^{-1} and 1150–1350 cm^{-1} for $C_{60}D_{36}$ and $C_{60}H_{36}$, respectively. We note that the Raman spectrum of $C_{60}D_{36}$ also contains several very weak peaks near 2900 cm^{-1}, which are coincident with C–H stretching modes of $C_{60}H_{36}$. This is related to the isotopic contamination of the samples under study, which contain a small amount of $C_{60}H_{36}$ caused by initial isotopic contamination of the deuterium provider (AlD_3) used for high-pressure synthesis.

The shift of the C–H stretching and bending modes is related to the isotopic effect on the vibrational frequencies caused by substitution of hydrogen by deuterium. The isotopic shift of the stretching mode frequencies can be estimated from the formula

$$\frac{\Omega_H}{\Omega_D} \sim \left(\frac{M_{\Omega D}}{M_{\Omega H}}\right)^{1/2}, \quad (1)$$

where Ω_H and Ω_D are the respective frequencies of the $C_{60}H_{36}$ and $C_{60}D_{36}$ molecules and $M_{\Omega H}$ and $M_{\Omega D}$ are the reduced masses involved in the vibrations. The large difference between the masses of the hydrogen atom and the C_{60} molecule indicates that the C–H stretching mode is mainly related to displacements of the hydrogen atom along the bond direction, whereas the C_{60} molecule remains practically stationary. Thus, the isotopic shift of the C–H stretching mode is expected to be close to the square root of the deuterium-to-hydrogen mass ratio. The same is also expected for the isotopic shift of the C–H bending mode. The frequencies and isotopic shifts of a number of intense modes related to the fullerene cage vibrations and to the C–H stretching vibrations are tabulated in Table 2. The largest isotopic shifts $[\Omega_H/\Omega_D]^2$ were observed for the C–H stretching modes that vary in the region 1.73–1.79. These values are close to the mass ratio of deuterium and hydrogen, $M_D/M_H = 2$. The isotopic shift for the modes related to the fullerene cage vibrations is small with respect to the C–H stretching modes and varies within 1–1.032. The difference in the isotopic shifts between the stretching modes and the fullerene cage modes is related to the fact that the hydrogen atoms essentially do not partici-

Table 1. Frequencies and intensities of the observed and calculated Raman peaks in $C_{60}H_{36}$

Experiment					Theory		
present work			ref. [12]	ref. [13]	ref. [12]		
no.	cm^{-1}	intensity*	cm^{-1}	cm^{-1}	cm^{-1}	$\sigma^{\dagger}$	isomer
1	86.0	vw	85				
2	101.1	w		128			
3	165.7	m	136				
4	176.6	vs	175	180	176	7	T
5	192.3	s			194	7	D_{3d}
6	196.9	vs			198	14	T_h
7	206.6	vs		207	206	21	T
8	212.4	vs	211		214	22	T_h
9	230.6	s			229	18	D_{3d}
10	239.3	s		239			
11	245.3	m					
12	253.6	m					
13	261.3	m	264				
14	291.0	m					
15	298.7	s			294	11	T
16	305.5	s					
17	311.8	m	313				
18	317.8	s					
19	326.4	m			325	3	D_{3d}
20	339.6	s			341	14	D_{3d}
21	347.2	s			346	8	S_6
22	360.3	s			365	10	S_6
23	366.9	s			367	10	$D_{3d}(c-k)$
24	381.8	m			379	1	T
25	396.3	s	395	395	395	7	D_{3d}
26	404.3	m			404	2	D_{3d}
27	415.3	m			415	7	$D_{3d}(c-k)$
28	423.0	m			422	9	S_6
29	429.4	m			427	4	D_{3d}
30	443.0	s		444	442	3	T
31	448.4	vs	448				
32	458.6	vs		458	460	17	S_6
33	465.6	s			465	1	D_{3d}
34	473.2	s			473	1	T
35	484.4	vs	484	484	488	44	T
36	491.3	vs			496	14	S_6
37	501.9	s					
38	509.9	s			509	31	S_6
39	522.0	m			522	1	T
40	531.0	m					
41	537.0	m					
42	541.4	m			545	1	T
43	549.9	s					
44	554.9	m					

Table 1. (Contd.)

Experiment					Theory		
present work			ref. [12]	ref. [13]	ref. [12]		
no.	cm^{-1}	intensity*	cm^{-1}	cm^{-1}	cm^{-1}	$\sigma^{\dagger}$	isomer
45	565.0	s			569	14	T
46	573.1	m			570	11	$D_{3d}(c-k)$
47	577.6	m			579	30	S_6
48	580.6	m			581	7	S_6
49	585.7	s			586	6	D_{3d}
50	590.7	m					
51	596.3	m			597	52	$D_{3d}(c-k)$
52	622.6	w					
53	632.4	w			634	1	$D_{3d}(c-k)$
54	642.5	w			647	2	S_6
55	653.3	w			653	3	T
56	657.7	w					
57	662.0	w			663	1	D_{3d}
58	669.5	w			671	4	S_6
59	685.0	m					
60	696.3	w			699	11	D_{3d}
61	711.6	m					
62	720.6	m			719	4	T
63	731.6	m			731	36	D_{3d}
64	744.5	w					
65	752.3	w		753			
66	761.1	w					
67	774.2	w					
68	781.3	w					
69	791.3	m			796	2	T_h
70	795.9	m			796	4	D_{3d}
71	817.2	w					
72	832.4	w			828	1	T
73	849.5	m			852	7	S_6
74	861.7	w			860	2	T
75	869.2	m					
76	880.4	m			875	2	T
77	922.2	w			921	3	T_h
78	940.2	w			939	1	T
79	951.1	w			948	1	S_6
80	960.8	w			959	5	T_h
81	972.3	w			972	1	T
82	985.6	vw					
83	991.4	w			988	5	T_h
84	1015.1	w	1015	1008	1018	2	S_6
85	1032.4	m	1039	1039			
86	1054.1	w			1053	4	S_6
87	1064.0	m					

Table 1. (Contd.)

Experiment					Theory		
present work			ref. [12]	ref. [13]	ref. [12]		
no.	cm^{-1}	intensity*	cm^{-1}	cm^{-1}	cm^{-1}	$\sigma^{\dagger}$	isomer
88	1073.2	w			1073	2	T_h
89	1088.6	w			1087	1	S_6
90	1125.2	w			1126	1	T_h
91	1154.0	m	1154		1154	16	$D_{3d}(c-k)$
92	1173.9	m			1172	18	T
93	1181.3	m			1182	12	T
94	1189.7	m		1186	1191	2	D_{3d}
95	1207.7	m			1209	9	D_{3d}
96	1212.6	s	1212	1213	1213	44	T_h
97	1218.7	s			1217	23	S_6
98	1227.2	m			1228	5	T
99	1231.8	m			1232	1	S_6
100	1240.3	s			1238	36	T
101	1250.3	s			1252	22	$D_{3d}(c-k)$
102	1256.9	m			1258	19	S_6
103	1263.2	s		1262	1263	42	D_{3d}
104	1274.0	s	1276		1274	74	D_{3d}
105	1283.8	m			1284	12	$D_{3d}(c-k)$
106	1306.1	w			1304	2	S_6
107	1314.4	m			1313	29	T_h
108	1319.6	s			1318	22	$D_{3d}(c-k)$
109	1328.3	m			1326	33	S_6
110	1334.9	s			1330	27	S_6
111	1348.3	m					
112	1353.2	s					
113	1362.9	s			1363	9	D_{3d}
114	1369.4	m	1386		1370	11	S_6
115	1430.6	m		1402	1429	5	D_{3d}
116	1457.0	s	1462	1462			
117	1496.6	w	1508	1497			
118	1653.8	w					
119	1674.8	m			1671	401	S_6
120	1714.0	s	1714		1712	211	$D_{3d}(c-k)$
121	1724.5	w					
122	1739.4	s	1736		1739	333	$D_{3d}(c-k)$
123	2826.8	s	2830	2829			
124	2852.5	s	2853	2852	2856	55	D_{3d}
125	2911.9	s	2913	2911	2911	180	D_{3d}
126	2931.2	m				8	

Note: $D_{3d}(c-k)$ is the lowest energy isomer with D_{3d} symmetry [13].

* Intensity characterization: very weak (vw), weak (w), medium (m), strong (s), and very strong (vs).

† Raman cross section ($Å^4$/amu) [12].

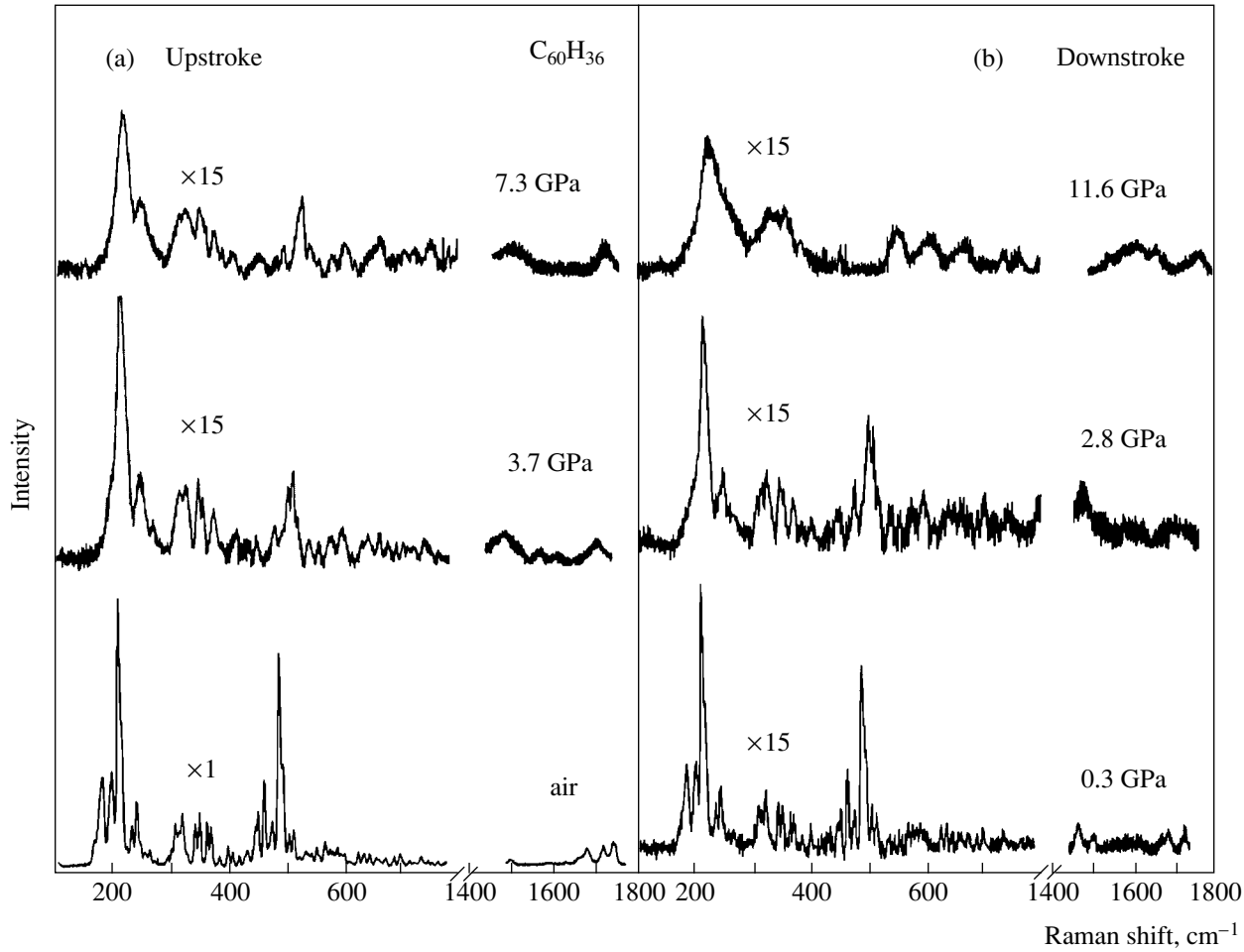


Fig. 2. Raman spectra of $C_{60}H_{36}$ recorded at room temperature during increasing (a) and decreasing (b) pressure cycle in the frequency range 100–1800 cm^{-1}. The frequency range containing the strong diamond Raman line is excluded.

pate in the cage vibrations. These results are very close to those obtained by means of surface-enhanced Raman scattering of hydrogen and deuterium chemisorbed on a diamond (100) surface [21]. The frequencies of the C–H and C–D stretching modes in [21] are 2830, 2865, 2928 cm^{-1} and 2102, 2165, 2195 cm^{-1}, respectively, and the isotopic shift varies within 1.32–1.35. Thus, the parameters of the C–H stretching modes of hydrofullerene could be similar to those of hydrogen, bonded with carbon atoms on the diamond surface. This similarity is related to the large difference between the masses of the hydrogen atom and the fullerene molecule or the carbon network in diamond. In addition, the C–H bonding in both cases takes place with sp^3 coordinated carbon atoms.

3.2. Pressure Behavior of Raman Spectra and Stability of $C_{60}H_{36}$ at High Pressure

The Raman spectra of $C_{60}H_{36}$ recorded in the regions 100–800 cm^{-1} and 1400–1750 cm^{-1} at room temperature and high pressure are shown in Fig. 2 for increasing (a) and decreasing (b) pressure runs, respectively. The Raman modes appearing in this frequency

Table 2. Molecular cage vibrations and C–H stretching modes with their isotopic shift ratio

$C_{60}H_{36}$		$C_{60}D_{36}$		Ω_H/Ω_D	$(\Omega_H/\Omega_D)^2$
Ω, cm^{-1}	Intensity*	Ω, cm^{-1}	Intensity*		
206.6	vs	203.4	vs	1.016	1.032
212.4	vs	209.7	vs	1.013	1.026
464.2	vs	464.2	vs	1	1
484.4	vs	484.4	vs	1	1
2826.8	s	2113.6	s	1.337	1.789
2852.5	s	2169.7	s	1.315	1.728
2911.9	s	2209.3	s	1.318	1.737

* Intensity characterization: strong (s) and very strong (vs).

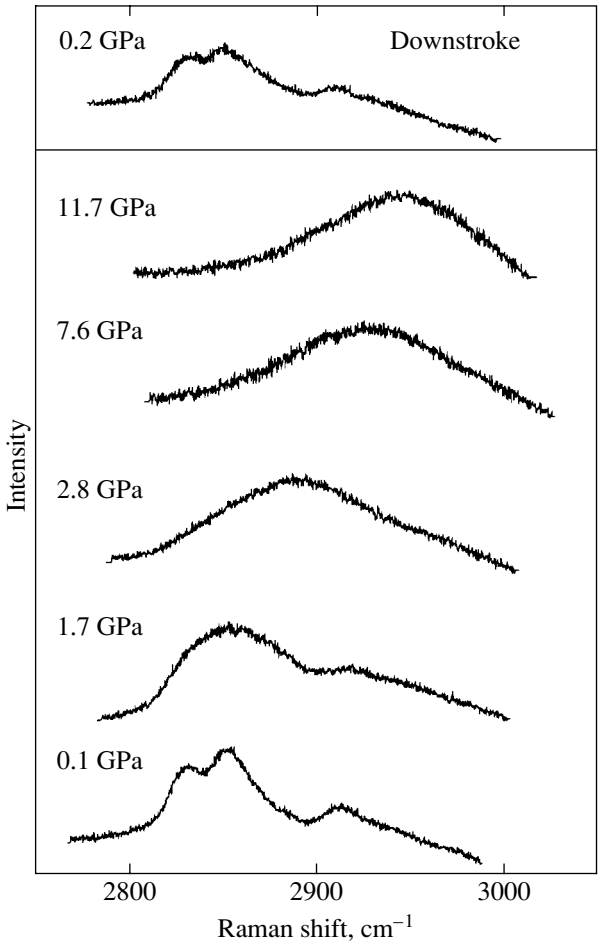


Fig. 3. Raman spectra of $C_{60}H_{36}$ recorded in the frequency range of the C–H stretching mode at ambient temperature and various pressures for increasing pressure. The spectrum in the upper frame, at 0.2 GPa, is recorded upon pressure release.

region are related mainly to the vibrations of the fullerene molecular cage. The spectral region where the strong diamond vibration is located is excluded. As can be seen from Fig. 2, a rich Raman spectrum, with well-defined structures in the low-frequency region, can easily be followed with pressure. On the contrary, the structures in the spectral region 1400–1750 cm^{-1} are broad and weak, but still they can be traced up to the highest pressure studied. At first glance, the application of pressure seems to have the expected effect in the Raman spectrum, i.e., an overall positive shift in the frequencies of the Raman modes and a relative increase in their line widths. But the situation differs significantly for the part of the Raman spectrum containing the hydrogen stretching vibrations. As can be seen in Fig. 3, where the C–H stretching vibrations are presented, the initially well-separated structures in the corresponding spectra are gradually washed out, becoming a very broad structure for pressures higher than approximately 2.0 GPa. We note that despite a considerable broadening of the Raman structures, the pressure effect is fully reversible upon releasing pressure, as can be seen from Fig. 2 and the top frame in Fig. 3.

The pressure dependence of the Raman frequencies is shown in Figs. 4 and 5 for pressure up to 12 GPa and at room temperature. The different open (closed) symbols correspond to pressure increase (decrease) in different pressure runs, while the solid lines through the experimental points represent linear least-square fittings. The shaded areas near the pressure approximately 0.6 and 6.0 GPa denote the pressure regions where the change in the slope of the pressure dependence or the disappearance of some Raman peaks occur. The parameters of linear least-square fittings and a tentative mode assignment of the observed Raman modes are compiled in Table 3. In the mode assignment column, we indicate the isomeric form to which this mode most likely belongs. The data related to pristine C_{60} are also included in Table 3 for comparison. The pressure dependence presented in Figs. 4 and 5 and numerical data in Table 1 show that all the observed modes have positive pressure coefficients and at least four Raman peaks disappear for pressures higher than approximately 6 GPa. Furthermore, the majority of Raman modes exhibit a change in the slope of their pressure dependence at approximately 0.6 and 6 GPa. These peculiarities were observed for both increasing and decreasing pressure runs, and therefore the pressure dependence of the Raman mode frequencies exhibits fully reversible behavior.

The pressure dependence of the stretching C–H vibrations, where a dramatic broadening occurs with increasing pressure, is the most striking pressure effect on the Raman spectra of $C_{60}H_{36}$. In our opinion, the abnormal broadening of the initially sharp peaks in the Raman spectra of $G_{60}H_{36}$ may be related to the presence of various isomers in the samples under study. Numerical calculations in [12] show that there are many identical Raman modes related to various isomers of $C_{60}H_{36}$ with very close frequencies. We believe that any difference in their pressure coefficients, even small, may result in additional pressure-induced broadening of these peaks, which gradually obscures the initially sharp Raman structure. A similar broadening of the Raman spectra at high pressure was also observed in the isomeric mixture of C_{84} fullerene samples [22]. It is important to note that this kind of broadening is expected to be reversible with pressure, and this behavior is indeed observed in our experiments. Bearing in mind that the number of the main C–H Raman modes of $C_{60}H_{36}$ does not change with pressure, we have fitted the experimental data in this region by keeping the same number of peaks at any pressure. The pressure behavior obtained in this way is displayed in Fig. 6, in which open (closed) circles denote increasing (decreasing) pressure runs. Despite the crudeness of the procedure, the pressure dependences of the C–H Raman peak positions show a behavior that is compatible with that

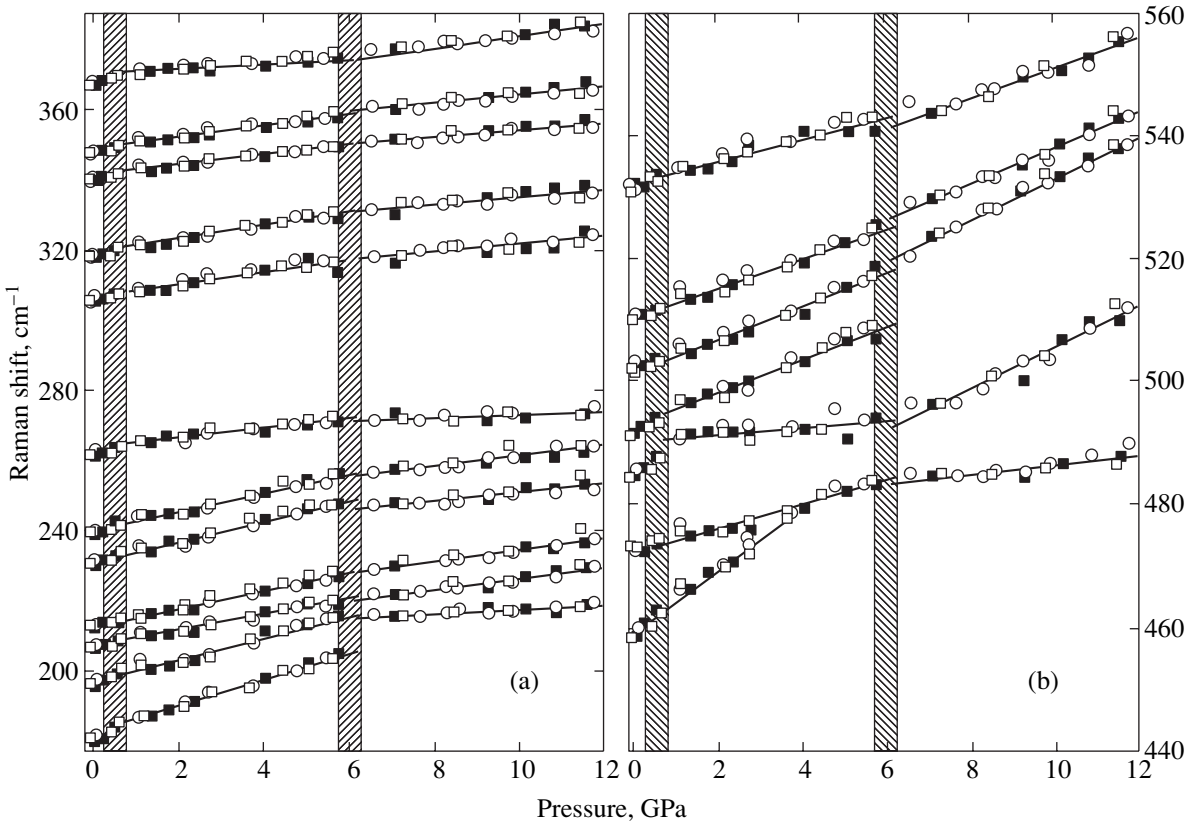


Fig. 4. Pressure dependence of Raman-active modes of $C_{60}H_{36}$ in the frequency ranges 180–390 cm^{-1} (a) and 440–560 cm^{-1} (b). Different open (closed) symbols correspond to increasing (decreasing) pressure in different pressure runs. Solid lines represent linear least-square fits. The shaded areas at $P \sim 0.6$ and approximately 6.0 GPa denote the regions of possible phase transitions.

of the Raman modes of the fullerene molecular cage up to approximately 6 GPa. In particular, these peaks exhibit an overall positive shift up to approximately 6 GPa as well as changes in the slopes of the pressure behavior of the Raman mode frequencies at approximately 0.6 GPa. However, their pressure dependence is different for pressures above 6 GPa, where a softening in the C–H stretching vibrations is observed. It is important to note that even if we follow the pressure dependence of the "center-of-gravity" frequency of the overall C–H Raman band region, we also find a change in the slope of the pressure dependence at approximately 6 GPa. This dependence is shown in Fig. 6 by solid stars, which represent both the pressure increase and decrease runs.

The observed peculiarities in the pressure dependence of the Raman modes of $C_{60}H_{36}$ may be understood by invoking the corresponding behavior of pristine C_{60} and C_{70} at high pressure. It is known that under hydrostatic pressure at room temperature, C_{60} transforms, at 0.4 GPa or at 259 K, from the FCC-structure, where the C_{60} molecules are orientationally disordered due to chaotic rotations, to the SC-structure, where the molecular rotations are partially ordered [23, 24]. Similarly, under pressure, C_{70} undergoes an orientational ordering phase transition from an FCC structure to a rhombohedral phase at approximately 0.35 GPa or at 280 K [25, 26]. Furthermore, Hall *et al.* [9] found that the distortion of the molecules brought about by hydrogenation along with the hydrogen atoms bonded around the equator gives the molecule a strongly oblate shape. The BCC-structure allows effective packing of oblate spheroids if the polar axes of the molecules are aligned. The longer second-nearest-neighbor distance then prevents close approaches of the equatorial hydrogen atoms. If this alignment actually occurs, then there should be a tendency to form a tetragonal structure with a ratio $c/a < 1$. Therefore, they predict that at sufficiently low temperatures, the BCC-structure must transform into a body-centered tetragonal structure. Thus, keeping in mind the pressure-induced orientational ordering of C_{60} and C_{70} and the finding by Hall *et al.* [9], we might speculate that the observed peculiarity at approximately 0.6 GPa in the pressure dependence of Raman modes can be assigned to an orientation ordering structural phase transition.

The changes in the pressure dependence of the Raman frequencies at approximately 6 GPa, especially the behavior of the C–H stretching modes, could be attributed to a possible phase transition in which the hydrogen bonds may be involved. It is known that the intermolecular and intramolecular distances in fullerene

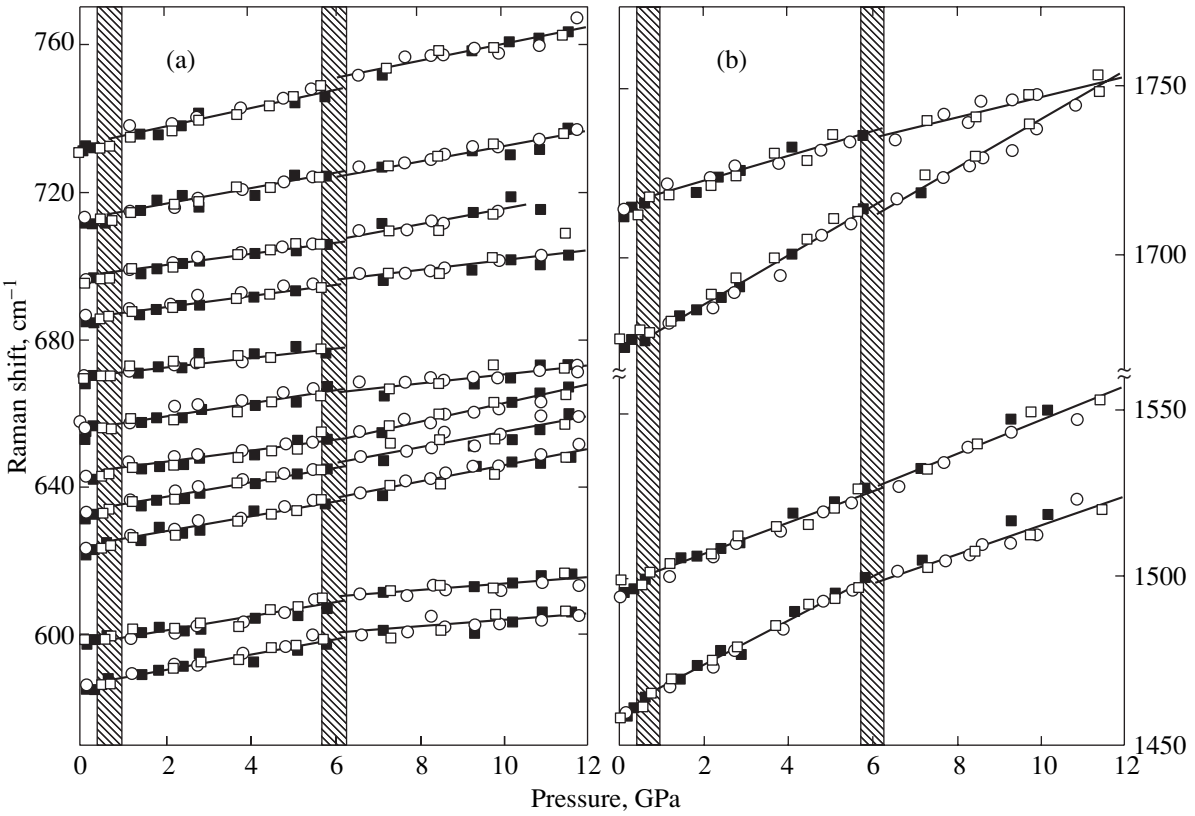


Fig. 5. Pressure dependence of Raman-active modes of $C_{60}H_{36}$ in the frequency ranges 570–770 cm^{-1} (a) and 1450-1770 cm^{-1} (b). Different open (closed) symbols correspond to increasing (decreasing) pressure in different pressure runs. Solid lines represent linear least-square fits. The shaded areas at $P \sim 0.6$ and approximately 6.0 GPa denote the regions of possible phase transitions.

become comparable at sufficiently high pressure. Therefore, the hydrogen atoms in the $C_{60}H_{36}$ molecule, which initially form terminal C–H bonds, can interact with carbon atoms belonging to neighboring molecular cages. This interaction may result in the formation of the bridging C–H–C hydrogen bonds when the decrease in the intermolecular distances becomes appropriate. As a result, the bulk modulus of the crystal increases, which is manifested by a decrease in the slope of the pressure dependence of the Raman frequencies. In addition, the formation of the bridging hydrogen bonds results in a certain elongation of the involved terminal C–H bonds. The increase in the C–H terminal bond lengths results in mode softening, which increases with a further increase of pressure. The presence of the bridging C–D–C modes in the $C_{60}D_x$ samples was found recently even at ambient pressure by means of NMR investigation of deuterofullerene [27]. Similar pressure behavior of hydrogen bonds was also found in other molecular materials that exhibit a softening in their C–H stretching mode frequency under pressure [28].

3.3. Pressure Behavior of the Photoluminescence Spectra of $C_{60}H_{36}$

The photoluminescence spectra of $C_{60}H_{36}$ at normal pressure and various temperatures are depicted in Fig. 7. The spectrum at room temperature contains two broad peaks and two shoulders, located near the lower and higher energy sides of the spectrum. The structure of the PL spectrum is reminiscent of that of pristine C_{60} at room temperature, but the intensity of luminescence is considerably higher. The onset of the spectrum is located near 2.5 eV, which is higher than the onset of the PL spectrum in pristine C_{60} by approximately 1 eV [29]. As the temperature drops to 80 K and below to the liquid helium temperature, the PL spectrum becomes more resolved, as shown in Figs. 7b and 7c. The PL spectrum at the lowest temperature (10 K) contains a number of sharp and well-resolved peaks, located near its onset. The relatively high PL intensity in $C_{60}H_{36}$, in comparison to pristine C_{60}, is mainly associated with small luminescence quantum yield in pristine C_{60} due to the dipole-forbidden transitions from the lowest exited state [30]. The fine structure of the PL spectrum in $C_{60}H_{36}$ at low temperature resembles the well-developed structure in the PL spectrum of the C_{60} single crystals at the liquid helium temperature. This structure is related to the shallow defect levels [31], while the fine structure of the PL spectrum in the $C_{60}H_{36}$ may be related to the abundant isomer composition of the samples under study.

Table 3. The phonon frequency, assignment and their pressure coefficients for the $C_{60}H_{36}$ Raman-active modes. The corresponding values for C_{60} are also included for comparison

Mode	$C_{60}H_{36}$				C_{60}^a	
		0–0.6 GPa	0.6–6 GPa	6–12 GPa	0.4–2.4 GPa	
	ω, cm^{-1}	$\partial\omega_i/\partial P$, cm^{-1}/GPa	$\partial\omega_i/\partial P$, cm^{-1}/GPa	$\partial\omega_i/\partial P$, cm^{-1}/GPa	ω, cm^{-1}	$\partial\omega_i/\partial P$, cm^{-1}/GPa
$T_g(T)$	180.6	5.6 ± 0.1	6.4 ± 0.1			
$E_g(T_h)$	196.0	4.7 ± 0.6	3.1 ± 0.4	0.6 ± 0.1		
$E_g(T)$	206.6	2.7 ± 0.3	2.2 ± 0.1	1.6 ± 0.1		
$T_g(T_h)$	212.6	2.3 ± 0.3	2.6 ± 0.1	1.7 ± 0.2		
$E_g(D_{3d})$	230.6	3.5 ± 0.1	2.9 ± 0.1	1.2 ± 0.2		
	239.1	3.2 ± 0.2	2.6 ± 0.2	1.5 ± 0.2		
	261.1	2.1 ± 0.1	1.3 ± 0.1	0.4 ± 0.2	$272H_g(1)$	3.2
	305.2	2.2 ± 0.2	1.8 ± 0.2	1.2 ± 0.2	$294\omega_1$	2.5
	317.8	3.1 ± 0.1	1.9 ± 0.1	1.0 ± 0.2		
$E_g(D_{3d})$	339.6	2.1 ± 0.1	1.4 ± 0.1	0.9 ± 0.1		
$A(S_6)$	347.1	2.7 ± 0.1	1.6 ± 0.1	1.1 ± 0.1	$345\omega_2$	2.9
$E_g(T_h)$	366.7	3.2 ± 0.2	0.5 ± 0.2	1.7 ± 0.3		
$A_g(T)$	484.2	4.6 ± 0.2	0.5 ± 0.2	3.3 ± 0.2		
$A_g(S_6)$	491.0	3.3 ± 0.1	2.7 ± 0.1		$495H_g(2)$	4.2
	501.8	2.0 ± 0.1	2.7 ± 0.1	3.3 ± 0.1		
$E_g(S_6)$	530.6	2.7 ± 0.2	1.7 ± 0.2	2.4 ± 0.2	$522\omega_5$	1.0
$E_g(D_{3d})$	585.7	2.6 ± 0.2	2.0 ± 0.2	0.9 ± 0.2		
$E_g(S_6)$	598.6	2.3 ± 0.2	1.9 ± 0.2	0.9 ± 0.2		
$T_g(T_h)$	622.6	2.9 ± 0.2	2.0 ± 0.2	2.3 ± 0.2	$624\omega_1$	1.5
$E_g(D_{3d})$	632.5	2.4 ± 0.1	2.0 ± 0.1	2.1 ± 0.3		
$E_g(D_{3d})$	642.5	1.9 ± 0.2	1.5 ± 0.2	2.5 ± 0.2		
$T_g(T)$	655.8	1.2 ± 0.2	1.7 ± 0.2	1.2 ± 0.2		
$E_g(T_h)$	669.3	1.9 ± 0.1	1.3 ± 0.1			
$A_{1g}(D_{3d})$	695.6	1.8 ± 0.1	1.6 ± 0.1	2.0 ± 0.4		
$A_{1g}(D_{3d})$	730.9	1.4 ± 0.2	2.5 ± 0.2	2.3 ± 0.2	$729\omega_8$	–2.9
$T_g(T)$	1459.5	9.0 ± 0.2	6.6 ± 0.2	4.4 ± 0.4	$1467A_g(2)$	5.5
$E_g(S_6)$?	1494.5	9.6 ± 0.3	4.8 ± 0.3	5.1 ± 0.4		
$A_g(S_6)$	1674.6	4.6 ± 0.3	7.2 ± 0.3	6.9 ± 0.4		
$A_g(T_h)$	1712.5	5.2 ± 0.3	3.6 ± 0.3	2.8 ± 0.5		

Note: Data taken from [35].

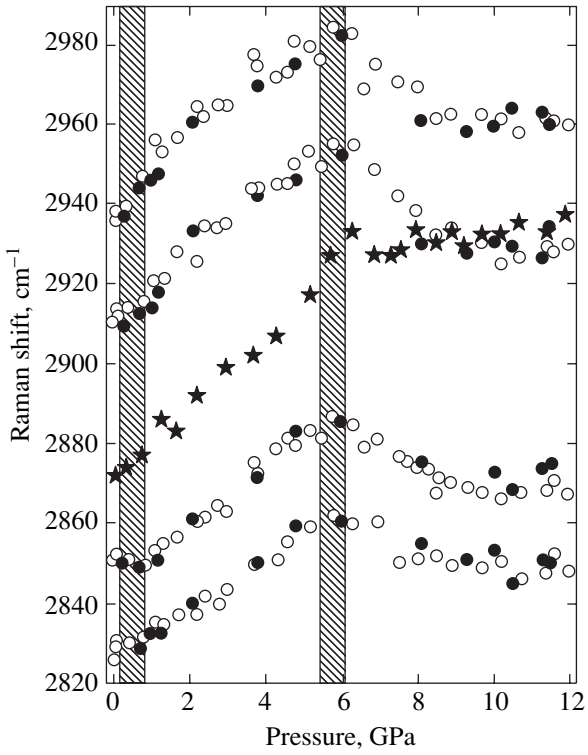


Fig. 6. Pressure dependence of the Raman-active modes of $C_{60}H_{36}$ in the range of the C–H stretching modes. Open (closed) circles correspond to the increasing (decreasing) pressure runs, while the peak positions were obtained by fitting Gaussians lineshapes to four experimental peaks at any pressure. The stars correspond to the frequency of the "center of gravity" of the C–H Raman band as a function of pressure.

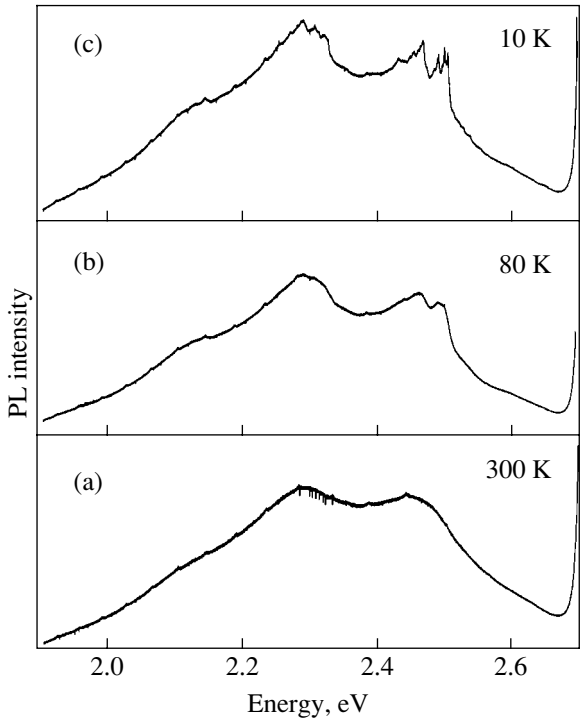


Fig. 7. Photoluminescence spectra of $C_{60}H_{36}$ at ambient pressure and various temperatures.

According to numerical local-density functional calculations of the electronic structure, the gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) is different for the five most stable isomers and varies between 3.84 and 3.91 eV [32]. These calculations are related to the isolated $C_{60}H_{36}$ molecule and their results should be compared with the experimental absorption spectra of the $C_{60}H_{36}$ solutions. The absorption spectra of the $C_{60}H_{36}$ solutions in methylene dichloride and benzene [13] show that the onset of the absorption spectrum is located between 2.96 and 3.24 eV. Concerning the HOMO–LUMO gap calculations for the solid $C_{60}H_{36}$, there are no available data to compare with the present experimental results. As is known, the calculations of the HOMO–LUMO gap for various molecular solids show that the gap is shifted to lower energies with respect to the isolated molecules due to the vapor-crystal shift of the electron energy spectrum. In any case, the onset of the photoluminescence spectrum in solid $C_{60}H_{36}$ observed in the present work is considerably smaller than the HOMO–LUMO gap for an isolated molecule [32].

The PL spectra of $C_{60}H_{36}$ taken at room temperature and various pressures are shown in Fig. 8. The initial spectrum taken at 0.5 GPa gradually broadens upon increase in pressure up to 12 GPa, while the shift of the spectrum is not apparent. The pressure-induced shift of the spectrum is unusually small as it follows from the pressure dependence of the band frequencies obtained by fitting Gaussian band shapes to the main peaks in the PL spectrum. The pressure dependence of the two main bands of the PL spectrum of $C_{60}H_{36}$ is shown in Fig. 9. The pressure coefficients for these bands are close to zero at pressure up to approximately 6.5 GPa, but they increase in the absolute value at higher pressures, to –7.5 and –9 meV/GPa, respectively. The pressure behavior of the electronic states in $C_{60}H_{36}$ is not typical of molecular crystals, whose electronic levels usually exhibit large negative pressure shifts, rapidly decreasing with pressure [33]. It is known that the pressure-induced shift of the electronic levels in molecular crystals is negative for the majority of the materials that have a center of symmetry, whereas it may be positive for materials in which the molecules do not have a center of symmetry [34]. The samples under study contain the *T* isomer in abundance, which does not have a center of symmetry. This means that at least a part of the PL spectrum related to this isomer may have a positive pressure-induced shift, while at the same time, we have a negative pressure-induced shift originating from isomers having a center of symmetry. The unusual pressure behavior up to 6.5 GPa may be associated with mutual compensation of the opposite shifts from the two parts of the luminescence spectrum, originating from electronic states of various isomers. At higher pressure, however, the luminescence related to isomers that have a center of symmetry dominates: their elec-

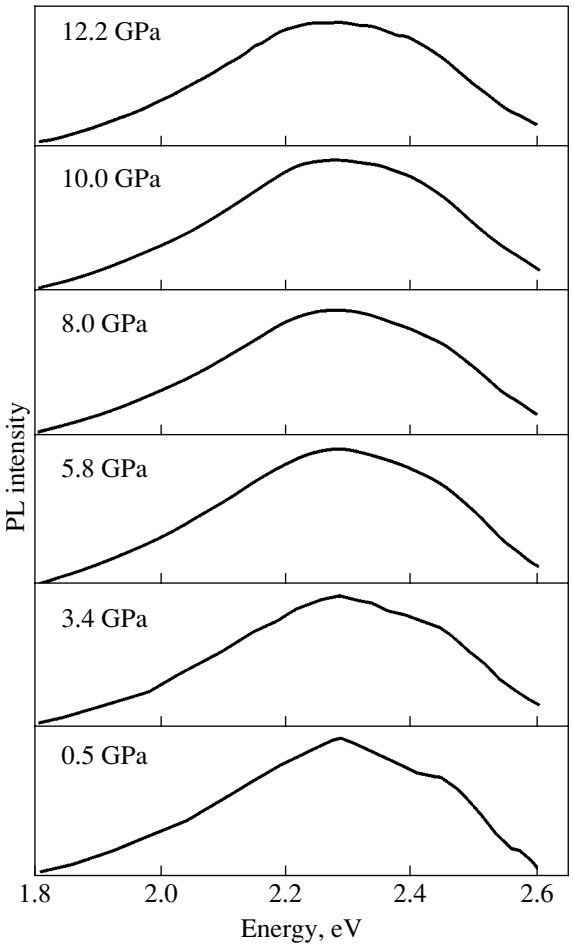


Fig. 8. Photoluminescence spectra of $C_{60}H_{36}$ at ambient temperature and various pressures.

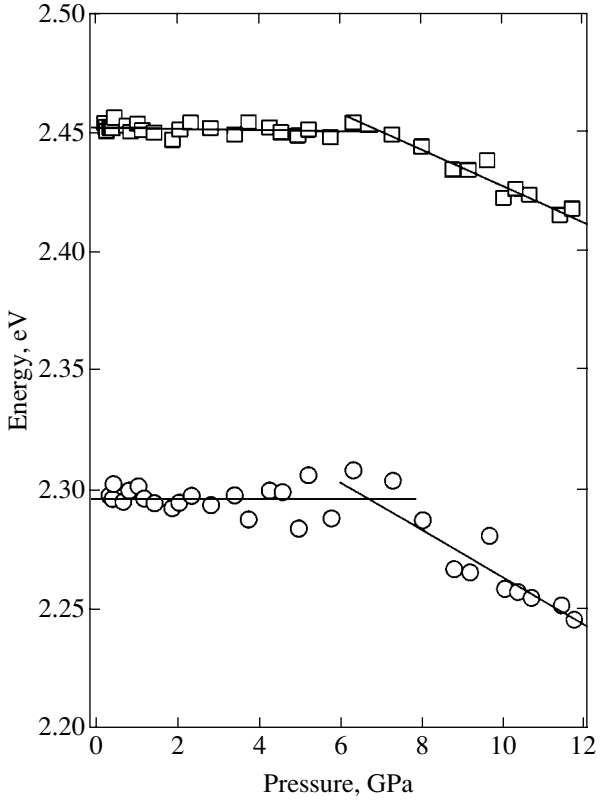


Fig. 9. Pressure dependence of the two main bands in the PL spectrum of $C_{60}H_{36}$.

tronic states are downshifted in energy to the lowest positions, and we therefore have an overall negative pressure shift.

4. CONCLUSIONS

The Raman spectrum of hydrofullerene $C_{60}H_{36}$ prepared by high-pressure hydrogenation has a very rich structure and contains about five times more peaks than that of transfer hydrogenated $C_{60}H_{36}$. The comparison of experimental Raman peaks with the results of calculation of the molecular dynamics shows the presence of five principal isomers in the samples under study. The majority of the experimentally observed Raman peaks belong to the S_6, T, and D_{3d} isomers. The micro-Raman probing of several samples, prepared under different reaction parameters, shows that the homogeneity of the samples depends strongly on the reaction temperature. The Raman spectrum of deutero-fullerene $C_{60}D_{36}$ prepared by the same method is in general similar to that of $C_{60}H_{36}$. The important difference between the two spectra is a large isotopic shift of the C–D stretching modes with respect to the corresponding C–H ones. The isotopic shift in the modes associated with the fullerene molecular cage is very small. This is a strong indication that the hydrogen atoms do not play an important role in the $C_{60}H_{36}$ molecular cage vibrations.

The pressure behavior of the optical Raman and PL spectra of $C_{60}H_{36}$ is not typical of fullerene-based materials because they become rather diffuse, even at relatively small pressure. The pressure dependence of the phonon frequencies is reversible with pressure and exhibits peculiarities at approximately 0.6 and 6 GPa. The first peculiarity is probably related to a phase transition from the initial orientationally disordered BCC-structure to an orientationally ordered structure. The peculiarity at approximately 6 GPa may be related to a pressure-driven enhancement of the C–H interaction between the hydrogen and carbon atoms belonging to neighboring molecular cages.

The PL spectrum of $C_{60}H_{36}$ is shifted to higher energy by about 1 eV with respect to that of pristine C_{60}. The spectrum at room temperature consists of two broad peaks and becomes more structured at 10 K. The pressure-induced shift of the PL spectrum of $C_{60}H_{36}$ is close to zero up to 6.5 GPa, while at higher pressure, a negative pressure shift was observed. The unusual pressure behavior of the PL spectrum is related to the isomer composition of the high-pressure hydrogenated fullerene samples.

ACKNOWLEDGMENTS

The authors thank S. Assimopoulos for assistance and I. 0. Bashkin for providing the $C_{60}H_{36}$ samples. K.P.M. acknowledges the hospitality of the Aristotle University of Thessaloniki. Support by the Russian Foundation for Basic Research, project no. 03-02-16011, and by the Russian State Research Program "Physical Properties of Carbon-based Nanostructures and Development of New Electronic Devices," grant no. 541-02, is gratefully acknowledged.

REFERENCES

1. A. C. Dillon, K. M. Jones, T. A. Bekkedahl, *et al.*, Nature **386**, 377 (1997).
2. A. Rathna and J. Chandrasekhar, Chem. Phys. Lett. **206**, 217 (1993).
3. L. D. Book and G. E. Scuseria, J. Phys. Chem. **98**, 4283 (1994).
4. M. I. Attalla, A. M. Vassallo, B. N. Tattam, and J. Hamma, J. Phys. Chem. **97**, 6329 (1993).
5. I. O. Baskin, A. I. Kolesnikov, V. E. Antonov, *et al.*, Mol. Mater. **10**, 265 (1998).
6. R. A. Assink, J. E. Schirber, D. A. Loy, *et al.*, J. Mater. Res. **7**, 2136 (1992).
7. R. E. Haufler, J. Conceicao, P. Chibante, *et al.*, J. Phys. Chem. **94**, 8634 (1990).
8. C. Jin, R. Hettich, R. Compton, *et al.*, J. Phys. Chem. **98**, 4215 (1994).
9. L. E. Hall, D. R. McKenzie, M. I. Attalla, *et al.*, J. Phys. Chem. **97**, 5741 (1993).
10. R. L. Davis, D. R. McKenzie, L. E. Hall, *et al.*, ISIS Annual Report, Vol. 2: Rutherford Appleton Laboratory Report RAL-94-050 (1994), p. A 200.
11. M. Bühl, W. Thiel, and U. Schneider, J. Am. Chem. Soc. **117**, 4623 (1995).
12. R. Bini, J. Ebenhoch, M. Fanti, *et al.*, Chem. Phys. **232**, 75 (1998).
13. R. V. Bensasson, T. J. Hill, E. J. Land, *et al.*, Chem. Phys. **215**, 111 (1997).
14. A. D. Darwish, A. K. Abdul-Sada, G. J. Langley, *et al.*, J. Chem. Soc., Perkin Trans. 2, 2359 (1995).
15. Y. Iwasa, T. Arima, R. M. Fleming, *et al.*, Science **264**, 1570 (1994).
16. S. K. Konovalov and B. M. Bulychev, Zh. Neorg. Khim. **37**, 2640 (1992).
17. A. Jayaraman, Rev. Sci. Instrum. **57**, 1013 (1986).
18. D. Barnett, S. Block, and G. J. Piermarini, Rev. Sci. Instrum. **44**, 1 (1973).
19. D. S. Bethune, G. Meijer, W. C. Tang, *et al.*, Chem. Phys. Lett. **179**, 181 (1991).
20. B. W. Clare and D. L. Kepert, J. Mol. Struct.: THEOCHEM **315**, 71 (1994).
21. K. Ushizawa, M. N.-Gamo, Y. Kikuchi, *et al.*, Phys. Rev. B **60**, R5165 (1999).
22. I. Margiolaki, S. Margadona, K. Prassides, *et al.*, Physica B (Amsterdam) **318**, 372 (2002).
23. G. A. Samara, J. E. Schirber, B. Morosin, *et al.*, Phys. Rev. Lett. **67**, 3136 (1991).
24. K. P. Meletov, G. Kourouklis, D. Christofilos, and S. Ves, Phys. Rev. B **52**, 10090 (1995).
25. C. Christides, I. M. Thomas, T. J. Dennis, and K. Prassides, Europhys. Lett. **22**, 611 (1993).
26. A. Lundin, A. Soldadov, and B. Sunquist, Europhys. Lett. **30**, 469 (1995).
27. V. P. Tarasov, Y. B. Muravlev, V. N. Fokin, and Y. M. Shulga, Appl. Phys. A **78**, 1001 (2004).
28. E. Katoh, H. Yamawaki, H. Fujihisa, *et al.*, Phys. Rev. B **59**, 11244 (1999).
29. K. P. Meletov, D. Christofilos, S. Ves, and G. A. Kourouklis, Phys. Status Solidi B **198**, 553 (1996).
30. P. A. Lane, L. S. Swanson, Q.-X. Ni, *et al.*, Phys. Rev. Lett. **68**, 887 (1992).
31. W. Guss, J. Feldman, E. O. Göbel, *et al.*, Phys. Rev. Lett. **72**, 2644 (1994).
32. B. I. Dunlap, D. W. Brenner, and G. W. Schriver, J. Phys. Chem. **98**, 1756 (1994).
33. K. P. Meletov, V. K. Dolganov, O. V. Zharikov, *et al.*, J. Phys. I **2**, 2097 (1992).
34. A. S. Davydov, *Theory of Molecular Excitons* (Nauka, Moscow, 1968; Plenum, New York, 1971).
35. K. P. Meletov, G. A. Kourouklis, D. Christophilos, and S. Ves, JETP **81**, 798 (1995).

Journal of Experimental and Theoretical Physics, Vol. 100, No. 4, 2005, pp. 775–783.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 4, 2005, pp. 877–885.

ELECTRONIC PROPERTIES OF SOLIDS

The Coordinate of a Singular Point of the Time Correlation Functions for a Heteronuclear Spin System of a Crystal

V. E. Zobov[a,*] and M. A. Popov[b]

[a]*Kirenskiĭ Institute of Physics, Siberian Division, Russian Academy of Sciences, Akademgorodok, Krasnoyarsk, 660036 Russia*

**e-mail: rsa@iph.krasn.ru*

[b]*Krasnoyarsk State University, Krasnoyarsk, 660041 Russia*

Received November 5, 2004

Abstract—The singularities of the time autocorrelation functions (ACFs) for a heteronuclear spin system of a crystal are investigated. Exact expressions are obtained for ten moments of the spectra of ACFs in the approximation of a self-consistent fluctuating field (SCFF) with arbitrary axial symmetry. These expressions are applied to determine the coordinate of the lowest singular point of these functions on the imaginary-time axis for a spin system with a dipole–dipole interaction (DDI). The leading corrections to this coordinate due to the correlation of local fields in real crystals are calculated. These corrections are determined by lattice sums with triangles of four bonds and pairs of four bonds. Numerical values of the coordinate are obtained for a LiF crystal in a magnetic field directed along three crystallographic axes. An increase in the coordinate of the singular point, which follows from the theory and leads to a faster falloff of the wings of the ACF spectra, qualitatively agrees with experiment. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Nuclear magnetic systems with controllable states whose variations can be observed by an NMR method [1] are of great interest for nonequilibrium statistical physics. Heteronuclear systems (i.e., systems that consist of nuclei with different Larmor frequencies in a strong magnetic field) open up new possibilities for studying relaxation processes. This fact was demonstrated in [2] with an example of cross relaxation between subsystems. It is also important that heteronuclear systems are encountered more frequently than homonuclear systems and that there are many methods developed for studying them; the informativeness of these methods depends on the development level of the theory. Finally, systems consisting of nuclei with different resonance frequencies have recently attracted the attention of researchers in relation to the problem of addressing spins in quantum computation [3].

The modern dynamic theory of dense spin systems in solids is based on the concept of a time-fluctuating random local magnetic field [4–9] whose properties are close to those of a Gaussian random field. In heteronuclear systems, this field has several (according to the number of different types of nuclei) components. This fact complicates the construction of a theory. The use of the self-consistency conditions [5] for the time correlation functions of the field and the spins has made it possible to construct a theory [10, 11] that qualitatively explains many experimental data. This theory implies that the spin correlation functions have singularities on the imaginary-time axis that are responsible for the exponential wings of the spectra of these functions, which are observed by magnetic-resonance methods. An important consequence of this result is that the wings of the spectra of different types of nuclei are universal because the coordinates of the singular points of the time correlation functions, which are coupled due to the interaction, must coincide. A comparison with experiment has shown that the correlation of local fields weakens their fluctuation; in particular, it increases the coordinate of the singular point. In [11], such a correlation was taken into account phenomenologically.

In the present paper, we develop a microscopic approach to the calculation of correlation phenomena in the theory of a self-consistent fluctuating field (SCFF). The validity of this approach has recently been demonstrated by a simpler example of a homonuclear system [12]. First of all, we obtain general expressions for moments up to the tenth order inclusive after generalizing a diagrammatic series for the memory function [6] to the heteronuclear case. Then, based on these moments, we calculate the coordinate of the lowest singular point of the correlation functions on the imaginary-time axis. Finally, we determine a correction to the moments due to the correlation of local fields and apply them to calculate the corresponding shift in the coordinate of the singular point.

1063-7761/05/10004-0775$26.00

2. EQUATIONS FOR AUTOCORRELATION FUNCTIONS

Equations for the autocorrelation functions (ACFs) of a spin precessing in an anisotropic Gaussian random field were derived in [6]. If we take into account that, in the case of a heteronuclear spin system, such equations should be written out for spins of each type, we arrive at the system of integral equations

$$\frac{d}{dt_1}\Gamma_{\alpha q}(t) = -\int_0^t G_{\alpha q}(t-t_1)\Gamma_{\alpha q}(t_1)dt, \qquad (1)$$

where the subscript α denotes the spin projections x, y, and z and q enumerates the subsystems. The memory functions $G_{\alpha q}(t)$ are represented as series in irreducible dressed skeleton diagrams with increasing number of vertices:

$$G_{\alpha q}(t) = \sum_{m=1}^{\infty} G_{\alpha q}^{2m}(t). \qquad (2)$$

For the case of a field with arbitrary anisotropy in a homonuclear system, all diagrams with 2, 4, 6, and 8 vertices are presented in [6]. In the heteronuclear case, the form of a diagram remains the same; however, in the explicit expressions for these diagrams, one should associate with the zz lines (dashed lines) a sum over contributions to the longitudinal field rather than a single term:

$$g_{zq}(t) = \sum_p \Delta_{qp}^2 \Gamma_{zp}(t). \qquad (3)$$

The xx and yy lines correspond to a single contribution as before, but this contribution is different for different types of nuclei:

$$g_{xq}(t) = \Delta_{xq}^2 \Gamma_{xq}(t), \quad g_{yq}(t) = \Delta_{yq}^2 \Gamma_{yq}(t).$$

Here, we express the correlation functions of a Gaussian random field in terms of the time-dependent spin ACFs $\Gamma_{xq}(t)$, $\Gamma_{yq}(t)$, and $\Gamma_{zq}(t)$ in a self-consistent way. At high temperatures, the ACF of the α component of the spin located at site i of the lattice is given by

$$\Gamma_{\alpha q}(t) = \frac{\mathrm{Sp}\{\exp(i\mathcal{H}t)I_{i\alpha}^{(q)}\exp(-i\mathcal{H}t)I_{i\alpha}^{(q)}\}}{\mathrm{Sp}\{(I_{i\alpha}^{(q)})^2\}}, \qquad (4)$$

where $\mathcal{H}$ is the Hamiltonian of the secular part of the dipole–dipole interaction (DDI) [1, 2] in a strong magnetic field. The mean squares of different contributions to the longitudinal field are

$$\Delta_{qp}^2 = I^{(p)}(I^{(p)}+1)\frac{4}{3}\sum_k b_{ik}^{(qp)^2},$$
$$b_{ij}^{(qp)} = \frac{\gamma_q\gamma_p\hbar}{2r_{ij}^3}(1-3\cos^2\theta_{ij}), \qquad (5)$$

where θ_{ij} is the angle between the internuclear vector $\mathbf{r}_{ij}$ and the direction of the static magnetic field. In the axially symmetric case, we have

$$\Delta_{xq}^2 = \Delta_{yq}^2 = \Delta_{qq}^2/4. \qquad (6)$$

Each term of series (2) is expressed in terms of a multiple time integral of the products of ACFs [6].

Table 1. Exact values of the ACF moments in a LiF crystal in the SCFF approximation when the magnetic field is applied along three crystallographic axes ($X_{2n}^{(q)} = M_{2nX}^{(q)}\Delta_{\mathrm{FF}}^{-2n}$ and $Z_{2n}^{(q)} = M_{2nZ}^{(q)}\Delta_{\mathrm{FF}}^{-2n}$)

	$H_0 \parallel [111]$		$H_0 \parallel [110]$		$H_0 \parallel [100]$	
	F	Li	F	Li	F	Li
X_2	1.5565	0.243175	3.2059	0.573075	10.9709	2.126295
X_4	7.2621246	0.2062389	30.122726	1.1550313	357.05369	14.396791
X_6	60.240334	0.4462115	476.06509	5.7148923	19329.186	193.62608
X_8	785.97935	2.6795506	10916.055	73.047633	1469814.4	5748.5288
X_{10}	15296.507	36.221363	342734.27	2018.0911	144788238	429421.26
Z_2	0.5	0.07275	0.5	0.07275	0.5	0.07275
Z_4	1.9315	0.0433208	3.5809	0.0913213	11.3459	0.3173148
Z_6	17.917389	0.0669163	67.435923	0.3356018	739.17481	4.1540520
Z_8	277.83143	0.1998922	2104.5260	2.4024558	80039.257	96.558982
Z_{10}	6249.6440	1.1415668	92520.012	31.813793	12125388	3612.1025

ACFs (4) can be expanded in power series,

$$\Gamma_{\alpha q}(t) = \sum_{n=0}^{\infty} \frac{(-1)^n M_{2n\alpha}^{(q)} t^{2n}}{(2n)!}. \quad (7)$$

It is well known [1] that $M_{2n\alpha}^{(q)}$ is a moment on the order of $2n$ of the spectral density of the corresponding ACF. From similar equations for the homonuclear case [6], after appropriate modifications according to (3), we obtained recurrence equations for the moments for the heteronuclear case. Henceforth, we will assume for definiteness that there are only two types of spins in the system. In the Appendix, we present expressions for the moments of the tenth order inclusive for a general axially symmetric case. As an example, we calculated the moments for a LiF crystal (see Table 1); the contributions Δ_{qp}^2 (5) for this crystal were taken from Table 2.

Equations (1) were derived for an interaction with arbitrary magnetic anisotropy. The application of an axially symmetric Hamiltonian of the DDI essentially improves the convergence of the series for the memory function [10–15]. Therefore, it is expedient to transform Eqs. (1) in order to maximally take into consideration the longitudinal component of the local field and minimally take into account the transverse component. In this approximation, we obtain the following system of nonlinear integral equations for the ACFs of a LiF crystal:

$$\Gamma_{z\mathrm{L}}(t) = 1 - \frac{1}{2}\Delta_{\mathrm{LL}}^2 \times \int_0^t\int_0^{t'} dt'dt''\Gamma_{x\mathrm{L}}^2(t'-t'')\Gamma_{z\mathrm{L}}(t''),$$

$$\Gamma_{x\mathrm{L}}(t) = \Gamma_{A\mathrm{L}}(t) - \frac{k_L}{4}\Delta_{\mathrm{LL}}^2 \int_0^t dt'\Gamma_{A\mathrm{L}}(t-t') \times \int_0^{t'} dt''\Gamma_{x\mathrm{L}}(t'-t'')\Gamma_{z\mathrm{L}}(t'-t'')\Gamma_{x\mathrm{L}}(t''),$$

$$\Gamma_{z\mathrm{F}}(t) = 1 - \frac{1}{2}\Delta_{\mathrm{FF}}^2 \times \int_0^t\int_0^{t'} dt'dt''\Gamma_{x\mathrm{F}}^2(t'-t'')\Gamma_{z\mathrm{F}}(t''),$$

$$\Gamma_{x\mathrm{F}}(t) = \Gamma_{A\mathrm{F}}(t) - \frac{k_\mathrm{F}}{4}\Delta_{\mathrm{FF}}^2 \int_0^t dt'\Gamma_{A\mathrm{F}}(t-t') \times \int_0^{t'} dt''\Gamma_{x\mathrm{F}}(t'-t'')\Gamma_{z\mathrm{F}}(t'-t'')\Gamma_{x\mathrm{F}}(t''), \quad (8)$$

Table 2. Mean squares of homo- and heteronuclear contributions to the longitudinal local fields, renormalization parameters, and the coordinates of the singular point of the ACF in LiF for a magnetic field directed along three crystallographic axes

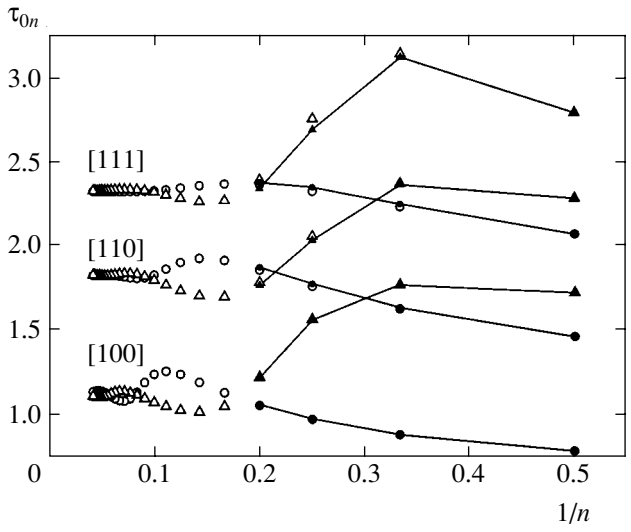
	$H_0 \parallel [111]$	$H_0 \parallel [110]$	$H_0 \parallel [100]$
Δ_{FF}^2, $(\mathrm{rad/s})^2$	$2838 \cdot 10^6$	$2581 \cdot 10^6$	$1809 \cdot 10^6$
$\Delta_{\mathrm{FL}}^2/\Delta_{\mathrm{FF}}^2$	0.3065	1.9559	9.7209
$\Delta_{\mathrm{LF}}^2/\Delta_{\mathrm{FF}}^2$	0.0613	0.3912	1.94442
$\Delta_{\mathrm{LL}}^2/\Delta_{\mathrm{FF}}^2$	0.1455	0.1455	0.1455
$\tau_0(\lambda_q = 1)\Delta_{\mathrm{FF}}$	2.372	1.843	1.127
$\tau_0(\lambda_q^2 = 5/4)\Delta_{\mathrm{FF}}$	2.263	1.783	1.113
λ_F	1.101	1.0911	1.0855
λ_L	1.099	1.0896	1.0850
$\tau_0(\lambda_q)\Delta_{\mathrm{FF}}$	2.33	1.82	1.12
$\tau_0(\lambda_q)\sqrt{M_{2\mathrm{F}}}$	3.72	3.74	3.87
$\delta\tau_c/\tau_0$	0.158	0.215	0.161
$\tau_0(\lambda_q) + \delta\tau_c$, μs	51	44	31
$\tau_c(\lambda_q)\sqrt{M_{2\mathrm{F}}}$	4.31	4.53	4.50

where

$$\Gamma_{A\mathrm{L}}(t) = \exp\left\{-\lambda_\mathrm{L}^2\Delta_{\mathrm{LL}}^2\int_0^t\int_0^{t'} dt'dt''\Gamma_{z\mathrm{L}}(t'') - \Delta_{\mathrm{LF}}^2\int_0^t\int_0^{t'} dt'dt''\Gamma_{z\mathrm{F}}(t'')\right\},$$

$$\Gamma_{A\mathrm{F}}(t) = \exp\left\{-\Delta_{\mathrm{FL}}^2\int_0^t\int_0^{t'} dt'dt''\Gamma_{z\mathrm{L}}(t'') - \lambda_\mathrm{F}^2\Delta_{\mathrm{FF}}^2\int_0^t\int_0^{t'} dt'dt''\Gamma_{z\mathrm{F}}(t'')\right\} \quad (9)$$

is the ACF of a spin rotating in the local field that has only a longitudinal component, similar to the function used in Anderson's model [4]. Unlike the previous work [11], we introduce a renormalization parameter for the longitudinal local field into Eq. (9). The value of



Ratios of moments (12) for the ACF of the x components of the spins of fluorine (circles) and lithium (triangles) nuclei in a LiF crystal when the magnetic field is applied along three crystallographic axes indicated in the figure. The results obtained by the moments of solutions to Eqs. (8) and (9) are indicated by open symbols, and the results obtained by the moments from Table 1 are indicated by closed symbols connected by straight lines.

this parameter is determined from the moments. The correctness of the second moments is guaranteed by the strict relation between the parameters k_q and λ_q that enter Eqs. (8) and (9):

$$k_q = 5 - 4\lambda_q^2.$$

Let us determine the fourth moment of the solution to Eqs. (8) and (9):

$$\frac{M_{4x}^{(q)}}{\Delta_{qq}^4} = 2\lambda_q^2 - \frac{5}{4}\lambda_q + \frac{15}{4} + \frac{\Delta_{qp}^2}{\Delta_{qq}^2}\left(3\lambda_q + \frac{15}{4} + \frac{\Delta_{pp}^2}{2\Delta_{qq}^2}\right) + 3\left(\frac{\Delta_{qp}^2}{\Delta_{qq}^2}\right)^2.$$

From the equality of this moment to its exact value (A1), we determine the parameters

$$\lambda_q = \frac{15}{6} - \frac{3D_q}{4} + \frac{1}{4}\left(9D_q^2 + \frac{47}{2}D_q + \frac{161}{16}\right)^{1/2}, \quad (10)$$

where $D_F = \Delta_{FL}^2/\Delta_{FF}^2$ and $D_L = \Delta_{LF}^2/\Delta_{LL}^2$. The values of the parameters calculated by formula (10) are presented in Table 2.

The solutions of Eqs. (8) and (9) have singular points on the imaginary-time axis. The principal parts of the ACFs in the neighborhoods of these points are given by

$$\Gamma_{zq}(t) \approx \frac{C_{zq}}{(\tau_0 + it)^{\zeta_q}}, \quad \Gamma_{xq}(t) \approx \frac{C_{xq}}{(\tau_0 + it)^{\chi_q}}, \quad \Gamma_{Aq}(t) \approx \frac{C_{Aq}}{(\tau_0 + it)^{\delta_q}} \quad (q = \mathrm{F}, \mathrm{Li}). \quad (11)$$

The exponents were determined in [10] by the Painelvé method for three orientations of a LiF crystal. For the F nuclei, all exponents equal 2, whereas, for the Li nuclei, only the exponents for the orientation [100] are equal to 2; for two other orientations, the exponents are as follows: $\delta = \chi = 0.123$ for [111] and $\delta = \chi = 0.784$ for [110] (we do not need the exponent ζ).

For a known value of the exponent of the lowest singular point, its coordinate, equal to the convergence radius with respect to moments (7), can be evaluated as the limit of the sequence of relations

$$\tau_{0n}^2 = \frac{M_{2(n-1)\alpha}^{(q)}\Gamma(2n+\chi)}{M_{2n\alpha}^{(q)}\Gamma(2n-2+\chi)}, \quad (12)$$

as $n \longrightarrow \infty$, where $\Gamma(x)$ is the gamma function. These sequences are shown in the figure. The calculations are performed by the exact values of ten moments from Table 1 and by 50 moments of the solutions to approximate Eqs. (8) and (9) for the values of the parameters λ_q given in Table 2. In the homonuclear case, the convergence of the sequence of relations is better [6]; this allowed us to determine τ_0 by the first ten moments to a sufficiently high degree of accuracy. In the heteronuclear case, the convergence deteriorates because the interaction between lithium nuclei is weak ($\Delta_{LL}^2 = 0.15\Delta_{FF}^2$). Therefore, it takes some time for the system of Li nuclei to adjust to the system of F nuclei. The figure shows that the first terms in the sequence of the ratios of moments of the solution to system (8), (9) are close to the ratios of exact moments. The approximate equations have allowed us to follow up how the ratios of moments pass to the limit (see Table 2). This approach represents the development of the simple estimate of [6]. Applying it to the homonuclear case for $\lambda = 1.105$, we arrive at the value $\tau_0\Delta_Z = 2.48$, which was determined earlier by ten moments. Note that an estimate for τ_0 was obtained in [10] by using Eqs. (8) and (9) for $\lambda = 1$ (which is also shown in Table 2). The variation of λ_q from 1 to 1.1 leads to variation of the coordinate τ_0 by less than 2%, which is indicative of the accuracy of its determination.

3. CALCULATION OF A CORRECTION TO THE COORDINATE OF A SINGULAR POINT OF THE CORRELATION FUNCTIONS

The SCFF approximation corresponds to the limit $d \longrightarrow \infty$. As is shown in [12] by an example of homonuclear systems, for $d = 3$, the coordinate of the singular point of the ACF on the imaginary-time axis increases due to the correlation of local fields,

$$\tau_c = \tau_0 + \delta\tau_c, \tag{13}$$

and the main contribution to $\delta\tau_c$ is given by the corrections to the moments that can be represented by trees of double bonds with a built-in pair of fourfold interaction or triangle of four bonds. In the heteronuclear case, the degree of correlation is characterized by the following ratios of lattice sums:

$$S_2/(S_1)^2, \quad S_2'/(S_1)^2, \quad S_3/(S_1)^2, \quad S_3'/(S_1S_1'), \tag{14}$$

where

$$S_1 = \sum_j b_{ij}^2, \quad S_2 = \sum_j b_{ij}^4, \quad S_3 = \sum_{k,j} b_{kj}^2 b_{ik} b_{ji} \tag{15}$$

with summation over the sites occupied by nuclei of one type with a nucleus at site *i*. In the primed sums, the summation is performed over the sites occupied by nuclei of another type. The numerical values of ratios (14) for three orientations of a LiF crystal are presented in [11]. The small value of these ratios corresponds to the real smallness parameter. For convenience, we introduce a formal parameter ε; in finite formulas, we set this parameter equal to zero. Contributions with lattice sums (14) are already contained in the fourth moment (A.3). The corresponding decrease in M_{4x} can be ascribed to the SCFF if one reduces λ_q. For example, for the orientation [110], we obtain $\lambda_F = 0.94$ and $\lambda_L = 0.75$. Such a variation leads to an increase in the coordinate τ_0 by about 2%. As is shown in [12, 16] in the homonuclear case, the incorporation of such fragments of (14) into large trees of bonds corresponding to higher order moments produces a more significant effect. Such contributions can be determined from Eqs. (1) for the ACFs.

Assuming that the correction $\varepsilon\delta\tau_c$ is small, we estimate it by taking a simplified version of equations in which the *zz* interactions are predominant. Take

$$\Gamma_{\alpha q}(t) = \Gamma_{\alpha q0}(t) - \varepsilon\Gamma_{\alpha q1}(t)$$

and substitute it into an integral equation with appropriate correction terms for the memory function

$$G_{\alpha q}(t) = G_{\alpha q0}(t) - \varepsilon G_{\alpha q1}(t).$$

In view of the form of the equations, it is more convenient to pass to an equation for the squared transverse ACF:

$$\Gamma_{xq}^2(t) = Y_q(t) = Y_{q0}(t) - \varepsilon Y_{q1}(t). \tag{16}$$

For the first correction $Y_{F1}(t)$, we find the equation

$$\begin{aligned}
\frac{d}{dt}Y_{F1}(t) &= 2\langle FF\rangle Y_{F1}(t)\int_0^t \Gamma_{zF0}(t_1)dt_1 \\
&+ 2\langle FL\rangle Y_{F1}(t)\int_0^t \Gamma_{zF0}(t_1)dt_1 \\
&+ \frac{4}{5}\langle FF\rangle Y_{F0}(t)\int_{0\,0\,0}^{t\,t_1\,t_2} Y_{F1}(t_3)dt_1dt_2dt_3 \\
&+ \frac{4}{5}\langle FL\rangle Y_{F0}(t)\int_{0\,0\,0}^{t\,t_1\,t_2} Y_{L1}(t_3)dt_1dt_2dt_3 \\
&+ 2\langle FL\rangle^2 S_2' Y_{F0}(t)\int_0^t \Gamma_{zL0}(t_1)dt_1\int_{0\,0}^{t\,t_1} \Gamma_{zL0}(t_2)dt_1dt_2 \\
&+ 2S_2\langle FF\rangle^2 Y_{F0}(t)\int_0^t \Gamma_{zF0}(t_1)dt_1\int_{0\,0}^{t\,t_1} \Gamma_{zF0}(t_2)dt_1dt_2 \\
&+ \frac{4}{5}\langle FF\rangle^2(S_2+S_3)Y_{F0}(t)\int_{0\,0\,0}^{t\,t_1\,t_2} Y_{F0}(t_3)dt_1dt_2dt_3 \\
&+ \frac{4}{5}\langle LF\rangle\langle LL\rangle S_3' Y_{F0}(t)\int_{0\,0\,0}^{t\,t_1\,t_2} Y_{L0}(t_3)dt_1dt_2dt_3 \\
&+ \langle FF\rangle^3\frac{8}{5}\left(\frac{3}{2}S_2+S_3\right)Y_{F0}(t) \\
&\times\int_{0\,0\,0}^{t\,t_1\,t_2} Y_{F0}(t_3)dt_1dt_2dt_3\int_{0\,0}^{t_3\,t_4}\Gamma_{zF0}(t_5)dt_4dt_5 \\
&+ \langle LL\rangle^2\langle FL\rangle\frac{8}{5}(S_2+S_3)Y_{F0}(t) \\
&\times\int_{0\,0\,0}^{t\,t_1\,t_2} Y_{L0}(t_3)dt_1dt_2dt_3\int_{0\,0}^{t_3\,t_4}\Gamma_{zL0}(t_5)dt_4dt_5 \\
&+ \langle LL\rangle\langle FL\rangle\langle LF\rangle\frac{4}{5}(S_2'+2S_3')Y_{F0}(t) \\
&\times\int_{0\,0\,0}^{t\,t_1\,t_2} Y_{L0}(t_3)dt_1dt_2dt_3\int_{0\,0}^{t_3\,t_4}\Gamma_{zF0}(t_5)dt_4dt_5 \\
&+ \langle FF\rangle^2\langle FL\rangle\frac{8}{5}S_3' Y_{F0}(t)
\end{aligned} \tag{17}$$

$$\times \int_0^t \int_0^{t_1} \int_0^{t_2} Y_{F0}(t_3)dt_1dt_2dt_3 \int_0^{t_3} \int_0^{t_4} \Gamma_{zL0}(t_5)dt_4dt_5 - R_F(t),$$

in which we introduced dimensionless imaginary time $t' = -it(5\Delta_{FF}^2/4)^{1/2}$; omitted the prime; denoted

$$\langle FF \rangle = 1, \quad \langle FL \rangle = \frac{4\Delta_{FL}^2}{5\Delta_{FF}^2},$$

$$\langle LL \rangle = \frac{\Delta_{LL}^2}{\Delta_{FF}^2}, \quad \langle LF \rangle = \frac{4\Delta_{LF}^2}{5\Delta_{FF}^2},$$

and denoted ratios (14) by S_2, S_2', S_3, S_3'. Equation (17) differs from that considered in the homonuclear case [12] by the contributions of nuclei of different types and by the change in the form of the correction $R_q(t)$ (because lithium nuclei have spin 3/2) due to the permissible fourfold interaction of nearest neighbors:

$$R_F(t) = 2S_2\langle FF \rangle^2 Y_{F0}(t) \int_0^t \Gamma_{zF0}(t-t_1)dt_1$$

$$\times \int_0^{t_1} \int_0^{t_2} \Gamma_{zF0}(t_3)dt_2dt_3$$

$$+ 2S_2'\langle FL \rangle^2 Y_{F0}(t) \int_0^t \Gamma_{zL0}(t-t_1)dt_1$$

$$\times \int_0^{t_1} \int_0^{t_2} \left[1 + \frac{16}{15}\Gamma_{zL0}(t_1-t_2)\right]\Gamma_{zL0}(t_3)dt_2dt_3,$$

$$R_L(t) = 2S_2'\langle LF \rangle^2 Y_{L0}(t) \int_0^t \Gamma_{zF0}(t-t_1)dt_1$$

$$\times \int_0^{t_1} \int_0^{t_2} \Gamma_{zF0}(t_3)dt_2dt_3$$

$$+ 2S_2\langle LL \rangle^2 Y_{L0}(t) \int_0^t \Gamma_{zL0}(t-t_1)dt_1$$

$$\times \int_0^{t_1} \int_0^{t_2} \left[1 + \frac{16}{15}\Gamma_{zL0}(t_1-t_2)\right]\Gamma_{zL0}(t_3)dt_2dt_3.$$

A change in the remaining part of Eq. (17) when passing to $Y_{L1}(t)$ reduces to a change in the subscripts $F \longrightarrow L$ and $L \longrightarrow F$. For the functions $Y_{q0}(t)$ and $\Gamma_{zq0}(t)$ of zeroth-order approximation, we use a simple set of four equations that is obtained from (8) for $\lambda_q^2 = 5/4$. Moreover, we neglect the time variation ($\Gamma_{zq}(t'') \sim \Gamma_{zq}(0) = 1$) compared with $(\Gamma_{xq}(t' - t''))^2$ in the integrand; as is shown in [12], this leads to a slight decrease in the moments but significantly simplifies the calculations. For comparison, the coordinate of the singular point $\tau_0(\lambda_q^2 = 5/4)\Delta_{FF}$ of the solution to such a system is shown in Table 2. We had to simplify the equation in order to increase the numerical positions up to 50. In turn, such a large mantissa is required to sum up the moments that strongly differ in magnitude.

Using the equations, we calculate the moments of functions (16) up to $n = 70$. The coordinate of the singular point is determined from the ratio of moments:

$$\tau_c^2 = \tau_0^2 \lim_{n\to\infty} \frac{1 - \varepsilon Y_{F2(n-1)}^{(1)}/Y_{F2(n-1)}^{(0)} + \ldots}{1 - \varepsilon Y_{F2n}^{(1)}/Y_{F2n}^{(0)} + \ldots}.$$

Then,

$$2\frac{\delta\tau_c}{\tau_0} = \lim_{n\to\infty}\left(\frac{Y_{F2n}^{(1)}}{Y_{F2n}^{(0)}} - \frac{Y_{F2(n-1)}^{(1)}}{Y_{F2(n-1)}^{(0)}}\right). \tag{18}$$

Extrapolating the ratios obtained, we determine the values presented in Table 2.

Let us compare the results for hetero- and homonuclear systems. In the homonuclear case, we have $\tau_0\Delta_Z = 2.48$ for any orientation, whereas, in the heteronuclear case, Table 2 shows that $\tau_0\Delta_{FF}$ decreases by a factor of 2 under rotation from [111] to [100]. Using the units of the total moment of the NMR spectrum of fluorine nuclei,

$$M_{2F} = \Delta_{FF}^2(1 + X_2^{(F)}),$$

we obtain $\tau_0\sqrt{M_{2F}} = 3.72$ for the [111] orientation. This result coincides with that in the homonuclear case, where the coordinate increases to 3.87 under the rotation to the [100] orientation. Such an increase is associated with an increase in the contribution of the heteronuclear zz interaction to M_{2F}. In the homonuclear case, a similar increase in the coordinate of the singular point of the ACF with the ratio Δ_Z^2/Δ_X^2 was observed in [17].

Let us pass to the correction $\delta\tau_c$. We calculated $\delta\tau_c/\tau_0$ for a homonuclear face-centered cubic lattice by the formula obtained in [12] for three orientations and obtained the following results: 0.17 for [111], 0.37 for [110], and 0.22 for [100]. A comparison of these results with the values presented in Table 2 shows that, in both cases, the correction attains its maximal value for the [110] orientation and is primarily associated with the large value of the parameter $S_2/(S_1)^2 = 0.225$. The addi-

tion of a heteronuclear interaction reduces the correction and smoothes out its dependence on orientation.

Thus, we have obtained, theoretically, the required increase in the coordinate of the singular point of an ACF due to the correlation of local fields, which was revealed in [11] by analyzing experimental spectra from the viewpoint of the SCFF theory. A quantitative comparison of the theory and experiment requires that one should determine a variation in the preexponential factor due to the correlation in the motion of spins. An appreciable effect of this factor was also shown in [11], which testifies to the fact that the observed detuning from the center of the spectrum is less than the mathematical asymptotics. Finally, publications do not contain all experimental conditions that are necessary for a successful comparison of the results.

In conclusion, note that we have performed calculations for a LiF crystal when a magnetic field is directed along three crystallographic axes. The theory allows us to obtain results for other heteronuclear systems and orientations. To this end, one should substitute the lattice sums and the contributions to the squared local fields into the formulas and perform the calculations described in this paper.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 02-02-17463

APPENDIX

Here, we present expressions for the moments up to the tenth order for a system consisting of two types of spins for an axially symmetric case $\Delta_{xq}^2 = \Delta_{yq}^2$ with an arbitrary relation between Δ_{xq}^2 and Δ_{qq}^2 (below, for definiteness, we will use subscripts "F" and "L" in place of q and p). The moments are calculated by the recurrence relations obtained from the equations of the homonuclear case [6] after the above-mentioned modifications. The results are exact in the SCFF approximation, which corresponds to infinite-dimensional lattices:

$$M_{2X}^{(F)} = \Delta_{FF}^2 + \Delta_{FL}^2 + \Delta_{XF}^2,$$

$$M_{4X}^{(F)} = 3\Delta_{FL}^4 + 3\Delta_{FF}^4 + (4\Delta_{XF}^2 + 6\Delta_{FF}^2 + 2\Delta_{XL}^2)\Delta_{FL}^2 + 5\Delta_{XF}^4 + 6\Delta_{FF}^2\Delta_{XF}^2,$$

$$M_{6X}^{(F)} = 15\Delta_{FL}^6 + 15\Delta_{FF}^6 + (21\Delta_{XF}^2 + 45\Delta_{FF}^2 + 30\Delta_{XL}^2)\Delta_{FL}^4 + 51\Delta_{XF}^6 + 73\Delta_{FF}^2\Delta_{XF}^4 + 55\Delta_{FF}^4\Delta_{XF}^2 + (45\Delta_{XF}^4 + 45\Delta_{FF}^4 + 10\Delta_{XL}^4 + 76\Delta_{FF}^2\Delta_{XF}^2 + 30\Delta_{FF}^2\Delta_{XL}^2 + 4\Delta_{LF}^2\Delta_{XL}^2 + 4\Delta_{LL}^2\Delta_{XL}^2 + 18\Delta_{XF}^2\Delta_{XL}^2)\Delta_{FL}^2,$$

$$M_{8X}^{(F)} = 105\Delta_{FL}^8 + (144\Delta_{XF}^2 + 420\Delta_{FF}^2 + 420\Delta_{XL}^2)\Delta_{FL}^6 + (112\Delta_{XL}^2\Delta_{LL}^2 + 448\Delta_{XF}^4 + 112\Delta_{XL}^2\Delta_{LF}^2 + 630\Delta_{FF}^4 + 840\Delta_{FF}^2\Delta_{XL}^2 + 360\Delta_{XF}^2\Delta_{XL}^2 + 420\Delta_{XL}^4 + 988\Delta_{FF}^2\Delta_{XF}^2)\Delta_{FL}^4 + \{1678\Delta_{FF}^2\Delta_{XF}^4 + 92\Delta_{XL}^6 + 1544\Delta_{FF}^4\Delta_{XF}^2 + 420\Delta_{XL}^6 + 24\Delta_{XL}^2\Delta_{LL}^4 + (72\Delta_{XF}^2\Delta_{XL}^2 + 78\Delta_{XL}^4 + 48\Delta_{XL}^2\Delta_{LF}^2 + 112\Delta_{FF}^2\Delta_{XL}^2)\Delta_{LL}^2 + 112\Delta_{FF}^2\Delta_{XL}^2\Delta_{LF}^2 + 24\Delta_{XL}^2\Delta_{LF}^4 + 180\Delta_{XF}^2\Delta_{XL}^4 + 284\Delta_{XF}^4\Delta_{XL}^2 + 648\Delta_{FF}^2\Delta_{XF}^2\Delta_{XL}^2 + 70\Delta_{XL}^4\Delta_{LF}^2 + 80\Delta_{XF}^2\Delta_{XL}^2\Delta_{LF}^2 + 914\Delta_{XF}^6 + 280\Delta_{FF}^2\Delta_{XL}^4 + 420\Delta_{FF}^4\Delta_{XL}^2\}\Delta_{FL}^2 + 1470\Delta_{FF}^2\Delta_{XF}^6 + 861\Delta_{XF}^8 + 105\Delta_{FF}^8 + 1378\Delta_{FF}^4\Delta_{XF}^4 + 700\Delta_{FF}^6\Delta_{XF}^2,$$

$$M_{10X}^{(F)} = (1245\Delta_{XF}^2 + 4725\Delta_{FF}^2 + 6300\Delta_{XL}^2)\Delta_{FL}^8 + (2520\Delta_{XL}^2\Delta_{LL}^2 + 5856\Delta_{XF}^4 + 2520\Delta_{XL}^2\Delta_{LF}^2 + 9450\Delta_{FF}^4 + 18900\Delta_{FF}^2\Delta_{XL}^2 + 6060\Delta_{XF}^2\Delta_{XL}^2 + 12600\Delta_{XL}^4 + 15120\Delta_{FF}^2\Delta_{XF}^2)\Delta_{FL}^6 + \{36858\Delta_{FF}^2\Delta_{XF}^4 + 4830\Delta_{XL}^4\Delta_{LF}^2 + 8340\Delta_{XL}^6 + 37890\Delta_{FF}^4\Delta_{XF}^2 + 9450\Delta_{FF}^6 + 1080\Delta_{XL}^2\Delta_{LL}^4 + (2280\Delta_{XF}^2\Delta_{XL}^2 + 5190\Delta_{XL}^4 + 2160\Delta_{XL}^2\Delta_{LF}^2 + 5040\Delta_{FF}^2\Delta_{XL}^2)\Delta_{LL}^2 + 5040\Delta_{FF}^2\Delta_{XL}^2\Delta_{LF}^2 + 16676\Delta_{XF}^6 + 1080\Delta_{XL}^2\Delta_{LF}^4 + 8280\Delta_{XF}^2\Delta_{XL}^4 + 9140\Delta_{XF}^4\Delta_{XL}^2 + 2700\Delta_{FF}^2\Delta_{XF}^2\Delta_{XL}^2 + 2656\Delta_{XF}^2\Delta_{XL}^2\Delta_{LF}^2 + 18900\Delta_{FF}^2\Delta_{XL}^4 + 18900\Delta_{FF}^4\Delta_{XL}^2\}\Delta_{FL}^4 + \{27643\Delta_{XF}^8 + 1442\Delta_{XL}^8 + 57858\Delta_{XF}^6\Delta_{FF}^2 + 64744\Delta_{XF}^4\Delta_{FF}^4 + 35400\Delta_{XF}^2\Delta_{FF}^6 + 4725\Delta_{FF}^8 + 1080\Delta_{XL}^2\Delta_{FF}^2\Delta_{LF}^4 + 4140\Delta_{XL}^6\Delta_{FF}^2 + 2852\Delta_{XF}^2\Delta_{XL}^6 + 6300\Delta_{XL}^2\Delta_{FF}^6 + 7492\Delta_{XF}^6\Delta_{XL}^2 + 2520\Delta_{XL}^2\Delta_{FF}^4\Delta_{LF}^2 + 9940\Delta_{XF}^2\Delta_{XL}^4\Delta_{FF}^2 + 3150\Delta_{XL}^4\Delta_{FF}^2\Delta_{LF}^2 + 4352\Delta_{XF}^2\Delta_{XL}^2\Delta_{FF}^2\Delta_{LF}^2 \quad (A.1)$$

$$+ 20940\Delta_{XF}^2\Delta_{XL}^2\Delta_{FF}^4 + 1532\Delta_{XL}^6\Delta_{LF}^2$$
$$+ 240\Delta_{XL}^2\Delta_{LF}^6 + 240\Delta_{XL}^2\Delta_{LL}^6 + 756\Delta_{XL}^4\Delta_{LF}^4$$
$$+ 6300\Delta_{XL}^4\Delta_{FF}^4 + 4560\Delta_{XF}^4\Delta_{XL}^4$$
$$+ 2418\Delta_{XL}^4\Delta_{XF}^2\Delta_{LF}^2 + 984\Delta_{XF}^2\Delta_{XL}^2\Delta_{LF}^4$$
$$+ 2112\Delta_{XF}^4\Delta_{XL}^2\Delta_{LF}^2 + 18996\Delta_{XF}^4\Delta_{XL}^2\Delta_{FF}^2$$
$$+ (3976\Delta_{XF}^2\Delta_{XL}^2\Delta_{FF}^2 + 720\Delta_{XL}^2\Delta_{LF}^4$$
$$+ 1824\Delta_{XF}^4\Delta_{XL}^2 + 2520\Delta_{XL}^2\Delta_{FF}^4 + 2160\Delta_{XL}^2\Delta_{FF}^2\Delta_{LF}^2$$
$$+ 1768\Delta_{XL}^4\Delta_{LF}^2 + 3510\Delta_{XL}^4\Delta_{FF}^2$$
$$+ 1728\Delta_{XF}^2\Delta_{XL}^2\Delta_{LF}^2 + 1820\Delta_{XL}^6 + 2418\Delta_{XF}^2\Delta_{XL}^4)\Delta_{LL}^2$$
$$+ (1080\Delta_{XL}^2\Delta_{FF}^2 + 1012\Delta_{XL}^4 + 720\Delta_{XL}^2\Delta_{LF}^2$$
$$+ 744\Delta_{XF}^2\Delta_{XL}^2)\Delta_{LL}^4\}\Delta_{FL}^2 + 945\Delta_{FL}^{10} + 43989\Delta_{FF}^2\Delta_{XF}^8$$
$$+ 21847\Delta_{XF}^{10} + 945\Delta_{FF}^{10} + 48498\Delta_{FF}^4\Delta_{XF}^6$$
$$+ 33742\Delta_{FF}^6\Delta_{XF}^4 + 11385\Delta_{FF}^8\Delta_{XF}^2,$$

$$M_{2Z}^{(F)} = 2\Delta_{XF}^2,$$

$$M_{4Z}^{(F)} = 4\Delta_{XF}^2\Delta_{FL}^2 + 10\Delta_{XF}^4 + 4\Delta_{FF}^2\Delta_{XF}^2,$$

$$M_{6Z}^{(F)} = 24\Delta_{XF}^2\Delta_{FL}^4 + 92\Delta_{XF}^6 + 78\Delta_{FF}^2\Delta_{XF}^4$$
$$+ 24\Delta_{FF}^4\Delta_{XF}^2 + (70\Delta_{XF}^4 + 48\Delta_{FF}^2\Delta_{XF}^2 + 8\Delta_{XF}^2\Delta_{XL}^2)\Delta_{FL}^2,$$

$$M_{8Z}^{(F)} = 240\Delta_{XF}^2\Delta_{FL}^6 + (240\Delta_{XF}^2\Delta_{XL}^2 + 756\Delta_{XF}^4$$
$$+ 720\Delta_{XF}^2\Delta_{FF}^2)\Delta_{FL}^4 + (720\Delta_{XF}^2\Delta_{FF}^4 + 1768\Delta_{XF}^4\Delta_{FF}^2$$
$$+ 16\Delta_{XF}^2\Delta_{LF}^2\Delta_{XL}^2 + 16\Delta_{XF}^2\Delta_{LL}^2\Delta_{XL}^2 + 1532\Delta_{XF}^6$$
$$+ 248\Delta_{XF}^4\Delta_{XL}^2 + 240\Delta_{XF}^2\Delta_{FF}^2\Delta_{XL}^2 + 40\Delta_{XF}^2\Delta_{XL}^4)\Delta_{FL}^2$$
$$+ 1820\Delta_{XF}^6\Delta_{FF}^2 + 1012\Delta_{XF}^4\Delta_{FF}^4$$
$$+ 240\Delta_{XF}^2\Delta_{FF}^6 + 1442\Delta_{XF}^8,$$

$$M_{10Z}^{(F)} = 3360\Delta_{XF}^2\Delta_{FL}^8 + (6720\Delta_{XF}^2\Delta_{XL}^2 + 10956\Delta_{XF}^4$$
$$+ 13440\Delta_{XF}^2\Delta_{FF}^2)\Delta_{FL}^6 + (20160\Delta_{XF}^2\Delta_{FF}^4 \qquad \text{(A.2)}$$
$$+ 40580\Delta_{XF}^4\Delta_{FF}^2 + 896\Delta_{XF}^2\Delta_{LF}^2\Delta_{XL}^2$$
$$+ 896\Delta_{XF}^2\Delta_{LL}^2\Delta_{XL}^2 + 28752\Delta_{XF}^6 + 11516\Delta_{XF}^4\Delta_{XL}^2$$
$$+ 13440\Delta_{XF}^2\Delta_{FF}^2\Delta_{XL}^2 + 3360\Delta_{XF}^2\Delta_{XL}^4)\Delta_{FL}^4$$
$$+ \{13440\Delta_{XF}^2\Delta_{FF}^6 + 368\Delta_{XF}^2\Delta_{XL}^6$$
$$+ 896\Delta_{XF}^2\Delta_{XL}^2\Delta_{FF}^2\Delta_{LF}^2 + 48292\Delta_{XF}^4\Delta_{FF}^4$$
$$+ 46702\Delta_{XF}^8 + 840\Delta_{XF}^4\Delta_{XL}^2\Delta_{LF}^2$$
$$+ 72348\Delta_{XF}^6\Delta_{FF}^2 + 13788\Delta_{XF}^4\Delta_{XL}^2\Delta_{FF}^2$$
$$+ 6720\Delta_{XF}^2\Delta_{XL}^2\Delta_{FF}^4 + (192\Delta_{XF}^2\Delta_{XL}^2\Delta_{LF}^2 + 808\Delta_{XF}^4\Delta_{XL}^2$$
$$+ 312\Delta_{XF}^2\Delta_{XL}^4 + 896\Delta_{XF}^2\Delta_{XL}^2\Delta_{FF}^2)\Delta_{LL}^2$$
$$+ 2240\Delta_{XF}^2\Delta_{XL}^4\Delta_{FF}^2 + 9048\Delta_{XF}^6\Delta_{XL}^2$$
$$+ 96\Delta_{XF}^2\Delta_{XL}^2\Delta_{LL}^4 + 96\Delta_{XF}^2\Delta_{XL}^2\Delta_{LF}^4 + 280\Delta_{XF}^2\Delta_{XL}^4\Delta_{LF}^2$$
$$+ 2020\Delta_{XF}^4\Delta_{XL}^4\}\Delta_{FL}^2 + 18668\Delta_{XF}^4\Delta_{FF}^6 + 58138\Delta_{XF}^8\Delta_{FF}^2$$
$$+ 3360\Delta_{XF}^2\Delta_{FF}^8 + 44748\Delta_{XF}^6\Delta_{FF}^4 + 35492\Delta_{XF}^{10}.$$

Expressions for the moments of the nuclei of the second type are obtained by changing the subscripts $F \longrightarrow L$ and $L \longrightarrow F$.

For comparison, we present an exact expression for the fourth moment of the ACF $\Gamma_{xF}(t)$ (4) for a system consisting of nuclei of two types with a DDI for a real LiF crystal, which was derived from the results for a similar moment of the NMR spectrum [18], and an expression for the fourth moment of the ACF of a homonuclear system [19]:

$$\frac{M_{4X}^{(F)}}{\Delta_{FF}^4} = \frac{77}{16} - \frac{9S_2}{4S_1^2} + \frac{S_3}{2S_1^2}$$
$$+ \frac{\Delta_{FL}^2}{\Delta_{FF}^2}\left(7 + \frac{\Delta_{LL}^2}{2\Delta_{FF}^2} + \frac{S_3'}{2S_1S_1'}\left(1 - \frac{\Delta_{LL}^2}{\Delta_{FF}^2}\right)\right) \qquad \text{(A.3)}$$
$$+ \left(\frac{\Delta_{FL}^2}{\Delta_{FF}^2}\right)^2\left(3 - \frac{34S_2'}{(5S_1')^2}\right).$$

A similar moment for the Li nuclei is obtained by the change of the subscripts $F \longrightarrow L$ and $L \longrightarrow F$ and by the simultaneous change of the numerical coefficients: $9S_2/4S_1^2$ to $1.27S_2/S_1^2$ and $34S_2'/(5S_1')^2$ to $2S_2'/(S_1')^2$, which is associated with the difference in the spin quantum numbers of the F and Li nuclei.

REFERENCES

1. R. R. Ernst, G. Bodenhausen, and A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions* (Clarendon, Oxford, 1987; Mir, Moscow, 1990).
2. M. Goldmant, *Spin Temperature and Nuclear Magnetic Resonance in Solids* (Clarendon, Oxford, 1970; Mir, Moscow, 1972).

3. K. A. Valiev and A. A. Kokin, *Quantum Computers: Hopes and Reality* (Regul. Khaot. Din., Izhevsk, 2001) [in Russian].
4. P. W. Anderson and P. R. Weiss, Rev. Mod. Phys. **25**, 269 (1953).
5. M. Blume and J. Hubbard, Phys. Rev. B **1**, 3815 (1970).
6. V. E. Zobov, Teor. Mat. Fiz. **77**, 426 (1988); **84**, 111 (1990).
7. G. E. Karnaukh, A. A. Lundin, B. N. Provotorov, and K. T. Summanen, Zh. Éksp. Teor. Fiz. **91**, 2229 (1986) [Sov. Phys. JETP **64**, 1324 (1986)].
8. M. I. Bulgakov, A. D. Gul'ko, F. S. Dzheparov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **58**, 614 (1993) [JETP Lett. **58**, 592 (1993)].
9. B. N. Provotorov, T. P. Kulagina, and G. E. Karnaukh, Zh. Éksp. Teor. Fiz. **113**, 967 (1998) [JETP **86**, 527 (1998)].
10. V. E. Zobov and A. A. Lundin, Zh. Éksp. Teor. Fiz. **106**, 1097 (1994) [JETP **79**, 595 (1994)].
11. V. E. Zobov, A. A. Lundin, and O. E. Rodionova, Zh. Éksp. Teor. Fiz. **120**, 619 (2001) [JETP **93**, 542 (2001)].
12. V. E. Zobov and M. A. Popov, Zh. Éksp. Teor. Fiz. **124**, 89 (2003) [JETP **97**, 78 (2003)].
13. P. Borckmans and D. Walgraef, Phys. Rev. B **7**, 563 (1973).
14. A. A. Lundin, A. V. Makarenko, and V. E. Zobov, J. Phys.: Condens. Matter **2**, 10131 (1990).
15. V. E. Zobov, M. A. Popov, Yu. N. Ivanov, and A. I. Livshits, Zh. Éksp. Teor. Fiz. **115**, 285 (1999) [JETP **88**, 157 (1999)].
16. V. E. Zobov and M. A. Popov, Teor. Mat. Fiz. **136**, 463 (2003).
17. V. E. Zobov and O. V. Falaleev, Fiz. Tverd. Tela (Leningrad) **31**, 30 (1989) [Sov. Phys. Solid State **31**, 16 (1989)].
18. A. Losche, *Kerninduktion* (Wissenschaften, Berlin, 1957; Inostrannaya Literatura, Moscow, 1963).
19. K. S. J. Jensen and K. E. Hansen, Phys. Rev. B **13**, 1903 (1976).

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 100, No. 4, 2005, pp. 784–794.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 4, 2005, pp. 886–897.

STATISTICAL, NONLINEAR, AND SOFT MATTER PHYSICS

Synchronization of Chaotic Oscillator Time Scales¶

A. E. Hramov*, A. A. Koronovskii, and Yu. I. Levin

Saratov State University, Saratov, 410012 Russia
**e-mail: aeh@cas.ssu.runnet.ru*

Received June 18, 2004

Abstract—We consider chaotic oscillator synchronization and propose a new approach for detecting the synchronized behavior of chaotic oscillators. This approach is based on analysis of different time scales in the time series generated by coupled chaotic oscillators. We show that complete synchronization, phase synchronization, lag synchronization, and generalized synchronization are particular cases of the synchronized behavior called time-scale synchronization. A quantitative measure of chaotic oscillator synchronous behavior is proposed. This approach is applied to coupled Rössler systems. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Synchronization of chaotic oscillators is one of the fundamental phenomena in nonlinear dynamics. It occurs in many physical [1–6] and biological [7–9] processes. It seems to play an important role in the ability of biological oscillators, such as neurons, to act cooperatively [10–12].

Several different types of synchronization of coupled chaotic oscillators have been described theoretically and observed experimentally [13–16]. Complete synchronization implies the coincidence of states of coupled oscillators, $\mathbf{x}_1(t) \approx \mathbf{x}_2(t)$, with the difference between state vectors of coupled systems converging to zero in the limit as $t \longrightarrow \infty$ [17–20]. It occurs when interacting systems are identical. If the parameters of coupled chaotic oscillators slightly mismatch, the state vectors are close, $|\mathbf{x}_1(t) - \mathbf{x}_2(t)| \approx 0$, but differ from each other. Another type of synchronized behavior of coupled chaotic oscillators with slightly mismatched parameters is lag synchronization: this is the case where state vectors coincide with each other after a time shift, $\mathbf{x}_1(t + \tau) = \mathbf{x}_2(t)$. As the coupling between oscillators increases, the time lag τ decreases and the synchronization regime tends to the complete synchronization described in [21–23]. Generalized synchronization [24–26], introduced for drive-response systems, means that there is some functional relation between coupled chaotic oscillators, i.e., $\mathbf{x}_2(t) = \mathbf{F}[\mathbf{x}_1(t)]$.

We mention finally the phase synchronization regime. To describe the phase synchronization, the instantaneous phase $\phi(t)$ of a chaotic continuous time series is usually introduced [13–16, 27, 28]. The phase synchronization means the entrainment of phases of chaotic signals, with their amplitudes remaining chaotic and uncorrelated.

All synchronization types mentioned above are related to each other (see [1, 22, 24] for details), but the relation between them has not yet been completely clarified. For each type of synchronization, there are specific ways of detecting synchronized behavior of coupled chaotic oscillators. Complete synchronization can be detected by comparing system state vectors $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$, whereas lag synchronization can be determined by means of a similarity function [21]. The case of the generalized synchronization is more intricate because the functional relation $\mathbf{F}[\ldots]$ can be very complicated, but there are several methods for detecting synchronized behavior of coupled chaotic oscillators, such as the auxiliary system approach [29] or the method of nearest neighbors [24, 30].

Finally, phase synchronization of two coupled chaotic oscillators occurs if the difference between the instantaneous phases $\phi(t)$ of chaotic signals $\mathbf{x}_{1,2}(t)$ is bounded by some constant:

$$|\phi_1(t) - \phi_2(t)| < \text{const}. \tag{1}$$

It is possible to define the mean frequency of the chaotic signal,

$$\bar{\Omega} = \lim_{t\to\infty} \frac{\phi(t)}{t} = \langle \dot{\phi}(t) \rangle, \tag{2}$$

which is the same for both coupled chaotic systems; i.e., phase locking leads to frequency entrainment. We note that for the results to be correct, the mean frequency $\bar{\Omega}$ a of chaotic signal $\mathbf{x}(t)$ must coincide with the main frequency $\Omega_0 = 2\pi f_0$ of the Fourier spectrum (see [31] for details). There is no general way to introduce the phase for chaotic time series. There are several approaches that allow defining the phase for "good" systems with a simple topology of a chaotic attractor (the so-called phase coherent attractor), whose Fourier spectrum contains a single main frequency f_0.

¶ This article was submitted by the authors in English.

First of all, a plane in the system phase space may exist such that the projection of the chaotic attractor on it looks like a circular band. For such a plane, coordinates x and y can be introduced with the origin placed somewhere near the center of the chaotic attractor projection. The phase can then be introduced as an angle in this coordinate system [32, 21], but this requires all trajectories of the chaotic attractor projection on the (x, y) plane to revolve around the origin. A coordinate transformation can be sometimes used to obtain a proper projection [32, 13]. If the projections of chaotic trajectories on the plane $(\dot{x}, \dot{y})$ always rotate around the origin, velocities $\dot{x}$ and $\dot{y}$ can also be used; in some cases, this approach is more suitable [33, 34]. Another way to define the phase $\phi(t)$ of a chaotic time series $x(t)$ is to construct the analytical signal [14, 27] using the Hilbert transform. Moreover, the Poincaré secant surface can be used to introduce the instantaneous phase of a chaotic dynamical system [14, 27]. Finally, the phase of a chaotic time series can be introduced by means of the continuous wavelet transform [35], but the appropriate wavelet function and its parameters should be chosen [36].

All these approaches give correct results for "good" systems with well-defined phase, but fail for oscillators with nonrevolving trajectories. Such chaotic oscillators are often called "systems with ill-defined phase" or "systems with the funnel attractor." Introducing the phase via the above-mentioned approaches usually leads to incorrect results for a system with ill-defined phase [31]. Therefore, the phase synchronization of such systems can be usually detected by means of indirect indications [32, 37] and measurements [33].

In this paper, we propose a new approach for detecting the synchronization between two coupled chaotic oscillators. The main idea of this approach consists in analyzing the system behavior on different time scales, which allows us to consider different cases of synchronization from a universal standpoint [38]. Using the continuous wavelet transform [39–42], we introduce the continuous set of time scales s and the instantaneous phases $\phi_s(t)$ associated with them. In other words, $\phi_s(t)$ is a continuous function of time t and time scale s. As we show in what follows, if two chaotic oscillators demonstrate any type of synchronized behavior mentioned above, the time series $\mathbf{x}_{1,2}(t)$ generated by these systems involve time scales s that are necessarily correlated and satisfy the phase locking condition

$$|\phi_{s1}(t) - \phi_{s2}(t)| < \text{const}. \tag{3}$$

In other words, complete, lag, phase, and generalized synchronizations are the particular cases of the synchronous coupled chaotic oscillator behavior called "time-scale synchronization."

The structure of this paper is as follows. In Section 2, we discuss the continuous wavelet transform and the method of time scales s, and define the phases $\phi_s(t)$ associated with them. In Section 3, we consider the phase synchronization of two coupled Rössler systems. We demonstrate the application of our method and discuss its relation to traditional approaches. Section 4 deals with synchronization of two coupled Rössler systems with funnel attractors. In this case, the traditional methods for introducing the phase fail and it is impossible to detect the phase synchronization regime. Synchronization between systems can be revealed here only by means of indirect measurements (see [33] for details). We demonstrate the efficiency of our method for such cases and discuss the correlation between phase, lag, and complete synchronizations. In Section 5, we apply our method to the unidirectional coupled Rössler systems with phase-coherent attractors in which generalized synchronization is observed. The quantitative measure of synchronization is described in Section 6. The conclusions are presented in Section 7.

2. CONTINUOUS WAVELET TRANSFORM

The continuous wavelet transform is a powerful tool for analyzing the behavior of nonlinear dynamical systems. In particular, the continuous wavelet analysis has been used for the detection of synchronization of chaotic oscillations in the brain [35, 43, 44] and chaotic laser array [45]. It has also been used to detect the basic frequency of oscillations in nephron autoregulation [46]. We propose to analyze the dynamics of coupled chaotic oscillators by considering system behavior at different time scales s, each of which is characterized by its own phase $\phi_s(t)$. In defining the continuous set of instantaneous phases $\phi_s(t)$, the continuous wavelet transform is a convenient mathematical tool.

We consider the continuous wavelet transform of a chaotic time series $x(t)$,

$$W(s, t_0) = \int_{-\infty}^{\infty} x(t)\psi^*_{s, t_0}(t)dt, \tag{4}$$

where $\psi_{s, t_0}(t)$ is the wavelet function related to the mother-wavelet function $\psi_0(t)$ as

$$\psi_{s, t_0}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t - t_0}{s}\right). \tag{5}$$

The time scale s corresponds to the width of the wavelet function $\psi_{s, t_0}(t)$, t_0 is the shift of the wavelet along the time axis, and the "*" symbol in (4) denotes complex conjugation. We note that the time scale s is typically used instead of the Fourier-transform frequency f and can be considered as the quantity inversed to it.

The Morlet wavelet [47]

$$\psi_0(\eta) = \frac{1}{\sqrt[4]{\pi}}\exp(j\Omega_0\eta)\exp\left(-\frac{\eta^2}{2}\right) \tag{6}$$

has been used as a mother-wavelet function. The choice of the parameter value $\Omega_0 = 2\pi$ provides the relation $s =$

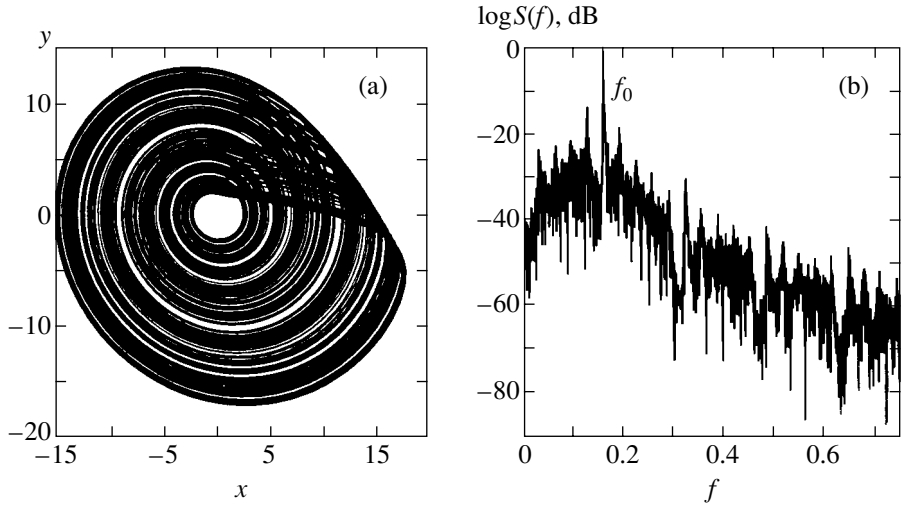


Fig. 1. (a) A phase coherent attractor and (b) the Fourier spectrum for the first Rössler system (10). The coupling parameter ε between the oscillators is zero.

$1/f$ between the time scale s of the wavelet transform and the frequency f of the Fourier transform.

The wavelet surface

$$W(s, t_0) = |W(s, t_0)| \exp[j\phi_s(t_0)] \quad (7)$$

describes the system dynamics on every time scale s at the time instant t_0. The value of $|W(s, t_0)|$ indicates the presence and intensity of the time scale s mode in the time series $x(t)$ at the time instant t_0. The quantities

$$E(s, t_0) = |W(s, t_0)|^2 \quad (8)$$

and

$$\langle E(s) \rangle = \int |W(s, t_0)|^2 dt_0 \quad (9)$$

are the instantaneous and integral energy distributions on time scales, respectively.

The phase $\phi_s(t) = \arg W(s, t)$ is naturally introduced for every time scale s. This means that the behavior of each time scale s can be described by means of its own phase $\phi_s(t)$. If two interacting chaotic oscillators are synchronized, the corresponding time series $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ involve scales s correlated with each other. This correlation can be detected by examining condition (3), which must be satisfied for synchronized time scales.

3. PHASE SYNCHRONIZATION OF TWO RÖSSLER SYSTEMS

We first consider two coupled Rössler systems with slightly mismatched parameters [27, 28],

$$\begin{aligned} \dot{x}_{1,2} &= -\omega_{1,2} y_{1,2} - z_{1,2} + \varepsilon(x_{2,1} - x_{1,2}), \\ \dot{y}_{1,2} &= \omega_{1,2} x_{1,2} + a y_{1,2}, \\ \dot{z}_{1,2} &= p + z_{1,2}(x_{1,2} - c), \end{aligned} \quad (10)$$

where $a = 0.165$, $p = 0.2$, and $c = 10$. The parameters $\omega_{1,2} = \omega_0 \pm \Delta$ determine the parameter detuning and ε is the coupling parameter ($\omega_0 = 0.97$, $\Delta = 0.02$). It was shown [21] that the phase synchronization is observed for these control parameter values and a coupling parameter of $\varepsilon = 0.05$.

In this case, the phase of the chaotic signal can be easily introduced in one of the ways mentioned above, because the phase coherent attractor with rather simple topological properties is realized in the system phase space. The attractor projection on the (x, y) plane resembles the smeared limit cycle where the phase point always rotates around the origin (Fig. 1a). The Fourier spectrum $S(f)$ contains the basic frequency peak $f_0 \approx 0.163$ (see Fig. 1b), which coincides with the mean frequency $\bar{f} = \bar{\Omega}/2\pi$ determined from the instantaneous phase $\phi(t)$ dynamics (2). Therefore, the phase synchronization regime can be detected in two coupled Rössler systems (10) by means of traditional approaches without complications.

When the coupling parameter ε is equal to 0.05, the phase synchronization between chaotic oscillators is observed. Phase locking condition (1) is satisfied and the mean frequencies $\bar{\Omega}_{1,2}$ are entrained. Hence, the time scales $s_0 \approx 6$ of both chaotic systems corresponding to the mean frequencies $\bar{\Omega}_{1,2}$ should be correlated with each other. Correspondingly, the phases $\phi_{s1,2}(t)$ associated with these time scales s should be locked and condition (3) should be satisfied. The time scales that are nearest to the time scale s_0 should also be correlated, but the interval of the correlated time scales depends on the coupling strength. At the same time, there should be time scales that remain uncorrelated. These uncorrelated time scales cause a difference between chaotic oscillations of coupled systems.

Figure 2 illustrates the behavior of different time scales for two coupled Rössler systems (10) with phase

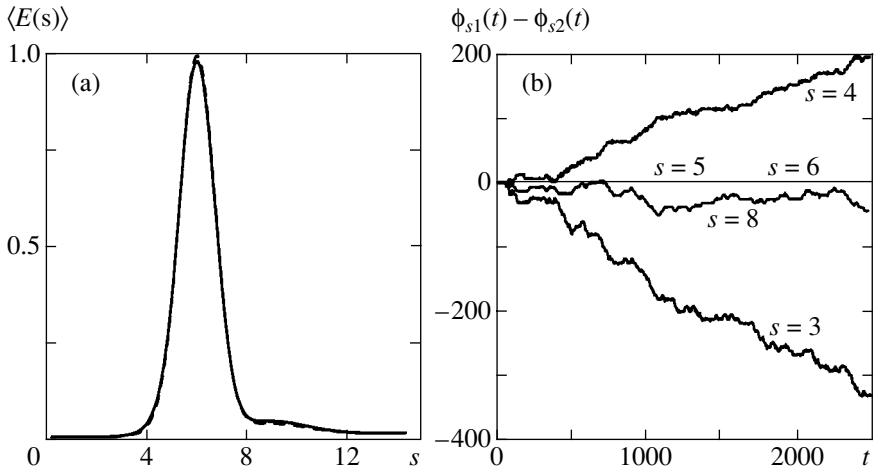


Fig. 2. (a) Wavelet power spectra $\langle E(s)\rangle$ for the first (solid line) and the second (dashed line) Rössler systems (10). (b) The dependence of the phase difference $\phi_{s1}(t) - \phi_{s2}(t)$ on time t for different time scales s. The coupling parameter between the oscillators is $\varepsilon = 0.05$. Phase synchronization for two coupled chaotic oscillators is observed.

coherent attractors. It is clear that the phase difference $\phi_{s1}(t) - \phi_{s2}(t)$ for scales $s_0 = 6$ is bounded, and therefore time scales $s_0 = 6$ corresponding to the main frequency f_0 of the Fourier spectrum are synchronized. It is important to note that the wavelet power spectra $\langle E_{1,2}(s)\rangle$ that are close to each other (Fig. 2a) and time scales s characterized by a large value of energy (e.g., $s = 5$) close to the main time scale $s_0 = 6.0$ are also correlated. There are also time scales that are not synchronized, for example, $s = 3.0$, $s = 4.0$ (Fig. 2b).

Therefore, phase synchronization of two coupled chaotic oscillators with phase coherent attractors manifests itself as a synchronous behavior of the time scales s_0 (and time scales s close to s_0) corresponding to the chaotic signal mean frequency $\bar{\Omega}$.

4. SYNCHRONIZATION OF TWO RÖSSLER SYSTEMS WITH FUNNEL ATTRACTORS

We consider a more complicated example where it is impossible to correctly introduce the instantaneous phase $\phi(t)$ of the chaotic signal $\mathbf{x}(t)$. It is clear that in such cases, the traditional methods of detecting phase synchronization fail and it is necessary to use the other techniques, e.g., indirect measurements [33]. On the contrary, our approach gives correct results and allows detection of the synchronization between chaotic oscillators as easily as before.

As an illustration, we consider two nonidentical coupled Rössler systems with funnel attractors (Fig. 3),

$$\begin{aligned} \dot{x}_{1,2} &= -\omega_{1,2}y_{1,2} - z_{1,2} + \varepsilon(x_{2,1} - x_{1,2}), \\ \dot{y}_{1,2} &= \omega_{1,2}x_{1,2} + ay_{1,2} + \varepsilon(y_{2,1} - y_{1,2}), \\ \dot{z}_{1,2} &= p + z_{1,2}(x_{1,2} - c), \end{aligned} \qquad (11)$$

where ε is a coupling parameter and $\omega_1 = 0.98$, $\omega_2 = 1.03$. The control parameter values have been selected analogously to [33] as $a = 0.22$, $p = 0.1$, and $c = 8.5$. We note that under these control parameter values, none of the methods mentioned above allows defining the phase of the chaotic signal correctly in entire range of the coupling parameter ε variation. Therefore, nobody can determine by means of direct measurements whether the synchronization regime occurs for several values of ε. On the other hand, our approach permits easy detection of synchronization between the coupled oscillators under consideration for all values of the coupling parameter.

In [33], it was shown by means of indirect measurements that for a coupling parameter value of $\varepsilon = 0.05$, synchronization of two coupled Rössler systems (11) occurs. Our approach based on the analysis of the dynamics of different time scales s gives analogous results. The behavior of the phase difference $\phi_{s1}(t) - \phi_{s2}(t)$ for this case is presented in Fig. 4b. One can see that phase locking occurs for time scales $s = 5.25$, which are characterized by the largest energy value in the wavelet power spectra $\langle E(s)\rangle$ (Fig. 4a).

We note that the phase difference $\phi_{s1}(t) - \phi_{s2}(t)$ is also bounded at the time scales close to $s = 5.25$. We can say that the time scales $s = 5.25$ (and close to them) of two oscillators are synchronized with each other. At the same time, other time scales (e.g., $s = 4.5$, 6.0) remain uncorrelated. For such time scales, phase locking was not observed (see Fig. 4b).

It is clear that the mechanism of synchronization of coupled chaotic oscillators is the same in both cases considered in Sections 3 and 4. The synchronization phenomenon is caused by the existence of time scales s in system dynamics correlated with each other. Therefore, there is no reason to divide the considered synchronization examples into different types.

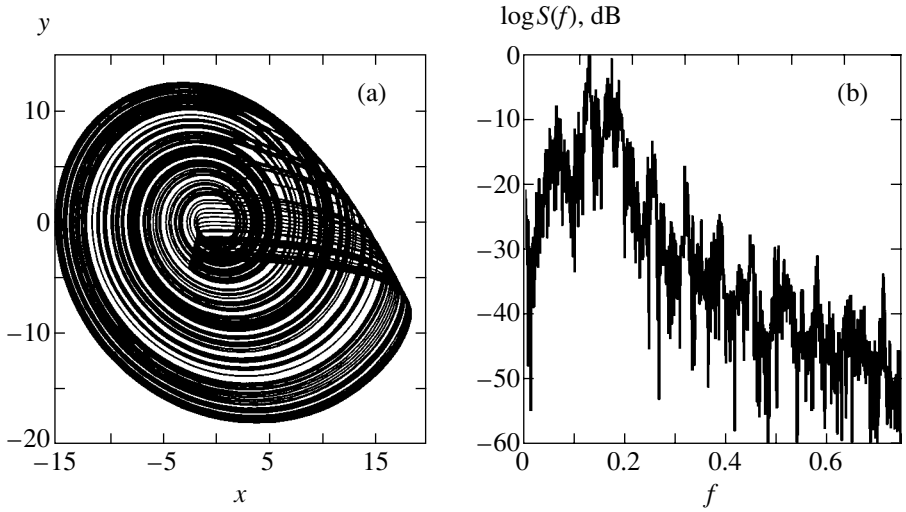


Fig. 3. (a) A phase picture and (b) the power spectrum of oscillation for the first Rössler system (11). The coupling parameter ε is equal to zero.

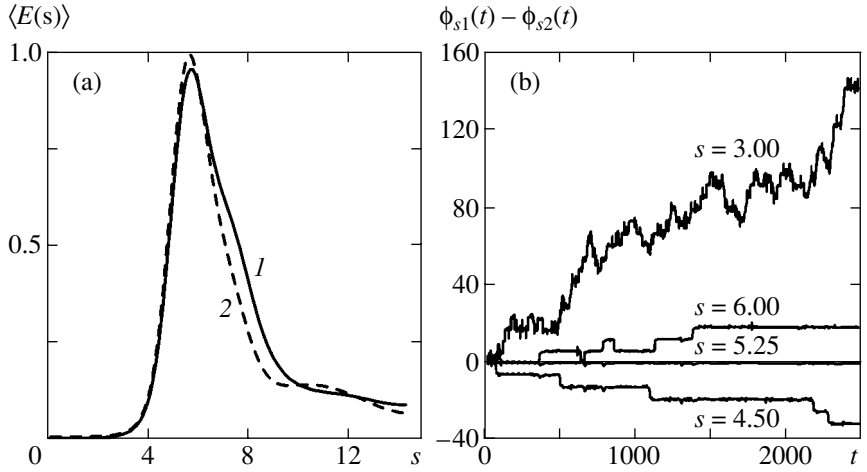


Fig. 4. (a) The normalized energy distribution in the wavelet spectrum $\langle E(s)\rangle$ for the first (line *1*) and the second (line *2*) Rössler systems (11); (b) the phase difference $\phi_{s1}(t) - \phi_{s2}(t)$ for two coupled Rössler systems. The value of the coupling parameter is selected as $\varepsilon = 0.05$. The time scales $s = 5.25$ are correlated with each other and synchronization is observed.

It has been shown [21] that there is a certain relation between phase, lag, and complete synchronizations for chaotic oscillators with slightly mismatched parameters. With the increase of the coupling strength, the systems undergo the transition from unsynchronized chaotic oscillations to phase synchronization. With a further increase in the coupling, lag synchronization is observed. As the coupling parameter increases further, the time lag decreases and both systems tend to have the complete synchronization regime.

We consider the dynamics of different time scales s of two nonidentical coupled Rössler systems (11) when the coupling parameter value increases. If there is no phase synchronization between the oscillators, their dynamics remain uncorrelated for all time scales s. Figure 5 illustrates the dynamics of two coupled Rössler systems when the coupling parameter ε is sufficiently small ($\varepsilon = 0.025$). The power spectra $\langle E(s)\rangle$ of the wavelet transform for Rössler systems differ from each other (Fig. 5a), but the maximum values of the energy correspond approximately to the same time scale s in both systems. It is clear that the phase difference $\phi_{s1}(t) - \phi_{s2}(t)$ is not bounded for almost all time scales (Fig. 5b). One can see that the phase difference $\phi_{s1}(t) - \phi_{s2}(t)$ increases for time scale $s = 3.0$, but decreases for $s = 4.5$. This means that there should be a time scale $3.0 < s^* < 4.5$ at which the phase difference remains bounded. This time scale s^* plays the role of a point separating the time scale areas with the phase difference increasing and decreasing, respectively. In this case, the measure of time scales at which the phase difference remains bounded is zero and we cannot talk about the synchronous behavior of coupled chaotic oscillators (see also Section 6).

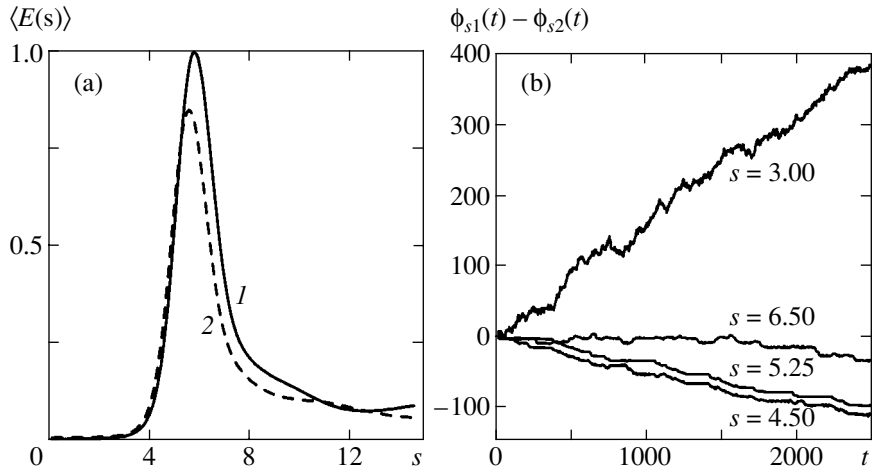


Fig. 5. (a) The normalized energy distribution in the wavelet spectrum $\langle E(s)\rangle$ for the first (line *1*) and the second (line *2*) Rössler systems; (b) the phase difference $\phi_{s1}(t) - \phi_{s2}(t)$ for two coupled Rössler systems. The value of the coupling parameter is selected as $\varepsilon = 0.025$. There is no phase synchronization between the systems.

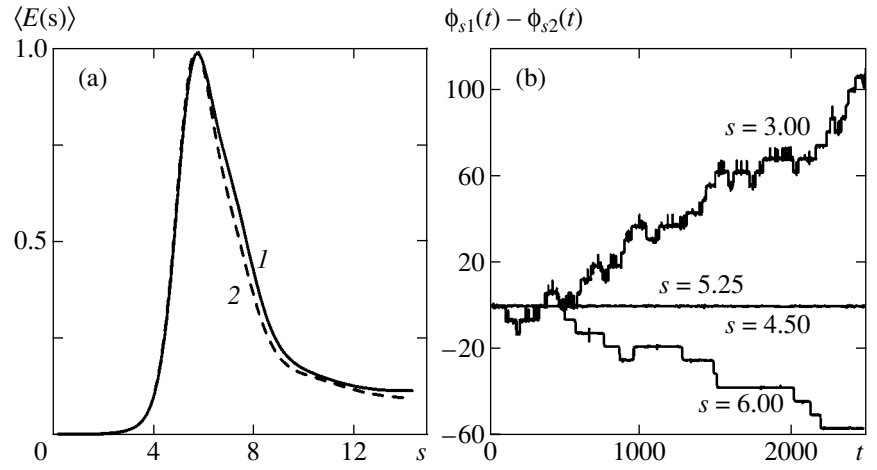


Fig. 6. (a) The normalized energy distribution in the wavelet spectrum $\langle E(s)\rangle$ for the first (line *1*) and the second (line *2*) Rössler systems; (b) the phase difference $\phi_{s1}(t) - \phi_{s2}(t)$ for two coupled Rössler systems. The value of coupling parameter is selected as $\varepsilon = 0.07$.

As soon as any of the time scales of the first chaotic oscillator becomes correlated with another time scale of the second oscillator (e.g., when the coupling parameter increases), phase synchronization occurs (see Fig. 4). The time scales s characterized by the largest value of energy in the wavelet spectrum $\langle E(s)\rangle$ are more likely to be correlated first. The other time scales remain uncorrelated as before. The phase synchronization between chaotic oscillators leads to phase locking (3) at the correlated time scales s.

As the parameter of coupling between the chaotic oscillators increases, more and more time scales become correlated and one can say that the degree of synchronization grows. Therefore, with the further increase of the coupling parameter value (e.g., $\varepsilon = 0.07$) in coupled Rössler systems (11), the time scales that were uncorrelated before become synchronized (Fig. 6b). It is evident that the time scales $s = 4.5$ are synchronized in comparison with the previous case ($\varepsilon = 0.05$, Fig. 4b) when these time scales were uncorrelated. The number of time scales s demonstrating phase locking increases, but there are nonsynchronized time scales as before (e.g., the time scales $s = 3$ and $s = 6$ remain nonsynchronized).

The occurrence of lag synchronization [21] between oscillators means that all time scales are correlated. Indeed, the lag synchronization condition $x_1(t - \tau) \approx x_2(t)$ implies that $W_1(s, t - \tau) \approx W_2(t, s)$ and therefore $\phi_{s1}(t-\tau) \approx \phi_{s2}(t)$. In this case, phase locking condition (3) is obviously satisfied for all time scales s. For instance, when the coupling parameter of chaotic oscillators (11) becomes sufficiently large ($s = 0.25$), lag synchronization of two coupled oscillators occurs. In this case, the power spectra of the wavelet transform coincide with

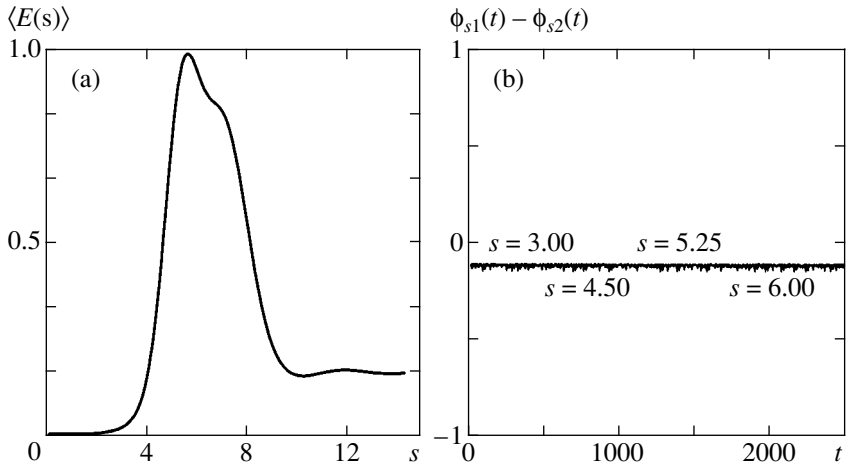


Fig. 7. (a) The normalized energy distribution in the wavelet spectrum $\langle E(s)\rangle$ for the Rössler system; (b) the phase difference $\phi_{s1}(t) - \phi_{s2}(t)$ for two coupled Rössler systems. The value of the coupling parameter is selected as $\varepsilon = 0.25$. Lag synchronization is observed, all time scales are synchronized.

each other (see Fig. 7a) and phase locking takes place for all time scales s (Fig. 7b). We note that the phase difference $\phi_{s1}(t) - \phi_{s2}(t)$ is not equal to zero in the case of lag synchronization. It is clear that this difference depends on the time lag τ.

A further increase of the coupling parameter leads to a decrease of the time lag τ [21]. Both systems tend to have the complete synchronization regime $x_1(t) \approx x_2(t)$, and hence the phase difference $\phi_{s1}(t) - \phi_{s2}(t)$ tends to be zero for all time scales.

The dependence of the synchronized time scale range $[s_m; s_b]$ on the coupling parameter is shown in Fig. 8. The range $[s_m; s_b]$ of synchronized time scales appears at $\varepsilon \approx 0.039$. The appearance of the synchronized time scale range corresponds to the phase synchronization regime. As the coupling parameter value increases, the range of synchronized time scales expands until all time scales become synchronized. Synchronization of all time scales means the presence of the lag synchronization regime.

We can therefore say that the time-scale synchronization is the most general synchronization type unifying (at least) phase, lag, and complete synchronization regimes.

5. GENERALIZED SYNCHRONIZATION REGIME

We consider another type of synchronized behavior, the so-called generalized synchronization. It has been shown above that phase, lag, and complete synchronizations are naturally related to each other and the synchronization type depends on the number of synchronized time scales. The details of the relations between phase and generalized synchronizations are not at all clear. There are several works [1, 22] dealing with the problem of how phase and generalized synchronizations are correlated with each other. For instance, it has been reported in [22] that two unidirectional coupled Rössler systems can demonstrate the generalized synchronization, while the phase synchronization has not been observed. This case can easily be explained by means of the time scale analysis. The equations of the coupled Rössler system are

$$\begin{aligned}
\dot{x}_1 &= -\omega_1 y_1 - z_1,\\
\dot{y}_1 &= \omega_1 x_1 + a y_1,\\
\dot{z}_1 &= p + z_1(x_1 - c)\\
\dot{x}_2 &= -\omega_2 y_2 - z_2 + \varepsilon(x_1 - x_2),\\
\dot{y}_2 &= \omega_2 x_2 + a y_2,\\
\dot{z}_2 &= p + z_2(x_2 - c),
\end{aligned} \tag{12}$$

where $\mathbf{x}_1 = (x_1, y_1, z_1)^T$ and $\mathbf{x}_2 = (x_2, y_2, z_2)^T$ are the respective state vectors of the first (drive) and the second (response) Rössler systems. The control parameter values are chosen as $\omega_1 = 0.8$, $\omega_2 = 1.0$, $a = 0.15$, $p =$

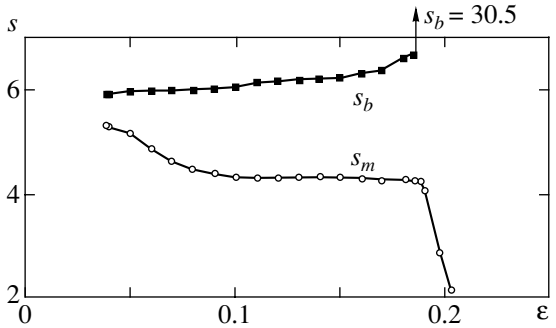


Fig. 8. The dependence of the synchronized time scale range $[s_m; s_b]$ on the coupling strength ε for two coupled Rössler systems (11) with funnel attractors.

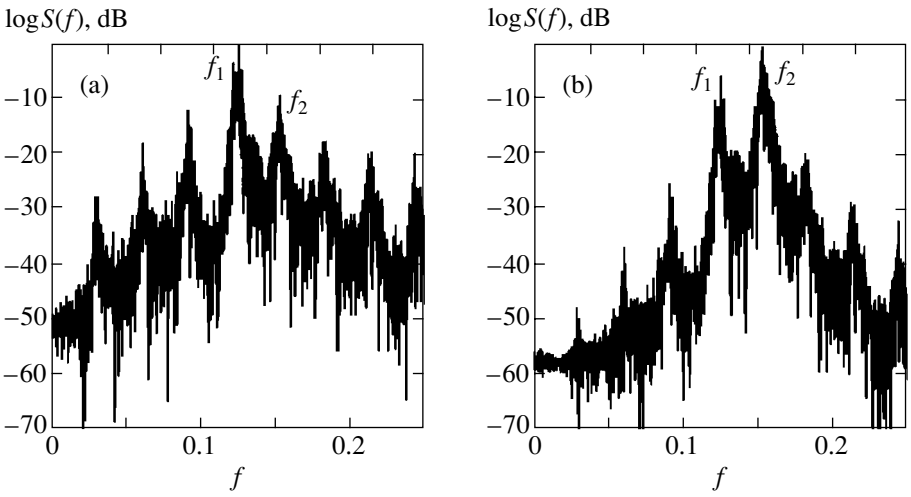


Fig. 9. Fourier spectra for (a) the first (drive) and (b) the second (response) Rössler systems (12). The coupling parameter is ε = 0.2. Generalized synchronization occurs.

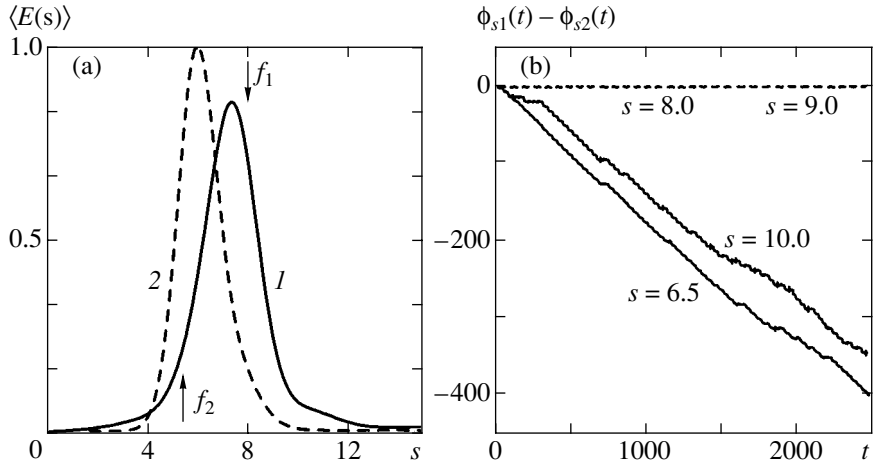


Fig. 10. (a) Normalized energy distribution in the wavelet spectrum $\langle E(s)\rangle$ for the first (line *1*) and the second (line *2*) Rössler systems. The time scales indicated with arrows correspond to the frequencies $f_1 = 0.125$ and $f_2 = 0.154$; (b) the phase difference $\phi_{s1}(t) - \phi_{s2}(t)$ for two coupled Rössler systems. Generalized synchronization is observed.

0.2, $c = 10$, and $\varepsilon = 0.2$. Generalized synchronization occurs in this case (see [22] for details). The time scale analysis explains why it is impossible to detect phase synchronization in system (12) despite generalized synchronization being observed.

We consider Fourier spectra of coupled chaotic oscillators (Fig. 9). There are two main spectral components with the frequencies $f_1 = 0.125$ and $f_2 = 0.154$ in these spectra. The analysis of the behavior of time scales shows that both the time scales $s_1 = 1/f_1 = 8.0$ of the coupled oscillators corresponding to the frequency f_1 and time scales close to s_1 are synchronized, while the time scales $s_2 = 1/f_2 \approx 6.5$ and those close to this value do not demonstrate synchronous behavior (Fig. 10b).

The source of such behavior of time scales becomes clear from the wavelet power spectra $\langle E(s)\rangle$ of both systems (see Fig. 10a). The time scale s_1 of the drive Rössler system is characterized by a large value of energy, while the part of energy associated with this scale of the response system is quite small. Therefore, the drive system dictates its own dynamics at the time scale s_1 to the response system. The opposite situation occurs for the time scales s_2 (Fig. 10a). The drive system cannot dictate its dynamics to the response system because the part of energy associated with this time scale is small in the first Rössler system and large enough in the second one. Therefore, time scales s_2 are not synchronized.

Thus, the generalized synchronization of the unidirectional coupled Rössler systems appears as the time scale synchronized dynamics, similarly to other synchronization types. It is also clear why the phase synchronization was not observed in this case. Figure 9 shows that the instantaneous phases $\phi_{1,2}(t)$ of chaotic

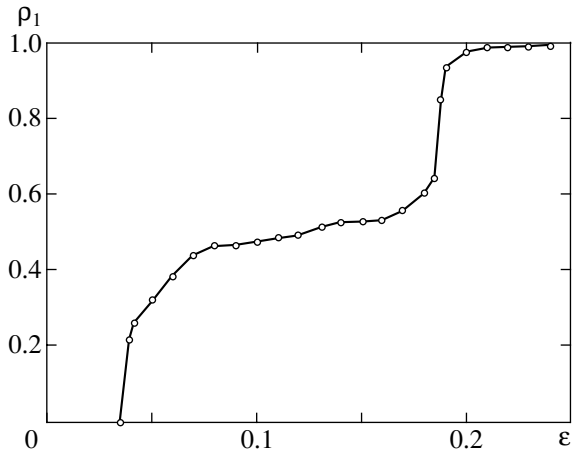


Fig. 11. Dependence of the synchronization measure ρ_1 for the first Rössler system (11) on the coupling strength ε. The measure ρ_2 for the second Rössler oscillator behaves in a similar manner (not shown in the figure).

signals $\mathbf{x}_{1,2}(t)$ introduced by means of traditional approaches are determined by both frequencies, f_1 and f_2, but only the spectral components with the frequency f_1 are synchronized. Therefore, observation of the instantaneous phases $\phi_{1,2}(t)$ does not allow detection of phase synchronization in this case, although the synchronization of time scales takes place.

Thus, one can see that there is a close relation between different types of chaotic oscillator synchronization. According to the results mentioned above, we can say that phase, lag, complete, and generalized synchronizations are particular cases of time-scale synchronization. Therefore, it is possible to consider different types of synchronized behavior from the universal standpoint. Unfortunately, it is not clear how one can distinguish phase synchronization[1] and generalized synchronization using only the results obtained from the analysis of the time scale dynamics.

6. MEASURE OF SYNCHRONIZATION

From the examples given above, we can see that any type of synchronous behavior of coupled chaotic oscillators leads to the occurrence of synchronized time scales. Therefore, the measure of synchronization can be introduced. This measure ρ can be defined as the part of the wavelet spectrum energy associated with the synchronized time scales,

$$\rho_{1,2} = \frac{1}{E_{1,2}}\int_{s_m}^{s_b}\langle E_{1,2}(s)\rangle ds, \qquad (13)$$

where $[s_m; s_b]$ is the range of time scales for which condition (1) is satisfied and

$$E_{1,2} = \int_0^\infty \langle E_{1,2}(s)\rangle ds \qquad (14)$$

is the total energy of the wavelet spectrum. This measure ρ is zero for nonsynchronized oscillations, and unity for the complete and lag synchronization regimes. If the phase synchronization regime is observed, ρ takes a value between zero and unity depending on the part of energy associated with the synchronized time scales. Hence, the synchronization measure ρ makes it possible not only to distinguish synchronized and nonsynchronized oscillations, but also to characterize the degree of time scale synchronization quantitatively.

Figure 11 presents the dependence of time-scale synchronization measure ρ_1 for the first Rössler oscillator of system (11) considered in Section 4 on the coupling parameter ε. It is clear that the part of the energy associated with the synchronized time scales grows monotonically with increasing coupling strength. Similar results have been obtained for the generalized synchronization of two coupled Rössler systems considered in Section 5.

We have already mentioned that when the coupled oscillators do not demonstrate synchronous behavior, there are time scales s^* at which the phase difference $\phi_{s1}(t) - \phi_{s2}(t)$ is bounded. Such time scales play the role of points separating the time scale areas where the phase difference increases and decreases, respectively (see also Section 4). Nevertheless, the presence of such time scales does not mean the occurrence of chaotic synchronization, because the part of energy associated with them is equal to zero. Therefore, the synchronization measure ρ of such oscillations is zero, and the dynamical regime realized in the system in this case should be classified as nonsynchronous.

7. CONCLUSIONS

Summarizing this work, we note several principal aspects. First, we have proposed considering the time scale dynamics of coupled chaotic oscillators. It allows us to examine the different types of behavior of coupled oscillators (such as complete synchronization, lag synchronization, phase synchronization, generalized synchronization, and nonsynchronized oscillations) from the universal standpoint. In this case, time scale synchronization is the most common type of synchronous coupled chaotic oscillator behavior. Therefore, the other types of synchronous oscillations (phase, lag, complete, and generalized) may be considered particular cases of time-scale synchronization. The quantitative characteristic ρ of the synchronization measure has also been introduced. It is important to note that our method (with insignificant modifications) can also be

[1] We mean here that phase synchronization between chaotic oscillators occurs if the instantaneous phase $\phi(t)$ of the chaotic signal may be correctly introduced by means of traditional approaches and phase locking condition (1) is satisfied.

applied to dynamical systems synchronized by the external (e.g., harmonic) signal.

Second, the traditional approach for the phase synchronization detecting based on the introduction of the instantaneous phase $\phi(t)$ of the chaotic signal is suitable and correct for such time series characterized by the Fourier spectrum with a single main frequency f_0. In this case, the phase ϕ_{s0} associated with the time scale s_0 corresponding to the main frequency f_0 of the Fourier spectrum coincides approximately with the instantaneous phase $\phi(t)$ of the chaotic signal introduced by means of the traditional approaches (see also [36]). Indeed, because the other frequencies (the other time scales) do not play a significant role in the Fourier spectrum, the phase $\phi(t)$ of the chaotic signal is close to the phase $\phi_{s0}(t)$ of the main spectral frequency f_0 (and the main time scale s_0, respectively). It is obvious that in this case, the mean frequencies $\bar{f} = \langle \dot{\phi}(t) \rangle / 2\pi$ and $\bar{f}_{s0} = \langle \dot{\phi}_{s0}(t) \rangle / 2\pi$ should coincide with each other and with the main frequency f_0 of the Fourier spectrum (see also [31]),

$$\bar{f} = \bar{f}_{s0} = f_0. \quad (15)$$

If the chaotic time series is characterized by the Fourier spectrum without a single basic frequency (like the spectrum shown in Fig. 3b), the traditional approaches fail. One has to consider the dynamics of the system at all time scales, but this cannot be done by means of the instantaneous phase $\phi(t)$. On the contrary, our approach based on the analysis of time scale dynamics can be used for both types of chaotic signals.

Finally, our approach can be easily applied to the experimental data because it does not require any a priori information on the dynamical systems considered. Moreover, in several cases, the influence of noise can be reduced by means of the wavelet transform (see [39, 48, 49] for details). We believe that our approach will be useful and effective for the analysis of physical, biological, physiological, and other data, such as described in [9, 10, 36, 35].

ACKNOWLEDGMENTS

We thank D. I. Trubetskov, V. S. Anishchenko, and T. E. Vadivasova for valuable discussions. We also thank S. V. Eremina for language support.

This work was supported by the CRDF (grant no. REC-006), the Russian Foundation for Basic Research (project no. 02-02-16351), and the Scientific Program "Universities of Russia." We also thank the Dynasty Foundation.

REFERENCES

1. U. Parlitz, L. Junge, and W. Lauterborn, Phys. Rev. E **54**, 2115 (1996).
2. D. Y. Tang, R. Dykstra, M. W. Hamilton, and N. R. Heckenberg, Phys. Rev. E **57**, 3649 (1998).
3. E. Allaria, F. T. Arecchi, A. D. Garbo, and R. Meucci, Phys. Rev. Lett. **86**, 791 (2001).
4. C. M. Ticos, E. Rosa, W. B. Pardo, *et al.*, Phys. Rev. Lett. **85**, 2929 (2000).
5. E. Rosa, W. B. Pardo, C. M. Ticos, *et al.*, Int. J. Bifurcations Chaos **10**, 2551 (2000).
6. D. I. Trubetskov and A. E. Hramov, J. Commun. Technol. Electron. **48**, 105 (2003).
7. P. A. Tass, M. G. Rosenblum, J. Weule, *et al.*, Phys. Rev. Lett. **81**, 3291 (1998).
8. V. S. Anishchenko, A. G. Balanov, N. B. Janson, *et al.*, Int. J. Bifurcations Chaos **10**, 2339 (2000).
9. M. D. Prokhorov, V. I. Ponomarenko, V. I. Gridnev, *et al.*, Phys. Rev. E **68**, 041913 (2003).
10. R. C. Elson, A. I. Selverston, R. Huerta, *et al.*, Phys. Rev. Lett. **81**, 5692 (1998).
11. N. F. Rulkov, Phys. Rev. E **65**, 041922 (2002).
12. P. A. Tass, Phys. Rev. Lett. **90**, 088101 (2003).
13. A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences* (Cambridge Univ. Press, Cambridge, 2001).
14. A. Pikovsky, M. Rosenblum, and J. Kurths, Int. J. Bifurcations Chaos **10**, 2291 (2000).
15. V. S. Anishchenko and T. E. Vadivasova, J. Commun. Technol. Electron. **47**, 117 (2002).
16. V. S. Anishchenko, V. Astakhov, A. Neiman, *et al.*, *Nonlinear Dynamics of Chaotic and Stochastic Systems. Tutorial and Modern Developments* (Springer, Heidelberg, 2001).
17. L. M. Pecora and T. L. Carroll, Phys. Rev. Lett. **64**, 821 (1990).
18. L. M. Pecora and T. L. Carroll, Phys. Rev. A **44**, 2374 (1991).
19. K. Murali and M. Lakshmanan, Phys. Rev. E **49**, 4882 (1994).
20. K. Murali and M. Lakshmanan, Phys. Rev. E **48**, R1624 (1993).
21. M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, Phys. Rev. Lett. **78**, 4193 (1997).
22. Z. Zheng and G. Hu, Phys. Rev. E **62**, 7882 (2000).
23. S. Taherion and Y. C. Lai, Phys. Rev. E **59**, R6247 (1999).
24. N. F. Rulkov, M. M. Sushchik, L. S. Tsimring, and H. D. I. Abarbanel, Phys. Rev. E **51**, 980 (1995).
25. L. Kocarev and U. Parlitz, Phys. Rev. Lett. **76**, 1816 (1996).
26. K. Pyragas, Phys. Rev. E **54**, R4508 (1996).
27. M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, Phys. Rev. Lett. **76**, 1804 (1996).
28. G. V. Osipov, A. S. Pikovsky, M. G. Rosenblum, and J. Kurth, Phys. Rev. E **55**, 2353 (1997).
29. H. D. I. Abarbanel, N. F. Rulkov, and M. M. Sushchik, Phys. Rev. E **53**, 4528 (1996).
30. L. M. Pecora, T. L. Carroll, and J. F. Heagy, Phys. Rev. E **52**, 3420 (1995).

31. V. S. Anishchenko and T. E. Vadivasova, J. Commun. Technol. Electron. **49**, 69 (2004).
32. A. Pikovsky, M. Rosenblum, G. Osipov, and J. Kurths, Physica D **104**, 219 (1997).
33. M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, Phys. Rev. Lett. **89**, 264102 (2002).
34. G. V. Osipov, B. Hu, C. Zhou, *et al.*, Phys. Rev. Lett. **91**, 024101 (2003).
35. J. P. Lachaux, E. Rodriguez, M. V. Quyen, *et al.*, Int. J. Bifurcations Chaos **10**, 2429 (2000).
36. R. Q. Quiroga, A. Kraskov, T. Kreuz, and P. Grassberger, Phys. Rev. E **65**, 041903 (2002).
37. A. S. Pikovsky, M. G. Rosenblum, and J. Kurths, Europhys. Lett. **34**, 165 (1996).
38. A. A. Koronovskii and A. E. Hramov, Pis'ma Zh. Éksp. Teor. Fiz. **79**, 391 (2004) [JETP Lett. **79**, 316 (2004)].
39. A. A. Koronovskii and A. E. Hramov, *Continuous Wavelet Analysis and Its Applications* (Fizmatlit, Moscow, 2003) [in Russian].
40. I. Daubechies, *Ten Lectures on Wavelets* (SIAM, Philadelphia, 1992; RKhD, Moscow, 2001).
41. G. Kaiser, *A Friendly Guide to Wavelets* (Springer, New York, 1994).
42. B. Torresani, *Continuous Wavelet Transform* (Savoire, Paris, 1995).
43. A. Lutz, D. Rudrauf, D. Cosmelli, *et al.*, Neurophysiol. Clin. **32**, 157 (2002).
44. M. L. V. Quyen, J. Martinerie, C. Adam, and F. J. Varela, J. Neurosci. Methods **111**, 83 (2001).
45. D. J. De Shazer, R. Breban, E. Ott, and R. Roy, Phys. Rev. Lett. **87**, 044101 (2001).
46. O. V. Sosnovtseva, A. N. Pavlov, E. Mosekilde, and N.-H. Holstein-Rathlou, Phys. Rev. E **66**, 061909 (2002).
47. A. Grossman and J. Morlet, SIAM J. Math. Anal. **15**, 273 (1984).
48. C. Torrence and G. P. Compo, Bull. Am. Meteorol. Soc. **79**, 61 (1998).
49. V. A. Gusev, A. A. Koronovskiy, and A. E. Hramov, Pis'ma Zh. Tekh. Fiz. **29** (18), 61 (2003) [Tech. Phys. Lett. **29**, 775 (2003)].

Journal of Experimental and Theoretical Physics, Vol. 100, No. 4, 2005, pp. 795–803.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 4, 2005, pp. 898–906.

STATISTICAL, NONLINEAR, AND SOFT MATTER PHYSICS

Inhomogeneous Threshold Director Reorientation in a Planar Nematic Flexoelectric Cell with Finite Anchoring Energy

M. F. Ledney and I. P. Pinkevich

Taras Shevchenko Kiev National University, Kiev, 03680 Ukraine
e-mail: ledney@univ.kiev.ua

Received June 24, 2004

Abstract—The dependence of the threshold parameters and the period of the electric-field-induced spatially periodic reorientation of the director in a flexoelectric nematic liquid crystal (NLC) on the anchoring conditions at the surface of a planar NLC cell has been studied. The threshold electric field and the corresponding wavenumber of the periodic structure of the director field have been numerically calculated for arbitrary values of the anchoring energy. In the case of strong anchoring, the corresponding analytical expressions are obtained in a single-constant approximation. A decrease in the azimuthal anchoring energy leads to an increase in the intervals of possible values of the flexoelectric parameter ν and the ratio K_2/K_1 of the Frank elastic constants. A decrease in the polar anchoring energy leads to narrowing of these intervals as compared to the case of infinitely strong anchoring at the NLC cell surface. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Considerable interest in the physics of phenomena in liquid crystal cells under the action of external fields—in particular, in the process of threshold reorientation of the director of a nematic liquid crystal (NLC) in the applied electric field (known as the Fréedericksz effect)—are related to the wide use of such cells in various electrooptical devices [1–3]. Although the threshold reorientation of the NLC director is a volume effect, important characteristics such as the threshold current and the degree of the director reorientation significantly depend on the interaction between the liquid crystal and the cell surface. The effect of the cell surface on these characteristics is so strong that even a spontaneous Fréedericksz transition (in fact, stimulated by a change in conditions at this surface) may take place [4–7]. One of the most important parameters determining the conditions at the NLC cell surface and influencing the NLC director behavior is the anchoring energy. The model of infinitely strong anchoring offers the simplest case, while allowance for a finite anchoring energy significantly complicates description of this system.

It was established that, under certain conditions, a threshold reorientation of the NLC director is accompanied by the formation of a periodic spatial structure in the plane of the NLC cell. Previously, Bobylev *et al.* [8, 9] described this phenomenon in a planar-oriented cell of a flexoelectric NLC with absolutely rigid (infinitely strong anchoring) boundary conditions. Romanov and Sklyarenko [10] studied the influence of the surface conditions on the electric-field-induced threshold periodic structure of the director field, but only in a homeotropically oriented cell of a flexoelectric NLC. The values of the threshold electric field and the spatial period in the director reorientation were determined as functions of the anchoring energy and the flexoelectric coefficients of the NLC. Recently, Barbero and Lelidis [11] studied the possibility of the formation of periodic structures of flexoelectric origin in a homeotropic NLC cell.

Lonberg and Meyer [12] showed that a periodic spatial structure of the director field can arise in a homeotropic NLC cell with absolutely rigid boundary conditions even in the absence of the flexoelectric polarization, depending only on the ratio of the Frank elastic constants K_1 and K_2. When $K_2/K_1 < r_o \approx 0.3$, the system exhibits a Fréedericksz transition with the formation of a periodic spatial structure, whereas only a homogeneous Fréedericksz transition is possible for $K_2/K_1 > r_o$. It was demonstrated [13–15] that the character of the periodic structure formed in this geometry strongly depends on the anchoring energy.

Simões *et al.* [16, 17] theoretically and experimentally investigated periodic structures appearing in the cells of lyotropic NLCs when the external magnetic field exceeded a threshold for the Fréedericksz transition. The influence of the elastic constant K_{24} on the spontaneous periodic distortions in a planar NLC cell was studied in [18–21], while a relationship between K_{24} and the parameters of a periodic structure formed during the Fréedericksz transition in the external magnetic field was analyzed in [22].

1063-7761/05/10004-0795$26.00

The aim of this study was to consider the effect of the anchoring energy on the periodic spatial structure of the director field in a planar flexoelectric NLC cell.

2. EQUATIONS DESCRIBING THE DIRECTOR FIELD

Let us consider a plane-parallel flexoelectric NLC cell bounded by the planes $z = -L/2$ and $z = +L/2$ with an initial planar director orientation along the x axis. The cell is exposed to an external homogeneous electric field $\mathbf{E} = (0, 0, E)$. The free energy of this NLC cell can be expressed as

$$F = F_{el} + F_E + F_d + F_S,$$

$$F_{el} = \frac{1}{2}\int_V \{K_1(\mathrm{div}\mathbf{n})^2 + K_2(\mathbf{n}\cdot\mathrm{curl}\mathbf{n})^2 + K_3[\mathbf{n}\times\mathrm{curl}\mathbf{n}]^2\}dV,$$

$$F_E = -\frac{\varepsilon_a}{8\pi}\int_V (\mathbf{n}\cdot\mathbf{E})^2 dV, \qquad (1)$$

$$F_d = -\int_V \{e_1(\mathbf{n}\cdot\mathbf{E})\mathrm{div}\mathbf{n} + e_3([\mathrm{curl}\mathbf{n}\times\mathbf{n}]\cdot\mathbf{E})\}dV,$$

$$F_S = -\frac{W_\varphi}{2}\int_{S_{1,2}} \cos^2\varphi dS - \frac{W_\theta}{2}\int_{S_{1,2}} \cos^2\theta dS,$$

$$W_\varphi > 0, \quad W_\theta > 0.$$

Here, $\mathbf{n}$ is the NLC director; F_{el} is the Frank elastic energy; F_E and F_d are the anisotropic and flexoelectric contributions to the energy of NLC interaction with the electric field; F_S is the free surface energy of the NLC; $\varepsilon_a = \varepsilon_\parallel - \varepsilon_\perp > 0$ is the anisotropy of the static permittivity; e_1 and e_3 are the flexoelectric coefficients; W_θ and W_φ are the polar and azimuthal anchoring energies at the cell surface, respectively; and θ and φ are the director tilt angles in the xz and xy planes, respectively.

It should be noted that, since we are interested in the influence of a finite anchoring energy of the NLC at the cell surface, the surface free energy in Eqs. (1) has been written in terms of the simplest (but still most frequently used) Rapini–Papoular model [23]. According to this, the energy of the NLC–surface interaction is proportional to square of the angle between the director and its easy axis on the cell surface. However, we have also taken into account that a change in the surface energy can be different for director rotation relative to the easy axis in the azimuthal and polar directions [24].

Since the threshold reorientation of the director in the planar geometry leads to the appearance of a periodic spatial structure along the y axis [8, 9], we seek the director in the following form:

$$\mathbf{n} = \mathbf{i}\cos\theta(y,z)\cos\varphi(y,z) + \mathbf{j}\cos\theta(y,z)\sin\varphi(y,z) + \mathbf{k}\sin\theta(y,z), \qquad (2)$$

where $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ are the unit vectors of the Cartesian coordinate system.

In the case of small distortions of the director field ($|\varphi|, |\theta| \ll 1$), minimization of the free energy (1) with respect to (θ, φ gives the set of stationary equations

$$r\frac{\partial^2\theta}{\partial y^2} + \frac{\partial^2\theta}{\partial z^2} + \epsilon E^2\theta + (1-r)\frac{\partial^2\varphi}{\partial y\partial z} + eE\frac{\partial\varphi}{\partial y} = 0,$$
$$\frac{\partial^2\varphi}{\partial y^2} + r\frac{\partial^2\varphi}{\partial z^2} + (1-r)\frac{\partial^2\theta}{\partial y\partial z} - eE\frac{\partial\theta}{\partial y} = 0 \qquad (3)$$

and the corresponding boundary conditions

$$\left[\left(\frac{W_\theta}{K_1} \mp e_o E\right)\theta \pm \left(\frac{\partial\theta}{\partial z} + \frac{\partial\varphi}{\partial y}\right)\right]_{z=\pm L/2} = 0,$$
$$\left[\frac{W_\varphi}{K_1}\varphi \pm r\left(\frac{\partial\varphi}{\partial z} - \frac{\partial\theta}{\partial y}\right)\right]_{z=\pm L/2} = 0, \qquad (4)$$

where

$$\epsilon = \frac{\varepsilon_a}{4\pi K_1}, \quad r = \frac{K_2}{K_1},$$
$$e = \frac{e_1 - e_3}{K_1}, \quad e_o = \frac{e_1 + e_3}{K_1}.$$

Taking into account the symmetry of Eqs. (3), a solution to this system can be found in the following form:

$$\theta(y,z) = \cos(qy)\theta_1(z), \quad \varphi(y,z) = \sin(qy)\varphi_1(z), \qquad (5)$$

where the functions $\theta_1(z)$ and $\varphi_1(z)$ meet the condition

$$\begin{pmatrix} \frac{d^2}{dz^2} - rq^2 + \epsilon E^2 & (1-r)q\frac{d}{dz} + eEq \\ eEq - (1-r)q\frac{d}{dz} & r\frac{d^2}{dz^2} - q^2 \end{pmatrix} \times \begin{pmatrix} \theta_1(z) \\ \varphi_1(z) \end{pmatrix} = 0. \qquad (6)$$

Substituting in these relations

$$\begin{pmatrix} \theta_1(z) \\ \varphi_1(z) \end{pmatrix} = e^{\lambda z}\begin{pmatrix} \theta_{10} \\ \varphi_{10} \end{pmatrix}, \qquad (7)$$

we obtain a homogeneous system of two algebraic equations for determining the unknown coefficients θ_{10} and φ_{10}. This system has a nontrivial solution provided that

$$\begin{aligned}&(\lambda^2 - q^2) + \epsilon E^2(\lambda^2 - q^2)\\ &-\frac{1}{r}[\epsilon E^2 q^2(1-r) + (eEq)^2] = 0, \quad r \neq 0.\end{aligned} \tag{8}$$

Solving Eq. (8), we obtain $\lambda = \pm ip_1$ and $\pm p_2$, where p_1 and p_2 are real quantities given by the formulas

$$\begin{aligned}p_1 &= \left\{\frac{1}{2}\left[\epsilon E^2 + \left((\epsilon E^2)^2\right.\right.\right.\\ &\left.\left.\left. + 4q^2\epsilon E^2\frac{1-r+1/\nu}{r}\right)^{1/2}\right] - q^2\right\}^{1/2},\\ p_2 &= \left\{q^2 - \frac{1}{2}\left[\epsilon E^2 - \left((\epsilon E^2)^2\right.\right.\right.\\ &\left.\left.\left. + 4q^2\epsilon E^2\frac{1-r+1/\nu}{r}\right)^{1/2}\right]\right\}^{1/2},\end{aligned} \tag{9}$$

and $\nu = \epsilon/e^2$.

Equations (6) yield

$$\frac{\varphi_{10}}{\theta_{10}} = \frac{eEq - (1-r)q\lambda}{q^2 - r\lambda^2}.$$

Taking this into account, the total solution of Eqs. (6) can be written as

$$\begin{aligned}\theta_1(z) &= a_1\cos(p_1 z) + a_2\sin(p_1 z)\\ &+ b_1\cosh(p_2 z) + b_2\sinh(p_2 z),\\ \varphi_1(z) &= a_1(\alpha_1\cos(p_1 z) - \beta_1\sin(p_1 z))\\ &+ a_2(\alpha_1\sin(p_1 z) + \beta_1\cos(p_1 z))\\ &+ b_1(\alpha_2\cosh(p_2 z) + \beta_2\sinh(p_2 z))\\ &+ b_2(\alpha_2\sinh(p_2 z) + \beta_2\cosh(p_2 z)),\end{aligned} \tag{10}$$

where

$$\alpha_1 = \frac{eEq}{q^2 + rp_1^2}, \quad \alpha_2 = \frac{eEq}{q^2 - rp_2^2}, \tag{11}$$

$$\beta_1 = -\frac{(1-r)qp_1}{q^2 + rp_1^2}, \quad \beta_2 = -\frac{(1-r)qp_2}{q^2 - rp_2^2}, \tag{12}$$

and a_i, b_i ($i = 1, 2$) are arbitrary constants whose values are determined from the boundary conditions (4).

3. DEPENDENCE OF THE DIRECTOR REORIENTATION THRESHOLD ON THE AZIMUTHAL ANCHORING ENERGY

Let us assume that the polar anchoring energy at the cell surface is infinitely large ($W_\theta = \infty$), while the azimuthal energy W_φ may have an arbitrary value. In this case, the boundary conditions (4) take the following form:

$$\begin{aligned}&\theta_1\big|_{z = \pm L/2} = 0,\\ &\left[\frac{W_\varphi}{K_1}\varphi_1 \pm r\frac{d\varphi_1}{dz}\right]_{z = \pm L/2} = 0.\end{aligned} \tag{13}$$

Substituting solution (10) into the boundary conditions (13), we obtain a homogeneous system of four algebraic equations for determining the unknown coefficients a_i, b_i ($i = 1, 2$). This system has a nontrivial solution provided that

$$\begin{aligned}&\left[\beta_2\cot\frac{p_1 L}{2}\left(\frac{W_\varphi}{K_1} + rp_2\coth\frac{p_2 L}{2}\right)\right.\\ &\left. + \beta_1\coth\frac{p_2 L}{2}\left(\frac{W_\varphi}{K_1} + rp_1\cot\frac{p_1 L}{2}\right)\right]\\ &\times\left[\beta_2\left(\frac{W_\varphi}{K_1}\coth\frac{p_2 L}{2} + rp_2\right)\right.\\ &\left. - \beta_1\left(\frac{W_\varphi}{K_1}\cot\frac{p_1 L}{2} - rp_1\right)\right]\\ &+\left[\alpha_2\cot\frac{p_1 L}{2}\left(\frac{W_\varphi}{K_1}\coth\frac{p_2 L}{2} + rp_2\right)\right.\\ &\left. - \alpha_1\coth\frac{p_2 L}{2}\left(\frac{W_\varphi}{K_1}\cot\frac{p_1 L}{2} - rp_1\right)\right]\\ &\times\left[\alpha_1\left(\frac{W_\varphi}{K_1} + rp_1\cot\frac{p_1 L}{2}\right)\right.\\ &\left. - \alpha_2\left(\frac{W_\varphi}{K_1} + rp_2\coth\frac{p_2 L}{2}\right)\right] = 0.\end{aligned} \tag{14}$$

Solving Eq. (14), we find the electric field E as a function of the parameter q. A minimum on the $E(q)$ curve determines the threshold (critical) field E_c for the onset of instability.

In the general case, Eq. (14) admits only numerical solution. It should be noted that a necessary condition for the spatially periodic reorientation of the director is

$$\left.\frac{dE}{dq}\right|_{q = 0} < 0. \tag{15}$$

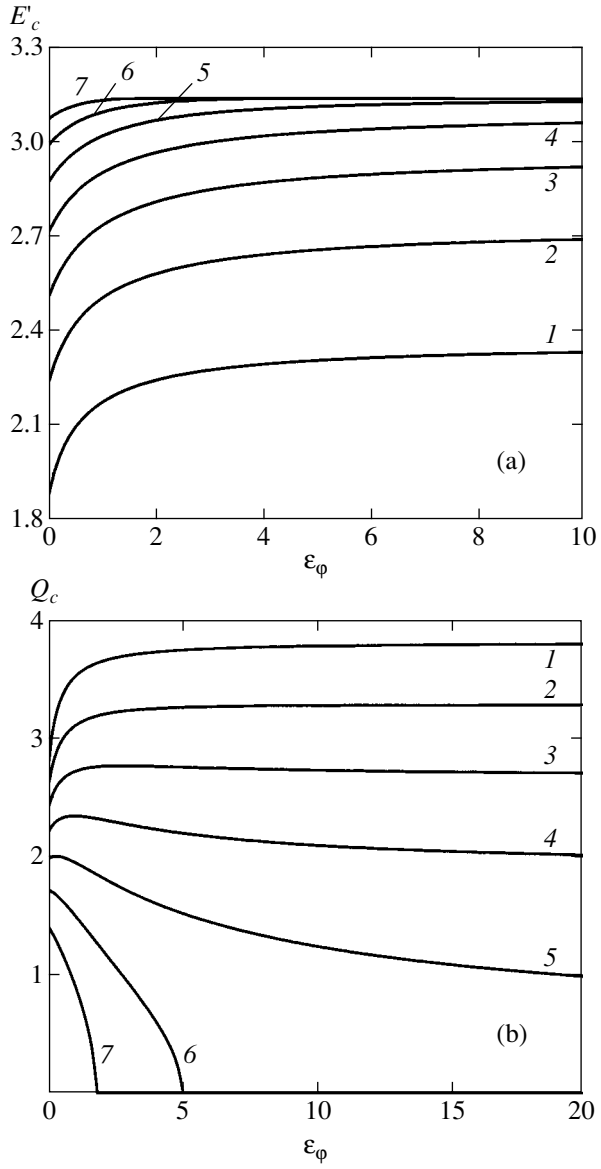


Fig. 1. Plots of (a) the dimensionless critical field E_c' and (b) the dimensionless critical wavenumber Q_c versus azimuthal anchoring energy ε_φ in the absence of flexoelectric polarization ($\nu = \infty$) for $r = 0.1$ (*1*), 0.15 (*2*), 0.2 (*3*), 0.25 (*4*), 0.3 (*5*), 0.35 (*6*), and 0.4 (*7*).

Differentiating Eq. (14) with respect to q, we obtain an expression for $dE/dq|_{q=0}$ and, taking into account condition (15), we arrive at the following relation between the NLC parameters r, ν, and $\varepsilon_\varphi = W_\varphi L/K_1$ in an NLC featuring a periodic spatial structure of the director field:

$$(1-r)^2 - \frac{\pi^2}{8}\left(1 - 2r + \frac{1}{\nu} + \frac{4r}{\nu\varepsilon_\varphi}\right)\left(1 + \frac{2r}{\varepsilon_\varphi}\right) < 0. \quad (16)$$

(a) NLC cell in the absence of flexoelectric polarization. In this case, Eq. (14) simplifies to

$$\beta_2 \cot\frac{p_1 L}{2}\left(\frac{W_\varphi}{K_1} + r p_2 \coth\frac{p_2 L}{2}\right) + \beta_1 \coth\frac{p_2 L}{2}\left(\frac{W_\varphi}{K_1} + r p_1 \cot\frac{p_1 L}{2}\right) = 0. \quad (17)$$

Solving this equation, we determine the values of the critical electric field E_c and the corresponding wavenumber q_c for arbitrary values of the dimensionless azimuthal anchoring energy ε_φ and the ratio r of the Frank elastic constants.

Figure 1 shows the calculated dependences of the dimensionless critical electric field $E_c' = \sqrt{\epsilon}\, E_c L$ and the corresponding dimensionless wavenumber $Q_c = q_c L$ on the azimuthal anchoring energy ε_φ for various values of the parameter r. As can be seen, critical electric field increases with the azimuthal anchoring energy ε_φ.

As can be seen, the period $\lambda_c = 2\pi/q_c$ of the spatial structure of the director field decreases with increasing ε_φ for $r \lesssim 0.2$ and increases for greater values of this parameter. The values of λ_c tend to a certain finite level as $\varepsilon_\varphi \longrightarrow \infty$, provided that

$$r < r_o = 1 - \frac{\pi^2}{8} + \frac{\pi}{8}\sqrt{\pi^2 - 8} \approx 0.3.$$

For each r from the interval $r_o < r < 0.5$, relation (16) gives a certain limiting (threshold) azimuthal anchoring energy

$$\varepsilon_{\varphi th}(r) = \frac{2\pi^2 r(1-2r)}{8(1-r)^2 - \pi^2(1-2r)}. \quad (18)$$

For $\varepsilon_\varphi < \varepsilon_{\varphi th}$, the NLC cell exhibits a Fréedericksz transition with the formation of a periodic spatial structure, while for $\varepsilon_\varphi > \varepsilon_{\varphi th}$ only a Fréedericksz transition with a uniform (along the y axis) director distribution can take place. For values of parameter $r \geq 0.5$, only a homogeneous Fréedericksz transition is possible in agreement with the results of Oldano [13].

Figure 2 shows the plots of the threshold electric field E_c' and the dimensionless wavenumber Q_c versus the ratio r of the Frank elastic constants for various values of the azimuthal anchoring energy ε_φ. As can be seen, both the critical electric field and the corresponding period λ_c of the arising structure of the director field increase with the parameter r. For each given value of the azimuthal anchoring energy ε_φ, relation (16) gives a certain threshold ratio of the Frank elastic constants

$$r_{th}(\varepsilon_\varphi) = \left\{1 - \frac{\pi^2}{8} + \frac{\pi^2}{8\varepsilon_\varphi} + \left[\left(1 - \frac{\pi^2}{8} + \frac{\pi^2}{8\varepsilon_\varphi}\right)^2 - \left(1 + \frac{\pi^2}{2\varepsilon_\varphi}\right)\left(1 - \frac{\pi^2}{8}\right)\right]^{1/2}\right\}\left(1 + \frac{\pi^2}{2\varepsilon_\varphi}\right)^{-1}. \quad (19)$$

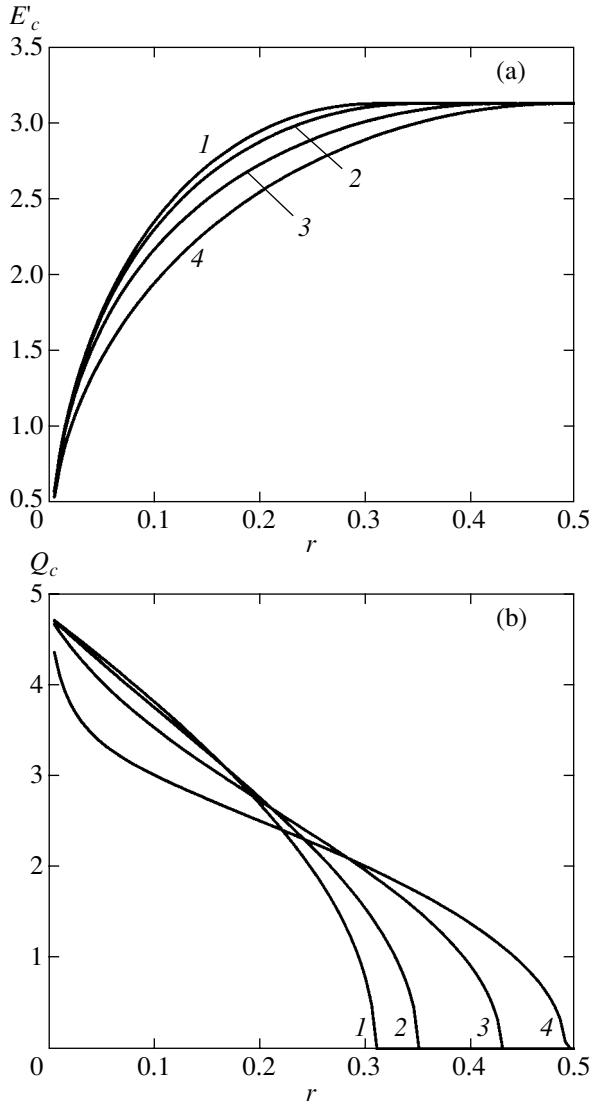


Fig. 2. Plots of (a) the critical field E'_c and (b) the critical wavenumber Q_c versus parameter r in the absence of flexoelectric polarization ($\nu = \infty$) for the azimuthal anchoring energy $\varepsilon_\varphi = 50$ (*1*), 5 (*2*), 1 (*3*), and 0.1 (*4*).

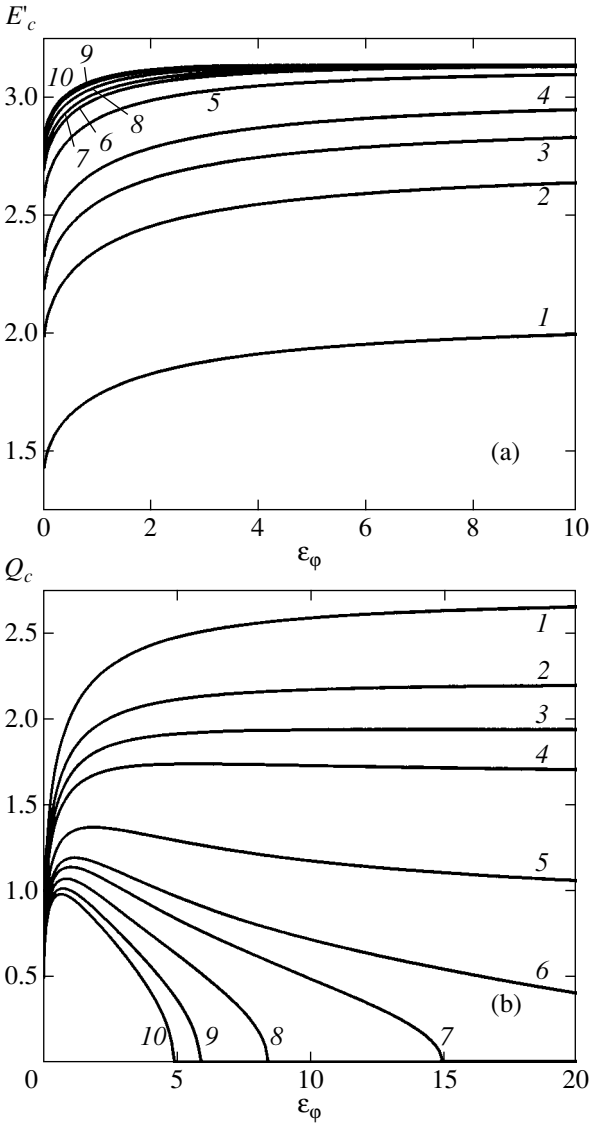


Fig. 3. Plots of (a) the critical field E'_c and (b) the critical wavenumber Q_c versus azimuthal anchoring energy ε_φ for $r = 0.75$ and $\nu = 0.2$ (*1*), 0.5 (*2*), 0.7 (*3*), 0.9 (*4*), 1.5 (*5*), 2.0 (*6*), 2.2 (*7*), 2.5 (*8*), 2.8 (*9*), and 3.0 (*10*).

For $r < r_{th}$, the NLC cell exhibits a Fréedericksz transition with the formation of a periodic spatial structure, while for $r > r_{th}$ only a homogeneous Fréedericksz transition is possible in agreement with the results obtained in [13, 15]. In the limiting case of the infinitely strong anchoring ($\varepsilon_\varphi \longrightarrow \infty$), formula (19) shows that the threshold ratio of the Frank elastic constants is $r_{th} = r_o$ [12].

(b) NLC cell with flexoelectric polarization. In the case of strong (but not infinitely strong) azimuthal anchoring ($\varepsilon_\varphi \gg 1$), solving Eq. (14) in a single constant approximation ($K_1 = K_2$) yields the critical electric field

$$E_c = E(q_c) = \frac{2\pi}{|e|(1+\nu)L}\left(1 - \frac{1-\nu}{\varepsilon_\varphi}\right). \qquad (20)$$

The corresponding wavenumber is

$$q_c = \frac{\pi}{L}\sqrt{\frac{1-\nu}{1+\nu}}\left(1 + \frac{1}{\varepsilon_\varphi}\frac{3\nu-1}{1-\nu}\right),$$

for $\nu < 1$ and $\varepsilon_\varphi(1-\nu) \ll 1$, and this wavenumber is

$$q_c = \frac{\pi}{\sqrt{2}L}\sqrt{\frac{4}{\varepsilon_\varphi} + 1 - \nu},$$

for $|1-\nu| \ll 1$.

For arbitrary values of the dimensionless azimuthal anchoring energy ε_φ and the $K_1 \neq K_2$, Eq. (14) was solved by numerical methods. Figure 3 shows the cal-

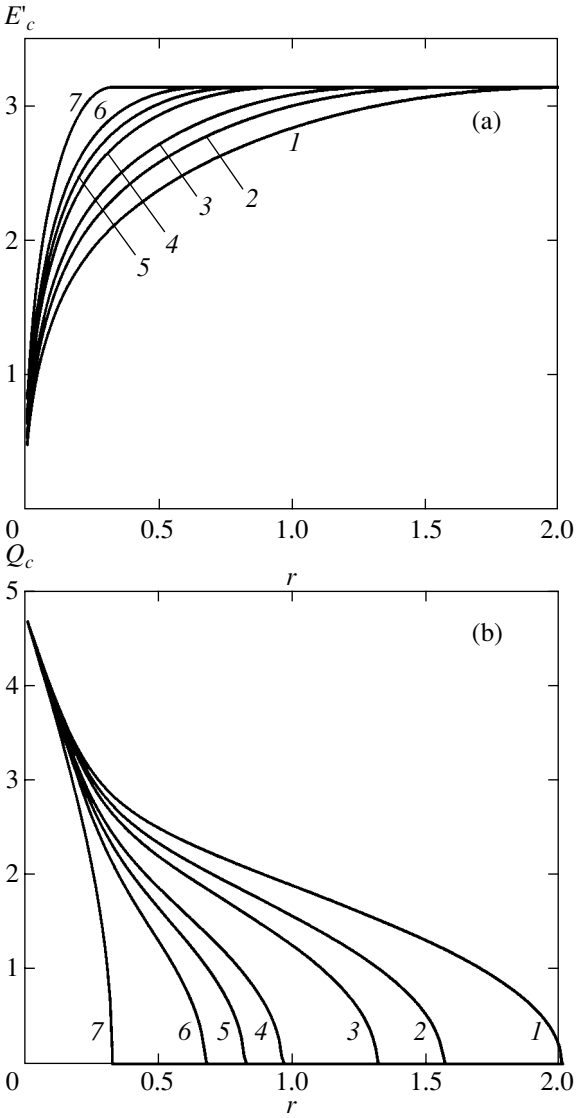


Fig. 4. Plots of (a) the critical field E'_c and (b) the critical wavenumber Q_c versus parameter r for $\varepsilon_\varphi = 10$ and $\nu = 0.5$ (*1*), 0.7 (*2*), 0.9 (*3*), 1.5 (*4*), 2.0 (*5*), 3.0 (*6*), and ∞ (*7*).

culated dependences of the dimensionless critical electric field E'_c and the corresponding wavenumber Q_c on the azimuthal anchoring energy ε_φ for various values of the flexoelectric parameter ν. For certainty, the ratio of the Frank elastic constants was assumed to be $r = 0.75$. As expected, the critical electric field grows with increasing azimuthal anchoring energy ε_φ and the flexoelectric parameter ν. However, the period λ_c of the spatial structure of the director field can be a nonmonotonic function of ε_φ (Fig. 3b). As the flexoelectric parameter ν increases, the critical period λ_c grows. For each ν value obeying the condition

$$\nu > \frac{\pi^2}{8(1-r)^2 - \pi^2(1-2r)},$$

there is a certain limiting (threshold) azimuthal anchoring energy which monotonically decreases with increasing ν and, as $\nu \longrightarrow \infty$, tends to the $\varepsilon_{\varphi th}(r)$ value given by formula (18).

Figure 4 shows the plots of the critical electric field E'_c and the dimensionless wavenumber Q_c versus the ratio r of the Frank elastic constants for several values of the flexoelectric parameter ν at an azimuthal anchoring energy of $\varepsilon_\varphi = 10$. Both the critical electric field and the corresponding period λ_c of the arising structure of the director field increase with the parameter r. As can be seen from Fig. 4b, the threshold ratio of the Frank elastic constants decreases with increasing ν and, in the limiting case corresponding to the absence of the flexoelectric polarization (curve *7*), acquires the value given by formula (19).

It should be noted that, as the parameter r increases, the critical electric field E_c tends to the limiting value $E_c(r_{th}) = \pi/\sqrt{\epsilon}L$, which depends neither on the azimuthal anchoring energy ε_φ nor on the flexoelectric parameter ν.

As can be seen from Figs. 1–4, the domain of existence of a periodic structure of the director field with respect to both r (the ratio of the Frank elastic constants) and ν (flexoelectric parameter) values expands at a finite value of the azimuthal anchoring energy ε_φ as compared to the case of infinitely strong anchoring.

4. DEPENDENCE OF THE DIRECTOR REORIENTATION THRESHOLD ON THE POLAR ANCHORING ENERGY

Now let us consider the opposite case, whereby the azimuthal polar anchoring energy at the cell surface is infinitely large ($W_\varphi = \infty$), while the polar anchoring energy W_θ is arbitrary. In this case, the boundary conditions (4) take the following form:

$$\left[\left(\frac{W_\theta}{K_1} \mp e_o E\right)\theta_1 \pm \frac{d\theta_1}{dz}\right]_{z = \pm L/2} = 0, \qquad \varphi_1\big|_{z = \pm L/2} = 0. \tag{21}$$

Substituting solution (10) into boundary conditions (21), we obtain a homogeneous system of four algebraic equations for determining the unknown coefficients a_i, b_i ($i = 1, 2$). The condition that this system has a nontrivial solution yields a determinant equation that determines the dispersion dependence $E(q)$, which is not presented here due to its being very cumbersome. The domain of the existence of a periodic spatial structure of the director field is determined, in accordance with condition (15), by the following inequality:

$$A(1-r)^2 + B(1-r) - \frac{1}{2}Cu\left(1 - 2r + \frac{1}{\nu}\right) - D < 0, \tag{22}$$

which is valid for all values of the polar (dimensionless) anchoring energy $\varepsilon_\theta = W_\theta L/K_1$. In this relation,

$$A = 2\frac{\varepsilon_\theta + 2u\cot u}{u\cot u}, \quad B = \frac{8}{\sqrt{\nu\nu_o}}(1 - u\cot u),$$

$$C = \left(\frac{\varepsilon_\theta}{\sin^2 u} + 2\right)\frac{\varepsilon_\theta + 2u\cot u}{u} - \frac{4u}{\nu_o \sin^2 u}$$

$$- 2\frac{\varepsilon_\theta \cot u - 2u}{u}\left(\cot u - \frac{u}{\sin^2 u}\right),$$

$$D = \frac{1}{\nu}\left(\frac{8u}{\nu_o}\cot u - \varepsilon_\theta \cot u\frac{\varepsilon_\theta + 2u\cot u}{u}\right.$$

$$\left. - (\varepsilon_\theta + 2)\frac{\varepsilon_\theta \cot u - 2u}{u}\right),$$

$$\nu_o = \frac{\epsilon}{e_o^2},$$

and u is the minimum positive root of the transcendental equation

$$\frac{4u^2}{\nu_o}\cot u = (\varepsilon_\theta \cot u - 2u)(\varepsilon_\theta + 2u\cot u).$$

(a) NLC cell in the absence of flexoelectric polarization. In this case, the equation for determining the critical field is as follows:

$$\beta_2\left(\frac{W_\theta}{K_1}\cot\frac{p_1 L}{2} - p_1\right) + \beta_1\left(\frac{W_\theta}{K_1}\coth\frac{p_2 L}{2} + p_2\right) = 0. \quad (23)$$

Figure 5a shows the calculated dependences of the wavenumber Q_c of the periodic structure of the director field on the polar anchoring energy ε_θ for various values of the parameter r, which were obtained by solving Eq. (23). As can be seen, the period $\lambda_c = 2\pi/q_c$ monotonically decreases with increasing ε_θ. For each value of the parameter r in the interval $r < 0.5$, relation (22) gives a certain threshold polar anchoring energy $\varepsilon_{\theta th}(r)$ such that for $\varepsilon_\theta < \varepsilon_{\theta th}$, the NLC cell exhibits a homogeneous Fréedericksz transition, while for $\varepsilon_\theta > \varepsilon_{\theta th}$, a Fréedericksz transition with the formation of a periodic spatial structure takes place. For $r > 0.5$, in this case (as well as for the infinite polar and a finite azimuthal anchoring energy), only a homogeneous Fréedericksz transition is possible in the NLC cell.

Figure 5b shows the plots of the wavenumber Q_c on the ratio r of the Frank elastic constants for various values of the polar anchoring energy ε_θ. The critical value of r according to relation (22) is

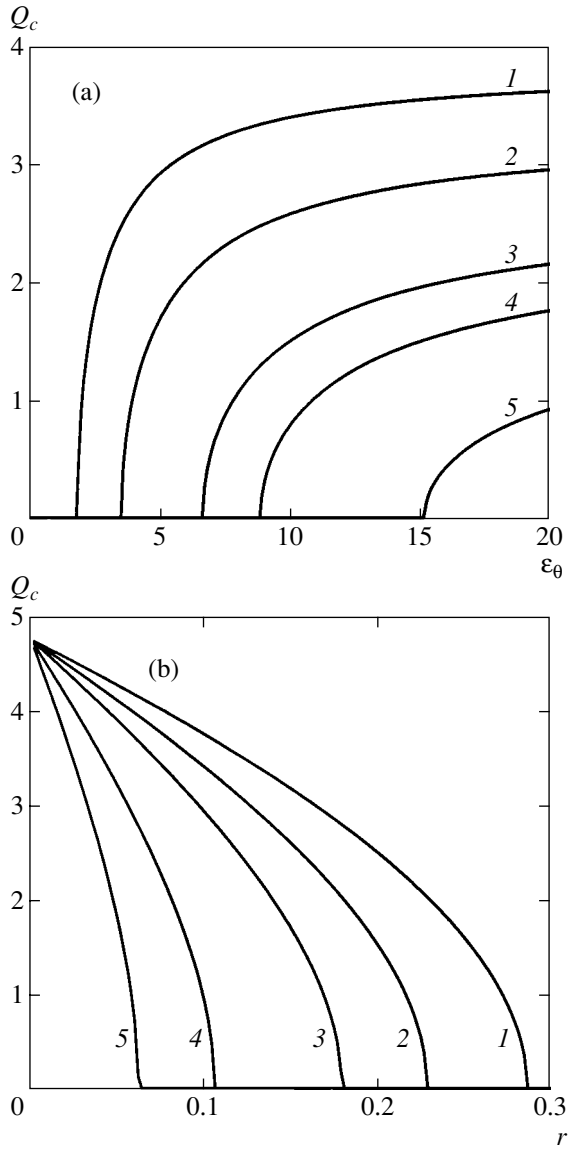


Fig. 5. Plots of the critical wavenumber Q_c in the absence of flexoelectric polarization (a) versus polar anchoring energy ε_θ for $r = 0.1$ (*1*), 0.15 (*2*), 0.2 (*3*), 0.22 (*4*), and 0.25 (*5*) and (b) versus parameter r for $\varepsilon_\theta = 50$ (*1*), 10 (*2*), 5 (*3*), 2 (*4*), and 1 (*5*).

$$r_{th}(\varepsilon_\theta) = 1 - \frac{u_o^2}{2\sin^2 u_o} - \frac{u_o^2}{\varepsilon_\theta} + \left[\left(\frac{u_o^2}{2\sin^2 u_o} + \frac{u_o^2}{\varepsilon_\theta}\right)^2 - \frac{u_o^2}{2\sin^2 u_o} - \frac{u_o^2}{\varepsilon_\theta}\right]^{1/2}, \quad (24)$$

where u_o is the minimum positive root of the equation $\cot u_o = 2u_o/\varepsilon_\theta$. For $r < r_{th}$, the NLC cell exhibits a Fréé-

dericksz transition with the formation of a periodic spatial structure, while for $r > r_{th}$ only a homogeneous Fréedericksz transition is possible.

Similar to the case of a finite azimuthal anchoring energy, the critical electric field monotonically grows with increasing polar anchoring energy ε_θ and parameter r. However, in the latter case, E_c tends to different limiting values $E_c(r_{th}) = 2u_o/\sqrt{\epsilon}L$ dependent on the polar anchoring energy ε_θ.

As can be seen from Fig. 5b, the finite polar anchoring energy ε_θ leads to narrowing of the domain of existence of a periodic structure of the director field with respect to the parameter r as compared to the case of the infinitely strong anchoring.

(b) NLC cell with flexoelectric polarization. In the case of strong polar anchoring ($\varepsilon_\theta \gg 1$), the critical electric field in a single constant approximation is

$$E_c = E(q_c) = \frac{2\pi}{|e|(1+\nu)L}\left(1 - \frac{1+\nu}{\varepsilon_\theta}\right). \qquad (25)$$

The corresponding wavenumber is

$$q_c = \frac{\pi}{L}\sqrt{\frac{1-\nu}{1+\nu}}\left(1 - \frac{1}{\varepsilon_\theta}\frac{1+\nu}{1-\nu}\right)$$

for $\nu < 1$ and $\varepsilon_\theta(1-\nu) \gg 1$, and this wavenumber is

$$q_c = \frac{\pi}{\sqrt{2}L}\sqrt{1 - \nu - \frac{4}{\varepsilon_\theta}}$$

for $|1-\nu| \ll 1$.

For arbitrary values of the polar anchoring energy ε_θ and the parameter r, the problem in this case (as well as in that considered above) has to be solved by numerical methods. The results of numerical calculations show that the critical electric field E_c in a flexoelectric NLC with $r \neq 1$ grows monotonically with increasing ε_θ and r for all values of the flexoelectric parameter ν. The period λ_c of the spatial structure of the director monotonically decreases with increasing ε_θ and monotonically increases with increasing r, also for all values of ν. The threshold value of the polar anchoring energy increases with the parameter ν for all r, in contrast to the case of finite azimuthal anchoring. The threshold value of r monotonically decreases with increasing ν and approaches (as $\nu \longrightarrow \infty$) a finite value $r_{th}(\varepsilon_\theta)$ determined by formula (24).

In a typical NLC with parameters of $\varepsilon_a = 0.2$, $r = 0.4$, and $K_1 = 5.5 \times 10^{-7}$ dyn and infinitely strong anchoring ($W_\varphi = W_\theta = \infty$) in the absence of flexoelectric polarization, the cell exhibits only the Fréedericksz transition with a homogeneous director distribution along the y axis. However, if the azimuthal anchoring energy is finite, for example, $\varepsilon_\varphi = 0.1$ (but $W_\theta = \infty$), the NLC cell exhibits a Fréedericksz transition with the formation of a periodic spatial structure with $\lambda_c \approx 4.6L$, whereas a finite polar anchoring with $\varepsilon_\theta = 0.1$ (but $W_\varphi = \infty$) still admits only the homogeneous Fréedericksz transition. By the same token, a flexoelectric NLC with $\nu = 0.5$ and $\nu_o = 0.1$ (corresponding to the flexoelectric coefficients e_1 and e_3 within $(0.7–2.5) \times 10^{-4}$ dyn$^{-1/2}$ [3]) shows a spatially periodic director field with $\lambda_c \approx 2.3L$ for infinitely strong anchoring and with $\lambda_c \approx 3.4L$ for $\varepsilon_\varphi = 0.1$ (but $W_\theta = \infty$), while exhibiting only a homogeneous Fréedericksz transition for $\varepsilon_\theta = 0.1$ (but $W_\varphi = \infty$). Thus, the period of the spatial structure of the director field depends much stronger on the azimuthal anchoring (this circumstance can be used for evaluating this energy), whereas the critical field in the general case depends more significantly on the polar anchoring.

In concluding, it should be emphasized that, in NLC cells with planar director, the finite anchoring significantly influences both the critical field and the period of a spatially periodic structure of the director field, as well as the intervals of the flexoelectric parameter ν and the ratio of the Frank elastic constants admitting the formation of such periodic structure. In the case of a finite azimuthal anchoring, the domain of existence of the periodic director field with respect to r and ν expands, whereas for a finite polar anchoring this domain contracts as compared to the case of infinitely strong anchoring.

REFERENCES

1. P. de Gennes, *The Physics of Liquid Crystals* (Clarendon, Oxford, 1974; Mir, Moscow, 1977).
2. L. M. Blinov, *Electro-Optical and Magneto-Optical Properties of Liquid Crystals* (Nauka, Moscow, 1978; Wiley, New York, 1983).
3. A. S. Sonin, *Introduction to the Physics of Liquid Crystals* (Nauka, Moscow, 1983) [in Russian].
4. A. N. Chuvyrov, Kristallografiya **25**, 326 (1980) [Sov. Phys. Crystallogr. **25**, 188 (1980)].
5. L. M. Blinov and A. A. Sonin, Zh. Éksp. Teor. Fiz. **87**, 476 (1984) [Sov. Phys. JETP **60**, 272 (1984)].
6. V. G. Nazarenko and O. D. Lavrentovich, Phys. Rev. E **49**, 990 (1994).
7. I. P. Pinkevich and M. F. Lednei, Proc. SPIE **2651**, 167 (1996).
8. Yu. P. Bobylev and S. A. Pikin, Zh. Éksp. Teor. Fiz. **72**, 369 (1977) [Sov. Phys. JETP **45**, 195 (1977)].
9. Y. P. Bobylev, V. G. Chigrinov, and S. A. Pikin, J. Phys. Colloq. **40**, C3-331 (1979).
10. V. P. Romanov and G. K. Sklyarenko, Zh. Éksp. Teor. Fiz. **116**, 543 (1999) [JETP **89**, 288 (1999)].
11. G. Barbero and I. Lelidis, Phys. Rev. E **67**, 061708 (2003).

12. F. Lonberg and R. B. Meyer, Phys. Rev. Lett. **55**, 718 (1985).
13. C. Oldano, Phys. Rev. Lett. **56**, 1098 (1986).
14. W. Zimmermann and L. Kramer, Phys. Rev. Lett. **56**, 2655 (1986).
15. E. Miraldi, C. Oldano, and A. Strigazzi, Phys. Rev. A **34**, 4348 (1986).
16. M. Simões, A. J. Palangana, and L. R. Evangelista, Phys. Rev. E **54**, 3765 (1996).
17. A. J. Palangana, M. Simões, L. R. Evangelista, and A. A. Arrotéia, Phys. Rev. E **56**, 4282 (1997).
18. V. M. Pergamenshchik, Phys. Rev. E **47**, 1881 (1993).
19. V. M. Pergamenshchik, Phys. Rev. E **61**, 3936 (2000).
20. A. L. Alexe-Ionescu, G. Barbero, and I. Lelidis, Phys. Rev. E **66**, 061705 (2002).
21. G. Barbero, L. R. Evangelista, and I. Lelidis, Phys. Rev. E **67**, 051708 (2003).
22. A. D. Kiselev and V. Yu. Reshetnyak, Mol. Cryst. Liq. Cryst. **321**, 133 (1998).
23. A. Rapini and M. Papolar, J. Phys. Colloq. **30** (C4), 54 (1969).
24. W. Zhao, C.-X. Wu, and M. Iwamoto, Phys. Rev. E **65**, 031709 (2002).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 4, 2005, pp. 804–810.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 4, 2005, pp. 907–914.

STATISTICAL, NONLINEAR, AND SOFT MATTER PHYSICS

Orientational Instability in a Nematic Liquid Crystal in a Decaying Poiseuille Flow

S. V. Pasechnik[a,*], A. P. Krekhov[b], D. V. Shmeleva[a], I. Sh. Nasibullaev[b], and V. A. Tsvetkov[a]

[a] *State Academy of Instrument-Making and Informatics, Moscow, 107846 Russia*
[b] *Institute of Molecular and Crystal Physics, Ural Research Center, Russian Academy of Sciences, Ufa, 450075 Russia*
**e-mail: s-p-a-s-m@hotmail.ru*
Received August 9, 2004

Abstract—The results of studies of orientational dynamics and instability in an MBBA nematic liquid crystal in a decaying Poiseuille flow are considered. The experiments were made on a wedge cell with a gap width varying in a direction perpendicular to the flow. Confining surfaces ensured homeotropic adhesion of the nematic to the surface. Above a certain critical value of the initial pressure drop, a uniform orientational instability is observed, which corresponds to the emergence of the director from the plane of the flow. The dependence of the critical pressure drop on the local thickness of the liquid crystal layer and on the external destabilizing electric field is determined. Simulation of nematodynamics equations is carried out. The results of theoretical calculations are in qualitative and quantitative agreement with the experimental data. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Nematic liquid crystals are anisotropic liquids characterized by orientational ordering of molecules. These bodies can serve as a model system for studying universal regularities of rheological behavior of various complex liquids such as liquid-crystal polymers, lamellar phases of solutions of surfactants, and melts of block-copolymers. Specific features in the behavior of a liquid crystal in a flow are determined by anisotropy of viscoelastic properties and by the interaction between the field of velocity **v** and the average local orientation of molecules, which is described by a unit vector (director) **n**. Orientational instabilities in a nematic flow have been studied most comprehensively for the case when the initial orientation of the director is perpendicular to the plane of the flow [1–3]. Theoretical analysis of a steady-state Poiseuille flow with a homeotropic orientation of the director at the boundary surfaces of the layer (vector **n** is perpendicular to the surface) shows that an increase in the gradient of pressure applied along the layer above a certain critical value leads to uniform orientational instability accompanied by the emergence of the director from the plane of the flow [4]. An analogous instability, which was predicted for an oscillating Poiseuille flow [5], was observed earlier and studied experimentally at frequencies from 1 to 20 Hz [6]. At small amplitudes of a decaying Poiseuille flow in a liquid crystal with an initial homeotropic orientation, the director field in the plane of the flow is distorted [7]. However, the stability of such a state upon an increase in the initial pressure drop has not been studied experimentally as yet.

Here, we report on the results of a study of the orientational behavior and instabilities in an MBBA (*n*'-metoxybenzylidene-*n*-butylaniline) nematic liquid crystal under the action of a decaying Poiseuille flow. The effect of a destabilizing electric field on the orientational instability threshold is investigated.

2. EXPERIMENT

The experimental cell is shown schematically in Fig. 1. A capillary with a wedge gap was formed by glass plates with inner surfaces coated with a thin conducting SnO_2 layer, which made it possible to apply an electric field to the liquid crystal layer. The treatment of the surfaces with chromolane ensured a homeotropic (perpendicular to the surface) orientation of the nematic on the substrates. The main feature of the cell was wedge-shaped with a local layer width h varying along the y axis. The linearity of the $h(y)$ dependence and the absolute values of local width h were monitored from the variation of the phase difference between the ordinary and extraordinary rays caused by a decrease in the alternating voltage (U_0 = 45 V, a frequency of 5 kHz) applied to the MBBA layer to zero [8]. The absolute error in determining the local width h was approximately 2–3 μm. Prior to the experiment, the cell was mounted vertically and was filled with the liquid crystal so that the material filled the capillary, filling channels, and a part of expansion vessels (cylindrical pipes of diameter D). Decaying Poiseuille flow (along the x axis) was produced because of introduction of the crystal into one of expansion vessels. The initial pressure

drop ΔP_0 created in this case and proportional to the initial difference ΔH_0 in the levels of the liquid crystal was calculated to within 5% from the mass of the crystal introduced in the cell and the diameter of the expansion vessels. The experiments were carried out at temperature $T = 22 \pm 0.5$°C.

For a small wedging,

$$(h_{\max} - h_0)/A \approx 0.002,$$

the capillary can be treated as a set of channels having different widths and parallel to the x axis, to which the same pressure gradient $\Delta P/L$ is applied. In addition, in view of the large aspect ratio of the cell,

$$\frac{A}{(h_{\max} + h_0)/2} \approx 800, \quad \frac{L}{(h_{\max} + h_0)/2} \approx 80,$$

we can expect that a plane Poiseuille flow along the x axis is formed in the capillary (except for the boundary regions at the ends of the cell); this is confirmed by observations of movement of small impurity particles (2–4 μm in diameter) added to the nematic.

The intensity $I(t)$ of light with a wavelength of 628 nm (He–Ne laser) transmitted (along the z axis) through the capillary was detected from an area of diameter $D = 0.3$ mm by a photodiode and recorded in digital form (with the help of A to D converter) on the hard disk of a computer. Two versions of positions of the polarizer and the analyzer were used in the experiment: crossed polaroids oriented at an angle of $\alpha = 45°$ to the direction of the flow (geometry *a*) and at an angle of $\alpha = 0°$ (geometry *b*). Geometry *b* made it possible to detect the emergence of the director from the xz plane of the flow. Shadow images of the cell in crossed polaroids in geometries *a* and *b* were recorded simultaneously with the help of a digital camera.

3. EXPERIMENTAL RESULTS

For small initial pressure drops ($\Delta P_0 \le 6$ Pa) in the entire range of local cell thicknesses $h_0 < h < h_{\max}$, no changes in the intensity of transmitted light were detected in the *b* geometry; consequently, the director preserved its orientation in the plane of the flow. In the *a* geometry, this regime corresponds to the shadow image of the cell, consisting of dark and bright fringes arranged along the direction of the flow and formed as a result of interference of the ordinary and extraordinary rays. The phase delay appears as a result of a change in the refractive index, which is in turn associated with a deviation of the director from the initial homeotropic orientation. The dynamics of the interference fringes is as follows: the formation of the fringe structure in the region of large local thicknesses begins immediately after the emergence of the initial pressure difference; after this, the system of fringes moves towards smaller thicknesses. In the steady-state decaying Poiseuille flow, the fringes slowly move towards larger thicknesses.

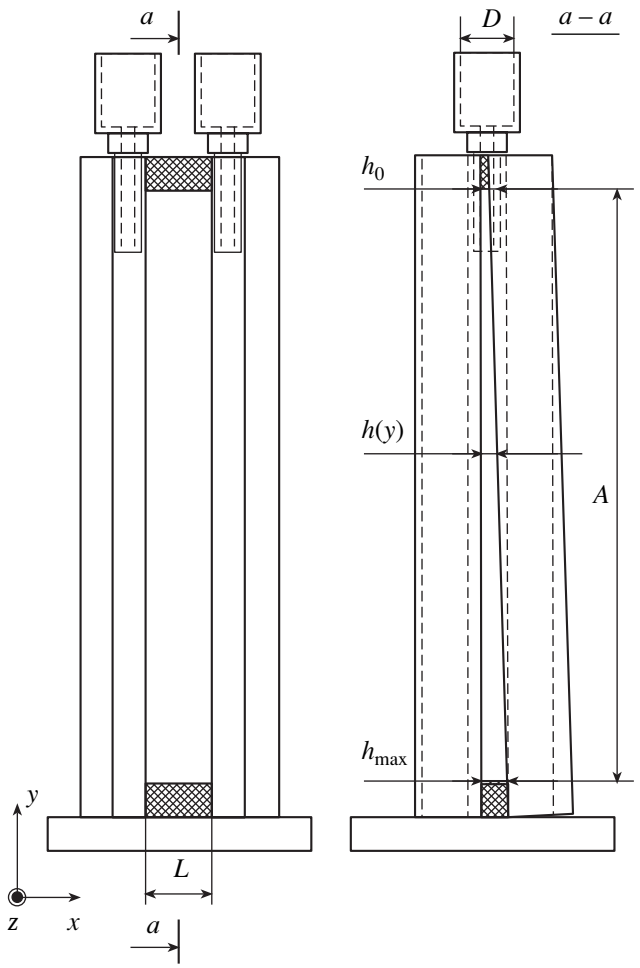


Fig. 1. Geometry of a wedge-shaped cell: $A = 10$ cm, $L = 1$ cm, $h_{\max} = 210$ μm, $h_0 = 33$ μm, and $D = 1.5$ cm.

Recording the intensity of transmitted light locally in the *a* geometry,

$$I(t) = I_0 \sin^2[\delta(t)/2],$$

where I_0 is the input intensity, we found that the phase lag $\delta(t)$ decreases exponentially with time (curve *1* in Fig. 2). For small deviations of the director (in the plane of the flow) from the initial homeotropic orientation, we can derive the following time dependence of the phase lag [7]:

$$\begin{aligned} \delta(t) &= \delta_0 \exp(-t/\tau_\delta), \quad \tau_\delta = \eta_{\text{hom}}/\rho g k_0, \\ k_0 &= A(h_{\max} + h_0)(h_{\max}^2 + h_0^2)/3\pi D^2 L, \end{aligned} \tag{1}$$

where δ_0 is the maximal value of the phase lag at the

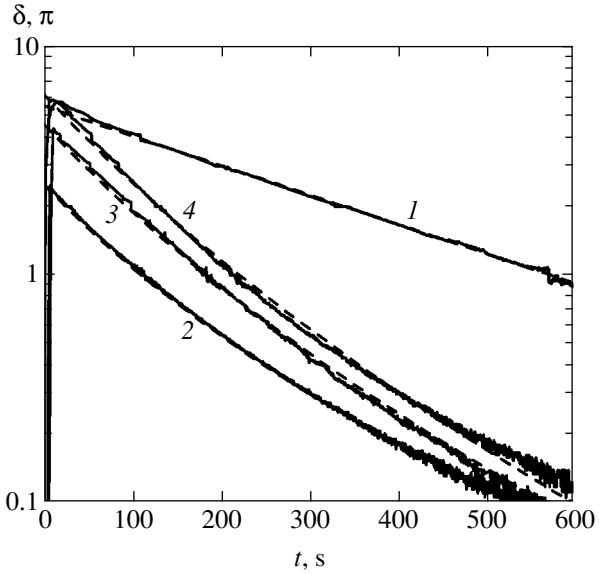


Fig. 2. Time dependence of the phase lag $\delta(t)$. Experimental data (solid curves) and results of simulation (dashed curves): $\Delta P_0 = 1.5$ Pa, $h = 164$ μm (*1*), $\Delta P_0 = 9.4$ Pa, $h = 70$ μm (*2*), $\Delta P_0 = 12.8$ Pa, $h = 70$ μm (*3*), and $\Delta P_0 = 15.5$ Pa, $h = 70$ μm (*4*).

instant of stabilization of the decaying flow,

$$\eta_{\text{hom}} = \frac{-\alpha_2 + \alpha_4 + \alpha_5}{2}$$

is the viscosity of the homeotropically oriented liquid crystal, ρ is the density of the nematic, g is the acceleration due to gravity, and k_0 is a quantity constant for the given cell and depending on the geometrical size. The experimental data for $\delta(t)$ are correctly described by dependence (1) using η_{hom} as a fitting parameter. The value of $\eta_{\text{hom}} = 0.16 \pm 0.02$ Pa s obtained for MBBA is in good agreement with the results of independent measurements [9, 10]. Our experiments suggest a new simple and reliable technique for measuring the viscosity coefficients of nematic liquid crystals, which is based on the recording of the time dependence of the phase lag in a decaying Poiseuille flow. If the cells used in experiments ensure the planar boundary conditions (the director is oriented parallel to the substrates in the plane of the flow), it is possible to measure the viscosity

$$\eta_{\text{plan}} = \frac{\alpha_3 + \alpha_4 + \alpha_6}{2}.$$

With increasing initial pressure drop ($\Delta P_0 > 6$ Pa), the signal intensity $I(t)$ of transmitted light in the *b* geometry, which is recorded in the range of large thicknesses of the cell, exhibits two peaks (curve *2* in Fig. 3a), indicating the emergence of the director from the plane of the flow. Figure 3b shows the theoretical dependences $I(t)$ in geometries *a* and *b* as well as the angle $\phi_m(t)$ of deviation of the director from the plane of the flow at the center of the layer. The nonlinear nematodynamics equations [11] for a planar layer, when the director and the velocity are functions of coordinate z and time t [12], were solved numerically using the material parameters for MBBA [10, 13]. The intensity of transmitted light was calculated using the Jones matrix method [14, 15]. Angle ϕ_m characterizes the orientation of the director at the center of the layer:

$$\mathbf{n}_m = (0, \sin\phi_m, \cos\phi_m).$$

The director distribution at the instant corresponding to $\phi_m = 60°$ is shown in Fig. 3c. The first peak of the $I(t)$ signal in the *b* geometry is associated with the emergence of the director from the plane of the flow in the case of a large initial pressure drop. As the pressure drop $\Delta P(t)$ decreases below the threshold value, the director returns to the plane of the flow (second peak on the $I(t)$ curve in the *b* geometry) and relaxes over long time periods to the uniform homeotropic orientation.

The transition associated with the emergence of the director from the plane of the flow is observed most clearly in the shadow image of the cell (Fig. 4). In the *b* geometry (Fig. 4b), the shadow image is (in the increasing order of the local layer thickness) dark field I in the range of smaller thicknesses, light fringe II, and the low-intensity region III. In the *a* geometry (Fig. 4a), two regions can be clearly distinguished on the shadow image: region I + II, corresponding to relatively small thicknesses, in which wide interference fringes parallel to the direction of the flow are observed, and region III, corresponding to large thicknesses, where narrow interference fringes are transformed into wide fringes.

The polarization and optical analysis, as well as a comparison of microphotographs (Fig. 4) with the time dependences of transmitted light (Fig. 3a), make it possible to unambiguously identify all regions on the shadow image of the cell: region I (the director is in the plane of the flow, the azimuth angle ϕ_m of deviation of the director at the center of the layer is zero); region II (the director emerges from the plane of the flow ($0 < \phi_m < 20°$), and region III (the director is oriented almost perpendicularly to the plane of the flow, $\phi_m \longrightarrow 90°$).

The recording of shadow images and the intensity of transmitted light in the *b* geometry made it possible to clarify the nature of formation of the region corresponding to the emergence of the director from the plane of the flow. Immediately after the application of the initial pressure gradient (for 10–15 s) in the range of large values of local thicknesses of the nematic layer, light fringe II is formed, which subsequently moves along the y axis towards smaller thicknesses over a time of approximately 30–40 s. After approximately 50–60 s, the position of the light fringe stabilizes, and the fringe begins to move slowly in the direction of increasing layer thickness. This stage corresponds to a steady-state decaying Poiseuille flow. The boundary y_b

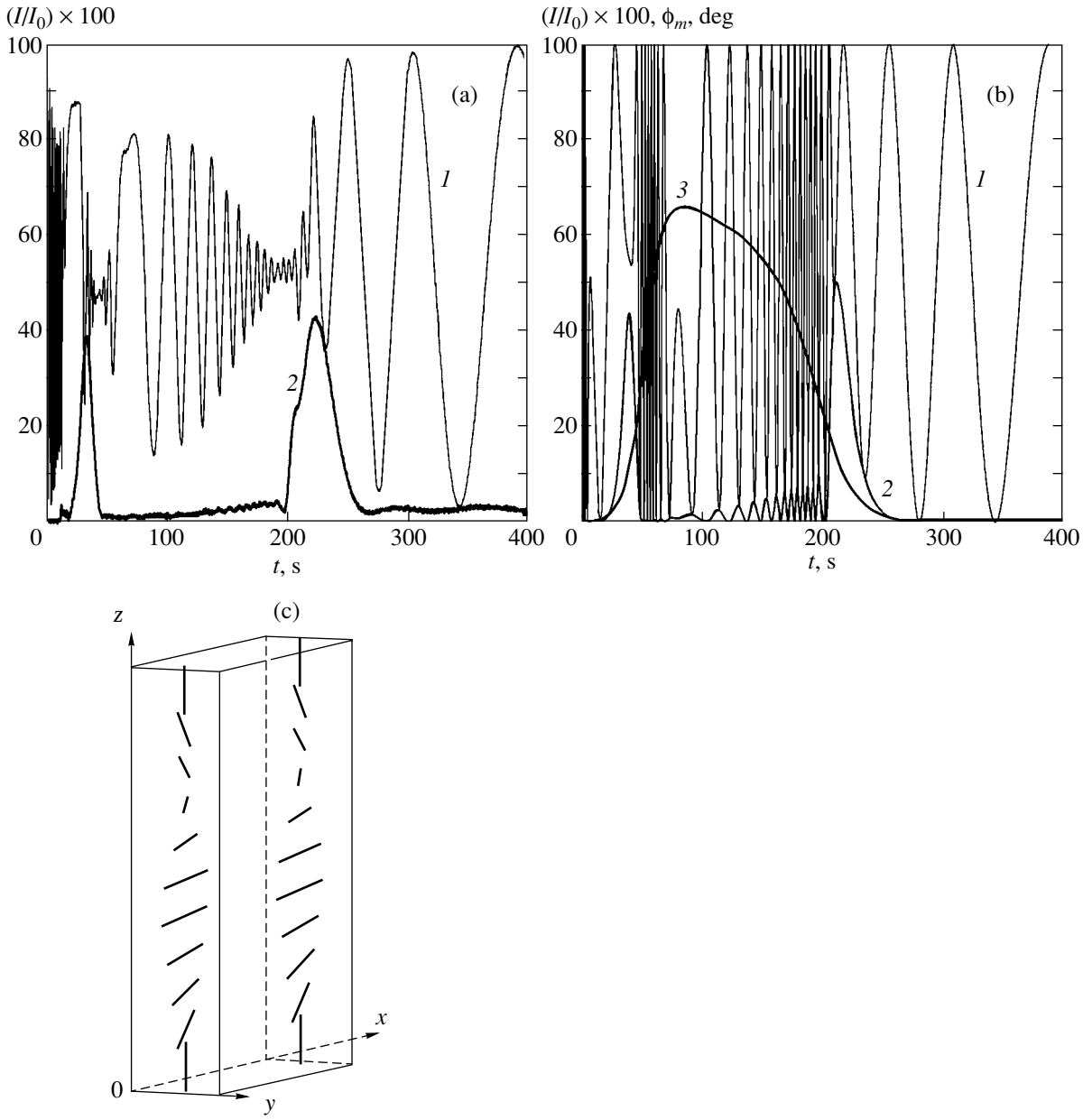


Fig. 3. Time dependences of the intensity of transmitted light, $I(t)$, in geometry *a* (curves *1*) and *b* (curves *2*) and of the angle of deviation of the director from the plane of the flow at the center of the layer, $\phi_m(t)$ (curve *3*). (a) Experimental data: $\Delta P_0 = 20$ Pa, $h = 86$ μm; (b) theoretical calculations: $\Delta P_0 = 21$ Pa, $h = 90$ μm; (c) schematic diagram of orientation of the director at the instant corresponding to $\phi_m = 60°$.

between regions II and III can be seen most clearly, while the boundary between regions I and II becomes less clear as fringe II moves towards large thicknesses. The width of region II attains its minimal value when it begins its reverse motion and increases as fringe II moves towards large local thicknesses. The presence of two peaks in the $I(t)$ signal in the *b* geometry (Fig. 3a) is due to the fact that light fringe II passes twice through the point of observation.

At the stage corresponding to a steady-state decaying Poiseuille flow, we recorded the time dependence $y_b(t)$ of the position of the boundary between regions II and III since this boundary remains the clearest during the entire experiment. The simultaneous recording of the phase lag $\delta(t)$ in the *a* geometry in the range of small local widths of the liquid crystal layer makes it possible to reconstruct the time dependence of the pressure drop, $\Delta P(t)$. Since the director is oriented in the plane of the layer in the range of small local thicknesses of the layer (the absence of signal $I(t)$ in the *b* geometry) and deviations from the homeotropic orientation are small, the measured phase lag is proportional to the square of

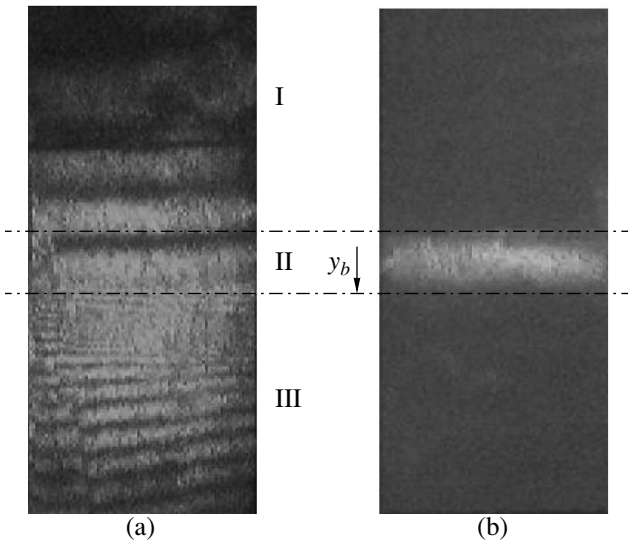


Fig. 4. Shadow images of the cell in crossed polaroids at instant $t = 30$ s for $\Delta P_0 = 15.5$ Pa: (a) in geometry *a*; (b) in geometry *b*.

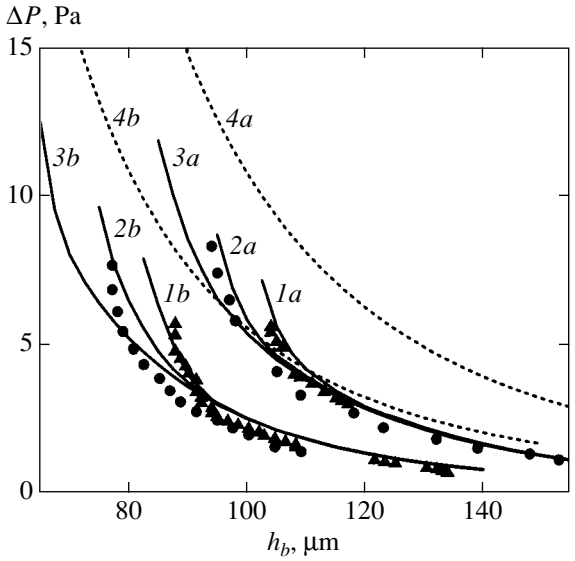


Fig. 5. The threshold pressure drop ΔP_c corresponding to the emergence of the director from the plane of the flow as a function of the local layer thickness h_b. Experimental data are shown by symbols and the results of calculations are given by curves. $U = 0$: $\Delta P_0 = 9.4$ Pa (▲ and *1a*), $\Delta P_0 = 12.8$ Pa (● and *2a*), $\Delta P_0 = 15.5$ Pa (■ and *3a*); $U = 3$ V: $\Delta P_0 = 8.7$ Pa (▲ and *1b*), $\Delta P_0 = 10.8$ Pa (● and *2b*), and $\Delta P_0 = 14.1$ Pa (■ and *3b*). Curves *4a* and *4b* correspond to calculations for the case of a steady-state Poiseuille flow for $U = 0$ and 3 V, respectively.

the pressure drop [16],

$$\delta(t) \sim [\Delta P(t)]^2,$$

which makes it possible to calculate $\Delta P(t)$ from the corresponding dependence $\delta(t)$. It should be noted that, for large initial pressure gradients, the time dependences $\delta(t)$ of the phase lag differ substantially from the simple exponential law (1) (curves *2*, *3*, and *4* in Fig. 2). This is due to the fact that, after the application of a large initial pressure gradient, the orientation of the director becomes almost perpendicular to the plane of the flow in the major part of the cell, and the director returns to the plane of the flow in the course of deceleration, after which it acquires the equilibrium homeotropic orientation; as a result, the effective viscosity of the nematic varies with time. For this regime of the flow, we can derive the following semi-empirical expression for $\delta(t)$:

$$\delta(t) = \delta_0 \exp[-t/\tau_\delta(t)],$$
$$\tau_\delta(t) = \frac{\eta_{hom} - (\eta_{hom} - \eta_{per})\exp(-t/\tau_0)}{\rho g k_0}, \quad (2)$$

with a slowly varying relaxation time $\tau_\delta(t)$. In expression (2),

$$\eta_{per} = \alpha_4/2$$

is the viscosity of the liquid crystal oriented perpendicularly to the plane of the flow and τ_0 is a fitting parameter. The relative variation of the relaxation time,

$$\frac{\tau_\delta(0)}{\tau_\delta(\infty)} = \frac{\eta_{per}}{\eta_{hom}},$$

is associated with the change in the orientation of the director from the orientation perpendicular to the plane of the flow to the homeotropic orientation. For MBBA at $T = 22°C$, we have $\eta_{per}/\eta_{hom} = 0.31$ [10]. It can be seen from Fig. 2 (curves *2*, *3*, and *4*) that the experimental data are correctly described by dependence (2) for various initial pressure gradients.

Thus, using the data on the time dependence $y_b(t)$ of the position of the boundary and the dependence $\Delta P(t)$ reconstructed from $\delta(t)$, we can associate the value of pressure drop with the position of boundary y_b recorded in the experiment, thus establishing the dependence of the critical pressure gradient ΔP_c, corresponding to the emergence of the director from the plane of the flow, on local thickness h_b of the liquid crystal layer. Figure 5 (curves marked by *a*) shows the $\Delta P_c(h_b)$ dependence obtained for various values of the initial pressure drop ΔP_0. It can be seen from the figure that, for large values of h, the curves obtained for different values of ΔP_0 almost coincide. This is due to the fact that large thicknesses correspond to large time intervals following the application of the initial pressure drop, when a quasi-stationary flow sets in the cell. The velocity varies slowly with time and the director can follow the variation of pressure. Accordingly, the position of the boundary y_b in such a flow regime is determined only by the current value of ΔP and does not depend on ΔP_0.

The results of analysis of the effect of an electric field applied along the z axis on the threshold for the emergence of the director from the plane of the flow are given by curves marked by b in Fig. 5. The reduction of the critical value $\Delta P_c(h_b)$ is due to the fact that the electric field exerts a destabilizing effect (in addition to the flow) on the homeotropically oriented MBBA layer (negative anisotropy of permittivity).

Figure 5 also shows the theoretical $\Delta P_c(h_b)$ dependences obtained from the results of simulation of the nonlinear nematodynamics equations [11] for the case when the director and velocity are functions of coordinate z and time t [12] using the material parameter of MBBA [10, 13]. For a number of values of thickness of a planar layer of the liquid crystal, the orientational dynamics of the director was calculated using the experimentally determined time dependence $\Delta P(t)$ of the pressure drop. At the initial instant, the director is oriented homeotropically and emerges from the plane of the flow when the value of ΔP exceeds a certain critical value (depending on the thickness). As the pressure decays, the director returns to the plane of the flow. The threshold value ΔP_c corresponds to the instantaneous value $\Delta P(t)$ at which the return of the director to the plane of the flow is registered. The results are in good agreement with the experimental data considering that the wedge-shape cell was simulated in numerical calculations by a set of planar capillaries, taking into account experimental errors in determining ΔP_c and h_b.

Figure 5 also shows for comparison the dependences $\Delta P_c^{\mathrm{st}}(h)$ of the critical pressure drop corresponding for the emergence of the director from the plane of the flow, which were calculated for a steady-state Poiseuille flow (curves *4*). The critical value of the pressure drop for a steady-state Poiseuille flow,

$$\Delta P_c^{\mathrm{st}} \sim 1/h^3,$$

systematically exceeds the corresponding values of ΔP_c for a decaying flow. This is due to the fact that the return of the director to the plane of the decaying flow occurs upon a decrease in pressure $\Delta P(t)$ below ΔP_c^{st} over a finite time (on the order of the director relaxation time), during which the pressure continues to decrease.

The wedge-shaped structure of the cell enabled us to observe the behavior of a liquid crystal in a wide range of thicknesses in the same experiment. It was found that at thicknesses exceeding the threshold values for the emergence of the director from the plane of the flow, regions (domains) of various shape and size were formed. A characteristic feature of these domains is their long lifetime. For example, for a layer having a thickness of about 130 μm and $\Delta P_0 \approx 25$ Pa, the time of relaxation of the director to the original homeotropic orientation in these domains is longer than 3 h, which exceeds the characteristic time of flow decay (<10 min) and the time of restoration of the initial orientation outside these domains.

In polarized monochromatic light, in the *a* geometry, the interference pattern in the domains is similar to that described above, but inclined interference fringes in this case move at a much lower velocity. In the *b* geometry, these domains have an average illuminance higher than that of neighboring regions of the cell. At times exceeding the flow decay time, the domains are successfully visualized against the background of the dark field corresponding to the homeotropic orientation and are separated from the latter by a bright threadlike boundary (domain wall). The boundary of the regions is also observed for the position of both polaroids parallel to the direction of the flow. In this case, it separates two regions with equal illuminances.

The domains were formed over approximately 1–2 minutes after the generation of the flow in the form of one or several formations oriented along the flow. The area of the domains is the larger, the higher the initial pressure drop ΔP_0. For large values of the initial pressure gradient, a single region occupying a substantial part of the cell was formed.

It should be noted that domains are formed for thicknesses $h > h_b$, i.e., at velocities exceeding the threshold value for the emergence of the director from the plane of the flow. Consequently, the formation of domains can be treated as the result of secondary hydrodynamic instabilities developing against the background of a strongly deformed structure of the layer associated with the primary instability. In the central regions of the layer, for large pressure drops, a distribution close to that perpendicular to the plane of the flow is formed for the director as a result of primary instability (Fig. 3c). For such an orientation, the same mechanisms affect the stability of the director as those considered in [2, 17] for the initial orientation perpendicular to the plane of the flow. However, the mechanism of formation of these domains with anomalously long lifetime remains unclear and requires further investigations.

4. CONCLUSIONS

Orientational instability formed in a nematic liquid crystal under the action of a decaying Poiseuille flow and accompanied by the emergence of the director from the plane of the flow is detected experimentally and studied for the first time. The development of instability is visualized in a wedge-shaped cell in the form of a sharp boundary (domain wall) separating spatial regions in which the orientation of the director is in the plane of the flow and outside this plane. Analysis of the dynamics of motion of the domain wall in a decaying flow made it possible to determine the dependence of the threshold pressure drop on the thickness of the liquid crystal layer. It is found that the additional action of an electric field on the nematic layer with a negative

anisotropy of the permittivity lowers the threshold for the emergence of the director from the plane of the flow. The results of theoretical calculations of the critical pressure drop are in quantitative agreement with the experimental data. The analysis of the orientational behavior of the director in a decaying Poiseuille flow has demonstrated the possibility of determining the viscosity coefficients of the nematic to a high degree of accuracy from the data on the optical response for small initial pressure gradients.

ACKNOWLEDGMENTS

The authors are grateful to O.S. Tarasov for fruitful discussions of the results and for a critical review of the article.

This study was supported by the Russian Foundation for Basic Research (project no. 02-02-17435), Federal Target Program "Integration" (grant no. B0065), INTAS (grant no. YSF 01/1-188), and DAAD (grant no. 325 A/03/01330).

REFERENCES

1. P. Pieranski and E. Guyon, Phys. Rev. A **9**, 404 (1974).
2. E. Guyon and P. Pieranski, J. Phys. Colloq. **36**, C1-203 (1975).
3. E. Dubois-Violette and P. Manneville, in *Pattern Formation in Liquid Crystals*, Ed. by A. Buka and L. Kramer (Springer, New York, 1996), p. 91.
4. I. Zuniga and F. M. Leslie, J. Non-Newtonian Fluid Mech. **33**, 123 (1989).
5. A. P. Krekhov and L. Kramer, J. Phys. II **4**, 677 (1994).
6. P. Toth, A. P. Krekhov, L. Kramer, and J. Peinke, Europhys. Lett. **51**, 48 (2000).
7. S. V. Pasechnik, V. A. Tsvetkov, A. V. Torchinskaya, and D. O. Karandashov, Mol. Cryst. Liq. Cryst. **366**, 165 (2001).
8. W. De Jeu, *Physical Properties of Liquid Crystalline Materials* (Gordon and Breach, New York, 1980; Mir, Moscow, 1982).
9. H. Kneppe and F. Schneider, Mol. Cryst. Liq. Cryst. **65**, 23 (1981).
10. H. Kneppe, F. Schneider, and N. K. Sharma, J. Chem. Phys. **77**, 3203 (1982).
11. P.-G. de Gennes, *The Physics of Liquid Crystals* (Clarendon, Oxford, 1974; Mir, Moscow, 1977).
12. A. P. Krekhov and L. Kramer, Phys. Rev. E **53**, 4925 (1996).
13. W. H. de Jeu, W. A. P. Claassen, and A. M. J. Spruijt, Mol. Cryst. Liq. Cryst. **37**, 269 (1976).
14. T. J. Scheffer and J. Nehring, J. Appl. Phys. **56**, 908 (1984).
15. P. Yeh and C. Gu, *Optics of Liquid Crystal Displays* (Wiley, New York, 1999).
16. S. V. Pasechnik and A. V. Torchinskaya, Mol. Cryst. Liq. Cryst. **331**, 341 (1999).
17. I. Janossy, P. Pieranski, and E. Guyon, J. Phys. (Paris) **37**, 110 (1976).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 100, No. 4, 2005, pp. 811–819.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 4, 2005, pp. 915–924.

STATISTICAL, NONLINEAR, AND SOFT MATTER PHYSICS

Dynamics of Hamiltonian Systems under Piecewise Linear Forcing

V. V. Vecheslavov

Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia
e-mail: vecheslavov@inp.nsk.su

Received August 5, 2004

Abstract—A two-parameter family of smooth Hamiltonian systems perturbed by a piecewise linear force is analyzed. The systems are represented both as maps and as dynamical systems. Currently available analytical and numerical results concerning the onset of chaos and global diffusion in such systems are reviewed. Dynamical behavior that has no analogs in the class of systems with analytic Hamiltonians is described. A comparison with the well-studied dynamics of a driven pendulum is presented, and essential differences in dynamics between smooth and analytic systems are highlighted. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Two-dimensional canonical maps of the form

$$\bar{p} = p + Kf(x), \quad \bar{x} = x + \bar{p} \pmod 1, \qquad (1.1)$$

where K is the perturbation strength, have been widely used in nonlinear physics as very useful and informative models [1–3].

Map (1.1) is equivalent to an explicitly time-dependent Hamiltonian dynamical system perturbed by kicking [1–3]:

$$\begin{aligned} H(x, p, t) &= \frac{p^2}{2} + KV(x)\delta_1(t) \\ &= H_0(x, p) + H_1(x, t), \end{aligned} \qquad (1.2)$$

where $V(x) = -\int f(x)\,dx$ is a force potential and

$$\delta_1(t) = 1 + 2\sum_{n \geq 1} \cos(2\pi n t)$$

is the delta function of period 1.

The unperturbed term in (1.2),

$$H_0 = \frac{p^2}{2} + KV(x), \qquad (1.3)$$

represents the main (integer) resonance, while

$$H_1(x, t) = KV(x)(\delta_1(t) - 1) \qquad (1.4)$$

is treated as a perturbation with period $T = 1$ and frequency $\Omega = 2\pi/T = 2\pi$.

The system defined by resonant Hamiltonian (1.3) is integrable and does not exhibit any attributes of chaos. Its phase portrait has the following basic structure. There exists a saddle point, which should be classified as a distinct trajectory (unperturbed pendulum remains at this point for an indefinitely long time). Two oppositely directed separatrices emanate from the saddle point and loop back toward it in the long-time limit. They separate the regions where phase rotates (outside resonance) from those where phase oscillates (inside resonance). Note that both separatrices are combinations of two similar trajectories corresponding to forward and backward time evolution. It is well known that each separatrix is split by perturbation into two distinct trajectories that intersect at the so-called homoclinic points.

Generally, perturbed system (1.2) is not integrable, and its phase space is divided into chaotic and regular regions. One problem of practical importance is an overlap criterion for the chaotic regions, i.e., conditions for the onset of global chaos. These conditions critically depend on the properties of the potential function $V(x)$ of the main resonance.

The founders of the Kolmogorov–Arnold–Moser (KAM) theory noted from the very start that the onset of global chaos (the possibility of unbounded diffusion in overlapping chaotic regions) depends not only on the perturbation strength, but also on the smoothness of the system, which is well characterized in terms of its Fourier spectrum. In the case of an analytic perturbation, the Fourier amplitudes decay exponentially and there exists a critical value K_{cr} such that global chaos develops only when $K \gtrsim K_{cr}$. If $K \lesssim K_{cr}$, then chaos is localized in relatively narrow layers (which exist for any $K > 0$), and global diffusion is impossible for a conservative system having less than two degrees of freedom.

Essentially different dynamical behavior is observed when the *n*th Fourier amplitude of a smooth potential $V(x)$ decays as the inverse of n raised to a certain power $\beta + 1$ (see [4] and references therein). For a two-dimensional map, a simple estimate was used in [4] to show that there exists $K_{cr} > 0$ corresponding to the onset of global chaos if $\beta > \beta_{cr} = 3$. This critical degree of smoothness should be verified in numerical experiments. A rigorous proof was found only for $\beta_{cr} = 5$ (see [5], where it was also conjectured that the correct value is $\beta_{cr} = 4$).

Map (1.1) has been analyzed for both analytic and smooth forcing functions. For example, the analytic function $f(x) = \sin(2\pi x)$ is associated with Chirikov's standard map. Extensive studies of the standard map and its dynamical analog (driven pendulum) have significantly contributed to progress in modern nonlinear dynamics.

Moreover, dynamical behavior of a new type not observed under analytic forcing has been discovered for systems with smooth potential functions (see discussion below).

This paper focuses on a Hamiltonian in which a smooth (sinusoidal) perturbation is replaced by a piecewise linear force (sawtooth pulse train) with $\beta = 2$ (see Section 2). Since the dynamical behavior corresponding to this particular degree of smoothness has long remained unclear, a brief background review is presented here.

Even early numerical experiments on systems with $\beta < \beta_{cr} = 3$ demonstrated that both global diffusion and trajectories restricted to bounded regions of the phase space can be observed in long-time simulations [6, 7]. Whereas those observations provided indirect evidence of suppressed or weakened diffusion, it was established in [8] that the phase space of a symmetric, piecewise linear two-dimensional (SPL2D) map with $\beta = 2$ contains global invariant curves with both irrational and rational winding numbers (see also [9]). Global invariant curves span the entire phase domain, which precludes unbounded diffusion in action.

It was originally shown in [8] that the invariant curves with rational winding numbers include persistent integer- and fractional-resonance separatrices. Of particular importance is the fact that the system remains nonintegrable, i.e., global diffusion is blocked by persistent separatrices even if local dynamics is strongly chaotic.

For some obscure reason, Bullett's important and interesting study was not widely acknowledged at the time of publication. In a similar theorem independently proved much later by Ovsyannikov for the same SPL2D map [10], a countable set of K values was found for which integer-resonance separatrices persist despite local chaos. This finding motivated the systematic analysis of the SPL2D map and its modifications presented in [11–17]. Since Ovsyannikov's theorem had never been published, its complete statement was included in appendices to [12, 13] with the author's permission. Bullett's paper was not mentioned in my first publications on the subject [11, 12] for the mere reason that I was not aware of it at the time.

It should be stressed here that the analyses performed by Bullett and Ovsyannikov were focused on the invariant curves of new type, because the pairs of orbits that result from separatrix splitting exhibit chaotic behavior are not amenable to analytical methods (i.e., they can be studied only in numerical [11–17] or physical experiments).

Every invariant curve of an SPL2D map is associated with a specific value of K. In [11], these values were called critical numbers and denoted by $K_{Q,n}$, where Q is the resonance order ($Q = 1$ and $Q \geq 2$ correspond to integer and fractional resonances, respectively) and n is the index of a critical number. By Ovsyannikov's theorem, the critical number corresponding to an integer resonance is

$$K_{1,n} = \sin^2\frac{\alpha_n}{2}, \quad n = 1, 2, \ldots, \tag{1.5}$$

where α_n is the least positive root of the equation

$$\sqrt{2}\sin\frac{n\alpha}{2} = \cos\frac{\alpha}{2}. \tag{1.6}$$

In [11–14], the values of $K_{Q,n}$ were found numerically by using the fact that these numbers correspond to zero angle of intersection between the split trajectories at the central homoclinic point (the angle changes sign smoothly and discontinuously for odd and even n, respectively). It has been noted that, among all attributes of chaos, only this angle can be measured to arbitrary accuracy [14]. Note also that a vanishing angle of intersection between separatrices does not necessarily imply their persistence. In an example considered in [17], the trajectories that result from separatrix splitting are mutually tangent (have kinks) at the homoclinic point.

The critical numbers make up a Cantor set, and there must exist intervals of K corresponding to global diffusion. One of these intervals was found in [8]: $0.2295 < K < 0.2500$ for $d = 1/2$ (see below).

Further analysis showed that each global invariant curve of the type discovered by Bullett (including the persistent integer- and fractional-resonance separatrices) strongly modifies the structure of the phase plane in a finite neighborhood of the corresponding $K_{Q,n}$. Based on this observation, the term *virtual invariant curve* was coined. These virtual curves are associated with extremely complicated motions of smooth systems, which were called fractal diffusion in [15, 16]. Recent studies of these motions raised numerous questions that call for special analyses. Their discussion is left outside the scope of this review.

The important role played by the standard map and its dynamical analog (driven pendulum) in nonlinear dynamics is mentioned above. When developing the theory of stochastic layer in a driven pendulum, Chirikov introduced and widely used the Melnikov–Arnold integral to determine the amplitude of the separatrix map responsible for the generation of a stochastic layer by perturbation of any frequency [1].

Analysis of smooth systems driven by a piecewise linear force is facilitated if map (1.1) is studied in conjunction with a Hamiltonian dynamical system (see Section 2). The first analytical expressions for the corresponding Melnikov–Arnold integrals were obtained in [13, 17]. They characterize the specific dynamics of the systems, including the existence of both critical numbers and separatrices that persist despite local chaos. These integrals are used in Section 3 to elucidate the details of the onset of chaos.

In Sections 4 and 5, the results obtained for a dynamical system driven by a piecewise linear force are compared with the well-studied dynamics of a driven pendulum, and both analogies and differences in the dynamical behavior of these systems are highlighted.

2. HAMILTONIAN DYNAMICAL SYSTEM

In conjunction with (1.1), consider the biharmonically driven dynamical system described by a Hamiltonian of the form

$$H(x, p, t) = H_0(x, p) + U(x, t),$$
$$H_0(x, p) = \frac{p^2}{2} + \omega_0^2 V(x). \tag{2.1}$$

In the general case,

$$U(x, t) = \varepsilon_1 \cos(2\pi m_1 x - \Omega_1 t) + \varepsilon_2 \cos(2\pi m_2 x - \Omega_2 t) \tag{2.2}$$

is an asymmetric perturbation with integer m_1 and m_2 and $\varepsilon_1, \varepsilon_2 \ll 1$.

Both harmonics in (2.2) are assumed to be resonant. Furthermore, it is assumed that $\Omega_1 > 0$, and the first term is called the upper harmonic (the corresponding trajectory lies above the main-resonance separatrix in the phase plane). It is also assumed that $\Omega_2 < 0$, and the second term in (2.2) is called the lower harmonic accordingly.

In (2.1), the potential function

$$V(x) = \frac{1}{4} - \int f(x)dx$$

is associated with a piecewise linear force of period 1:

$$f(x) = \begin{cases} \dfrac{2x}{1-d}, & 0 \le x < \dfrac{1-d}{2}, \\ \dfrac{1-2x}{d}, & \dfrac{1-d}{2} \le x \le \dfrac{1+d}{2}, \\ \dfrac{2(x-1)}{1-d}, & 1+d < x < 1. \end{cases} \tag{2.3}$$

The corresponding Fourier series is [15]

$$V(x) = \frac{1}{4} + \sum_{n \ge 1} \frac{f_n}{2\pi n^{\beta+1}} \cos(2\pi n x),$$
$$f_n = -\frac{2}{\pi^2} \frac{\cos(n\pi)\sin(n\pi d)}{d(1-d)}, \tag{2.4}$$

where $\beta = 2$ is the degree of smoothness.

Varying the parameter d in expressions (2.3) and (2.4), one can analyze the dynamics of any system driven by a perturbation belonging to the family of sawtooth functions [8, 14]. To date, the system with $d = 1/2$ has been studied most thoroughly. Note that systems with $0 < d < 1$ exhibit the dynamical behavior associated with the existence of global invariant curves revealed and analyzed in [8, 10–17] if the parameter K in (1.1) satisfies the following condition found in [8]:

$$K \le K_B(d) = \frac{2d^2}{1+d}. \tag{2.5}$$

If $K > K_B$, then system (1.1), (2.3) has no global invariant curves.

In the limit case of forcing by a discontinuous sawtooth function ($d = 0$), the Fourier series expansion analogous to (2.4) is characterized by

$$f_n = -\frac{2}{\pi}\cos(n\pi), \quad \beta = 1. \tag{2.6}$$

This degree of smoothness is lower by unity than that corresponding to $0 < d < 1$, and both values of β are obviously less than $\beta_{cr} = 3$. In this case, the system is ergodic; i.e., there are no invariant curves, and global diffusion occurs at any $K > 0$. The regular motion corresponding to the opposite limit case of $d = 1$ [14] is of no interest for the present study.

The period of the force defined by (2.3) includes "elliptic" and "hyperbolic" intervals, where df/dx is negative and positive, respectively. The derivative is discontinuous across the boundaries of these intervals.

The motion along the upper unperturbed separatrix (when $\varepsilon_1 = \varepsilon_2 = 0$) is described by the following func-

tion of the dimensionless time $\psi = 2\omega_0 t$:

$$x_s(\psi_s) = \begin{cases} A_d \exp\dfrac{\psi_s}{\sqrt{2(1-d)}}, & -\infty < \psi_s < -\psi_{s,1}, \\ \dfrac{1}{2}\left(1 + \sqrt{d}\sin\dfrac{\psi_s}{\sqrt{2d}}\right), & -\psi_{s,1} \le \psi_s \le \psi_{s,1}, \\ 1 - A_d \exp\left(-\dfrac{\psi_s}{\sqrt{2(1-d)}}\right), & \psi_{s,1} < \psi_s < \infty. \end{cases} \quad (2.7)$$

The corresponding momentum is $p_s = \dot{x} = 2\omega_0 dx_s/d\psi_s$, where

$$\psi_{s,1} = \sqrt{2d}\arcsin\sqrt{d}, \qquad A_d = \frac{1-d}{2}\exp\frac{\psi_{s,1}}{\sqrt{2(1-d)}}. \quad (2.8)$$

The relative energy deviation from the unperturbed separatrix is $w = H_0/H_{0,s} - 1$, where $H_{0,s} = \omega_0^2/4$ is the value of the Hamiltonian on the separatrix. The period of motion in the vicinity of the separatrix can be calculated as

$$T_0(w) = 2T_{s,1} + \frac{1}{\omega_2}\ln\frac{4\sqrt{1-d}\cos(\omega_1 T_{s,1})}{|w|}, \quad (2.9)$$

where

$$\omega_1 = \omega_0\sqrt{\frac{2}{d}}, \quad \omega_2 = \omega_0\sqrt{\frac{2}{1-d}}, \qquad T_{s,1} = \frac{\arcsin\sqrt{d}}{\omega_1}. \quad (2.10)$$

The stochastic-layer width is calculated by using the relation between w and T_0 given by the inverse of (2.9):

$$w(T_0) = 4\sqrt{1-d}\cos(\omega_1 T_{s,1}) \times \exp(-\omega_2(T_0 - 2T_{s,1})). \quad (2.11)$$

Expressions (2.7)–(2.11) were derived in [17].

In the sections that follow, the onset of chaos in smooth systems of the type defined in this section is compared with the well-studied dynamics of the driven pendulum described by the Hamiltonian

$$H(x, p, t) = \frac{p^2}{2} + \cos x + U(x, t), \quad (2.12)$$

where the frequency of small-amplitude oscillation is set to unity and the perturbation term is given by (2.2) in the general case.

3. MELNIKOV–ARNOLD INTEGRAL FOR (2.1)

Generally, the stochastic layer that develops when the main-resonance separatrices are broken by perturbation consists of three parts: an upper part of width w_u (where the phase rotates so that $p > 0$), a middle one of width w_m (where the phase oscillates), and a lower one of width w_l (where the phase rotates so that $p < 0$). In the case of an asymmetric perturbation, the individual parts of the layer may have substantially different widths. Note that the upper and lower parts of the layer are mainly generated by the upper and lower resonant harmonics, respectively, while the middle one develops as a result of their combined effect [17]. The present analysis is mainly focused on the upper part of the stochastic layer.

Following the approach described in [1], consider the change in energy per oscillation half-period or rotation period relative to H_0:

$$\Delta H_0 = \int_{-\infty}^{\infty}\{H, H_0\}dt = -\varepsilon_1\int_{-\infty}^{\infty} p(t)\frac{\partial U}{\partial x}dt = 2\pi m\varepsilon_1\int_{-\infty}^{\infty} p(t)\sin(2\pi m_1 x - \tau - \tau_0)dt = 2\pi m_1\varepsilon_1\sin\tau_0 W_M,$$

where $\{,\}$ denotes the Poisson bracket, $\tau = \Omega_1 t$, and

$$W_M = -\int_{-\infty}^{\infty} p_s(t)\cos[2\pi m_1 x_s(t) - \Omega_1 t]dt \quad (3.1)$$

is the Melnikov–Arnold integral. Here, only the even function is retained in the expansion of $\sin(2\pi m_1 x - \tau - \tau_0)$ for a motion localized in the vicinity of the unperturbed separatrix.

Note that the normalization required to suppress the oscillation in (3.1) is provided by p_s: this factor asymptotically vanishes at both infinite limits of integration along the unperturbed separatrix (see details in [1, Section 4.4]).

The amplitude of the separatrix-map harmonic with frequency Ω is determined by the properties and behavior of the Melnikov–Arnold integral:

$$W = \pm\max|\bar{w} - w| = \pm\frac{\Delta H_0}{H_{0,s}} = \frac{8\pi m\varepsilon}{\omega_0^2}W_{MA}. \quad (3.2)$$

Rewriting integral (3.1) along unperturbed separatrix (2.7) in terms of the dimensionless time $\psi = 2\omega_0 t$, one finds that the contribution of the upper harmonic to

the motion in the upper part of the stochastic layer is determined by the expression

$$W_{\mathrm{MA}}(\lambda_1 > 0) = -\frac{1}{\sqrt{2}}\int_0^{\psi_{s,1}} \cos\frac{\psi}{\sqrt{2d}} \times \cos\left[\pi m_1\left(1+\sqrt{d}\sin\frac{\psi}{\sqrt{2d}}\right) - \lambda_1\psi\right]d\psi - A_d\sqrt{\frac{2}{1-d}}\int_{\psi_{s,1}}^{\infty} \exp\left(\frac{-\psi}{\sqrt{2(1-d)}}\right) \times \cos\left[2\pi m_1 A_d \exp\left(-\frac{-\psi}{\sqrt{2(1-d)}}\right) + \lambda_1\psi\right]d\psi, \quad (3.3)$$

where $\lambda_1 = \Omega_1/2\omega_0$ is an adiabaticity parameter [1].

Similarly, the contribution of the lower harmonic to the motion in the upper part of the stochastic layer can be expressed as

$$W_{\mathrm{MA}}(\lambda_2 < 0) = \frac{1}{\sqrt{2}}\int_0^{\psi_{s,1}} \cos\frac{\psi}{\sqrt{2d}} \times \cos\left[\pi m_2\left(1+\sqrt{d}\sin\frac{\psi}{\sqrt{2d}}\right) + |\lambda_2|\psi\right]d\psi + A_d\sqrt{\frac{2}{1-d}}\int_{\psi_{s,1}}^{\infty} \exp\left(\frac{-\psi}{\sqrt{2(1-d)}}\right) \times \cos\left[2\pi m_2 A_d \exp\left(-\frac{-\psi}{\sqrt{2(1-d)}}\right) - |\lambda_2|\psi\right]d\psi, \quad (3.4)$$

where $\lambda_2 = \Omega_2/2\omega_0$.

The total separatrix-map amplitude for the upper part of the layer is given by (3.2) as the sum of the contributions of all perturbing harmonics. In the next two sections, the expressions obtained here are used to analyze the behavior of system (2.1) under symmetric and asymmetric perturbations (2.2).

4. SYMMETRIC PERTURBATION

Perturbation (2.2) is symmetric if

$$\varepsilon_1 = \varepsilon_2 = \varepsilon, \quad m_1 = m_2 = m, \quad \Omega_1 = -\Omega_2 = \Omega. \quad (4.1)$$

First, recall that the frequency dependence of the onset of chaos for driven pendulum (2.12) under symmetric forcing can be examined by tentatively dividing the frequency domain into low-, intermediate-, and high-frequency intervals. An analysis presented in [1] showed that both separatrix-map amplitude W and energy width of the stochastic layer exponentially decrease with increasing frequency and the three parts of the layer tend to have equal widths as $\Omega \longrightarrow \infty$:

$$w_{\mathrm{u}} = |w_{\mathrm{m}}| = w_{\mathrm{l}} = \lambda W, \quad (4.2)$$

where $\lambda \equiv \Omega$ and $w = p^2/2 + \cos x - 1$ is the relative energy deviation from the unperturbed separatrix for pendulum (2.12).

In a recent analysis of the low-frequency asymptotic behavior in the limit of $\Omega \longrightarrow 0$ [18], it was found that the separatrix-map amplitude linearly increases with frequency, whereas the layer width is independent of frequency.

The asymptotic behavior is relatively simple in both limits ($\Omega \gg 1$ and $\Omega \ll 1$). The most difficult to analyze is the case of intermediate-frequency forcing, when neither a small nor large adiabaticity parameter can be introduced. The key difficulty here lies in the discontinuity of the stochastic-layer width as a function of W [19]. In modern theory, it is explained by successive breakdown of invariant curves characterized by irrational winding numbers with increasing W: as the curve separating the main stochastic layer from the nearest resonance of the separatrix map breaks down, the finite-sized region occupied by the resonance is added to the layer.

The so-called resonance invariants can be used to analyze chaotic behavior in this interval, since they adequately represent the topology of individual resonances. A technique for constructing resonance invariants of the first to third order (corresponding to the 1 : 1, 1 : 2, and 1 : 3 resonances) and examples of their use were presented in [20] for Chirikov's standard map and in [19] for a single-frequency separatrix map.

Now, consider the onset of chaos in a smooth system in the low- and high-frequency limits. The sum of the contributions of both harmonics to the Melnikov–Arnold integral for the upper part of the layer is

$$W_{\mathrm{MA}}(\lambda) = \sqrt{2}\int_0^{\psi_{s,1}} \cos\frac{\psi}{\sqrt{2d}} \times \sin\left[\pi m\left(1+\sqrt{d}\sin\frac{\psi}{\sqrt{2d}}\right)\right] \times \sin(\lambda\psi)d\psi + 2A_d\sqrt{\frac{2}{1-d}} \times \int_{\psi_{s,1}}^{\infty} \exp\left(\frac{-\psi}{\sqrt{2(1-d)}}\right) \times \sin\left[2\pi m A_d \exp\left(\frac{-\psi}{\sqrt{2(1-d)}}\right)\right]\sin(\lambda\psi)d\psi, \quad (4.3)$$

where $\lambda = \Omega/2\omega_0$. The first and second summands in this expression represent the contributions of the "ellip-

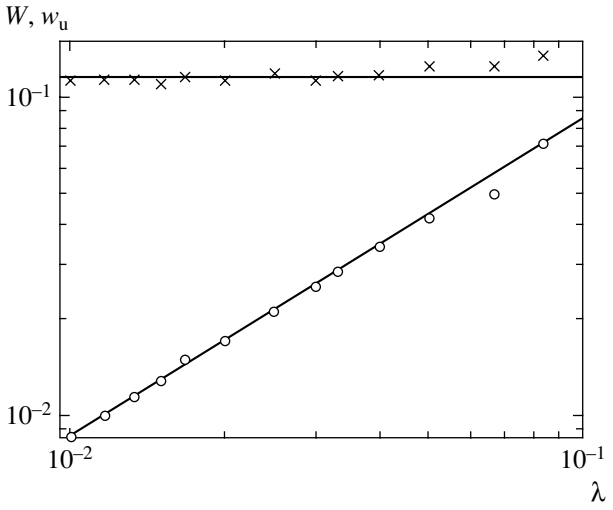


Fig. 1. Asymptotic behavior of system (2.1) under symmetric forcing at $\Omega \longrightarrow 0$: circles and crosses represent computed amplitudes of the separatrix map and widths of the upper part of the stochastic layer found by iterating the map; sloped line is Melnikov–Arnold integral (4.3) multiplied by a fitting factor of 0.75.

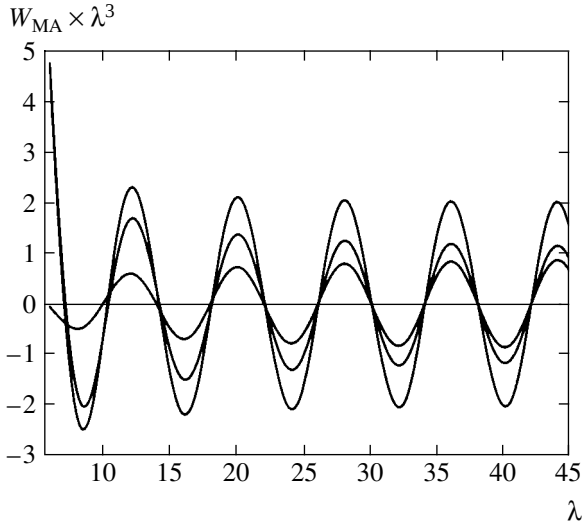


Fig. 2. Asymptotic behavior of system (2.1) under symmetric forcing at $\Omega \longrightarrow \infty$: curves with large, intermediate, and small amplitudes represent the joint contributions of the upper and lower harmonics, the contribution of the upper one, and that of the lower one, respectively.

tic" interval in (2.3) and the remaining part of the perturbation period, respectively.

First, consider the limit of $\lambda \longrightarrow 0$. Replacing $\sin(\lambda\psi)$ with $\lambda\psi$ and factoring λ out of the integral, one finds that the low-frequency Melnikov–Arnold integral (and, therefore, the separatrix-map amplitude W) linearly increases with frequency:

$$W \propto W_{\mathrm{MA}} \propto \lambda. \tag{4.4}$$

Note that the use of infinite integration limits in (4.3) does not change this result, because momentum asymptotically vanishes at the endpoints of the unperturbed separatrix (see above).

It was shown in [18] that the stochastic-layer width is independent of frequency if the separatrix-map amplitude linearly increases with frequency. Figure 1 illustrates this fact for symmetrically perturbed system (2.1), (2.2) with $\varepsilon_1 = \varepsilon_2 = 0.01$. Thus, the behavior of the system driven by a piecewise linear force is qualitatively similar to that of a driven pendulum.

System (2.1) is remarkable in that the corresponding separatrix map contains both single- and double-frequency harmonics when $\lambda \ll 1$. In the case of intermediate-frequency perturbation, the latter harmonic vanishes, i.e., the separatrix map contains only one harmonic. Recall that a driven pendulum exhibits opposite behavior: the double-frequency contribution is significant only under intermediate-frequency forcing and vanishes as $\lambda \longrightarrow 0$ or $\lambda \longrightarrow \infty$ [21].

In the limit of $\Omega \longrightarrow \infty$, a qualitatively different behavior is observed. In this case, both summands in (4.3) change sign, oscillating almost in antiphase with one another. The resultant function $W_{\mathrm{MA}}(\lambda)$ also oscillates and changes sign (see Fig. 2).

Asymptotic estimates for Melnikov–Arnold integrals (3.3) and (3.4) performed in the limit of $\lambda \longrightarrow \infty$ (see [17]) lead to the following results.

When $\lambda \gg 1$, the integrals are periodic functions of λ,

$$W_{\mathrm{MA}}(\lambda) \approx (-1)^m \frac{|\lambda|^{-3}}{4d}\sqrt{\frac{1-d}{2}} \times \left(1 \pm \pi|\lambda|^{-1}\sqrt{\frac{1-d}{2}}\right)\sin(\pi m d \mp |\lambda|\psi_{\mathrm{s},1}) \tag{4.5}$$

(upper sign corresponds to the upper harmonic), with the period

$$T_\lambda = \frac{2\pi}{\psi_{\mathrm{s},1}} = \frac{\pi}{\arcsin\sqrt{d}}\sqrt{\frac{2}{d}}. \tag{4.6}$$

Figure 2 shows the normalized Melnikov–Arnold integral $W^*_{\mathrm{MA}} = W_{\mathrm{MA}}\lambda^3$ calculated for $d = 1/2$ (symmetric sawtooth function). Similar graphs are obtained for $d \neq 1/2$, except for a phase shift between the harmonics. It is clear that expression (4.5) is qualitatively correct, even though numerical corrections are required to improve its accuracy.

In [13], a separatrix map with amplitudes given by the Melnikov–Arnold integral was iterated to show that relation (4.2) between the stochastic-layer width and the separatrix-map amplitude W holds for a smooth system as well,

$$w_{\mathrm{u}} \approx \lambda|W|, \tag{4.7}$$

even in the neighborhoods of the zeros of W. In particular, $0.94 \le w_u/\lambda|W| \le 1.3$ for $\lambda = \Omega/2\omega_0 > 10$.

The results obtained above for smooth dynamical system (2.1) can be associated with characteristics of map (1.1) in a straightforward manner as follows [13]. Define critical $\lambda_{1,n}$ ($n = 1, 2, \ldots$) as the zeros of the Melnikov–Arnold integral (the first index refers to the main resonance). Then, dynamical system (2.1) parameterized by λ is equivalent to map (1.1) parameterized by K, and vice versa, because the relation

$$K = \omega_0^2 = \left(\frac{\pi}{\lambda}\right)^2, \quad n = 1, 2, \ldots$$

can be used to calculate the critical numbers of the map by substituting $\lambda_{1,n}$. Indeed, the result of a comparison of

$$K^*_{1,n} = \left(\frac{\pi}{\lambda_{1,n}}\right)^2, \quad n = 1, 2, \ldots$$

with $K_{1,n}$ given by (1.5) and (1.6) can be accurately represented by the approximate formula

$$\frac{K_{1,n}}{K^*_{1,n}} \approx 1 + 0.676n^{-0.875}, \quad n = 1, 2, \ldots$$

Since each $K^*_{1,n}$ tends to the corresponding $K_{1,n}$ in the limit of $n \gg 1$, the results obtained here can be applied to map (1.1).

In the limit of $\Omega \gg 1$, the periodic behavior of the Melnikov–Arnold function obtained here, with an amplitude decreasing as λ^{-3}, contrasts with the monotonic behavior of the exponentially decreasing $W_{MA}(\lambda)$ for an analytic potential function. Moreover, the contribution of the lower harmonic to the upper part of the stochastic layer in the latter case is smaller by a factor of $\exp(-\pi|\lambda|)$ as compared to the contribution of the upper harmonic [1], whereas the contributions of the upper and lower harmonics are asymptotically equal for the system driven by a piecewise linear force. The drastic difference between systems of the two types discussed here may be explained by different location of the singularities of the Melnikov–Arnold integral: they lie on the real axis and in the complex plane in the cases of smooth potential and driven pendulum, respectively.

5. ASYMMETRIC PERTURBATION

Asymmetric perturbation (2.2) with Ω_1, $|\Omega_2| \gg 1$ was originally analyzed in [22, 23] for the driven pendulum described by Hamiltonian (2.12). In this case, it was found that the perturbed motion has secondary harmonics with amplitudes proportional to $\varepsilon_1\varepsilon_2$ and frequencies equal to the sum and difference of the primary frequencies:

$$\Delta\Omega_+ = \Omega_1 + \Omega_2, \quad \Delta\Omega_- = \Omega_2 - \Omega_1. \qquad (5.1)$$

Moreover, it was found that the secondary harmonics are much weaker than the primary ones when $\varepsilon_1, \varepsilon_2 \ll 1$.

Even the first numerical experiments revealed the seemingly surprising fact that these weak harmonics determine both the amplitude of the separatrix map and width of the chaotic layer under certain conditions.

In the example considered in [23], the system had the parameters $\varepsilon_1 = \varepsilon_2 = 0.075$, $\Omega_1 = 13$, and $\Omega_2 = -10$. The amplitude of the sum-frequency harmonic of the perturbed motion (with $\Delta\Omega_+ = 3$) was $\varepsilon \approx 4.5 \times 10^{-5}$, which is smaller than the primary-harmonic amplitude by a factor of about 1700. However, its contribution to the separatrix-map amplitude corresponding to the upper part of the layer (responsible for the onset of chaos) exceeded the total contribution of the primary harmonics by a factor of almost 400, while the individual parts of the layer had substantially different widths. This is explained by the exponential dependence of layer width on frequency at $\Omega \gg 1$, which leads to a dominant effect of weak low-frequency harmonics on the onset of chaos.

In smooth systems, a similar role is played by sum-frequency secondary harmonics, whereas the effect of difference-frequency harmonics is much weaker as compared to that in a driven pendulum [17].

The secondary-harmonic amplitudes are not known a priori, and no rigorous method for finding them is available to this day. In [22], a general approach to the problem was proposed, and approximate analytical estimates were obtained. In this approach (see also [17]), a change from the coordinate $x(t)$ and momentum $p(t)$ to the deviations from $x_s(t)$ (given by (2.7)) and $p_s(t)$ on the unperturbed separatrix is performed in (2.1) and (2.2). By assuming that the deviations are small and dropping the terms of order higher than two, it can be shown that the perturbation contains both the sum-frequency secondary harmonic

$$\varepsilon_+ \cos(2\pi m_+ x_s - \Delta\Omega_+ t) \qquad (5.2)$$

with

$$\varepsilon_+ = -2\pi^2\varepsilon_1\varepsilon_2 m_1 m_2\left(\frac{m_1}{\Omega_2^2} + \frac{m_2}{\Omega_1^2}\right),$$

and $m_+ = m_1 + m_2$, and the difference-frequency harmonic

$$\varepsilon_- \cos(2\pi m_- x_s - \Delta\Omega_- t) \qquad (5.3)$$

with

$$\varepsilon_- = -2\pi^2\varepsilon_1\varepsilon_2 m_1 m_2\left(\frac{m_1}{\Omega_2^2} - \frac{m_2}{\Omega_1^2}\right),$$

and $m_- = m_2 - m_1$. (These expressions were derived for $|\Omega_{1,2}| \gg 2\pi m_{1,2} p_{s,\max}$.)

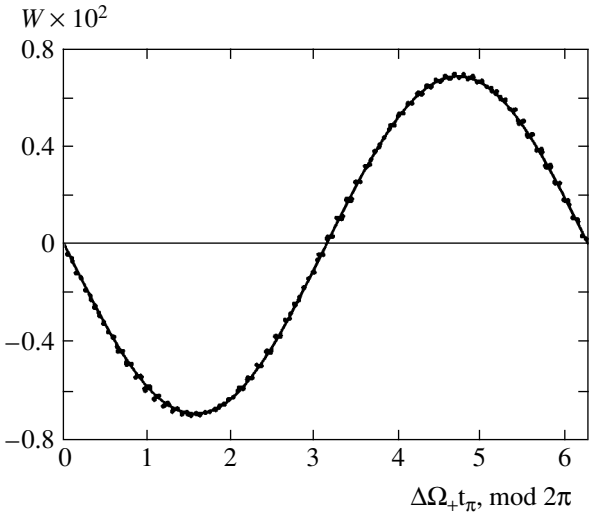


Fig. 3. Separatrix map for system (2.1) under asymmetric forcing defined by (5.5): symbols and curve represent numerical results and their least-squares approximation, respectively.

The analysis that follows is restricted to the most interesting case of sum-frequency harmonic. When the primary harmonics have almost equal frequencies (in absolute value), it has a low frequency. Therefore, one should find out if the system follows the behavior characteristic of the primary harmonics when $\lambda \ll 1$ under symmetric forcing (see preceding section).

As a first step, an analysis of system (2.1), (2.2) is performed for a high-frequency symmetric perturbation:

$$\omega_0^2 = 0.09, \quad \varepsilon_1 = \varepsilon_2 = 0.05, \quad m_1 = m_2 = 1, \quad \Omega_1 = 30.0, \quad \Omega_2 = -30. \tag{5.4}$$

The graph of the corresponding separatrix map is a perfectly sinusoidal curve with an amplitude of 1.65×10^{-4} (see [22] for details). By iterating the map, it has been shown that $w_u = |w_m| = w_l \approx 0.013$.

Now, suppose that a low-frequency secondary harmonic is generated by slightly changing the lower harmonic frequency:

$$\Omega_1 = 30.0, \quad \Omega_2 = -29.5, \quad \Delta\Omega_+ = 0.5. \tag{5.5}$$

Figure 3 shows the separatrix map calculated numerically for system (2.1), (5.5) at the instants t_π when the system passes through the stable equilibrium point at $x = 0.5$ (see [22]). Note that the sinusoidal curve obtained by least-squares fitting has an amplitude of 6.91×10^{-3} and, most importantly, the totally different frequency $\Delta\Omega_+ = 0.5$. Note also that $w_u \approx 0.50$, i.e., is greater by a factor of almost 40. A spectral analysis has shown that the total contribution of the primary harmonics is less than two percent. The system behaves as if the perturbation contains a single low-frequency harmonic rather than two high-frequency ones (as corroborated numerically in [23]).

It is clear that this dynamical behavior is qualitatively similar to that characteristic of pendulum dynamics: even a weak low-frequency secondary harmonic "outdoes" the primary ones as a factor contributing to the onset of chaos.

Additional analysis has shown that the separatrix-map amplitude scales linearly with frequency for a system characterized by a very low secondary-harmonic frequency, as in the case of symmetric perturbation, and the upper part of the stochastic layer has an almost frequency-independent width.

6. CONCLUSIONS

An analysis of a family of smooth Hamiltonian systems driven by a piecewise linear force performed as an extension of Bullett's pioneering study [8] has revealed quite unexpected dynamics.

Persistence of the separatrices of nonlinear resonances in a chaotic sea, complete suppression of diffusion in critical regimes, fractal diffusion in the neighborhoods of these regimes in the parameter space, periodic and power-law behavior of the Melnikov–Arnold integral as a function of frequency, and other phenomena have no analogs in the dynamics of systems with analytic Hamiltonian functions. The discovery of these phenomena challenges certain seemingly well-established views. In particular, it is commonly believed that the separatrices of nonlinear resonances with rational winding numbers are the least stable under forcing, whereas invariant curves with irrational winding numbers are the last to break down. As noted in the Introduction, a particularly wide variety of challenging questions are raised in the ongoing studies of fractal diffusion.

Continued research is required to answer these questions and find other systems exhibiting dynamical behavior of the type reviewed here.

ACKNOWLEDGMENTS

This work was supported, in part, by the Russian Academy of Sciences under the multidisciplinary scientific program "Mathematical methods in nonlinear mechanics."

REFERENCES

1. B. V. Chirikov, Phys. Rep. **52**, 263 (1979).
2. A. Lichtenberg and M. Lieberman, *Regular and Chaotic Dynamics*, 2nd ed. (Springer, Berlin, 1992).
3. G. M. Zaslavskiĭ and R. Z. Sagdeev, *Introduction to Nonlinear Physics* (Nauka, Moscow, 1988) [in Russian].
4. B. V. Chirikov, Chaos, Solitons and Fractals **1**, 79 (1991).

5. J. Moser, *Stable and Random Motion in Dynamical Systems* (Princeton Univ. Press, Princeton, 1973).
6. B. V. Chirikov, E. Keil, and A. Sessler, J. Stat. Phys. **3**, 307 (1971).
7. M. Hénon and J. Wisdom, Physica D (Amsterdam) **8**, 157 (1983).
8. S. Bullett, Commun. Math. Phys. **107**, 241 (1986).
9. M. Wojtkowski, Commun. Math. Phys. **80**, 453 (1981); Ergodic Theory Dyn. Syst. **2**, 525 (1982).
10. L. V. Ovsyannikov, private communication (1999).
11. V. V. Vecheslavov, Preprint No. 99-69, IYaF SO RAN (Inst. of Nuclear Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 1999).
12. V. V. Vecheslavov, Preprint No. 2000-27, IYaF SO RAN (Inst. of Nuclear Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 2000); nlin.CD/0005048.
13. V. V. Vecheslavov, Zh. Éksp. Teor. Fiz. **119**, 853 (2001) [JETP **92**, 744 (2001)].
14. V. V. Vecheslavov and B. V. Chirikov, Zh. Éksp. Teor. Fiz. **120**, 740 (2001) [JETP **93**, 649 (2001)].
15. V. V. Vecheslavov and B. V. Chirikov, Zh. Éksp. Teor. Fiz. **122**, 175 (2002) [JETP **95**, 154 (2002)].
16. B. V. Chirikov and V. V. Vecheslavov, Zh. Éksp. Teor. Fiz. **122**, 647 (2002) [JETP **95**, 560 (2002)].
17. V. V. Vecheslavov, Zh. Tekh. Fiz. **73** (9), 1 (2003) [Tech. Phys. **48**, 1079 (2003)].
18. V. V. Vecheslavov, Zh. Tekh. Fiz. **74** (5), 1 (2004) [Tech. Phys. **49**, 521 (2004)].
19. V. V. Vecheslavov, Zh. Tekh. Fiz. **72** (2), 20 (2002) [Tech. Phys. **47**, 160 (2002)].
20. V. V. Vecheslavov, Zh. Tekh. Fiz. **58** (1), 20 (1988) [Sov. Phys. Tech. Phys. **33**, 11 (1988)].
21. V. V. Vecheslavov, Zh. Éksp. Teor. Fiz. **125**, 399 (2004) [JETP **98**, 352 (2004)].
22. V. V. Vecheslavov, Zh. Éksp. Teor. Fiz. **109**, 2208 (1996) [JETP **82**, 1190 (1996)].
23. V. V. Vecheslavov, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 989 (1996) [JETP Lett. **63**, 1047 (1996)].

Translated by A. Betev