Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 1–8.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 5–13.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Spatio-Temporal Structure of the Muon Disk at $E_0 \geq 5 \times 10^{16}$ eV from Yakutsk EAS Array Data

A. V. Glushkov[a], L. G. Dedenko[b], M. I. Pravdin[a], and I. E. Sleptsov[a]

[a]*Shafer Institute of Space Physics and Aeronomy, Siberian Branch, Russian Academy of Sciences, Yakutsk, 677891 Russia*

[b]*Skobeltsin Institute of Nuclear Physics, Moscow State University, Moscow, 119992 Russia*

e-mail: a.v.glushkov@ikfia.ysn.ru

Received October 29, 2003

Abstract—We present the results of our study of the temporal structure of the muon disk at the Yakutsk array in extensive air showers with primary energies $E_0 \geq 5 \times 10^{16}$ eV at distances of 250–1500 m from the shower axis obtained using a large muon detector with an area of 184 m^2 and a detection threshold of $E_\mu \approx 0.5 \sec\theta$ GeV. We have found two components with different muon disk thicknesses that require significant revisions of our view of the development of extensive air showers. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Muons with energies of ~0.5–1.0 GeV are an important component of the extensive air showers (EASs) produced by ultrahigh-energy ($E_0 \geq 10^{15}$ eV) cosmic-ray particles. These are weakly absorbed in the atmosphere and are sensitive to the characteristics of the nuclear interactions during the development of an EAS and to the chemical composition of primary cosmic-ray (PCR) particles.

PCR composition is believed to be appreciably enriched with heavy nuclei at energies $3 \times 10^{15} < E_0 \leq 10^{17}$ eV (see, e.g., [1, 2]) and to rapidly change again toward protons at $10^{17} < E_0 \leq 10^{18}$ eV [3]. The composition is heaviest at $E_0 \approx 10^{17}$ eV. These results can be explained in terms of the diffusion model [4].

Muons with energies $E_\mu \geq 1.0$ GeV have been continuously studied at the Yakutsk EAS array since 1974. In our previous papers [5–8], we showed that air showers at $E_0 \geq (3–5) \times 10^{18}$ eV develop differently than those at lower energies. In our opinion [9–19], this difference may be attributable to new PCR particles of extragalactic origin.

A large muon detector with a threshold $E_\mu \approx 0.5 \sec\theta$ GeV (θ is the zenith angle) came into operation at the Yakutsk EAS array in November 1995. A preliminary analysis of the data obtained with this detector showed that it is highly efficient in studying the development of EASs [20, 21]. Below, we present some of the results of our study of the temporal structure of the muon disk. Based on the quark–gluon string (QGS) model [22], we compare the measured and calculated parameters.

2. THE LARGE MUON DETECTOR

The large muon detector (LMD) consists of 92 scintillation counters with an area of 2 m^2 similar to those used at the ground-based stations of the Yakutsk EAS array. These are arranged in six rows in a 26×12 m underground room located at a distance of 180 m from the array center. The main electronics was made in the CAMAC standard and placed in the same room, except for the IBM PC 486 control computer that was installed in the laboratory. This computer controls the proper operation of the electronic circuits, calibrates the detectors, and records and stores information. Information between the computer and the electronics in the underground room is exchanged by means of two serial drivers with six communication lines each.

The electronics was designed in such a way that the amplitude and arrival time of the signal in an air shower were measured from each counter, irrespective of the triggering of other detectors. An ATC (amplitude–time channel) unit was created for such measurements and installed on each scintillation counter. This unit consists of an amplitude channel (AC) to measure the number of muons and a time channel (TC) to measure the arrival time of the first particle. All ATCs are concentrated in eleven crates, and the signals from a FEU-125 photomultiplier tube are fed to the input of each ATC over an RK-75 cable from 20 to 70 m in length.

The amplitude channel operates with a tracking threshold (from 1 mV to 2 V) in a background load mode of 100–1000 events s^{-1}. Its dynamic range is ~10^3. The analog signal at the AC input is first delayed by about 100 ns (20 m of the RK-75 cable) for the time channel to generate a 2-μs-long square signal at the control AC input to integrate the charge q from the photomultiplier tube (PMT). In 2 μs, the linear transforma-

tion $q \longrightarrow T$ begins, and this duration is written into memory by a 10-MHz clock generator. Amplitude information is stored in the ATC memory until the arrival of the next signal from the PMT or until the central recorder ends the LMD polling (see below).

The time channel of the ATC, which consists of two synchronized channels, is used to accurately measure the time from the counter triggering (start) to the signal from the receiver of synchronization pulses of the main EAS array (stop). The start signal triggers the precise time channel (PTC) that measures the interval until the next clock of the 10-MHz reference generator, which is common to all ATCs. In turn, the PTC stop serves as the start of a coarse time channel (CTC) that counts the number of clocks of the reference generator until the main LMD stop generated by the receiver of synchronization pulses of the EAS array. The LMD signal stop is synchronized with the start of the next 10-MHz clock, which guarantees a simultaneous CTC stop for all of the counters triggered in an air shower. In the PTC, the short time interval (from 0 to 100 ns) is extended by approximately a factor of 100 by a time–time converter, and the same 10-MHz clock generator is used to produce a digital code. This scheme allows one to measure the entire interval until the internal LMD stop with an accuracy of ~5 ns and to achieve a relative accuracy of the same order of magnitude between different counters inside the basement.

The signal is fed to the time channel after preamplification and high-frequency front adjustment to a discriminator with a tracking threshold that triggers at a certain phase of the leading edge and makes the triggering of this channel independent of the input signal amplitude. Synchronization pulses of the EAS array are fed by a 10-kHz transmitter.

The total TC triggering time for an individual counter is defined as the sum of the CTC and PTC readings with an allowance made for the delay of the signal propagation over the cable from the PMT to the ATC and for the additional equipment delay.

Once the processing of the received signal has started, the ATC is disabled from repeated triggerings until the end of the processing in each channel (this time can change from 5 to 250 μs). If the Yakutsk array selects an EAS or records a local LMD triggering within 100 μs after the signal arrival, then the triggered ATCs are disabled to save information and to transfer data to the control computer.

Information about the last selected event is always stored in the ATC memory. It is cleared only when a new event arrives (if the unit has not been specially disabled on command). The ATC has a special trigger to memorize the counters that triggered in a given shower. It is set to a certain state in the presence of a signal at the ATC input when a shower is selected with a resolution time of 100 μs (participation bit) and is then polled when reading information.

A master selection unit (MSU) serves to synchronize the operation of all ATCs, to obtain the master of the main EAS array, to select local events, and to generate an interrupt signal for the computer. The signal on the shower recorded by the main EAS array is fed from the central recorder to the MSU over a special cable, and the MSU generates an interrupt for the computer that works with the LMD and sends a signal to all ATCs to generate a participation bit. Such an event is always recorded, irrespective of whether the LMD counters triggered or not.

In addition, the MSU itself selects local showers with energy $E_0 \sim 10^{15}$ eV when several counters trigger simultaneously within 2 μs. The number and arrangement of counters are determined by the adjustments of the adding circuits in each crate and in the MSU itself. A given type of selection can be permitted or forbidden by software. The control software can also simulate an artificial shower at the operator's command. It is used to test the equipment and check the connection between the computer and the remote electronics.

The MSU has memory to store 32 16-bit words into which the occurrence times of individual events are continuously written. A special ten-digit time counter (clocks) counts the number of clocks of the 10-MHz generator starting from the next array synchronization pulse. The clocks are zeroed when such a signal arrives. Ten digits are used to store the time, and the event flag is written into the remaining six digits. The events of this kind include the arrival of a synchronization pulse from the EAS array, whose time is written into memory before time counter reset (flag 0); the triggering of any detector from each row of 14–16 counters, the flag is specified by the row number from the first to the sixth digit; the termination of the pulse generated in the MSU in shower events to produce a participation bit in the ATC that disables further writing until the polling and the subsequent disabling by the control software (flag 0). This scheme allows one to obtain a time scan of events for several 100-μm periods until the time of shower selection and to control the operation of the time channel in individual ATCs.

A separate computer is used to record and accumulate the LMD data. In showers, the MSU generates an interrupt signal on which the event is written into a file on the hard disk of the computer. Each record of this file refers to one shower and has a structure that allows information from all LMD counters to be stored. To easily identify and combine the data from the LMD and the main EAS array in common showers, both recorders record the event time from UT (Universal Time) clocks. To this end, an input register is installed in a special crate connected to one of the serial driver lines, and the code from external UT clocks is fed to this register. The same code is simultaneously fed to the input register of the main recorder of the Yakutsk array.

In the intervals between showers, the LMD recording program accumulates check and calibration infor-

mation. It regularly measures the background and the amplitude spectra from all counters and writes these data into separate files on the hard disk. The latter are used for the amplitude calibration of the counters.

3. THE CHARACTERISTICS UNDER STUDY

We studied the temporal structure of the muon disk in EASs with energies $E_0 \geq 5 \times 10^{16}$ eV and zenith angles $\cos\theta \geq 0.7$. As an example, Fig. 1 shows the distribution of arrival delays T (curve *1*) for muons with a threshold energy $E_\mu \approx 0.5\sec\theta$ GeV relative to the plane front (the plane perpendicular to the EAS axis at the point of its intersection with the array plane) calculated using the QGS model for primary protons with $E_0 = 10^{18}$ eV and $\cos\theta \geq 0.9$ at $R = 630$ m from the shower axis. The mean muon density at this distance is $\rho_\mu(630) = 0.35$ m^{-2}, the mean delay is $\langle T\rangle = 156$ ns, and the standard deviation is $\sigma_T = 114.2$ ns. The full width at half-maximum of this pulse is $T_{1/2} = 107$ ns, and the width at 0.01 of its maximum is 475 ns. This is the total integration time (99%) of all muons in this specific event.

Curve *2* in Fig. 1 corresponds to the distribution obtained by integrating curve *1*. In fact, it reflects the pulse shape that would be recorded by an oscillograph at the output of an ideal detector. The width of this distribution at the level from 10 to 50% is $T15 = 62$ ns. This quantity is commonly measured in experiments and used as an EAS disk thickness parameter.

Of practical interest in measuring the particle density are not the values of T, but the relative delays

$$t_i = T_i - T_1, \tag{1}$$

where T_1 is the arrival time of the fastest muon from n triggered LMD counters in each individual shower, and T_i is the arrival time of the first muon to the ith counter. The reason is that the quantities in Eq. (1) reflect the actual time it takes to collect a given fraction of all particles at a chosen distance from the shower axis.

The relative delays (1) have an exponential distribution:

$$P(t) \approx \exp(-t/\lambda). \tag{2}$$

It uniquely depends on the mean density of the detected muons $\langle\rho_\mu(R)\rangle$ or, to be more precise, on the number of muons m that passed through an area S (in our case, $S = 2$ m^2), which for the mean

$$\langle m\rangle = \langle\rho_\mu(R)\rangle S$$

have the Poisson distribution

$$P_m(\langle m\rangle) = \langle m\rangle^m/m!\exp(-\langle m\rangle). \tag{3}$$

Histogram *3* in Fig. 1 represents the distribution of delays (1) obtained from curve *1* by the Monte Carlo method for $\langle m\rangle = 0.7$ and $n = 2$ triggered counters. The delay and its standard deviation that correspond to this histogram are $\langle t\rangle = 99$ ns and $\sigma_t = 115.8$ ns, respectively. The parameters $\langle t\rangle$ and $T15$ are related by

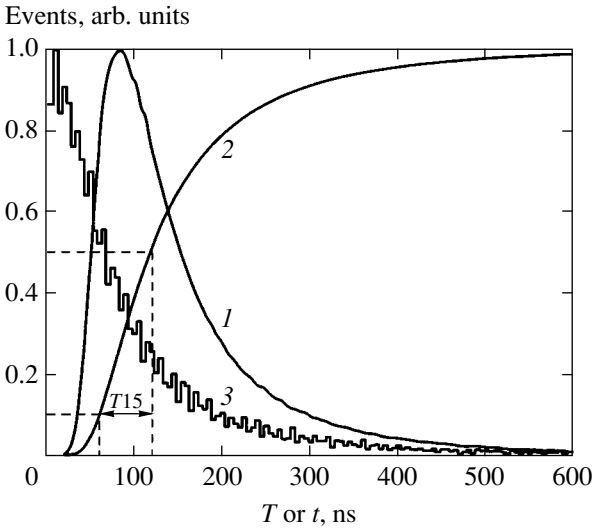


Fig. 1. The distributions of arrival delays for muons with energy $E_\mu \approx 0.5\sec\theta$ GeV at an observation level of $X = 1020\sec\theta$ g cm^{-2} calculated using the QGS model for primary protons with $E_0 = 10^{18}$ eV at $R = 630$ m from the axis of a shower with $\cos\theta \geq 0.9$: relative to the plane front (curve *1*); and relative to the first muon (curve *3*) at a density $\rho_\mu \leq 0.35$ m^{-2} and when two counters (with an area of 2 m^2) trigger in each of the 5000 showers; $T15$ is the rise time of curve *2* at the level from 10 to 50% when integrating curve *1*.

$$\langle t\rangle \approx 1.6(T15). \tag{4}$$

This relation may prove to be useful for estimating the signal integration time at the input of the amplitude converters when measuring the number of particles at various distances from the EAS axis.

Analysis shows that, in our case, the delays at $\rho_\mu \leq 1$ m^{-2} have exponential distributions (2) with similar parameters λ and $\langle t\rangle$. Thus, we can easily derive a relation to estimate the time T_η it takes to detect a fraction η of all muons:

$$T_\eta = -\langle t\rangle\ln(1-\eta). \tag{5}$$

It follows from Eq. (5) and from our calculations (see Fig. 1) that the effective thickness of the muon disk (95% of all particles) in showers with $E_0 \leq 10^{18}$ eV at $R \leq 630$ m from the EAS axis does not exceed 300 ns, and that 99.7% of all muons will be detected in the time $T_\eta = 600$ ns.

4. THE RESULTS

We analyzed the showers detected at the Yakutsk EAS array with the LMD during the period 1995–2002.

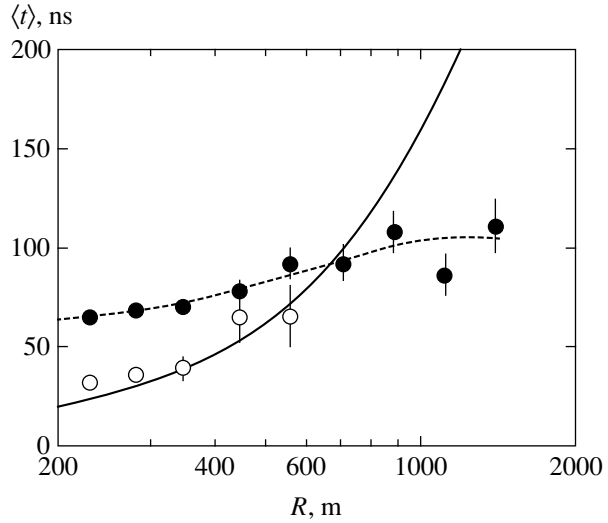


Fig. 2. Delays $\langle t \rangle$ in EASs with $\langle E_0 \rangle \approx 3 \times 10^{17}$ eV and $\langle \cos\theta \rangle \approx 0.95$ at various distances from the shower axis relative to the fastest muons in each shower: the muon densities are $\langle \rho_\mu \rangle \approx 0.45$ m^{-2} (filled circles) and $\rho_\mu \geq 2.5$ m^{-2} (open circles); the dashed curve represents the average behavior; the solid curve represents the calculations using the QGS model for primary protons.

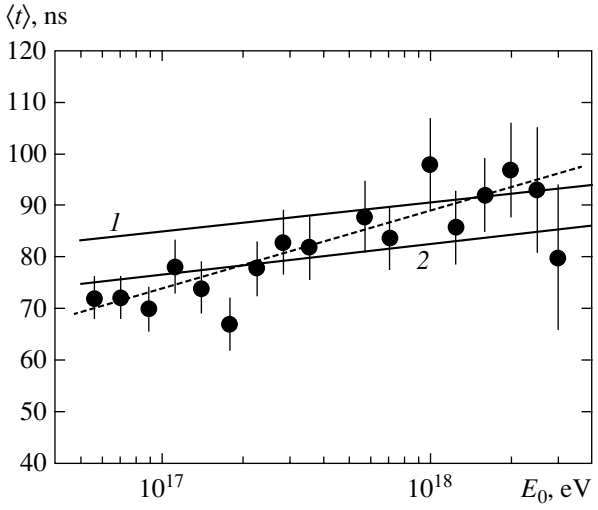


Fig. 3. Delay $\langle t \rangle$ versus E_0 in EASs with $\cos\theta \geq 0.8$ at R = 630 m from the shower axis: the filled circles indicate the experimental data at $\langle \rho_\mu \rangle \approx 0.45$ m^{-2} for the triggering of two LMD counters in each shower; the dashed curve represents the average behavior; the solid curves represent the calculations using the QGS model for primary protons (*1*) and iron nuclei (*2*).

The primary particle energy E_0 was determined from the relations

$$E_0 = (4.8 \pm 1.6) \times 10^{17} (\rho_{s,\,600}(0^\circ))^{1.00 \pm 0.02} \text{ [eV]}, \quad (6)$$

$$\rho_{s,\,600}(0^\circ) = \rho_{s,\,600}(\theta) \exp((\sec\theta - 1) \times 1020/\lambda_p) \text{ [m}^{-2}\text{]}, \quad (7)$$

$$\lambda_p = (450 \pm 44) + (32 \pm 15) \log(\rho_{s,\,600}(0^\circ)) \text{ [g/cm}^2\text{]}, \quad (8)$$

where $\rho_{s,\,600}(\theta)$ is the density of the charged particles measured by ground-based scintillation detectors at R = 600 m from the shower axis.

Below, we consider only the mean delays derived from parameters (1). Figure 2 shows $\langle t \rangle$ in showers with $10^{17} \leq E_0 \leq 10^{18}$ eV and $\cos\theta \geq 0.9$ at R = 250–1500 m from the EAS axis. The filled circles correspond to $\langle t \rangle$ during the triggering of two LMD counters when $m \leq 2$ ($\langle m \rangle \approx 0.9$) muons passed through them. The dashed curve represents the average behavior. The solid curve corresponds to the expected values of $\langle t \rangle$ calculated using the QGS model for primary protons.

We see that the measured dependence $\langle t \rangle$ is flatter than that predicted by the model. The difference between the dependences at R < 500 m stems from the fact that all muon densities were taken in the calculations without constraining them by the condition $m \leq 2$. This leads to a significant decrease in the "looseness" of the muon disk and, as a result, to a decrease in $\langle t \rangle$. This difference vanishes when $m \geq 5$ muons pass through LMD counters in the experiment (open circles). However, the discrepancy between the dependences at $R >$ 800 m remains difficult to explain.

Let us now analyze the thickness of the muon disk at $R \approx$ 500–800 m from the shower axis. There are several reasons for this choice of R. First, the LMD data are presented in this distance range most widely as a function of E_0 and θ. Second, since the muon density $\rho_{\mu,\,600}$ measured at a distance of 600 m from the EAS axis depends weakly on the zenith angle at $E_0 \leq$ (3–5) $\times$ 10^{18} eV [7, 8], it is another convenient parameter, along with (6), for estimating the primary particle energy. The following dependence in vertical showers was derived at the Yakutsk EAS array for muons with a threshold $E_\mu \geq 1.0$ GeV [8]:

$$E_0 = 2.4 \times 10^{18} (\rho_{\mu,\,600}(0^\circ))^{1.08 \pm 0.01} \text{ [eV]}. \quad (9)$$

Third, since there is agreement between our model and the experiment at $R \approx$ 500–800 m from the shower axis (see Fig. 2), calculations can be used to interpret the results obtained below.

We chose R = 630 m as a standard distance and reduced all our results to it. The lines in Figs. 3 and 4 indicate the expected (at this distance) delays $\langle t \rangle$ calculated using the QGS model for primary protons (*1*) and iron nuclei (*2*) after the triggering of two LMD counters when $m \leq 2$ muons passed through them: in showers with $\cos\theta \geq 0.8$ as a function of E_0 (Fig. 3) and in show-

ers with $10^{17} \leq E_0 < 10^{18}$ eV as a function of $\sec\theta$ (Fig. 4). The calculations satisfy the relation

$$\langle t\rangle \approx 104 + 5(\log(E_0/18) - \log A) - 140(\sec\theta - 1) + 0.194(R - 630) \quad (10)$$

for primary particles with an atomic weight A at $10^{16.7} \leq E_0 \leq 10^{18.3}$ eV and $\theta \leq 45°$ in the range of distances from the shower axis $500 \leq R \leq 1000$ m.

The filled circles correspond to the experimental data obtained for the same selection conditions for which the calculations were performed. The dashed curves represent the average behavior of the experimental data. We see that the measured and calculated values of $\langle t\rangle$ in Fig. 3 are consistent with the hypothesis about a mixed composition of the PCR particles at $E_0 \leq 10^{18}$ eV. This composition is significantly enriched with iron nuclei at $E_0 \approx 10^{17}$ eV and is close to the purely proton composition at $E_0 \approx 10^{18}$ eV. At first glance, this conclusion is in reasonably good agreement with the results obtained by many authors (see the Introduction). However, we see from Fig. 4 that the zenith-angle dependence of the experimental values of $\langle t\rangle$ is in poorer agreement with the calculations for oblique EASs ($\theta > 35°$–$40°$), where the mean thickness of the muon disk proved to be much larger than that expected from the QGS model.

To figure out this problem and the above discrepancy between theory and experiment in Fig. 2, we analyzed the measured delays (1) in more detail. Their distribution at $E_0 \leq (3–5) \times 10^{17}$ eV turned out to differ markedly from the purely exponential distribution (2). This is clearly seen from Fig. 5a, which shows the experimental results for a sample of 477 showers with $10^{16.7} \leq E_0 \leq 10^{17.0}$ eV and $\cos\theta \geq 0.8$ after the triggering of two LMD counters when $m \leq 2$ muons passed through them. The filled circles correspond to an integral distribution that may be represented as

$$N(\geq t) = N_1\exp(-t/\lambda_1) + N_2\exp(-t/\lambda_2). \quad (11)$$

The first term of this sum (solid line) has $\lambda_1 \approx 105$ ns and includes about 50% of all events. It is not attributable to technical LMD operation factors but reflects the presence of showers in the sample under consideration with a wider muon distribution at $R = 630$ m from the EAS axis than that expected from model calculations. If we subtract these events from the integral distribution, then the second term (dashed line) with $\lambda_2 \approx 52$ ns remains. At $E_0 > (5–7) \times 10^{17}$ eV, the pattern looks different. In this case, delays (1) have a purely exponential form over the entire range of their measurements. This is clearly seen from Fig. 5b, which shows the experimental results for a sample of 154 showers with $10^{17.9} \leq E_0 \leq 10^{18.2}$ eV and $\cos\theta \geq 0.8$ after the triggering of two LMD counters when $m \leq 2$ muons passed

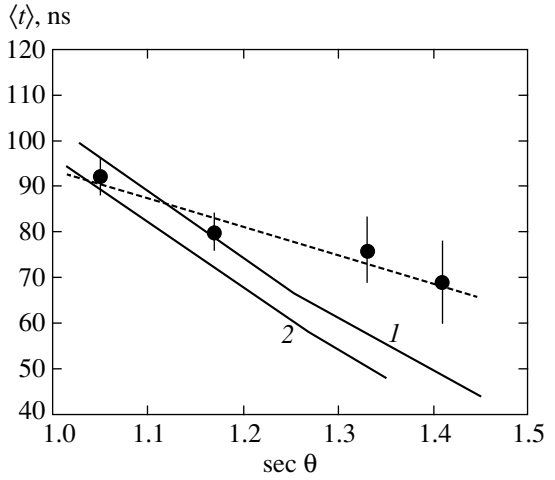


Fig. 4. Delay $\langle t\rangle$ versus $\sec\theta$ in EASs with $\langle E_0\rangle \approx 3 \times 10^{17}$ eV at $R = 630$ m from the shower axis. The notation is the same as that in Fig. 3.

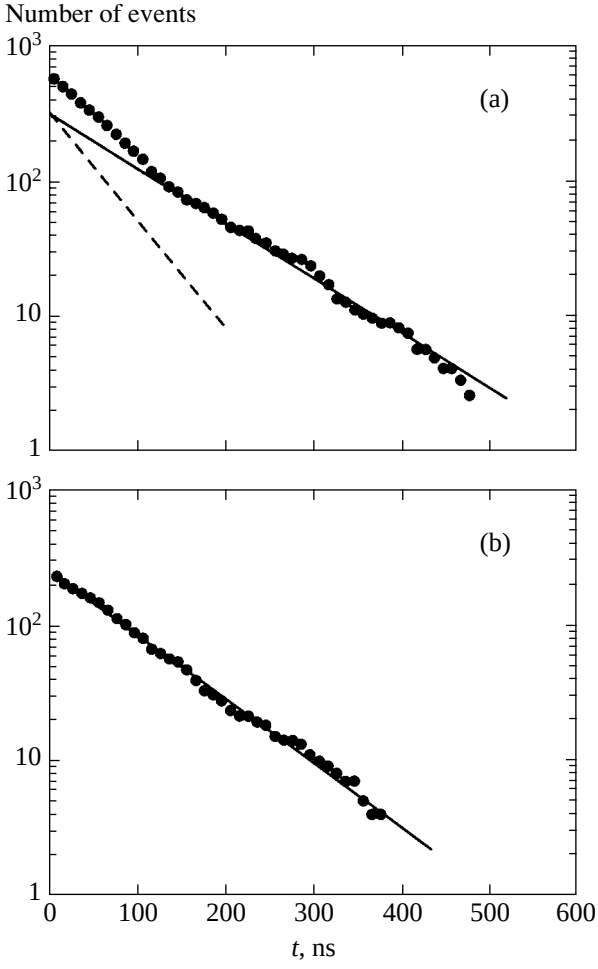


Fig. 5. The integral distribution of delays (1) at $R = 630$ m in EASs with $\langle\cos\theta\rangle \approx 0.9$ after the triggering of two LMD counters with $\langle\rho_\mu\rangle \approx 0.45$ m^{-2} in each shower for samples with $10^{16.7} \leq E_0 \leq 10^{17.0}$ eV (a) and $10^{17.9} \leq E_0 \leq 10^{18.2}$ eV (b): the solid line represents the first term in (11) with $\lambda_1 \approx 105$ ns; and the dashed line represents the second term in (11) with $\lambda_2 \approx 52$ ns.

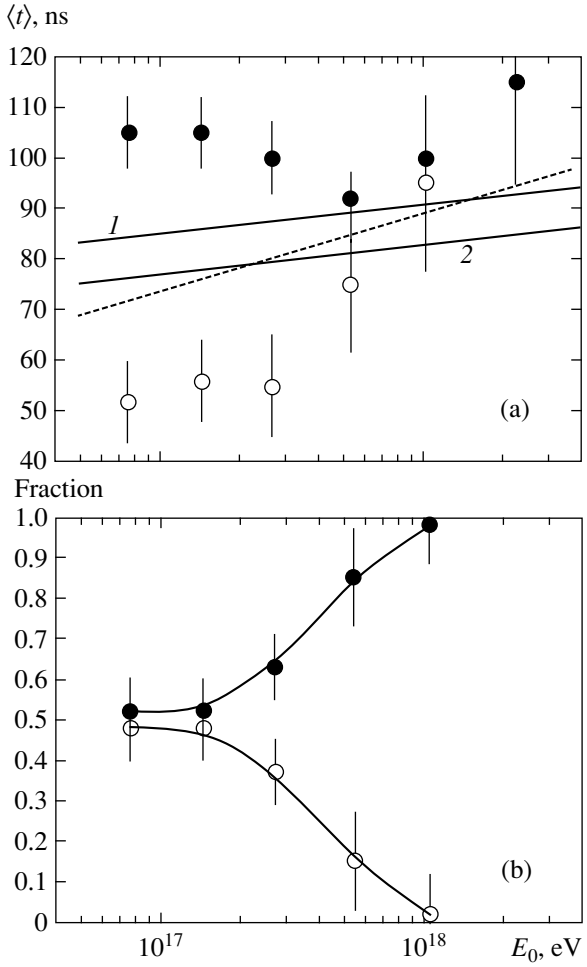


Fig. 6. Variations in the structural parameters λ_1 (filled circles) and λ_2 (open circles) of the integral distribution of delays (11) (a) and in the fraction of these components (b) with EAS energy (in $\Delta\log(E_0) = 0.3$ bins). The dashed and solid lines indicate the data from Fig. 3.

through them. The solid line corresponds to the distribution

$$N(\geq t) = N_1 \exp(-t/100).$$

5. DISCUSSION

Figure 6 shows variations in the structural parameters of the delay spectrum (11) with EAS energy in $\Delta\log(E_0) = 0.3$ bins. At $E_0 \leq (5\text{–}7) \times 10^{17}$ eV, delays (1) have a stable two-component distribution (11) with distinctly different parameters λ_1 (filled circles) and λ_2 (open circles). This can be seen from Fig. 6a, which, for comparison, also shows the model calculations and the average dependence of the experimental data (dashed line) presented in Fig. 3. Figure 6b shows the fraction of the two components. All of the results apply to $R \approx 630$ m and showers with $\langle\cos\theta\rangle \approx 0.9$.

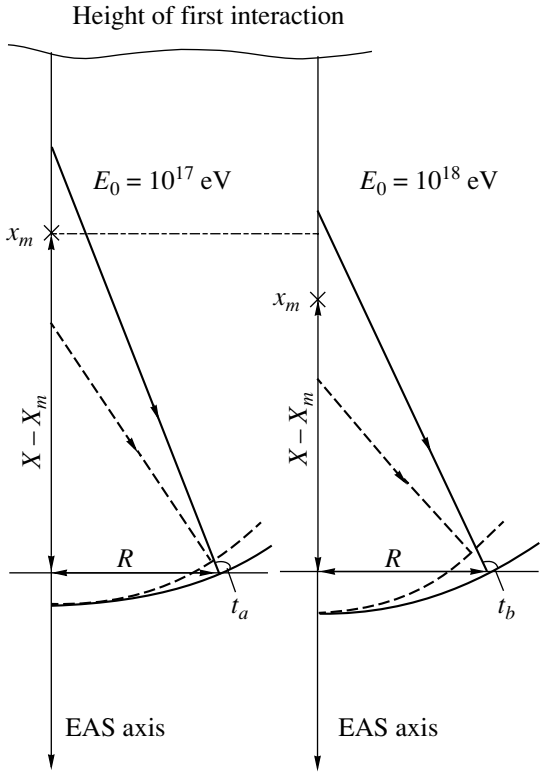


Fig. 7. The formation of relative delays in EASs with different primary energies.

Before we interpret these results, let us consider how the delays $\langle t\rangle$ are formed in general terms. Our calculations indicate that these are determined mainly by the muons that arrive from heights near the EAS maximum. The delay $\langle t\rangle$ decreases with decreasing primary energy E_0, because the depth of the shower maximum X_m recedes from the observation level X (for Yakutsk, $X = 1020\sec\theta$). This is clearly seen from Fig. 7, where the relative delays $t_a < t_b$ due to the deterioration of the geometrical muon collection factor. At fixed E_0, $X - X_m$ increases with zenith angle, causing the difference between the muon delays to decrease (see Fig. 4).

Let us revert to the data presented in Fig. 6 and try to understand their physical meaning. Given the formation mechanism of delays (1) described above, these may be assumed to be attributable to two types of cascade curves with distinctly different values of X_m. One refers to showers for which the depth of the maximum is much higher in the atmosphere than that expected from the QGS model for any composition of the primary particles (from protons to iron nuclei). The fraction of these showers at energies $E_0 \approx (4\text{–}10) \times 10^{17}$ eV rapidly decreases (see Fig. 6b).

These events may be assumed to be attributable to primary particles of a new nature. In [19], we analyzed the arrival directions of PCR particles with energies $E_0 \approx 10^{16.9\text{–}17.2}$ eV. We showed that a significant fraction

of these (about 50%) form clusters with a small-scale cellular structure and are probably produced by neutral particles of extragalactic origin. In [19], we also showed that these particles have a very short range for the first nuclear interaction ($\Lambda_1 \approx 3.3 \times 10^{-2}$ g cm^{-2}). These probably disappear after the first interaction, giving way to a normal cascade of secondary particles in EAS development; otherwise, the showers from them would differ greatly from ordinary showers and would easily reveal themselves. Due to such a short range Λ_1, the showers from enigmatic neutral particles would accelerate EAS development with a higher maximum of the cascade curve than that for primary protons. One of the components, indicated in Fig. 6 by open circles, may roughly reflect the contribution from these particles.

So far, it is hard to tell what these particles are. It may well be that these are neutrinos. The calculations performed in [23] indicate that EASs that are in many ways similar to the showers from PCR particles of ordinary composition can be formed under certain condition for the increase in the neutrino–nucleon interaction cross section ($\sigma_{\nu N}$) at ultrahigh energies. The hypothesis about ultrahigh-energy neutrinos that interact with relic neutrinos near Earth to produce Z-bozon showers was considered in [24, 25].

These neutral PCR particles may also prove to be neutral pions [26]. The possibility that stable pions can exist in the composition of ultrahigh-energy cosmic-ray particles was shown in this paper. The calculations were performed using the QGS model [22] by taking into account the Landau–Pomeranchuk–Migdal effect [27] in terms of the Coleman–Glashow hypothesis about the very weak violation of the Lorentz invariance. Of particular importance here is the fact that the reactions of the interaction between pions and cosmic microwave background photons are kinematically forbidden, which allows the Greisen–Zatsepin–Kuzmin paradox [28, 29] to be resolved.

The second group of data presented in Fig. 6 (filled circles) refers to showers for which the depth of the maximum is probably much lower in the atmosphere than that expected from the QGS model for primary protons. The fraction of these events at energies $E_0 \approx (4–10) \times 10^{17}$ eV rapidly decreases (see Fig. 6b). The cause of this disagreement between theory and experiment is not yet clear. The imperfectness of the QGS model may be responsible for this disagreement. However, it may well be that these events belong to long-range showers for which the depth of the maximum "sinks" deep into the atmosphere compared to ordinary EASs. Such showers are experimentally observed at the Tian Shan array at $E_0 > 2 \times 10^{16}$ eV [30]. These are explained by the formation of an appreciable fraction of charmed particles during the EAS development that penetrate deep into the matter without interaction and, thus, significantly shift the maximum of the cascade curve to the observation level.

6. CONCLUSIONS

In our view, the results shown in Figs. 3 and 6 are very instructive. On the one hand, they are roughly consistent with the hypothesis on the mixed composition of the PCR particles in the energy range under consideration, where it rapidly changes from dominant iron nuclei at $E_0 \approx 10^{17}$ eV to a purely proton composition at $E_0 \approx 10^{18}$ eV. On the other hand, a more careful analysis of the same experimental data yields results (see Fig. 6) that reveal a completely different pattern of EAS development. We do not rule out the situation where the superposition of two components in Fig. 6 can be camouflaged as showers from a mixed PCR composition. This may be why the problem of the origin of the first knee in the comic-ray spectrum at $E_0 \geq 3 \times 10^{15}$ eV has not yet been solved since it was detected more than 40 years ago by a team from Moscow State University [31]. Note that most of the methods for determining PCR composition are indirect. These are based on a comparison of various observational EAS characteristics with those calculated from models of EAS development with a particular assumed PCR composition.

A large number of experiments have been carried out over time, but there is no clear understanding of the EAS phenomenon as yet. Revealing its nature would in many ways contribute to solving the problem of the origin of cosmic-ray particles at ultrahigh energies (up to ~10^{20} eV). In [9–19], we showed that some of the PCR particles at $E_0 > 5 \times 10^{16}$ eV have a small-scale ordered structure related to the distribution of extragalactic pointlike PCR sources, which probably generate neutral particles. These conclusions are consistent with our results (regarding the possible existence of extragalactic neutral particles). However, this problem requires more serious theoretical astrophysical studies whose results would agree with the experimentally measured nuclear physical characteristics of EAS development; these studies are beyond the scope on the traditional views of PCR composition.

ACKNOWLEDGMENTS

This work was performed with financial support for the Yakutsk EAS array by the Russian Ministry of Science (reg. no. 01-30), which was included in the "List of Unique Research and Experimental Facilities of National Importance," and with financial support of the Federal Scientific and Technical Program "Research and Development on Priority Lines of Development of Science and Technology" for 2002–2006 (contract no. 40.014.1.1.1110).

REFERENCES

1. V. I. Vishnevskaya, N. N. Kalmykov, G. V. Kulikov, *et al.*, Yad. Fiz. **62**, 300 (1999) [Phys. At. Nucl. **62**, 265 (1999)].

2. H. Ulrich, T. Antoni, W. D. Apel, *et al.*, in *Proceedings of 27th ICRC* (Hamburg, 2001), p. 97.
3. T. Abu-Zayyad, K. Belov, D. J. Clay, *et al.*, astro-ph/0010652 (2000).
4. V. S. Ptuskin, S. I. Rogavaya, V. N. Zirakashvili, *et al.*, Astron. Astrophys. **268**, 726 (1993).
5. A. V. Glushkov, I. T. Makarov, E. S. Nikiforova, *et al.*, Yad. Fiz. **58**, 1265 (1995) [Phys. At. Nucl. **58**, 1186 (1995)].
6. A. V. Glushkov, I. T. Makarov, M. I. Pravdin, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 145 (2000) [JETP Lett. **71**, 97 (2000)].
7. A. V. Glushkov, M. I. Pravdin, I. E. Sleptsov, *et al.*, Yad. Fiz. **63**, 1557 (2000) [Phys. At. Nucl. **63**, 1477 (2000)].
8. A. V. Glushkov, M. I. Pravdin, I. E. Sleptsov, *et al.*, Yad. Fiz. **65**, 1346 (2002) [Phys. At. Nucl. **65**, 1313 (2002)].
9. A. V. Glushkov, Pis'ma Zh. Éksp. Teor. Fiz. **48**, 513 (1988) [JETP Lett. **48**, 555 (1988)].
10. A. V. Glushkov, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 355 (2001) [JETP Lett. **73**, 313 (2001)].
11. A. V. Glushkov and I. E. Sleptsov, Izv. Ross. Akad. Nauk, Ser. Fiz. **65**, 437 (2001).
12. A. V. Glushkov and M. I. Pravdin, Zh. Éksp. Teor. Fiz. **119**, 1029 (2001) [JETP **92**, 887 (2001)].
13. A. V. Glushkov and M. I. Pravdin, Pis'ma Astron. Zh. **27**, 577 (2001) [Astron. Lett. **27**, 493 (2001)].
14. A. V. Glushkov, Pis'ma Zh. Éksp. Teor. Fiz. **75**, 3 (2002) [JETP Lett. **75**, 1 (2002)].
15. A. V. Glushkov, Izv. Ross. Akad. Nauk, Ser. Fiz. **66**, 1599 (2002).
16. A. V. Glushkov and M. I. Pravdin, Pis'ma Astron. Zh. **28**, 341 (2002) [Astron. Lett. **28**, 296 (2002)].
17. A. V. Glushkov, Pis'ma Astron. Zh. **29**, 172 (2003) [Astron. Lett. **29**, 142 (2003)].
18. A. V. Glushkov and M. I. Pravdin, Yad. Fiz. **66**, 886 (2003) [Phys. At. Nucl. **66**, 854 (2003)].
19. A. V. Glushkov, Yad. Fiz. **66**, 1292 (2003) [Phys. At. Nucl. **66**, 1252 (2003)].
20. A. V. Glushkov, V. B. Kosarev, I. T. Makarov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 361 (1998) [JETP Lett. **67**, 383 (1998)].
21. A. V. Glushkov, L. G. Dedenko, V. B. Kosarev, *et al.*, in *Proceedings of 26th ICRC* (Salt Lake City, 1999), Vol. 1, p. 387.
22. A. B. Kaĭdalov, K. A. Ter-Martirosyan, and Yu. M. Shabel'skiĭ, Yad. Fiz. **43**, 1282 (1986) [Sov. J. Nucl. Phys. **43**, 822 (1986)].
23. L. Anchordoqui, H. Goldberg, T. McCauley, *et al.*, hep-ph/0011097.
24. D. Fargion, B. Mele, and A. Salis, Astrophys. J. **517**, 725 (1999).
25. T. J. Weiler, Astropart. Phys. **11**, 303 (1999).
26. L. G. Dedenko, T. M. Roganova, G. F. Fedorova, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **78**, 131 (2003) [JETP Lett. **78**, 101 (2003)].
27. A. B. Migdal, Phys. Rev. **103**, 1811 (1956).
28. K. Greisen, Phys. Rev. Lett. **2**, 748 (1966).
29. G. T. Zatsepin and V. A. Kuz'min, Pis'ma Zh. Éksp. Teor. Fiz. **4**, 78 (1966) [JETP Lett. **4**, 53 (1966)].
30. P. F. Bejl, R. U. Bejsembaev, N. G. Vildanov, *et al.*, in *Proceedings of 28th ICRC* (Tsukuba, 2003), Vol. 1, p. 9.
31. G. V. Kulikov and G. B. Khristiansen, Zh. Éksp. Teor. Fiz. **35**, 635 (1958) [Sov. Phys. JETP **8**, 441 (1959)].

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 101–108.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 120–128.

FLUIDS

Fine Structure of Angular Selectivity of Diffraction Efficiency of Holograms Recorded on Materials Polymerized by IR Radiation

R. S. Akopyan[a,*], A. V. Galstyan[a], and T. V. Galstian[b]

[a] *Yerevan State University, Yerevan, 375025 Armenia*
[b] *Optics, Photonics, and Laser Center, Laval University, G1K7P4, Quebec, Canada*
**e-mail: rhakob@server.physdep.r.am*

Received November 24, 2003

Abstract—Asymmetry in the angular selectivity of the diffraction efficiency observed in polymer-dispersed liquid-crystal holograms polymerizing under the action of IR radiation is studied experimentally and explained theoretically. The theory of coupled waves developed for anisotropic media successfully describes basic diffraction properties of the anisotropic holographic gratings obtained. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Diffraction of light from volume phase gratings based on multilayered or holographic materials has been studied in detail (see, for example, [1]). In particular, holographic polymer systems have been thoroughly studied in recent decades [2–4] in connection with their possible application in various optical systems used for data recording [5], detecting acoustic waves [6], obtaining flat displays as well as displays with a required curvature [7, 8], beam cleanup [9], holographic focusing [10], etc. In contrast to classical recording materials such as photographic films, photoresistors, or gelatins, these systems permit recording with a high resolution in real time with low irradiation energies and do not require a wet process of film development. Hologram fixation in these materials is ensured by polymerization (initiated in illuminated regions) and molecular diffusion; as a result, the refractive index experiences periodic modulation.

In some cases, it is required to control the diffraction efficiency, for example, for obtaining electrooptically controlled holographic multipliers, data storage units, and lenses with a dynamically variable focal length. In this connection, the interest in holographic polymer-dispersed liquid crystals (PDLCs) has increased in recent years [11]. This interest is due to strong optical anisotropy and the possibility to control birefringence in LC drops, which makes it possible to easily control the optical properties of parts made of such materials.

Holographic PDLC materials are formed by illuminating homogeneous light-sensitive monomer–LC mixtures with two coherent interfering waves. In the course of polymerization, as the number of monomers decreases in the illuminated areas, monomers from dark regions diffuse to illuminated areas. For PDLCs, LC molecules are chosen that can mix with the initial monomer solution but do not form mixtures with a polymer or partly polymerized solution. As a result, phase separation of the LC and the polymer takes place, producing 3D LC domains (drops) predominantly in dark regions [12]. In some special cases, well-separated planar zones of the polymer and the LC with a submicrometer resolution are formed [13]. In all these systems, the initial homogeneous mixture is completely separated into polymer-enriched regions and the LC-enriched regions; the refractive index of the material experiences a periodic perturbation in this case. The molecular-orientation distribution in each drop (or layer) of a LC is controlled by the elastic deformation energy associated with the coupling conditions at the walls of drops, pressure, and variation of the drop shape and size. Since the LC density in dark regions is higher than in illuminated regions, the effective refractive index in these regions is mainly determined by the LC orientation. Illuminated regions are more saturated with the polymer and the total refractive index of these regions is close to that of the polymer. For this reason, modulation of the refractive index and the diffraction efficiency of such PDLC gratings strongly depend on the size of LC drops, their density, shape, and LC orientation in the drops. Since these materials are strongly anisotropic, one of the main questions is understanding the role of LCs in the angular and polarization dependences of the characteristics of the obtained systems [14–16].

On the other hand, it is well known that the key parameter of 3D holographic gratings is the Bragg condition and the Bragg detuning. Although this condition for the above anisotropic systems can be obtained using rigorous numerical methods, a simple formula for estimating this condition as well as a moderate angular deviation from it would be very useful. Indeed, the

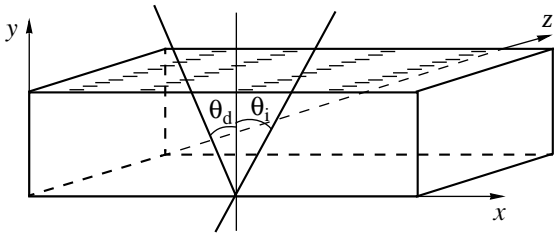


Fig. 1. Diffraction of light from a thick transmission hologram.

problem of phase deviation and the diffraction shift associated with this deviation is quite important. The most generally accepted technique for describing diffraction properties of thick oscillograms is the theory of coupled waves or the theory of coupled modes [17]. This theory was successfully generalized in [18] to the case of diffraction gratings produced from anisotropic materials. However, the final results are given in [18] only for photorefractive crystals. In addition, the angular dependence of diffraction efficiency is not derived in explicit form, which makes it impossible to use this dependence for explaining experimental results. It is therefore necessary to modify this theory for studying anisotropic diffraction in the case of the Bragg detuning. We performed specific theoretical calculations of an anisotropic diffraction grating to explain the specific behavior of recently discovered [19, 20] holographic PDLCs, which are sensitive to the near-IR spectral region (850 nm). In some cases, such gratings may have a diffraction efficiency above 95%. Since the above-mentioned wavelength is important for laser diodes and lasers with vertical resonators, it has numerous applications in the recording of holographic structures.

In many of the above-mentioned cases, asymmetry in the angular selectivity of holographic elements of this type is of special importance. Although this asymmetry is an experimentally established fact [12, 13], it is often disregarded. It should also be noted that our experimental and theoretical investigations make it possible to obtain information on the orientation of molecules in LC drops without resorting to complex microscopic methods.

The article has the following structure. In Section 2, the solution is given for a system of coupled wave equations in the unsalted-band approximation for purely phase-transmission 3D grating. A special case of phase deviation from the Bragg conditions in holographic PDLCs is considered in Section 3. Corresponding experimental results are described in Section 4. Conclusions are summarized in Section 5.

2. THEORETICAL TREATMENT

Let us consider a unsalted transmission diffraction grating of thickness d and period Λ. We direct the y axis along the normal to the grating surface and the x axis along the grating vector (Fig. 1). We assume that incident light with a wavelength λ_0 and a wave number $k_0 = 2\pi/\lambda_0$ is monochromatic and linearly polarized in vacuum. Light is incident on a PDLC at an angle θ_i measured in the sample. We consider the grating formed by the modulation of the refractive index alone; consequently, the permittivity tensor $\hat{\varepsilon}$ has no complex part. We assume that $\hat{\varepsilon}$ varies sinusiodally along the x axis and that the principal optical axis of the medium is also directed along the x axis. Consequently, we can write permittivity tensor $\hat{\varepsilon}$ in the form

$$\hat{\varepsilon} = \hat{\varepsilon}^0 + \hat{\varepsilon}^1 \cos(Kx), \tag{1}$$

where $\hat{\varepsilon}^0$ is the mean permittivity tensor and $\hat{\varepsilon}^1$ is the permittivity tensor whose components describe the modulation depth. The wave vector $\mathbf{K}$ is oriented along the x axis; the magnitude of this vector is $K = 2\pi/\Lambda$, where Λ is the grating period. Following [17], we consider the propagation of only two waves (incident and diffracted) for thick holograms. We can write the expression for the electric field in an anisotropic medium as the sum of two above-mentioned waves,

$$\begin{aligned}\mathbf{E}(\mathbf{r}, t) = [\mathbf{E}_i(\mathbf{r})\exp(i\mathbf{k}_i\cdot\mathbf{r}) + \mathbf{E}_d(\mathbf{r})\exp(i\mathbf{k}_d\cdot\mathbf{r})]\\ \times\exp(i\omega t) + \text{c.c.},\end{aligned} \tag{2}$$

where $\mathbf{E}_i(\mathbf{r})$ and $\mathbf{E}_d(\mathbf{r})$ are the complex amplitudes of the incident and diffracted waves, respectively. Using the Bragg condition $\mathbf{k}_i + K = k_d$, we can write

$$\exp(i\mathbf{k}_i\cdot\mathbf{r})\exp(i\mathbf{K}\cdot\mathbf{r}) = \exp(i\mathbf{k}_d\cdot\mathbf{r}),$$

$$\exp(i\mathbf{k}_d\cdot\mathbf{r})\exp(-i\mathbf{K}\cdot\mathbf{r}) = \exp(i\mathbf{k}_i\cdot\mathbf{r}).$$

Taking these relations into account as was done in [18], we substitute Eqs. (1) and (2) into the wave equation

$$\nabla\times(\nabla\times\mathbf{E}(\mathbf{r})) = k_0^2\hat{\varepsilon}\mathbf{E}(\mathbf{r}). \tag{3}$$

Eliminating the terms describing the propagation of light in an unperturbed medium and applying the slowly-varying-amplitude approximation, we arrive at the following system of coupled equations:

$$2i|k_i|\nabla E_i[\hat{\mathbf{k}}_i - \mathbf{e}_i(\mathbf{e}_i\cdot\hat{\mathbf{k}}_i)] = \frac{k_0}{2}AE_d, \tag{4}$$

$$\begin{aligned}2i|k_d|\nabla E_d[\hat{\mathbf{k}}_d - \mathbf{e}_d(\mathbf{e}_d\cdot\hat{\mathbf{k}}_d)]\\ -2k_iE_d(\mathbf{d}_d\cdot\mathbf{e}_d)^2\Delta = \frac{k_0}{2}AE_i.\end{aligned} \tag{5}$$

Here and below, we use the following notation: $\mathbf{e}_i = \mathbf{E}_i/E_i$ and $\mathbf{e}_d = \mathbf{E}_d/E_d$ are the unit vectors directed along

$\mathbf{E}_i$ and $\mathbf{E}_d$, respectively; $\mathbf{d}_{i,d}$ are the unit vectors along the corresponding electric induction vectors; $A = \mathbf{e}_i\hat{\varepsilon}^1\mathbf{e}_d = \mathbf{e}_d\hat{\varepsilon}^1\mathbf{e}_i$; and $\hat{\mathbf{k}}_{i,d}$ are the unit vectors along the wave vectors $\mathbf{k}_{i,d}$. Following [17], we define the phase detuning from the Bragg condition as

$$\Delta = \frac{k_d^2 - k_i^2}{2k_i}. \quad (6)$$

This approach is more convenient for comparing theoretical and experimental results. The expressions in the brackets in Eqs. (4) and (5) are vectors in the direction of energy propagation (Poynting vector); consequently, we can write

$$\begin{aligned}\hat{\mathbf{k}}_i - \mathbf{e}_i(\mathbf{e}_i\cdot\hat{\mathbf{k}}_i) &= g_i\mathbf{u}_i,\\ \hat{\mathbf{k}}_d - \mathbf{e}_d(\mathbf{e}_d\cdot\hat{\mathbf{k}}_d) &= g_d\mathbf{u}_d,\end{aligned} \quad (7)$$

where $\mathbf{u}_i$ and $\mathbf{u}_d$ are unit vectors in the direction of the Poynting vector and $g_{i,d}$ are cosines of the angles formed by the wave vectors and the Poynting vectors. In the plane-wave approximation, we assume that the electric field amplitude depends only on coordinate *y*. Thus, We can write Eqs. (4) and (5) for coupled waves in the form

$$E_i' = -i\chi_i E_d, \quad (8)$$

$$E_d' + 2i\frac{\xi}{d}E_d = -i\chi_d E_i, \quad (9)$$

where the derivative is taken only with respect to *y*; ξ is the parameter describing Bragg detuning; $\chi_{i,d}$ are the coupling constants for two waves,

$$\xi = \frac{dg_d k_i\Delta}{2k_d\cos\varphi_d}, \quad \chi_{i,d} = \frac{k_0 A}{4n_{i,d}g_{i,d}\cos\varphi_{i,d}}; \quad (10)$$

φ_i and φ_d are the angles between the normal to the surface of the PDLC (along the *y* axis) and the Poynting vector for the incident and diffracted beams, respectively; and $n_{i,d}$ is the mean refractive index for the incident and diffracted beams ($k_{i,d} = k_0 n_{i,d}$). We can find the electric field amplitude at the grating outlet by solving Eqs. (8) and (9) for coupled waves. The principal solution to this system can be written in the form

$$\begin{aligned}E_i &= E_i^{01}\exp(\gamma_1 y) + E_i^{02}\exp(\gamma_2 y),\\ E_d &= E_d^{01}\exp(\gamma_1 y) + E_d^{02}\exp(\gamma_2 y),\end{aligned} \quad (11)$$

where $E_{i,d}^{01}$ and $E_{i,d}^{02}$ are complex-valued constants. Substituting these expressions into the system of equations (8), (9), we obtain

$$\gamma_{1,2} = -i\frac{\xi}{d} \pm i\sqrt{\chi_i\chi_d + \frac{\xi^2}{d^2}}. \quad (12)$$

Taking into account the normalized initial conditions $E_i(0) = 1$ and $E_d(0) = 0$, as well as relation (12), we obtain the solution to system (8), (9) for diffracted waves with the *p* and *s* polarization in the form

$$E_d(d) = -i\sqrt{\frac{\chi_d}{\chi_i}}\frac{\sin\sqrt{\nu^2 + \xi^2}}{\sqrt{1 + \xi^2/\nu^2}}e^{i\xi}, \quad (13)$$

where $\nu = d\sqrt{\chi_i\chi_d}$ describes the modulation of the grating and ξ describes the phase detuning from the Bragg condition ($\xi = 0$ when light is incident at the Bragg angle).

Diffraction efficiency η is defined as the ratio of the normal (to the grating surface) components of the power flux of the diffracted and incident waves. Taking into account relation (13), we obtain the following expression for the diffraction efficiency:

$$\eta = \frac{\sin^2\sqrt{\xi^2 + \nu^2}}{1 + \xi^2/\nu^2}. \quad (14)$$

This expression has the same form as the corresponding expression in [17, 18]. The definition of ν coincides with that given in [18], while the phase detuning is defined as in [17].

3. HOLOGRAPHIC POLYMER-DISPERSE LC WITH PHASE DETUNING

To further develop the model and compare the theoretical results with experiment, we must make certain assumptions concerning a composite grating. We assume that the grating contains regions with a high concentration of LC drops, which are separated by solid polymer regions.

Measurements of scattering of *p*- and *s*-polarized light with normal incidence shows that scattering in the case when light is polarized along the wave vector of the grating (*p* polarization) is stronger than for the *s* polarization (perpendicular to the wave vector). It is well known that anisotropic LC molecules scatter light polarized along the director more effectively than light polarized across the director of LC molecules [21]. Thus, we have established that LC molecules in the drops are oriented parallel to the grating vector (see Fig. 1), which is confirmed by our analysis of diffraction efficiency reported here. Thus, *s*-polarized light has the effective refractive index determined by the perpendicular component of the LC permittivity. This is confirmed by our experiments with a polarization microscope and scanning with the help of an electron

microscope. Similar results on the orientation of LC molecules were obtained in [22, 23] by comparing the theoretical and experimental studies of diffraction efficiency of a slightly different PDLC. We also assume that polymer-enriched regions are almost free of LC drops. With allowance for the above-mentioned assumptions, the permittivity tensor for the PDLC has the form

$$\hat{\varepsilon} = \begin{pmatrix} \varepsilon_{xx} & 0 & 0 \\ 0 & \varepsilon_{yy} & 0 \\ 0 & 0 & \varepsilon_{zz} \end{pmatrix}, \tag{15}$$

where $\varepsilon_{xx} = \varepsilon_{\parallel}$ is the LC permittivity for light with the polarization vector directed along the grating vector and $\varepsilon_{yy} = \varepsilon_{zz} = \varepsilon_{\perp}$ is the permittivity for light with the polarization vector perpendicular to the grating vector. Thus, in Eq. (1) we set

$$\begin{aligned} \varepsilon^0_{\perp,\parallel} &= \varepsilon^{\mathrm{LC}}_{\perp,\parallel} c + \varepsilon_{\mathrm{pol}}(1-c), \\ \varepsilon^1_{\perp,\parallel} &= (\varepsilon^{\mathrm{LC}}_{\perp,\parallel} c - \varepsilon_{\mathrm{pol}})c, \end{aligned} \tag{16}$$

where c is the volume concentration of the LC, $\varepsilon^{\mathrm{LC}}_{\parallel}$ and $\varepsilon^{\mathrm{LC}}_{\perp}$ are the LC permittivity components for light polarized parallel and perpendicularly to the grating vector, and $\varepsilon_{\mathrm{pol}}$ is the permittivity of the polymer. For obtaining numerical estimates, we set $\varepsilon^{\mathrm{LC}}_{\parallel} = 2.95$ and $\varepsilon^{\mathrm{LC}}_{\perp} = 2.3$ for an E7 LC and $\varepsilon_{\mathrm{pol}} = 2.4$ for the photopolymer used in our experiments. The LC concentration in our experiments was on the order of 30%.

To study the angular selectivity of the diffraction efficiency, we must know the explicit dependences of all parameters on the angle of incidence in expression (14) for diffraction efficiency. It can easily be seen that the angles $\theta_{\mathrm{i,d}}$ between the normal to the grating surface (y axis) and the wave vectors $\theta_{\mathrm{i,d}}$ of the incident and diffracted light are connected with angles $\varphi_{\mathrm{i,d}}$ via the relations $\varphi_{\mathrm{i,d}} = \theta_{\mathrm{i,d}} + \arccos g_{\mathrm{i,d}}$. In the general case, the expressions for refractive indices $n_{\mathrm{i,d}}$ and the cosines of angles $g_{\mathrm{i,d}}$ between the wave vectors and the Poynting vectors have the form

$$\begin{aligned} g_{\mathrm{i,d}} &= \frac{\varepsilon^0_{\parallel}\sin^2\theta_{\mathrm{i,d}} + \varepsilon^0_{\perp}\cos^2\theta_{\mathrm{i,d}}}{\sqrt{(\varepsilon^0_{\parallel})^2\sin^2\theta_{\mathrm{i,d}} + (\varepsilon^0_{\perp})^2\cos^2\theta_{\mathrm{i,d}}}}, \\ n^2_{\mathrm{i,d}} &= \frac{\varepsilon^0_{\perp}\varepsilon^0_{\parallel}}{\varepsilon^0_{\parallel}\sin^2\theta_{\mathrm{i,d}} + \varepsilon^0_{\perp}\cos^2\theta_{\mathrm{i,d}}}, \end{aligned} \tag{17}$$

for *p*-polarized light and

$$g_{\mathrm{i,d}} = 1, \quad n^2_{\mathrm{i,d}} = \varepsilon^0_{\perp} \tag{18}$$

for *s*-polarized light. Definition (6) of Δ leads to the following expression:

$$\Delta = \frac{K^2 - 2k_0 K n_{\mathrm{i}}\sin\theta_{\mathrm{i}}}{2k_0 n_{\mathrm{i}}}. \tag{19}$$

Parameter $A = e_{\mathrm{i}}\hat{\varepsilon}^1 e_{\mathrm{d}} = e_{\mathrm{d}}\hat{\varepsilon}^1 e_{\mathrm{i}}$ can be written in the form

$$A = \varepsilon^1_{\perp}\sin\varphi_{\mathrm{i}}\sin\varphi_{\mathrm{d}} - \varepsilon^1_{\parallel}\cos\varphi_{\mathrm{i}}\cos\varphi_{\mathrm{d}} \tag{20a}$$

for the *p* wave and

$$A = \varepsilon^1_{\perp} \tag{20b}$$

for the *s* wave.

It can easily be verified that all expressions in (17), (19), and (20) are asymmetric relative to the Bragg angle θ_{B}, which can be determined from formula (19) by setting $\Delta = 0$. Consequently, the angular dependence of diffraction efficiency must also be asymmetric relative to the Bragg angle θ_{B}. It is important to note that $\theta_{\mathrm{i,d}}$ are angles in the sample; to compare the theoretical results with experimental data, we recalculated the results for the external angle.

In numerical calculations with a small Bragg detuning, we set $\theta_{\mathrm{i}} = \theta_{\mathrm{d}} = \theta$. The angular dependences of parameters $g_{\mathrm{i,d}}$ and the refractive index lead to a small correction to the asymmetry in the angular dependence of diffraction efficiency. The main contribution to the asymmetry comes from the asymmetric dependence of the Bragg detuning and the asymmetric dependence of modulation parameter A or ν. The former mechanism is isotropic (is also observed in isotropic gratings), while the latter mechanism is anisotropic (is present only in anisotropic gratings).

Numerical calculations were made with the help of MATHEMATICA 4.0. The parameters used in calculations for the E7 polymer and liquid crystal are given below. The wavelength of the test beam was $\lambda = 0.628$ μm, the lattice period was $\Lambda = 1.0$ μm, and the thickness was $d = 27.4$ μm. It can be seen from Fig. 2 (solid curve) that the angular dependence of diffraction efficiency is slightly asymmetric relative to the Bragg angle $\theta_{\mathrm{B}} = 18.3° = 0.32$ rad. The first left peak (η_{-1}) in Fig. 2 is obviously higher the first right peak (η_{+1}). We define parameter a as a measure of asymmetry as

$$a = 2\frac{\eta_{-1} - \eta_{+1}}{\eta_{-1} + \eta_{+1}}, \tag{21}$$

in this case, $a_p \approx 6\%$ for a *p*-polarized test wave in our experiments with the above-mentioned parameters. To calculate the diffraction efficiency for an *s*-polarized wave, we must substitute $\varepsilon_{\perp}$ for $\varepsilon_{\parallel}$ everywhere. Then the angular dependence of the modulation parameter disappears, and asymmetry (19) of the Bragg detuning

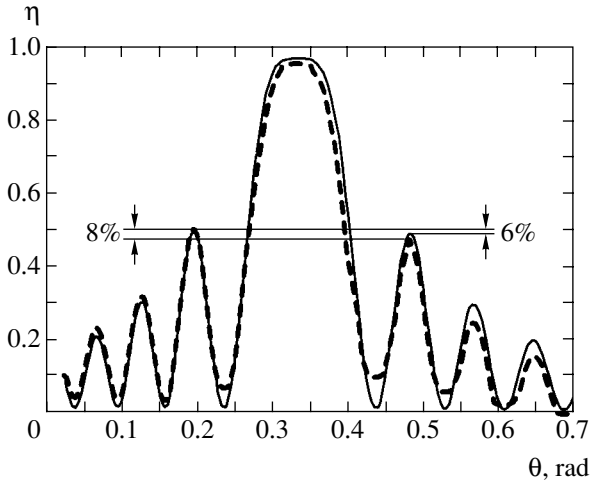


Fig. 2. Theoretical (solid curve) and experimental (dashed curve) dependences of the diffraction efficiency on the angle of incidence for a *p*-polarized wave.

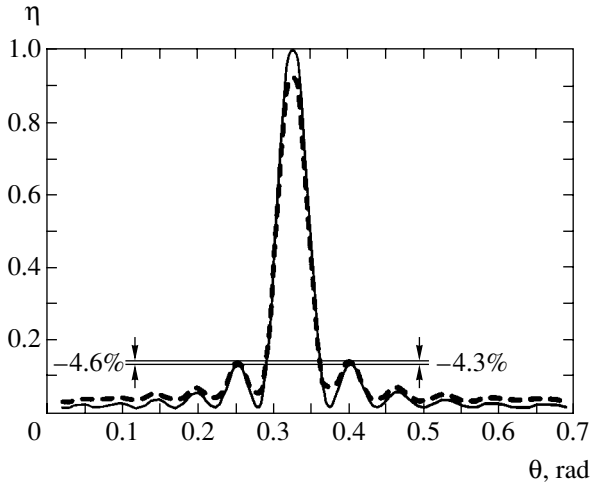


Fig. 3. Theoretical (solid curve) and experimental (dashed curve) dependences of the diffraction efficiency on the angle of incidence for an *s*-polarized wave.

remains the only mechanism of angular asymmetry in the diffraction efficiency (solid curve in Fig. 3). In accordance with definition (21), this asymmetry is negative and $a_s \approx -4.3\%$. It follows from the obtained results that the contribution to asymmetry from the anisotropic mechanism is much larger than from the isotropic mechanism. In addition, the contribution of the anisotropic mechanism is partly compensated by the isotropic mechanism of asymmetry.

It is interesting to consider a certain difference between the side peaks in the case of *p*- and *s*-polarized test waves. It should be noted that the side peaks are higher for the *p*-polarized test wave. It can be seen from relation (14) that the *m*th-order diffraction peak is given by

$$\eta_m = \frac{\nu^2}{\nu^2 + \xi^2} \text{ for } \sqrt{\nu^2 + \xi^2} = \frac{\pi}{2} + \pi m.$$

This means that the ratio of the neighboring peak heights is

$$\frac{\eta_{m+1}}{\eta_m} = \frac{(2m+1)^2}{(2m+3)^2}.$$

The central peak is obtained for $\xi = 0$ and $\nu = \pi/2 + \pi m$ ($m = 0, \pm 1, \pm 2, \ldots$). For our grating made of PDLC, we have $\nu_p = 7.43$ and $\nu_s = -1.44$ for $\theta = \theta_B$. It is for this reason that the value of *m* corresponding to the central and next peaks begins from zero for an *s*-polarized test beam and from two for a *p*-polarized test beam. As a result, $\eta_1/\eta_0 = 1/9$ for an *s*-polarized test beam and 25/49 for a *p*-polarized test beam.

One more (the third) mechanism of angular asymmetry in diffraction efficiency also exists. This mechanism is associated with absorption. The larger the angle of incidence on the grating, the longer the optical path of the beam and, hence, the stronger its absorption. This leads to a positive asymmetry. The allowance for absorption leads to the following correction: expression (14) for the diffraction efficiency must be multiplied by $\exp(-\alpha d/\cos\theta)$, where α is the absorption coefficient. However, this mechanism is beyond the scope of this paper since the absorption in our case is quite small ($\alpha = 35$ cm^{-1} and $\alpha d = 0.0875$) and the asymmetry in the diffraction efficiency due to absorption is insignificant.

Thus, we have established the existence of an anisotropic mechanism of asymmetry, which competes with the asymmetry of the Bragg detuning. Naturally, the contribution of the anisotropic mechanism strongly depends on the polarization of the incident beam, which makes it possible to control the value of asymmetry. Since the asymmetry has a positive value of approximately 6% in the case of *p* polarization (azimuth angle $\beta = 0$) and a negative value of about –4.3% in the case of *s* polarization (azimuth angle $\beta = 90°$), the asymmetry must vanish at a certain polarization. To analyze the dependence of the asymmetry on the azimuth angle of the incident polarization, we denote the intensities of diffracted *p*- and *s*-polarized waves by I_{dp} and I_{ds}, the intensities of incident *p*- and *s*-polarized waves by I_{ip} and I_{is}, and the angles between the normal to the surface of the PDLC (*y* axis) and the Poynting vector of the incident and diffracted waves with *p* and *s* polarizations by φ_{ip} and φ_{is}, φ_{dp}, and φ_{ds}, respectively. In accordance with the definition of diffraction efficiency, for *p* and *s* waves, we can write

$$\eta_p = \frac{I_{dp}(d) n_{dp} g_{dp} \cos\varphi_{dp}}{I_{ip}(0) n_{ip} g_{ip} \cos\varphi_{ip}},$$
$$\eta_s = \frac{I_{ds}(d) n_{ds} g_{ds} \cos\varphi_{ds}}{I_{is}(0) n_{is} g_{is} \cos\varphi_{is}}. \quad (22)$$

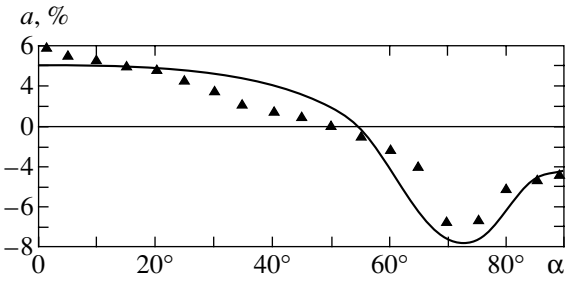


Fig. 4. Theoretical (solid curve) and experimental (triangles) dependences of the asymmetry in the angular selectivity of the diffraction efficiency on the azimuth angle of polarization.

Since

$$I_{ip} = E_{ip}E_{ip}^* = E_iE_i^*\cos^2\beta = I_i\cos^2\beta,$$

$$I_{dp} = E_{dp}E_{dp}^* = E_dE_d^*\cos^2\beta = I_d\cos^2\beta,$$
$$I_{is} = E_{is}E_{is}^* = E_iE_i^*\sin^2\beta = I_i\sin^2\beta, \quad (23)$$

$$I_{ds} = E_{ds}E_{ds}^* = E_dE_d^*\sin^2\beta = I_d\sin^2\beta,$$

where β is the azimuth angle of polarization of the incident beam and I_i and I_d are the total intensities of the incident and diffracted beams, we obtain the following expression for the intensity of the diffracted wave:

$$I_d = I_{ds} + I_{dp} = I_i(\eta_s f_s\sin^2\beta + \eta_p f_p\cos^2\beta), \quad (24)$$

where

$$f_p = \frac{n_{ip}g_{ip}\cos\varphi_{ip}}{n_{dp}g_{dp}\cos\varphi_{dp}}, \quad f_s = \frac{n_{is}g_{is}\cos\varphi_{is}}{n_{ds}g_{ds}\cos\varphi_{ds}}.$$

In the general case, the expression for the diffraction efficiency has the form

$$\eta = \frac{I_d(d)\cos\theta_d'}{I_i(0)\cos\theta_i'} = \frac{1}{f}\frac{I_d}{I_i}, \quad (25)$$

where $f = \cos\theta_i'/\cos\theta_d'$, θ_i' and θ_d' are the angles of incidence and diffraction outside the sample, which can be expressed in terms of the angles of incidence and diffraction θ_i and θ_d inside the sample with the help of Snell's formula. Substituting relation (24) into (25), we obtain the following expression for the dependence of the diffraction efficiency on the azimuth angle of polarization of the incident beam:

$$\eta = \frac{1}{f}(\eta_s f_s\sin^2\beta + \eta_p f_p\cos^2\beta). \quad (26)$$

This formula is used for plotting a series of curves describing the angular dependence of diffraction efficiency for the azimuth angle of polarization varying from 0 to 90° with a step of 5°; from these curves the dependence of the asymmetry on the angle of polarization was calculated using formula (21). The result is given by the solid curve in Fig. 4. It can be seen from the figure that the asymmetry decreases from the value $a = 6\%$, which corresponds to *p* polarization, to zero for a polarization angle of approximately 55°, attains its minimum value, and then corresponds to the case of an *s*-polarized test wave.

4. EXPERIMENTAL SETUP AND RESULTS

In our experiments, we used a photopolymerizable solution of a monomer and an LC sensitive in the near IR spectral region [20]. The material consisted of the following components: monomer DPEPA (d-pentaerythrol-penta-acrylate), monomer 2EEEA (2-ethoxy-ethoxy-ethyl-acrylate), liquid crystal E7, and a photosensitive system. The latter consisted of three elements: cyanine dye IR-140 with an absorption peak near 818 nm (with a linewidth of 100 nm), electron donor CBr_4, and initiator EDAAB (ethyl-dimethyl-aminobenzoate). This photopolymerizable solution was placed between two parallel glass plates (25 mm in width and length and 1 mm in thickness) with a 27.4-μm gasket.

It should be noted that the main five-duty monomer DPEPA produces a 3D network, while the single-duty monomer 2EEEA with a low viscosity can easily diffuse from dark to strongly illuminated regions. Illumination of the composite system of the photoinitiator by infrared light leads to irreversible photolysis and dye bleaching. Thus, the final absorption of the material is much lower even for samples with an initial optical density higher than two (at a wavelength of 823 nm). As a result, we can obtain a very effective (with an efficiency of more than 95%) reproducible holographic grating with a high resolution. An unsalted transmission grating was recorded using a standard interference setup. Two vertically polarized beams (*s* polarization) from a titanium:sapphire cw laser ($\lambda = 823$ nm) interfere at an angle of 48.6°; as a result, the most effective period on the order of 1 μm is formed [24].

The block diagram of the measuring setup is shown in Fig. 5. A test beam from a He–Ne laser ($\lambda = 628$ nm), which is polarized at 45° relative to the plane of incidence, passes through a quarter-wave plate and becomes circularly polarized. Then the beam passes through polarizer P fixed to a step motor SM2 and is program-controlled by computer PC through a CAMAC interface. The minimal angle through which the motor can rotate the polarizer is about 0.77°. Using this system, the required polarization of the incident beam can be programmed. Then light hits the holographic grating fixed to a step motor SM1. The minimal angle through which the motor can rotate the holographic grating is

approximately equal to 0.25°. From the two waves (transmitted and diffracted) emerging from the grating, the diffracted wave propagates along the straight line, while the transmitted wave reflected by the mirror passes through optomechanical gates G1 and G2 with open and closed states controlled by the computer as well. These two beams hit photodetector PM. The gates shutting down the diffracted wave with a period of 140 ms, open the transmitted wave and the photodetector measures its intensity; then the transmitted wave is shut down and the diffracted wave is opened. When step motor SM1 rotates the grating, the transmitted wave remains stationary, while the diffracted wave is displaced. For this purpose, lens L is installed between gate G1 and the grating, which maps an illuminated point of the grating on the photodetector; consequently, any beam (incident at any angle) hitting the lens necessarily falls on the photodetector. The setup measures first the angular dependence of diffraction efficiency for *p* polarization and then rotates polarizer P with step motor SM2 through 5° and measures again the angular dependence of diffraction efficiency until the *s* polarization is attained. We obtain a series of curves describing the angular dependence of diffraction efficiency as a function of the polarization. Each intensity measurement is performed ten times and the results are then averaged. The statistical error of the experiment amounts approximately to 0.01%.

Figures 2 and 3 show experimental results for the angular dependence of diffraction efficiency for *p*- and *s*-polarized incident waves (dashed curves). The experimentally obtained the Bragg angle is $\theta_B = 18.9° = 0.329$ rad. Calculating the asymmetry on the basis of formula (21) and taking into account experimental data, we obtain $a_p \approx 8\%$ for a *p*-polarized incident wave (see Fig. 2) and $a_s \approx -4.6\%$ for the *s*-polarized wave (see Fig. 3). These results are in good agreement with theoretical calculations.

Figure 4 shows the experimental curve describing the dependence of the asymmetry in the angular dependence of the diffraction efficiency on the polarization angle (triangles). As expected, the asymmetry vanishes at a certain angle of polarization. The experimental value of this angle is 50°. By varying the polarization, one can in fact control the value of asymmetry. The behavior of the experimental curve is in good agreement with the theoretical results.

It should be noted that for our parameters (d, Λ, etc.), the value of the quantity $Q = K^2\lambda d/2\pi n_0$ describing a thick holographic grating is equal to 67. However, the grating is also determined by one more parameter $\rho = 2\lambda^2/\Lambda^2 n_0 n_1$. In our case, $\rho = 14.499$ for a *p*-polarized wave and $\rho = 46.398$ for an *s*-polarized wave. According to [25], parameter ρ must satisfy the condition $\rho > 20$ for thick holographic gratings. Thus, for our parameters, we are completely in the Bragg regime for *s*-polarized waves and are close to this regime for *p*-polarized waves. Our calculations for thick holograms coincide with experiment to a high degree of accuracy.

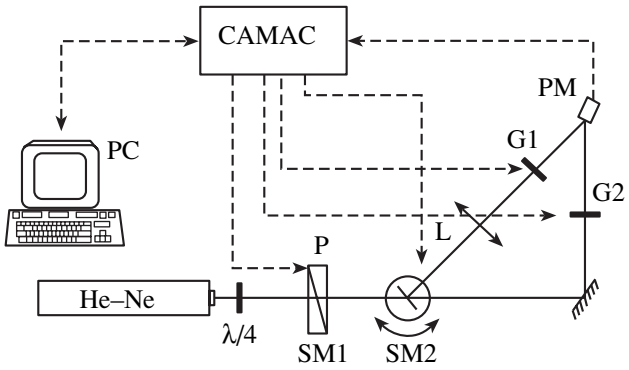


Fig. 5. Block diagram of the setup for measurement of angular and polarization dependences of diffraction gratings.

5. CONCLUSIONS

Thus, we have studied theoretically and experimentally a diffraction grating based on PDLCs, which is sensitive to the near IR spectral region. Analysis of light scattering reveals that the orientation of LC molecules in a polymer matrix is parallel to the wave vector of the grating. On this basis, a model is constructed for such an anisotropic grating. Comparison of our experiments and theoretical calculations confirms the validity of the distribution of the LC director described above. The asymmetry in the angular selectivity of diffraction efficiency is observed and theoretically explained. The dependence of the asymmetry on the polarization of the incident beam is studied as well as the possibility of controlling the asymmetry in the azimuth angle of polarization of incident radiation.

ACKNOWLEDGMENTS

We are grateful to S. Harbour, who kindly provided the experimental setup for recording our holograms.

This study was partly supported by the CRDF (grant no. AR2-2302-YE-02), and a collaborative grant of the Quebec government and NSERC (Nat. Science, Engin. & Reserch, Council of Canada).

REFERENCES

1. P. Yeh, *Optical Waves in Layered Media* (Wiley, New York, 1988), p. 150.
2. N. Noiret, C. Meyer, D. J. Lougnot, *et al.*, Pure Appl. Opt. **3**, 55 (1994).
3. G. Zhao and P. Mouroulis, J. Mod. Opt. **41**, 1929 (1994).
4. J. T. Sheridan and J. R. Lawrence, J. Opt. Soc. Am. A **17**, 1108 (2000).
5. H. Lee, X. Gu, and D. Psaltis, J. Appl. Phys. **65**, 2191 (1989).

6. D. A. Larson, T. D. Black, M. Green, *et al.*, J. Opt. Soc. Am. A **7**, 1745 (1990).
7. J. Qi, M. DeSarkar, G. T. Warren, and G. P. Crawford, J. Appl. Phys. **91**, 4795 (2002).
8. M. J. Escuti, J. Qi, and G. P. Crawford, Opt. Lett. **28**, 522 (2003).
9. P. Pilot and T. V. Galstian, Proc. SPIE **4087**, 1302 (2000).
10. F. Bouguin and T. V. Galstian, Proc. SPIE **4342**, 492 (2001).
11. T. J. Bunning, L. V. Natarajan, V. P. Tondiglia, *et al.*, Annu. Rev. Mater. Sci. **30**, 83 (2000).
12. R. L. Sutherland, L. V. Natarajan, V. P. Tondiglia, *et al.*, Chem. Mater. **5**, 1533 (1993).
13. R. Caputo, A. V. Sukhov, C. Umeton, *et al.*, Zh. Éksp. Teor. Fiz. **118**, 1374 (2000) [JETP **91**, 1190 (2000)].
14. J. J. Butler, M. A. Rodriguez, M. S. Malcuit, *et al.*, Opt. Commun. **155**, 23 (1998).
15. R. L. Sutherland, V. P. Tondiglia, L. V. Natarajan, *et al.*, Appl. Phys. Lett. **79**, 1420 (2001).
16. P. Pilot, Y. B. Boiko, and T. V. Galstian, Proc. SPIE **3638**, 26 (1999).
17. H. Kogelnik, Bell Syst. Tech. J. **48**, 2909 (1969).
18. G. Montemezzani and M. Zgonik, Phys. Rev. E **55**, 1035 (1997).
19. P. Pilot, Y. B. Boiko, and T. V. Galstian, Proc. SPIE **3635**, 143 (1999).
20. T. Galstian and A. Tork, U.S. Patent No. 6,398,981 (June 4, 2002).
21. P. de Gennes, *The Physics of Liquid Crystals* (Clarendon, Oxford, 1974; Mir, Moscow, 1977).
22. R. T. Pogue, L. V. Natarajian, V. P. Tondiglia, *et al.*, Proc. SPIE **3475**, 2 (1998).
23. J. J. Butler, M. S. Malcuit, and M. A. Rodriguez, J. Opt. Soc. Am. B **19**, 183 (2002).
24. P. Nagtegaele and T. V. Galstian, Synth. Met. **127**, 85 (2002).
25. M. G. Moharam and T. K. Gaylord, J. Opt. Soc. Am. **71**, 811 (1981).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 109–120.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 129–141.

SOLIDS
Structure

The Mechanism of Landau–Rumer Relaxation of Thermal and High-Frequency Phonons in Cubic Crystals of Germanium, Silicon, and Diamond

I. G. Kuleev* and I. I. Kuleev

Institute of Metal Physics, Ural Division, Russian Academy of Sciences, Yekaterinburg, 620219 Russia
**e-mail: kuleev@imp.uran.ru*

Received December 22, 2003

Abstract—Transverse phonon relaxation according to the Landau–Rumer mechanism is considered for an isotropic medium and crystals of germanium, silicon, and diamond possessing a cubic symmetry. The energy of elastic deformation caused by the anharmonicity of vibrations of the cubic crystal lattice is expressed via the second- and third-order moduli of elasticity. Using the known values of these elastic moduli, parameters determining the frequencies of the transverse phonon relaxation in the Landau–Rumer mechanism are evaluated for the germanium, silicon, and diamond crystals. It is shown that the dependence of the relaxation frequency on the wavevector of thermal and high-frequency phonons sharply differs from the classical Landau–Rumer relationship both in the isotropic medium and in the cubic crystals. It is established that the observed peculiarities in the relaxation frequency are related to the angular dependence of the probability of anharmonic scattering and the anisotropy of elastic properties of the germanium, silicon, and diamond crystals. A new method is proposed for the experimental determination of the relaxation frequency of high-frequency phonons as a function of the wavevector using the temperature dependence of the coefficient of absorption of high-frequency ultrasound. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In recent years, much attention has been devoted to investigation of the isotope effects manifested in the thermal conductivity of germanium, silicon, and diamond [1–8], The results of experimental investigations [2, 3] showed that the maximum values of thermal conductivity increase by one order of magnitude on the passage from germanium crystals with the natural isotope composition (^{Nat}Ge) to the samples highly (up to 99.99%) enriched with ^{70}Ge isotope. For silicon enriched (99.8588%) with ^{28}Si isotope, the maximum thermal conductivity exhibits a sixfold increase in comparison to natural crystals. Theoretical analysis of these results [9–11] showed that an important role in the lattice thermal conductivity of isotopically enriched, chemically pure germanium, silicon, and diamond crystals is played by the normal phonon–phonon scattering processes (*N* processes) characterized by conservation of the total momentum of colliding phonons. These processes, together with the boundary scattering, determine the maximum values of thermal conductivity in the isotopically enriched crystals [10, 11].

Calculations of the thermal conductivity within the framework of the relaxation method [1–13] usually employ expressions for the frequencies of phonon relaxation in the *N* processes obtained in the long-wavelength approximation,

$$z_{q\lambda} = \frac{\hbar\omega_{q\lambda}}{k_B T} \ll 1,$$

where $\omega_{q\lambda}$ is the frequency of the phonon with the wavevector $\mathbf{q}$ and the polarization λ. However, the main contribution to the lattice thermal conductivity is due to the thermal phonons with $z_{q\lambda} \approx 1$, and in the isotopically enriched crystals, due to the phonons with $z_{q\lambda} \approx 2$–3. In this approach, quantities determining the intensity of anharmonic scattering processes are the fitting parameters of the theory, determined from a comparison of theory and experiment [1–13]. Using the model of isotropic medium for evaluating the probability of anharmonic scattering processes also does not adequately describe cubic crystals of germanium, silicon, and diamond possessing a significant anisotropy of both second- and third-order elastic moduli.

In this paper, the transverse phonon relaxation is considered within the framework of the Landau–Rumer mechanism for the isotropic medium and crystals of germanium, silicon, and diamond possessing a cubic symmetry. In Section 2, we will obtain an expression for the energy of elastic deformation of a cubic crystal via the second- and third-order moduli of elasticity.

This energy determines the probabilities of various anharmonic three-photon scattering processes. In Section 3, the relaxation frequencies of the transverse thermal and high-frequency phonons in the isotropic medium and the cubic crystals are calculated in terms of the Landau–Rumer mechanism. In the crystals of germanium, silicon, and diamond, this is the principal mechanism of the transverse phonon relaxation. Since the elastic moduli for the germanium, silicon, and diamond crystals are known, this calculation provides an independent method for determining the relaxation frequencies of transverse phonons in the anharmonic scattering processes. The obtained results will be compared to the data obtained previously using the model of isotropic medium. Section 4 considers the long-wavelength limit, a transition to the model of the isotropic medium, and the dependence of the phonon relaxation frequency on the temperature and the wavevector for the germanium, silicon, and diamond crystals. It will be shown that the dependence of the transverse phonon relaxation frequency on the wavevector can be determined using the temperature dependence of the coefficient of absorption of high-frequency ultrasound.

2. THE ELASTIC ENERGY OF CUBIC CRYSTALS

The expression for the elastic energy of a cubic crystal, written to within the third-order terms in components of the strain tensor η_{ij}, is as follows [14, 15]:

$$\begin{aligned} W_k &= \frac{1}{2}c_{11}(\eta_{11}+\eta_{22}+\eta_{33}) \\ &+ c_{12}(\eta_{11}\eta_{22}+\eta_{22}\eta_{33}+\eta_{33}\eta_{11}) \\ &+ 2c_{44}(\eta_{12}^2+\eta_{23}^2+\eta_{13}^2) + c_{111}(\eta_{11}^3+\eta_{22}^3+\eta_{33}^3) \\ &+ 3c_{112}[\eta_{11}^2(\eta_{22}+\eta_{33}) \\ &+ \eta_{22}^2(\eta_{11}+\eta_{33}) + \eta_{33}^2(\eta_{11}+\eta_{22})] \\ &+ 6c_{123}\eta_{11}\eta_{22}\eta_{33} \\ &+ 12c_{144}(\eta_{11}\eta_{23}^2+\eta_{22}\eta_{13}^2+\eta_{33}\eta_{12}^2) \\ &+ 12c_{155}[(\eta_{11}+\eta_{22})\eta_{12}^2+(\eta_{22}+\eta_{33})\eta_{23}^2 \\ &+ (\eta_{11}+\eta_{33})\eta_{13}^2] + 48c_{456}\eta_{12}\eta_{23}\eta_{13}. \end{aligned} \tag{1}$$

Here, the third-order elastic moduli c_{ijk} are normalized according to Tucker and Rampton [16]. The passage to the Birch normalization c_{ijk}^{B} [14] is provided by the substitution

$$\begin{aligned} c_{111} &= c_{111}^{\mathrm{B}}, \quad 3c_{112} = c_{112}^{\mathrm{B}}, \quad 6c_{123} = c_{123}^{\mathrm{B}}, \\ 12c_{144} &= c_{144}^{\mathrm{B}}, \quad 12c_{155} = c_{155}^{\mathrm{B}}, \quad 24c_{456} = c_{456}^{\mathrm{B}}. \end{aligned}$$

Expression (1) can be converted to the following form, which is more convenient for subsequent calculations:

$$\begin{aligned} W_k &= \frac{1}{2}c_{11}\left(\sum_i \eta_{ii}\right)^2 + \frac{1}{2}(c_{11}-c_{12}-c_{44})\sum_i \eta_{ii}^2 \\ &+ c_{44}\sum_i \eta_{ik}^2 + \tilde{c}_{111}\sum_i \eta_{ii}^3 + 3\tilde{c}_{112}\sum_{i,k}\eta_{ii}^2\eta_{kk} \\ &+ c_{123}\left(\sum_i \eta_{ii}\right)^3 + 6c_{144}\sum_{i,k}\eta_{ii}^2\eta_{jk} \\ &+ 8c_{456}\sum_{i,k}\eta_{ik}\eta_{kj}\eta_{ji} + 12\tilde{c}_{155}\sum_{i,k}\eta_{ii}\eta_{ik}^2, \end{aligned} \tag{2}$$

where

$$\begin{aligned} \tilde{c}_{111} &= c_{111} - 3c_{112} + 2c_{123} \\ &+ 12c_{144} - 12c_{155} + 16c_{456}, \\ \tilde{c}_{112} &= c_{112} - c_{123} - 2c_{144}, \\ \tilde{c}_{155} &= c_{155} - c_{144} - 2c_{456}. \end{aligned} \tag{3}$$

The strain tensor components η_{ik} can be expressed via the distortion tensor components ξ_{ik} [17]:

$$\begin{aligned} \eta_{ik} &= \frac{1}{2}\left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i} + \sum_j \frac{\partial u_j}{\partial x_i}\frac{\partial u_j}{\partial x_k}\right) \\ &\equiv \frac{1}{2}\left(\xi_{ik} + \xi_{ki} + \sum_j \xi_{ji}\xi_{jk}\right). \end{aligned} \tag{4}$$

Substituting expression (4) into (2), we obtain

$$\begin{aligned} W_k &= \frac{1}{2}c_{11}\left(\sum_i \xi_{ii}\right)^2 + \frac{1}{2}(c_{11}-c_{12}-2c_{44}) \\ &\times \sum_i \xi_{ii}^2 + \frac{1}{4}c_{44}\sum_i (\xi_{ik}+\xi_{ki})^2 + \tilde{c}_{111}\sum_i \xi_{ii}^3 \\ &+ 3\tilde{c}_{112}\sum_{i,k}\xi_{ii}^2\xi_{kk} + c_{123}\left(\sum_i \xi_{ii}\right)^3 \\ &+ 3c_{144}\sum_{i,j,k}\xi_{ii}\xi_{jk}\xi_{kj} + \left(3c_{144}+\frac{1}{2}c_{12}\right) \\ &\times \sum_{i,j,k}\xi_{ii}\xi_{jk}^2 + 2c_{456}\sum_{i,j,k}\xi_{ik}\xi_{kj}\xi_{ji} + (c_{44}+6c_{456}) \\ &\times \sum_{i,j,k}\xi_{ik}\xi_{ji}\xi_{jk} + 3\tilde{c}_{155}\sum_{i,k}(\xi_{ii}\xi_{ik}^2 + 2\xi_{ii}\xi_{ik}\xi_{ki}) \\ &+ \left[3\tilde{c}_{155} + \frac{1}{2}(c_{11}-c_{12}-2c_{44})\right]\sum_{i,k}\xi_{ii}\xi_{ki}^2. \end{aligned} \tag{5}$$

Table 1. Elastic moduli of the first and second order of germanium, silicon, and diamond crystals

Crystal	Elastic modulus								
	c_{11}	c_{12}	c_{44}	c_{111}	c_{112}	c_{123}	c_{144}	c_{155}	c_{456}
Ge	1.289	0.483	0.671	–7.10	–3.89	–0.18	–0.23	–2.92	–0.53
Si	1.657	0.638	0.796	–8.25	–4.51	–0.64	0.12	–3.10	–0.64
C	10.76	1.25	5.758	–62.6	–22.6	1.12	–6.74	–28.6	–8.23
Ge	–	–	–	–1.18	–0.65	–0.03	–0.04	–0.49	–0.18
Si	–	–	–	–1.38	–0.75	–0.11	0.02	–0.52	–0.21
C	–	–	–	–10.43	–3.77	0.187	–1.123	–4.76	–2.74

Note: Thermodynamic quantities c_{ijk} in the first three lines are presented according to [17]; the values in the last three lines are given according to the normalization [16] adopted in this study (in units of 10^{12} dyn/cm^2).

The third-order terms with respect to ξ_{ik} in expression (4) correspond to the phonon–phonon interaction. It is expedient to pass from the elastic energy of a cubic crystal to the elastic deformation energy W_{is} of the isotropic medium. For this purpose, let us compare expression (5) to [16, Eq. (4.6)]. The condition $\Delta W = W_k - W_{is} = 0$ yields

$$\begin{aligned}
\Delta C &= c_{11} - c_{12} - 2c_{44} = 0,\\
\tilde{c}_{155} &= c_{155} - c_{144} - 2c_{456} = 0,\\
\tilde{c}_{112} &= c_{112} - c_{123} - 2c_{144} = 0,\\
\tilde{c}_{111} &= c_{111} - 3c_{112} + 2c_{123}\\
&+ 12c_{144} - 12c_{155} + 16c_{456} = 0.
\end{aligned} \tag{6}$$

Let us check how the conditions (6) are satisfied in the case of germanium, silicon, and diamond crystals using the results of measurements of their second- and third-order elastic moduli reported in [15, 16, 18–22] and summarized in Table 1. Our analysis showed (see Table 2) that neither second-order nor third-order elastic moduli of germanium, silicon, and diamond crystals satisfy relations (6): for these crystals, $\Delta C \sim c_{12}$, while the value of $\tilde{c}_{155}$ amounts to about one-fourth of c_{155} and is more than two times greater than c_{144}. The maximum discrepancy with the model of isotropic medium is observed for the elastic modulus $\tilde{c}_{111}$: this value not only significantly exceeds the other third-order elastic moduli of germanium, silicon, and diamond, but has a sign opposite to that of the c_{111} value. The third-order moduli $\tilde{c}_{111}$, $\tilde{c}_{112}$, and $\tilde{c}_{155}$ distinguish the cubic crystals from the isotropic medium. Therefore, the maximum deviation from this model must take place for the phonon relaxation frequencies containing contributions from the terms involving $\tilde{c}_{111}$ modulus.

Let us represent the displacement of a particle in the standard form [16],

$$u_i(x) = i\sum_{\mathbf{q}}\left(\frac{\hbar}{2\rho V\omega_{q\lambda}}\right)^{1/2} e_i^{\lambda} e^{i\mathbf{q}\cdot\mathbf{r}}(b_{q\lambda} - b^{\dagger}_{-q\lambda}), \tag{7}$$

where $b^{\dagger}_{-q\lambda}$ and $b_{q\lambda}$ are the operators of production and annihilation of phonons, ρ is the density, V is the normalization volume, and $\mathbf{e}$ is the polarization vector. Obviously, each ξ_{ij} value corresponds to the coefficient $e_i q_j$. Direct substitution of formula (7) into expression (5) yields the following relation for the energy density related to the anharmonicity of the crystal lattice vibrations:

$$\begin{aligned}
W = -i \sum_{\substack{\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_3\\ \lambda_1,\lambda_2,\lambda_3}} \left(\frac{\hbar}{2\rho V}\right)^{3/2} (\omega_{q_1\lambda_1}\omega_{q_2\lambda_2}\omega_{q_3\lambda_3})^{-1/2}\\
\times (b_{q_1\lambda_1} - b^{\dagger}_{-q_1\lambda_1})(b_{q_2\lambda_2} - b^{\dagger}_{-q_2\lambda_2})(b_{q_3\lambda_3} - b^{\dagger}_{-q_3\lambda_3})\\
\times \exp[i(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3)\cdot\mathbf{r}]V^{\lambda_1\lambda_2\lambda_3}_{\mathbf{q}_1\mathbf{q}_2\mathbf{q}_3},
\end{aligned} \tag{8}$$

where

$$\begin{aligned}
V^{\lambda_1\lambda_2\lambda_3}_{\mathbf{q}_1\mathbf{q}_2\mathbf{q}_3} &= 6\tilde{c}_{111}\sum_i e_{1i}q_{1i}e_{2i}q_{2i}e_{3i}q_{3i}\\
&+ 6c_{123}(\mathbf{e}_1\cdot\mathbf{q}_1)(\mathbf{e}_2\cdot\mathbf{q}_2)(\mathbf{e}_3\cdot\mathbf{q}_3)\\
&+ 6\tilde{c}_{112}\sum_i[(\mathbf{e}_1\cdot\mathbf{q}_1)e_{2i}q_{2i}e_{3i}q_{3i}\\
&+ (\mathbf{e}_2\cdot\mathbf{q}_2)e_{1i}q_{1i}e_{3i}q_{3i} + (\mathbf{e}_3\cdot\mathbf{q}_3)e_{1i}q_{1i}e_{2i}q_{2i}]\\
&+ 6c_{144}[(\mathbf{e}_1\cdot\mathbf{q}_1)(\mathbf{e}_2\cdot\mathbf{q}_3)(\mathbf{e}_3\cdot\mathbf{q}_2)\\
&+ (\mathbf{e}_2\cdot\mathbf{q}_2)(\mathbf{e}_1\cdot\mathbf{q}_3)(\mathbf{e}_3\cdot\mathbf{q}_1)
\end{aligned}$$

$$+ (\mathbf{e}_3 \cdot \mathbf{q}_3)(\mathbf{e}_1 \cdot \mathbf{q}_2)(\mathbf{e}_2 \cdot \mathbf{q}_1)]$$
$$+ (c_{12} + 6c_{144})[(\mathbf{e}_1 \cdot \mathbf{q}_1)(\mathbf{e}_2 \cdot \mathbf{e}_3)(\mathbf{q}_3 \cdot \mathbf{q}_2)$$
$$+ (\mathbf{e}_2 \cdot \mathbf{q}_2)(\mathbf{e}_1 \cdot \mathbf{e}_3)(\mathbf{q}_3 \cdot \mathbf{q}_1)$$
$$+ (\mathbf{e}_3 \cdot \mathbf{q}_3)(\mathbf{e}_1 \cdot \mathbf{e}_2)(\mathbf{q}_1 \cdot \mathbf{q}_2)]$$
$$+ 6c_{456}[(\mathbf{e}_1 \cdot \mathbf{q}_3)(\mathbf{e}_2 \cdot \mathbf{e}_1)(\mathbf{q}_3 \cdot \mathbf{q}_2)$$
$$+ (\mathbf{e}_1 \cdot \mathbf{q}_2)(\mathbf{e}_2 \cdot \mathbf{q}_3)(\mathbf{e}_3 \cdot \mathbf{q}_1)]$$
$$+ (c_{44} + 6c_{456})[(\mathbf{e}_1 \cdot \mathbf{q}_2)(\mathbf{q}_1 \cdot \mathbf{q}_3)(\mathbf{e}_2 \cdot \mathbf{e}_3)$$
$$+ (\mathbf{e}_2 \cdot \mathbf{q}_3)(\mathbf{q}_1 \cdot \mathbf{q}_2)(\mathbf{e}_1 \cdot \mathbf{e}_3)$$
$$+ (\mathbf{e}_3 \cdot \mathbf{q}_1)(\mathbf{q}_2 \cdot \mathbf{q}_3)(\mathbf{e}_1 \cdot \mathbf{e}_2)$$

$$+ (\mathbf{e}_1 \cdot \mathbf{q}_3)(\mathbf{q}_1 \cdot \mathbf{q}_2)(\mathbf{e}_2 \cdot \mathbf{e}_3)$$
$$+ (\mathbf{e}_2 \cdot \mathbf{q}_1)(\mathbf{q}_2 \cdot \mathbf{q}_3)(\mathbf{e}_1 \cdot \mathbf{e}_3)$$
$$+ (\mathbf{e}_3 \cdot \mathbf{q}_2)(\mathbf{q}_1 \cdot \mathbf{q}_3)(\mathbf{e}_1 \cdot \mathbf{e}_2)]$$

$$+ 6\tilde{c}_{155}\sum_i \{e_{1i}e_{2i}e_{3i}[q_{1i}(\mathbf{q}_2 \cdot \mathbf{q}_3)$$
$$+ q_{2i}(\mathbf{q}_1 \cdot \mathbf{q}_3) + q_{3i}(\mathbf{q}_1 \cdot \mathbf{q}_2)]$$
$$+ e_{1i}q_{1i}[e_{2i}q_{3i}(\mathbf{e}_3 \cdot \mathbf{q}_2) + e_{3i}q_{2i}(\mathbf{e}_2 \cdot \mathbf{q}_3)]$$
$$+ e_{2i}q_{2i}[e_{1i}q_{3i}(\mathbf{e}_3 \cdot \mathbf{q}_1) + e_{3i}q_{1i}(\mathbf{e}_1 \cdot \mathbf{q}_3)]$$
$$+ e_{3i}q_{3i}[e_{1i}q_{2i}(\mathbf{e}_2 \cdot \mathbf{q}_1) + e_{2i}q_{1i}(\mathbf{e}_1 \cdot \mathbf{q}_2)]\}$$

$$+ (6\tilde{c}_{155} + \Delta C)\sum_i q_{1i}q_{2i}q_{3i}[e_{1i}(\mathbf{e}_2 \cdot \mathbf{e}_3)$$
$$+ e_{2i}(\mathbf{e}_1 \cdot \mathbf{e}_3) + e_{3i}(\mathbf{e}_1 \cdot \mathbf{e}_2)].$$

Using relation (8), it is possible to study various mechanisms of the three-photon scattering processes in cubic crystals. Under conditions (6), we can obtain expressions for the energy density related to the anharmonicity of the crystal lattice vibrations in terms of the model of isotropic medium (see [16, Eq. (4.22)]. The new terms, involving coefficients $\tilde{c}_{111}$, $\tilde{c}_{112}$, and $\tilde{c}_{155}$, distinguish the cubic crystals from the isotropic medium. Integration over the normalization volume yields the matrix element $V^{\lambda_1\lambda_2\lambda_3}_{\mathbf{q}_1\mathbf{q}_2\mathbf{q}_3}$, the squared modulus of which determines the probability of phonon scattering in the anharmonic scattering processes. This element will be used for calculating the phonon relaxation frequencies in the N processes ($\delta_{\mathbf{q}1+\mathbf{q}2+\mathbf{q}3,0}$) in the cubic crystals of germanium, silicon, and diamond.

Table 2. Elastic moduli ΔC, $\tilde{c}_{111}$, $\tilde{c}_{112}$, and $\tilde{c}_{155}$ of germanium, silicon, and diamond crystals (in units of 10^{12} dyn/cm^2)

Elastic modulus	Crystal		
	Ge	Si	C
ΔC	−0.536	−0.574	–2.05
$\tilde{c}_{155}$	−0.095	−0.111	1.84
$\tilde{c}_{112}$	−0.542	−0.685	–1.71
$\tilde{c}_{111}$	3.26	3.72	1.05

3. LANDAU–RUMER MECHANISM FOR THERMAL PHONONS IN GERMANIUM, SILICON, AND DIAMOND

According to the Landau–Rumer mechanism [23] the relaxation of long-wavelength transverse phonons ($\hbar\omega_{qT} \ll k_B T$) consists in merging of transverse (T) and longitudinal (L) phonons with the formation of a new longitudinal phonon:

$$\omega_{q_1\mathrm{T}} + \omega_{q_2\mathrm{L}} = \omega_{q_3\mathrm{L}}.$$

An analysis of this relaxation occurring at a sufficiently low temperature yields an expression for the relaxation frequency [23]

$$\nu^{\mathrm{LR}}_{\mathrm{phN}} = B_{\mathrm{T}0} z_1 T^5, \quad z_1 = \frac{\hbar\omega_{q_1\mathrm{T}}}{k_B T}, \tag{9}$$

where $B_{\mathrm{T}0}$ is a coefficient depending on the second- and third-order elastic moduli, atomic masses, and lattice parameters. In the theory of lattice thermal conductivity, this coefficient is considered as the fitting parameter determined from a comparison of theory and experiment [1–13]. In [13, 16, 24], an analysis of the phonon relaxation according to the Landau–Rumer mechanism was restricted to the long-wavelength approximation ($q_1 \ll q_2, q_3$; $z_1 \ll 1$) and the model of the isotropic medium. The same approximations were used for the relaxation frequency of thermal phonons in the calculations of phonon thermal conductivity, although condition $z_1 \ll 1$ is not valid for thermal phonons. Estimates show that, in germanium and silicon crystals with natural isotope composition, the maximum contribution to the lattice thermal conductivity is due to the longitudinal phonons with $z_{\mathrm{L}} \approx 0.4$–$0.5$ and the transverse phonons with $z_{\mathrm{T}} \approx 2$, while for isotopically enriched crystals, the maximum contribution is due to the longitudinal phonons with $z_{\mathrm{L}} \approx 1$ and the transverse phonons with $z_{\mathrm{T}} \approx 4$. Obviously, the previous results [1–13] have to be refined so as to reject the long-wavelength approximation and the model of isotropic medium.

Below, we calculate the relaxation frequency $\nu^{\mathrm{TLL}}_{\mathrm{phN}}$ for the aforementioned cubic crystals. This will allow us to determine the coefficient $B_{\mathrm{T}0}$ using the known values of the second- and third-order elastic moduli and to refine the dependence of the relaxation frequency on the temperature and wavevector of thermal phonons.

According to [13], the expression for the phonon relaxation frequency can be written as

$$\nu_{\mathrm{phN}}(q_1, \lambda_1) = \frac{\pi\hbar^4}{(2\rho k_B T)^3}\frac{1}{V}$$

$$\times \sum_{\substack{\mathbf{q}_2, \mathbf{q}_3 \\ \lambda_2, \lambda_3}} \frac{\sinh(z_1/2)\delta_{\mathbf{q}_1+\mathbf{q}_2+\mathbf{q}_3, 0}}{z_1 z_2 z_3 \sinh(z_2/2)\sinh(z_3/2)} \left|V^{\lambda_1\lambda_2\lambda_3}_{\mathbf{q}_1\mathbf{q}_2\mathbf{q}_3}\right|^2 \tag{10}$$

$$\times [2\delta(\omega_{q_1\lambda_1} + \omega_{q_2\lambda_2} - \omega_{q_3\lambda_3})$$

$$+ \delta(\omega_{q_1\lambda_1} - \omega_{q_2\lambda_2} - \omega_{q_3\lambda_3})].$$

The transverse phonons can be involved only in the process of merging (reflected by the first term in square brackets), while the probability of decay (second term in square brackets) is zero because otherwise the law of energy conservation cannot be satisfied. The process of phonon merging obeys the following relations:

$$\mathbf{e}_1 \cdot \mathbf{q}_1 = 0, \quad \mathbf{e}_2 \cdot \mathbf{q}_2 = q_2, \quad \mathbf{e}_3 \cdot \mathbf{q}_3 = q_3,$$
$$\mathbf{e}_1 \cdot \mathbf{q}_3 = \mathbf{e}_1 \cdot \mathbf{q}_2, \quad \mathbf{q}_3 = \mathbf{q}_2 + \mathbf{q}_1. \tag{11}$$

Using expression (8) and taking into account relations (11), one can readily obtain a formula for the matrix element according to the Landau–Rumer mechanism:

$$V^{\mathrm{TLL}}_{\mathbf{q}_1\mathbf{q}_2\mathbf{q}_3} = A\frac{(\mathbf{e}_1 \cdot \mathbf{q}_2)}{q_2 q_3}[q_2^2 + (\mathbf{q}_1 \cdot \mathbf{q}_2)][q_1^2 + (\mathbf{q}_1 \cdot \mathbf{q}_2)]$$

$$+ \frac{1}{q_2 q_3}\sum_i \{12\tilde{c}_{155} e_{1i} q_{2i} q_{3i}[q_{1i}(\mathbf{q}_2 \cdot \mathbf{q}_3)$$

$$+ q_{2i}(\mathbf{q}_1 \cdot \mathbf{q}_3) + q_{3i}(\mathbf{q}_1 \cdot \mathbf{q}_2)] \tag{12}$$

$$+ (12\tilde{c}_{155} + \Delta C)[(\mathbf{e}_1 \cdot \mathbf{q}_2) q_{1i} q_{2i} q_{3i}(q_{1i} + q_{2i})$$

$$+ e_{1i} q_{1i} q_{2i} q_{3i}(\mathbf{q}_2 \cdot \mathbf{q}_3)]\}.$$

The first term in this expression corresponds to the isotropic scattering of phonons, and the second, to their anisotropic scattering (this term vanishes on the passage to the model of isotropic medium). For a cubic crystal,

$$A = A_{\mathrm{cub}} = c_{12} + 3c_{44} + 12c_{144} + 24c_{456}. \tag{13a}$$

Using relations (6) and passing to the isotropic case, we obtain the result of Tucker and Rampton [16]:

$$A_{\mathrm{is}} = A_1 + 3A_2 + A_3 + A_4$$
$$= \frac{3}{2}c_{11} - \frac{1}{2}c_{12} + 12c_{155}, \tag{13b}$$

where A_1–A_4 are the coefficients determined in [16]. In the general case, the relaxation frequency $\nu^{\mathrm{TLL}}_{\mathrm{phN}}$, as well as the intrinsic frequencies, depends on the direction of propagation of the transverse phonon relative to the crystallographic axes.

In what follows, we simplify the problem and restrict the consideration to one of the symmetric directions ([100], [001], [111], etc.). The z axis and the phonon wavevector $\mathbf{q}_1$ are also oriented along this direction, so that vector $\mathbf{e}_1$ occurs in the xy plane (for certainty, it coincides with the x axis). In this case, the condition $\mathbf{e}_1 \cdot \mathbf{q}_1 = 0$ can be supplemented by the conditions $e_{1i}q_{1i} = 0$, and the matrix element (12) can be written in terms of the angular variables θ_2 and φ_2 of the $\mathbf{q}_2$ vector:

$$V^{\mathrm{TLL}}_{\mathbf{q}_1\mathbf{q}_2\mathbf{q}_3} = 2\frac{q_1 q_2^3}{q_3}\sin\theta_2\cos\varphi_2\left(\frac{q_1}{2q_2} + \cos\theta_2\right)$$

$$\times\left\{A_{\mathrm{cub}}\left(1 + \frac{q_1}{q_2}\cos\theta_2\right)\right.$$

$$\left. + 12\tilde{c}_{155}\sin^2\theta_2\cos^2\varphi_2 + (12\tilde{c}_{155} + \Delta C)\right. \tag{14}$$

$$\left.\times\cos\theta_2\left(\frac{q_1}{q_2} + \cos\theta_2\right)\right\}.$$

Here,

$$\cos\theta_2 = s^* - \frac{q_1}{2q_2}(1 - s^{*2}), \quad s^* = \frac{s_{\mathrm{T}}}{s_{\mathrm{L}}}, \tag{14'}$$

and s_{T} and s_{L} are the sound velocities for the transverse and longitudinal phonons, respectively. In expression (10), the integrals with respect to the angular variables θ_2 and φ_2 can be readily calculated. Indeed, the integral over θ_2 is calculated using δ function and taking into account the law of energy conservation for the three-phonon scattering (14'), while the integral over φ_2 has nonzero contributions only due to the terms containing even powers of $\cos\varphi_2$. Eventually, we obtain the following expression the relaxation frequency according to the Landau–Rumer mechanism:

$$\nu^{\mathrm{TLL}}_{\mathrm{phN}} = T^5 z_1 B^{\mathrm{T}},$$
$$B^{\mathrm{T}} = \frac{k_B^5 s^{*2}(1 - s^{*2})}{16\pi\hbar^4\rho^3 s_{\mathrm{T}} s_{\mathrm{L}}^8}\frac{\sinh(z_1 ù\, h2)}{z_1/2} J_z(z_1, T), \tag{15}$$

Table 3. Parameters determining the transverse phonon relaxation frequencies in germanium, silicon, and diamond crystals

	B_{is}, 10^{24} $(\text{dyn/cm}^2)^2$	B, 10^{24} $(\text{dyn/cm}^2)^2$ [100]	B, 10^{24} $(\text{dyn/cm}^2)^2$ [111]	B_{T0}, $\text{s}^{-1}\,\text{K}^{-5}$ [100]	B_{T0}, $\text{s}^{-1}\,\text{K}^{-5}$ [111]
Ge	13.46	12.20	10.99	1.4	0.46
Si	11.55	10.79	9.74	0.12	0.045
C	1628	1787	1826	7.66×10^{-3}	5.01×10^{-3}

where

$$J_z(z_1, T) = \int_{z_{\min}}^{z_{DL}} dz F(z, z_1)\varphi_1(z, z_1) \times \left(b_2^2 + \frac{3}{2}b_2 b_3 + \frac{5}{8}b_3^2\right), \quad (15')$$

$$F(z, z_1) = \frac{z^4\left(1 + \frac{z_1}{2z}\right)^2}{\left(1 + \frac{z_1}{z}\right)^2 \sinh\left(\frac{z}{2}\right)\sinh\left(\frac{z_1 + z}{2}\right)},$$

$$\varphi_1(z, z_1) = \left[1 + \frac{z_1}{2z}\left(1 + \frac{1}{s^*}\right)\right]\left[1 + \frac{z_1}{2z}\left(1 - \frac{1}{s^*}\right)\right],$$

$$b_2 = A\varphi_2(z, z_1) + (\tilde{\tilde{c}}_{155} + \Delta C)s^{*2}\varphi_3(z, z_1),$$

$$b_3 = \tilde{\tilde{c}}_{155}(1 - s^{*2})\varphi_1(z, z_1),$$

$$\varphi_2(z, z_1) = 1 + \frac{z_1}{z} - \frac{1}{2}\left(\frac{z_1}{z}\right)^2\left(\frac{1}{s^{*2}} - 1\right),$$

$$\varphi_3(z, z_1) = \left[1 - \frac{1}{2}\frac{z_1}{z}\left(\frac{1}{s^{*2}} - 1\right)\right]\left[1 + \frac{z_1}{2z}\left(1 + \frac{1}{s^{*2}}\right)\right],$$

$$\tilde{\tilde{c}}_{155} = 12\tilde{c}_{155}, \quad z_1 = \frac{\hbar\omega_q T}{k_B T},$$

$$z_{\min} = \frac{1}{2}\left(\frac{1}{s^*} - 1\right)z_1, \quad z_{DL} = \frac{\hbar\omega_{DL}}{k_B T},$$

and ω_{DL} is the Debye frequency for the longitudinal phonons.

In the isotropic case, $\tilde{c}_{155} = \Delta C = 0$ and, hence, $b_3 = 0$. The expression for b_2 is as follows:

$$b_2 = A_{is}\varphi_2(z, z_1). \quad (16)$$

In the long-wavelength limit ($z_1/z \ll 1$) for cubic crystals, expressions (15) and (15') yield formula (9) with the coefficient B_{T0} given by the expression

$$B_{T0} = \frac{Bk_B^5(1 - s^{*2})}{16\pi\hbar^4\rho^3 s_T s_L^8}\int_0^{z_{DL}} \frac{z^4 dz}{\sinh^2(z/2)}. \quad (17)$$

For the temperatures significantly below the Debye temperature, this formula yields

$$B_{T0} = \frac{B\pi^3 k_B^5 s^{*2}(1 - s^{*2})}{15\hbar^4\rho^3 s_T s_L^8}, \quad (17')$$

where

$$B = A^2 + A\left[2s^{*2}(\tilde{\tilde{c}}_{155} + \Delta C) + \frac{3}{2}\tilde{\tilde{c}}_{155}(1 - s^{*2})\right] + \frac{3}{2}(\tilde{\tilde{c}}_{155} + \Delta C)\tilde{\tilde{c}}_{155}s^{*2}(1 - s^{*2}) + (\tilde{\tilde{c}}_{155} + \Delta C)^2 s^{*4} + \frac{5}{8}\tilde{\tilde{c}}_{155}^2(1 - s^{*2})^2.$$

Thus, the transverse phonon relaxation frequency according to the Landau–Rumer mechanism is retained in the long-wavelength limit: $\nu_{phN}^{TLL} \approx z_1 T^5$. However, in addition to the isotropic scattering (the first term in expression (17')) this mechanism also contains contributions due to the interference of the isotropic and anisotropic scattering (the second term in expression (17')) and the anisotropic scattering as such (the last three terms in expression (17')). Estimates show that contributions due to the isotropic scattering in the [100] direction for germanium and silicon amount to approximately 40 and 32%, respectively; the corresponding contributions of the interference term in these crystals are 46 and 49%, respectively; and the anisotropic scattering accounts for 14 and 19%, respectively.

The values of parameters B_{is}, B, and B_{T0} for the crystals of germanium, silicon, and diamond are presented in Table 3. For the [100] direction at $T = 10$ K and $z = 1$, the ν_{phN}^{TLL} values are $1.4 \times 10^5\ \text{s}^{-1}$ for germanium, $1.2 \times 10^4\ \text{s}^{-1}$ silicon, and $7.7 \times 10^3\ \text{s}^{-1}$ for diamond. The values of B_{is} for all these materials are approximately 10% greater than the coefficient B. The values of the fitting parameters B_{T0} used for the analysis of thermal conductivity in germanium and silicon crystals with various isotope compositions in [3, 9–11] (see [11, Table 2]) are smaller than the values according to the theory of elasticity, on the average by a factor of 50 for germanium and 46 for silicon (see Table 2). Since the relaxation frequency ν_{phN}^{TLL} for thermal phonons ($z_1 > 1$) is significantly lower than that according to the Landau–Rumer theory in the entire temperature interval $1 < T < 100$ K, the effective B_{T0} value for thermal phonons has to be also smaller than that according to the long-wavelength approximation (see the next section). However, even with allowance for these factors, the effective relax-

ation frequencies ν_{phN}^{TLL} obtained using the fitting parameters B_{T0} for germanium and silicon crystals with different isotope compositions [3, 9–11] are lower by one order of magnitude than the values obtained in our calculations within the framework of the theory of elasticity. It can be noted that the anisotropy of the relaxation frequency ν_{phN}^{TLL} in the germanium and silicon crystals is mostly due to the anisotropy of the second-order elastic modulus.

In the isotropic case, formula (17) yields the result of Tucker and Rampton [16]:

$$B_{is} = A_{is}^2 = \left(\frac{3}{2}c_{11} - \frac{1}{2}c_{12} + 12c_{155}\right)^2. \quad (18)$$

This expression is at variance with the result obtained for the isotropic medium by Maris (see [22, Eq. (210)]:

$$B_{T0} = \frac{\pi^3 k_B^5}{120\hbar^4 s_L^5}\frac{c_{11}-c_{44}}{c_{11}} \times\left(\frac{c_{111}-c_{112}+3c_{11}-c_{12}}{c_{11}}\right)^2. \quad (19)$$

A difference by a factor of two in numerical coefficients of formulas (19) and (17') is due to the fact that two relaxation processes of transverse phonons are taken into consideration in formula (10). However, the main difference between expressions (15)–(17) and formula (19) consists in the elastic modulus c_{111} entering into the latter formula. The point is that this elastic modulus accounts for the relaxation of longitudinal phonons and is not related to the transverse phonon relaxation according to the Landau–Rumer mechanism. Apparently, the introduction of the generalized Grüneisen parameter [24] and its averaging is adequate neither for the cubic crystals nor in the isotropic case. We have reproduced the main transformations involved in the calculation of the transverse phonon relaxation frequency in much detail, because even recently published papers (see, e.g. [25]) contain errors. For example, calculations of the thermal conductivity of Ge, Si, GaN, and C crystals in [25] employ relation (9) for the relaxation frequency ν_{phN}^{TLL}, in which the coefficient B_{T0} is given by the expression

$$B_{T0} \approx \frac{k_B^5 \gamma_T^2 V}{M\hbar^3 \rho s_T^5}, \quad (20)$$

where γ_T is the Grüneisen coefficient for the transverse phonons. A comparison of expressions (15)–(18) and (19) shows that formula (20) incorrectly describes polarization in the sound velocity entering into the denominator of B_{T0}. Use of this approximation for the relaxation frequency of thermal phonons in evaluation of the isotope effects in the thermal conductivity of Ge, Si, GaN, and C crystals [25] can hardly give reasonable results.

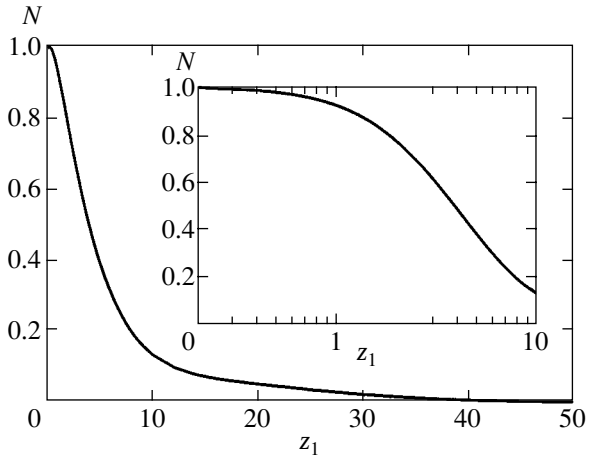


Fig. 1. Plots of the $N(z_1, T)$ versus reduced wavevector z_1 for $T = 1$ K and the sound velocities $s_L = 4.92 \times 10^5$ cm/s and $s_T = 3.55 \times 10^5$ cm/s.

4. RESULTS OF NUMERICAL ANALYSIS

Let us analyze the dependence of the relaxation frequency $\nu_{phN}^{TLL}(z_1, T)$ on the reduced wavevector z_1 and the temperature T according to formulas (15) and (15'). First, we will estimate the deviation from the Landau–Rumer formula (9) for the isotropic medium. For this purpose, we will construct a plot of the ratio of frequencies $\nu_{phN}^{TLL}(z_1, T)$ and $\nu_{phN}^{LR}(z_1, T)$ calculated using expressions (15)–(18):

$$N(z_1, T) = \frac{\nu_{phN}^{TLL}(z_1, T)}{\nu_{phN}^{LR}(z_1, T)} = \frac{15}{4\pi^4}\frac{\sinh(z_1/2)}{z_1/2}J_{is}(z_1, T), \quad (21)$$

where $J_{is}(z_1, T) = J_z(z_1, T)$ is calculated for $b_3 = 0$ and $b_2 = A_{is}\varphi_2(z, z_1)$.

As can be seen from Fig. 1, this ratio for small z_1 tends to unity and, accordingly, $\nu_{phN}^{TLL}(z_1, T) \longrightarrow \nu_{phN}^{LR}(z_1, T)$. However, as z_1 increases, the deviation of $N(z_1, T)$ from unity at a fixed temperature grows and, for $z_1 > 1$, we have $\nu_{phN}^{TLL}(z_1, T) \ll \nu_{phN}^{LR}(z_1, T)$ in the entire temperature interval $1 < T < 100$ K. As for the temperature dependence of $\nu_{phN}^{TLL}(z_1, T)$ in the region of $z_1 \ll 1$, we have $\nu_{phN}^{TLL}(z_1, T) \sim \nu_{phN}^{LR}(z_1, T) \sim z_1 T^5$; how-

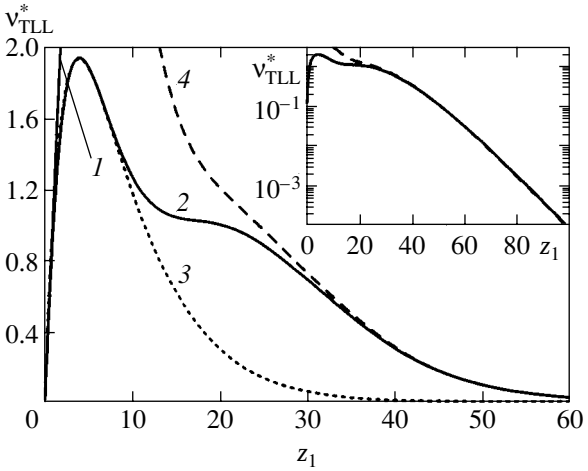


Fig. 2. Plots of ν^*_{TLL} versus reduced wavevector z_1 at T = 1 K calculated using various models for the isotropic case: (*1*) linear Landau–Rumer approximation; (*2*) this study; (*3*) approximation (26); (*4*) approximation (25); all calculations performed for $s_L = 4.92 \times 10^5$ cm/s and $s_T = 3.55 \times 10^5$ cm/s.

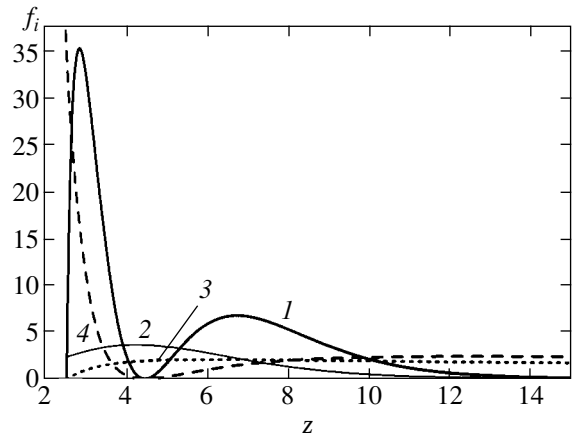


Fig. 3. Plots of the integrand functions versus reduced wavevector z: (*1*) $F^*(z, z_1)\varphi_1(z, z_1)\varphi_2^2(z, z_1)$; (*2*) $F^*(z, z_1)$; (*3*) $\varphi_1(z, z_1)$; (*4*) $\varphi_2^2(z, z_1)$. All calculations were performed for $z_1 = 13$ and the sound velocities $s_L = 4.92 \times 10^5$ cm/s and $s_T = 3.55 \times 10^5$ cm/s.

ever, at $T > 50$ K and $z_1 > 1$, the temperature dependence becomes less strong and the exponent is smaller than five.

Deviation of the relaxation frequency from the linear law (9) is most clearly manifested for the function $\nu^{TLL}_{phN}(z_1, T)$. Figure 2 shows plots of the quantities

$$\nu^*_{TLL}(z_1, T) = \frac{\nu^{TLL}_{phN}(z_1, T)}{B_{T0}T^5}, \quad \nu^{*LR}_{TLL}(z_1) = z_1 \qquad (22)$$

as functions of the reduced wavevector z_1. As can be seen, a linear approximation is possible only for $z_1 \le 1$ ($\nu^{*LR}_{TLL}(z_1)$, the Landau–Rumer approximation (9)). For $z_1 > 4$, the value of $\nu^*_{TLL}(z_1, T)$ first sharply decreases, then exhibits a plateau for $10 \le z_1 \le 20$, and finally shows a monotonic decrease with increasing z_1. Such a striking deviation from the classical behavior according to the Landau–Rumer approximation (9) was unexpected. Previous analysis [26, 27] showed that, for $z_1 \gg 1$ ($\hbar\omega_{qT} \gg k_BT$), the relaxation frequency $\nu^{TLL}_{phN}(z_1, T)$ exponentially decreases with increasing z_1 as

$$\nu^{TLL}_{phN}(z_1, T) \propto z_1^n \exp\left[-\frac{1}{2}\left(\frac{1}{s^*} - 1\right)z_1\right], \qquad (23)$$

where $n = 3$ according to [16, 24] and $n = 5$ according to [27]. Therefore, we expected only a single maximum in $\nu^*_{phN}(z_1, T)$ followed by a monotonic decay for $z_1 \gg 1$. The appearance of a new peculiarity, the plateau at $10 < z_1 < 20$, in the plot of the relaxation frequency versus wavevector requires additional thorough analysis.

Before proceeding with this analysis, it is necessary to make two important remarks. First, the observed peculiarities are related to the dependence of the matrix element (14) (transition probability) on the angle of phonon scattering. In contrast to [26, 27], this dependence was strictly taken into consideration. Second, the characteristic shape of the $\nu^*_{TLL}(z_1, T)$ curve can be realized only for sufficiently low temperatures. Indeed, the limiting frequency of the transverse phonons is restricted to the Debye frequency, which amounts to approximately 118 K for germanium and 210 K for silicon. Therefore, although the characteristic shape of the $\nu^*_{TLL}(z_1, T)$ curve is retained at high temperatures, the plateau in this curve at $T \sim 50$–100 K falls within a nonphysical region of the wavevectors exceeding the Debye wavevector: $z > z_{DT} = \hbar\omega_{DT}/k_BT$ (where ω_{DT} is the Debye transverse phonon frequency). However, at low temperatures (about 1–4 K for germanium and 1–10 K for silicon), the aforementioned peculiarities fall within the physical region of the phonon wavevectors ($z \le z_{DT}$).

Let us consider the factors responsible for the appearance of a plateau on the plot of relaxation frequency as a function of the phonon wavevector. The integrand in Eq. (15') for the isotropic case comprises a product of three functions, $F^*(z, z_1)\varphi_1(z, z_1)\varphi_2^2(z, z_1)$, where $F^*(z, z_1) = F(z, z_1)\sinh(z_1/2)$ and the functions $\varphi_1(z, z_1)$ and $\varphi_2(z, z_1)$ are related to the angular dependence of the scattering probability (see Eq. (14)). The dependences of these functions on z at a fixed value of z_1 are shown in Fig. 3. The function $F^*(z, z_1)$ reaches a

maximum at $z \approx 4$ and then exhibits exponential decay. The functions $\varphi_1(z, z_1)$ and $\varphi_2(z, z_1)$ tend to unity for $z_1 \ll z$ and are equal to zero when

$$z = z_{10} = z_{\min} = \alpha_{10} z_1, \quad \alpha_{10} = \frac{1}{2}\left(\frac{1}{s^*} - 1\right),$$
$$z = z_{20} = \alpha_{20} z_1, \quad \alpha_{20} = \frac{1}{2}\left[\sqrt{\frac{2}{s^{*2}} - 1} - 1\right]. \tag{24}$$

Evidently, the probabilities of phonon scattering by the angles θ_1 and θ_2 ($\cos\theta_1 = -1$, $\sin\theta_1 = 0$, $\cos\theta_2 = s^* - (1 - s^*)(2s^*\alpha_{20})^{-1}$) are zero, which leads to a nonmonotonic dependence of the scattering probability on the phonon wavevector (Fig. 3, curve *1*). The function $\varphi_2^2(z, z_1)$ quite rapidly increases on both sides of the point $z = z_{20}$. The value of z_{20} increases with the parameter z_1 and, for $z = z_{20} = 4$–5, the function $\varphi_2(z, z_1)$ becomes zero exactly in the region of maximum of the function $F^*(z, z_1)$. This leads to the appearance of inflection on the $\nu^*_{\mathrm{TLL}}(z_1, T)$ curve and the onset of plateau at $z_1 \approx 8(\sqrt{2/s^{*2} - 1} - 1)^{-1}$. Further increase in z_1 leads to a shift of the zero of the function $\varphi_2(z, z_1)$ toward the region of exponential decay in $F^*(z, z_1)$, whereby the values of $\nu^*_{\mathrm{TLL}}(z_1, T)$ slightly decrease (which corresponds to the plateau) until the zero of the function $\varphi_1(z, z_1)$ would pass through the maximum of $F^*(z, z_1)$, which corresponds to the second inflection on the $\nu^*_{\mathrm{TLL}}(z_1, T)$ curve (see Fig. 2). At $z_{\min} > 4$ or $z_1 > 8(1/s^* - 1)^{-1}$, the transverse phonon relaxation frequency occurs in the region of exponential decay. Thus, it is evident that the angular dependence of the phonon scattering probability should be taken into account in calculations of the phonon relaxation frequencies.

In the region of $z_1 \gg 1$ and $z_{\min}, z_{02} \gg 1$, the value of the preexponential factor can be estimated as follows. The main contribution to the integral is related to the region of the first peak, $z_{\min} < z < z_{20}$ (see Fig. 3). According to the theorem of the mean, functions smoothly varying at the middle point $\bar{z} = (z_{\min} + z_{20})/2$ can be placed before the integral over this region, while the remaining strongly varying part $\varphi_1\varphi_2^2\exp(-z)$ has to be precisely integrated. This yields

$$\nu^{\mathrm{TLL}}_{\mathrm{phN}}(z_1, T) = 7\left(\frac{\bar{z} + 1/2}{\bar{z} + 1}\right)^2$$
$$\times\left(\bar{z} + \frac{1}{2}\sqrt{\frac{2}{s^*} - 1} + 1\right)^2\left[\bar{z} + \frac{1}{2}\left(\frac{1}{s^*} + 1\right)\right] \tag{25}$$
$$\times\left[(\alpha_{20} - \alpha_{10})^2 z_1^3 - 4(\alpha_{20} - \alpha_{10}) z_1^2 + 6z_1\right]\exp(-\alpha_{10} z_1).$$

As can be seen from Fig. 2 (curve *4*), this approach provides a reliable estimate of $\nu^{\mathrm{TLL}}_{\mathrm{phN}}(z_1, T)$ for the high-frequency phonons. Thus, allowance for the angular dependence (14) of the phonon scattering probability leads, in contrast to the results obtained in [26, 27], to a preexponential factor in the form of third-order polynomial, which cannot be reduced to a simple power dependence (23).

In contrast to the isotropic case, the angular dependence of the squared matrix element for the cubic crystals is determined by a combination of three functions: $\varphi_1(z, z_1)$, $\varphi_2(z, z_1)$, and $\varphi_3(z, z_1)$. In this case, a change in positions of the zeros of these functions ($z_{10} = z_{\min}$, z_{20} and $z_{30} = (1/2)z_1(1/s^{*2} - 1)$) relative to the maximum of $F^*(z, z_1)$ in the course of increasing the phonon wavevector also determines peculiarities in the behavior of the relaxation frequency $\nu^*_{\mathrm{TLL}}(z_1, T)$. For the cubic crystals of germanium, silicon, and diamond, the probability of phonon scattering turns zero only at the lower integration limit $z_{10} = z_{\min}(\theta_2 = \pi)$. However, separate contributions to the scattering probability may also turn zero. For example, the isotropic contribution to the scattering is zero for $z = z_{20}$; the interference contribution and some of the terms corresponding to the anisotropic scattering of phonons (proportional to the functions $\varphi_2(z, z_1)$ and $\varphi_3(z, z_1)$) are zero at $z = z_{20}$ and $z = z_{30}$. The anisotropic part of the scattering probability proportional to b_3^2 is nonzero in the entire integration domain and accounts for a monotonically decreasing contribution to the phonon relaxation frequency. This contribution does not exceed 4% for germanium and 6% for silicon.

Let us consider the dependence of the phonon relaxation frequency on the reduced wavevector in the [100] direction of germanium, silicon, and diamond crystals. As can be seen from Fig. 4, these dependences qualitatively differ from those obtained in the case of the isotropic medium. The $\nu^*_{\mathrm{TLL}}(z_1, T)$ curves are nonmonotonic and display a minimum and the second maximum instead of the plateau. The appearance of new peculiarities (the minimum and the second maximum at $z \gg 1$) in the dependence of the phonon relaxation frequency on the phonon wavevector is of considerable interest from the standpoint of ultrasonic investigations and requires additional thorough investigation. Our analysis of all contributions to the relaxation frequency $\nu^*_{\mathrm{TLL}}(z_1, T)$ showed that the appearance of the second maximum is related to the cubic anisotropy of the crystals. The interference of the isotropic and anisotropic scattering (the second term in expression (17')), as well as the anisotropic scattering proper (the last three terms in expression (17')), accounts for the two-hump shape of the observed dependence of the phonon relaxation frequency on the reduced wavevector in the crystals of germanium and silicon (see Fig. 4, curves *3* and *4*). A comparison of the curves in Fig. 4 shows that the predominant contribution to the relaxation frequency in the case of germanium and silicon is due to the interfer-

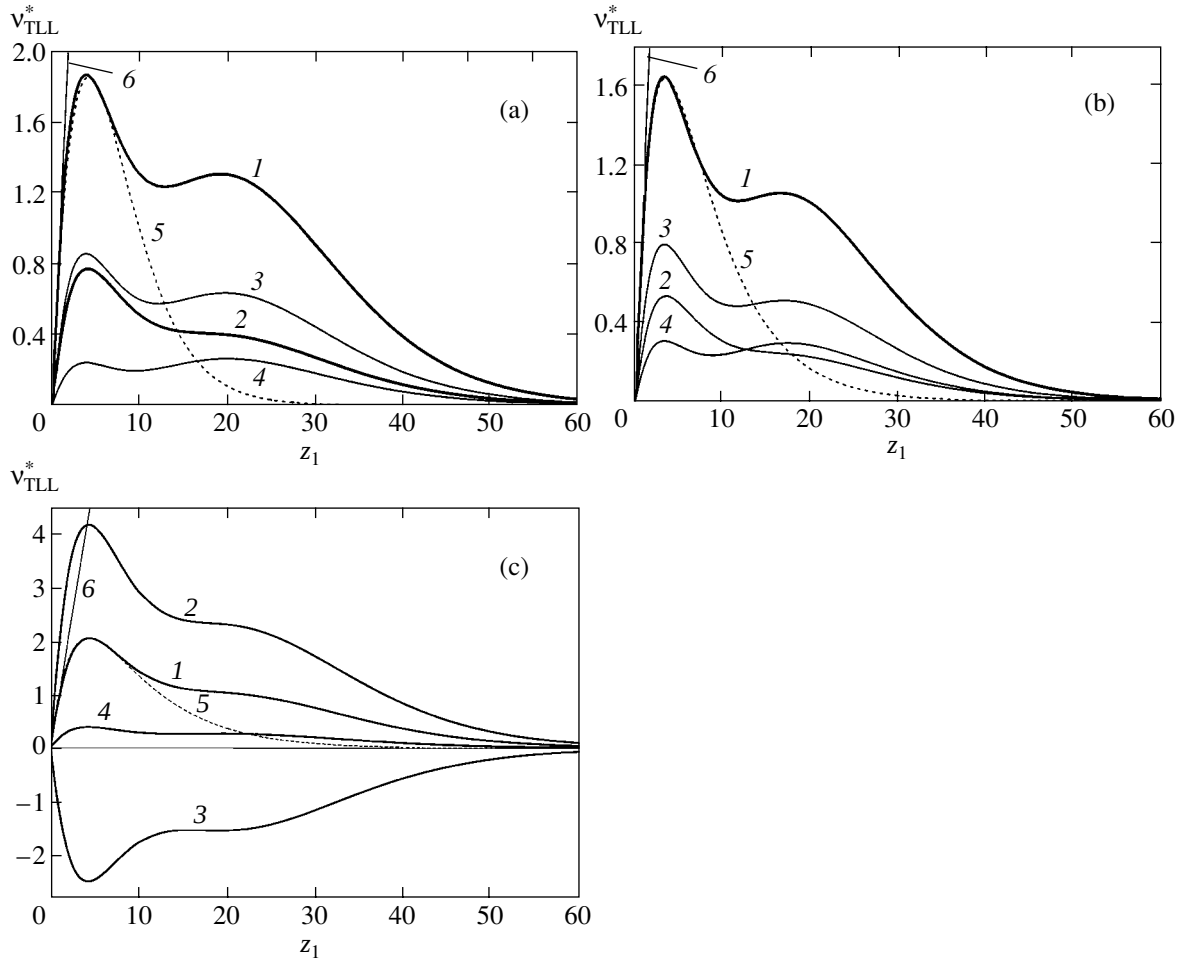


Fig. 4. Plots of the phonon relaxation frequency ν^*_{TLL} versus reduced wavevector z_1 in the [100] direction for (a) germanium, (b) silicon, and (c) diamond crystals at $T = 1$ K (*1*) and contributions due to the isotropic scattering (*2*), interference of the isotropic and anisotropic scattering (*3*), and anisotropic scattering (*4*); curve *5* shows the approximation by formula (26); curve *6* shows the linear Landau–Rumer approximation.

ence term, while the contribution of isotropic scattering is somewhat less significant. In diamond crystals, predominant contribution in the region of the first maximum is due to the isotropic scattering, while the interference scattering is negative and is one and a half times lower in absolute value; for this reason, we have a plateau instead of the second maximum as in the isotropic case. The contribution due to the anisotropic scattering (curves *4*) is rather significant for germanium and silicon and is half as small for diamond in the entire range of wavevectors. It should be noted that the dependence of the relaxation frequency $\nu^*_{TLL}(z_1, T)$ on the phonon wavevector in the [111] direction of germanium, silicon, and diamond crystals is similar to that in the isotropic case (see Fig. 2): the second maximum is missing and there is a region of slower decay instead of the plateau. Thus, the appearance of the second maximum on the $\nu^*_{TLL}(z_1, T)$ curves for the [100] direction in the crystals of germanium and silicon is related to the cubic anisotropy of these crystals.

As for the analysis of thermal conductivity of the germanium, silicon, and diamond crystals, the use of formulas (15) for $\nu^{TLL}_{phN}(z_1, T)$ is rather inconvenient and strongly complicates calculations. Since the realistic energies of photons important from the standpoint of thermal conductivity are limited from above on a level of $z_1 < 4$–5 (greater values are exponentially cut by the Planck distribution function), the phonon relaxation frequency in the region of the first maximum can be approximated by the expression

$$\nu^{TLL}_{phN}(z_1, T) \approx 0.65 B_{T0} T^5 z_1 \times [\exp(-\alpha_{10} z_1) + \exp(-\alpha_{20} z_1)]. \quad (26)$$

In the region of $0 < z_1 < 8$ (see Fig. 4), this approximation agrees well with formulas (15) for both the isotropic case and the cubic crystals of germanium, silicon,

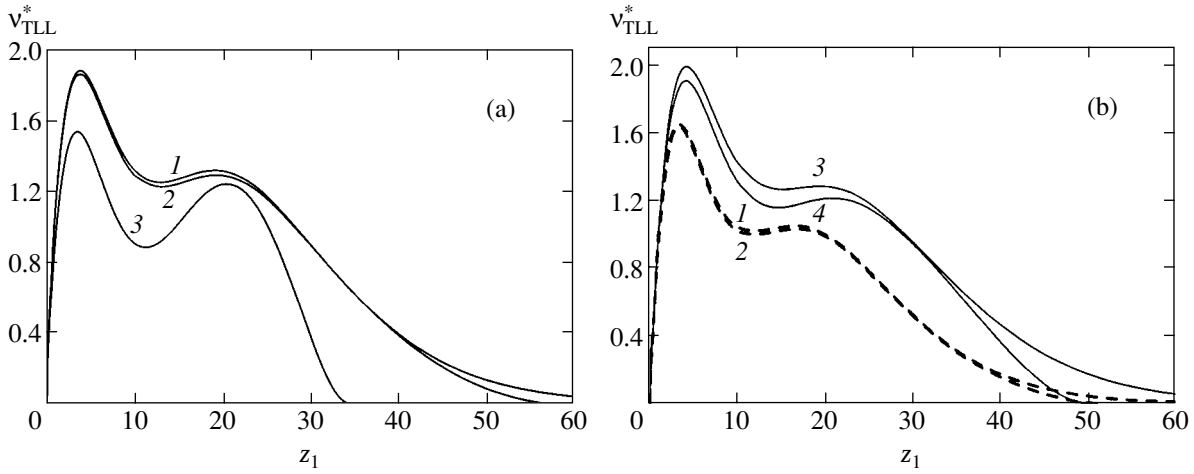


Fig. 5. Plots of the phonon relaxation frequency ν^*_{TLL} versus reduced wavevector z_1 in the [100] direction for (a) germanium and (b) silicon and diamond crystals at various temperatures: (a) germanium at T = 1–20 K (*1*), 30 K (*2*), and 50 K (*3*); (b) silicon at T = 1–40 K (*1*) and 50 K; diamond at T = 1–100 K (*3*) and 200 K (*4*).

and diamond. At least, this formula provides a better approximation than the Landau–Rumer formula (9) with a fitting parameter.

Let us also consider the possibility of experimental observation of the aforementioned features in the dependence of the transverse phonon relaxation frequency on the wavevector. As can be seen from Fig. 5, the relaxation frequency $\nu^*_{TLL}(z_1, T)$ for germanium, silicon, and diamond in the region of sufficiently low temperatures is practically a function of variable z_1 only. This function very slightly changes depending on the temperature: the positions of maxima and minima, as well as their absolute values, vary within less than 1% in the temperature interval $1 < T < 20$ K for germanium, $1 < T < 40$ K for silicon, and $1 < T < 100$ K for diamond. As is known (see, e.g., [18, 24], the coefficient of ultrasound absorption α_{TLL} is proportional to the phonon relaxation frequency:

$$\alpha_{TLL}(z, T) = \nu^{TLL}_{phN}(z, T)/2s_T. \quad (27)$$

Therefore, once the absorption coefficient of ultrasound with an energy of $\hbar\omega_{DT} \approx 10$ K is measured in the temperature interval from 0.1 to 50 K, it is possible to determine the behavior of $\nu^*_{TLL}(z_1, T)$ in the interval of wavevectors $0.2 < z_1 < 100$ as

$$\nu^*_{TLL}(z_1, T) = 2s_T\alpha_{TLL}/B_{T0}T^5. \quad (28)$$

For this purpose, it is necessary to transform the temperature dependence $\nu^*_{TLL}(z_1, T)$ at a fixed phonon frequency ω_{qT} into the function $z_{qT} = \hbar\omega_{qT}/k_BT$ at a fixed temperature, taking into account that, in the interval of low temperatures indicated above, the frequency $\nu^*_{TLL}(z_1, T)$ is practically a function only of z_1. In silicon and diamond crystals, the observation conditions are more favorable for transverse phonons in the terahertz range (1 THz ≈ 50 K). Thus, if the Landau–Rumer mechanism is the main process of the transverse phonon relaxation, the dependence of the relaxation frequency on the phonon frequency in germanium, silicon, and diamond crystals can be determined from the results of measurements of the ultrasound absorption coefficient. It can be recommended to perform such experiments with highly isotopically enriched, pure crystals of germanium, silicon, and diamond in order to reduce the phonon scattering on defects, which can mask the anharmonic phonon scattering processes at rather low temperatures.

5. CONCLUSIONS

Thus, we have considered the relaxation of transverse thermal and high-frequency phonons according to the Landau–Rumer mechanism in the isotropic medium and in the crystals of germanium, silicon, and diamond possessing a cubic symmetry. The elastic energy caused by the anharmonicity of vibrations of the cubic crystal lattice is expressed via the second- and third-order moduli of elasticity. Using the known values of these elastic moduli, parameters determining the frequencies of the transverse phonon relaxation according to the Landau–Rumer mechanism have been evaluated for the germanium, silicon, and diamond crystals. It is shown that the dependence of the relaxation frequency on the wavevector of thermal and high-frequency phonons sharply differs from the classical Landau–Rumer relationship both in the isotropic medium and in the cubic crystals studied. For the [100] crystallographic direction in the germanium, silicon, and diamond crystals, in contrast to the case of isotropic media, this dependence exhibits a substantially nonmonotonic character with two maxima; the second maximum falls within the region of high-frequency

phonons ($\hbar\omega_{qT} \geqslant k_B T$). It is established that peculiarities revealed in the dependence of the relaxation frequency of the phonon wavevector are related to the angular dependence of the probability of anharmonic scattering and the anisotropy of elastic properties of germanium, silicon, and diamond crystals. The values of scattering angles are determined for which the scattering probability is zero. We have also considered the possibility of experimentally determining the relaxation frequency for high-frequency phonons as a function of the phonon wavevector by measuring the temperature dependence of the coefficient of absorption of high-frequency ultrasound.

ACKNOWLEDGMENTS

The authors are grateful to A.P. Tankeev, I.F. Mirsaev, V.I. Ozhogin, A.V. Inyushkin, and A.N. Taldenkov for fruitful discussions and useful critical remarks.

This study was supported by the Presidential Program for Support of the Leading Scientific Schools in Russia (project no. NSh-1380.2003.2), the Youth project of the Ural Division, Russian Academy of Sciences (grant no. 15-02-04), the Dynasty Foundation, and the Moscow International Center for Fundamental Physics (MTsFFM).

REFERENCES

1. A. P. Zhernov and A. V. Inyushkin, *Isotope Effects in Solids* (Ross. Nauchn. Tsentr "Kurchatovskiĭ Institut," Moscow, 2001).
2. M. Asen-Palmer, K. Bartkowski, E. Gmelin, *et al.*, Phys. Rev. B **56**, 9431 (1997).
3. V. I. Ozhogin, A. V. Inyushkin, A. N. Taldenkov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 463 (1996) [JETP Lett. **63**, 490 (1996)].
4. T. Ruf, R. W. Henn, M. Asen-Palmer, *et al.*, Solid State Commun. **115**, 243 (2000).
5. W. S. Capinski, H. J. Maris, E. Bauser, *et al.*, Appl. Phys. Lett. **71**, 2109 (1997).
6. J. E. Graebner, M. E. Reiss, L. Seibles, *et al.*, Phys. Rev. B **50**, 3702 (1994).
7. J. R. Olson, R. O. Phol, J. W. Vandersande, *et al.*, Phys. Rev. B **47**, 14850 (1993).
8. Lanhua Wei, P. K. Kuo, R. L. Thomas, *et al.*, Phys. Rev. Lett. **70**, 3764 (1993).
9. A. P. Zhernov and D. A. Zhernov, Zh. Éksp. Teor. Fiz. **114**, 1757 (1998) [JETP **87**, 952 (1998)]; A. P. Zhernov, Fiz. Tverd. Tela (St. Petersburg) **41**, 1185 (1999) [Phys. Solid State **41**, 1079 (1999)].
10. I. G. Kuleev and I. I. Kuleev, Zh. Éksp. Teor. Fiz. **120**, 649 (2001) [JETP **93**, 568 (2001)].
11. I. G. Kuleev and I. I. Kuleev, Zh. Éksp. Teor. Fiz. **122**, 558 (2002) [JETP **95**, 480 (2002)].
12. R. Berman, *Thermal Conduction in Solids* (Clarendon, Oxford, 1976; Mir, Moscow, 1979).
13. B. M. Mogilevskiĭ and A. F. Chudnovskiĭ, *Thermal Conduction in Semiconductors* (Nauka, Moscow, 1972).
14. F. Birch, Phys. Rev. **71**, 809 (1947).
15. L. K. Zarembo and V. A. Krasil'nikov, *Introduction to Nonlinear Physical Acoustics* (Nauka, Moscow, 1966).
16. J. W. Tucker and V. W. Rampton, *Microwave Ultrasonics in Solid State Physics* (North-Holland, Amsterdam, 1972; Mir, Moscow, 1975).
17. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 7: *Theory of Elasticity*, 4th ed. (Nauka, Moscow, 1987; Pergamon, New York, 1986).
18. R. Truell, C. Elbaum, and B. B. Chick, *Ultrasonic Methods in Solid State Physics* (Academic, New York, 1969; Mir, Moscow, 1972).
19. T. Bateman, W. P. Mason, and H. J. McSkimin, J. Appl. Phys. **32**, 928 (1961).
20. H. J. McSkimin and P. Andreath, J. Appl. Phys. **35**, 3312 (1964).
21. M. H. Grimsditch, E. Anastassakis, and M. Cardona, Phys. Rev. B **18**, 901 (1978).
22. I. N. Frantsevich, F. F. Voronov, and S. A. Bakuta, *Elastic Constants and the Modula of Elasticity of Metals and Nonmetals: Handbook* (Naukova Dumka, Kiev, 1982).
23. L. Landau and J. Rumer, Phys. Z. Sowjetunion **11**, 18 (1937).
24. H. J. Maris, Phys. Acoust. **7**, 280 (1971).
25. D. T. Morelli, J. P. Hereman, and G. A. Slack, Phys. Rev. B **66**, 195394 (2002).
26. R. Orbach and L. A. Vredevoe, Physics (Long Island City, N.Y.) **1**, 91 (1964).
27. P. S. Zyryanov and G. G. Taluts, Zh. Éksp. Teor. Fiz. **49**, 1942 (1965) [Sov. Phys. JETP **22**, 1326 (1965)].
28. B. L. Timan, Fiz. Tverd. Tela (Leningrad) **6**, 950 (1964) [Sov. Phys. Solid State **6**, 732 (1964)].

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 121–132.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 142–154.

SOLIDS
Structure

Gibbs Free Energy and Thermodynamic Equilibrium Boundary Condition for a First-Order Phase Transition in an Isotropic Solid

A. P. Kochkin

Vereshchagin Institute of High-Pressure Physics, Russian Academy of Sciences, Troitsk, Moscow oblast, 142432 Russia

e-mail: kochkin@hppi.troitsk.ru

Received August 20, 2003

Abstract—For elastic deformation of arbitrary magnitude in an isotropic solid, Legendre conjugate strain variables are found and used to define the Gibbs free energy of a deformable solid. An additional thermodynamic equilibrium condition is found in the case when transition in a nonuniform strain field is incomplete and there exists an equilibrium boundary between phases. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

1.1. Phase equilibria and phase transitions in solids under high pressure (when the isotropic strain component is large) cannot be described if strain is treated as a small quantity. A straightforward, but tedious, mathematical analysis can be performed outside the scope of linear elasticity theory by applying the nonlinear theory of natural strain developed in [1–3], which is particularly well suited to dealing with states in the vicinity of the hydrostatic axis, $\sigma_{ij} = -p\delta_{ij}$, in the stress space. To examine the regions of phase coexistence in this space, thermodynamic potentials that are Legendre conjugate to energy and Helmholtz free energy (i.e., enthalpy and Gibbs free energy, respectively) are required. Their determination is one of the objectives of this study.

1.2. The symbolic notation used here is almost identical to that adopted in [1–3]. The most frequently employed symbols include the following: $\langle a\rangle = \mathrm{Sp}\, a = a_{ii}$; an element a of the space of rank 2 tensors $\mathbb{E}(3) \otimes \mathbb{E}(3)$ is denoted by $a\rangle$; its conjugate (i.e., its transpose a^{T} in matrix representation), by $\langle a$. Elasticity theory mostly deals with elements of the subspace $\mathrm{Sym}\mathbb{E}(3) \otimes \mathbb{E}(3)$ (symmetrized with respect to permutation of indices), for which $a^{\mathrm{T}} = a$, i.e., $a\rangle = \langle a$. Cursive notation (as $\mathscr{A}$) is used to denote operators represented by rank 4 tensors in this subspace. The unit tensor E is the Kronecker delta: $E_{ij} \equiv \delta_{ij}$. It is assumed that $\mathscr{A}E\rangle = \mathscr{A}\rangle$, $\mathscr{A}a\rangle = \mathscr{A}a$, and $\mathscr{A}\rangle_{ij} = A_{ijkk}$. It is obvious that $\mathscr{A}a\rangle = \langle a\mathscr{A}^t$, i.e., $(\mathscr{A}^t)_{ijkl} = A_{klij}$. When the indices of a tensor operator are omitted, the following notation is used

$$a_{ij}b_{kl} \longrightarrow a\rangle\langle b$$

for a dyadic product and

$$A_{ijkl} = \frac{\partial a_{ij}}{\partial b_{kl}} \longrightarrow \frac{\partial a}{\partial b}$$

for a derivative of a tensor with respect to a tensor, where the indices in the differentiation variable are placed rightmost to satisfy the rules of matrix multiplication. Different components of the same tensor contained in the same expression are distinguished by using primed variables and placing them rightmost, as in the following derivative of a scalar function:

$$A_{ijkl} = \frac{\partial^2 w}{\partial a_{ij}\partial a_{kl}} \longrightarrow \frac{\partial^2 w}{\partial a \partial a'}.$$

Similarly, the quadratic form based on the second derivative of a scalar function with respect to the same tensor may be written as

$$b_{ij}\frac{\partial^2 w}{\partial a_{ij}\partial a_{kl}}b_{kl} \longrightarrow \left\langle b\frac{\partial^2 w}{\partial a \partial a'}b'\right\rangle.$$

More complicated expressions are not considered here.

2. HELMHOLTZ FREE ENERGY AS A FUNCTION OF "EXTENSIVE" STRAIN VARIABLES

2.1. It was shown in [1] that the work required to create an infinitesimal elastic strain in an isotropic solid can be written as

$$\delta R = \oint dS_i \sigma_{ij}\delta x_j = \int dV \langle \sigma \delta s\rangle, \qquad (1)$$

where the integrals are, respectively, over the actual surface and volume of the strained solid (i.e., in Eule-

rian variables); σ is the stress tensor satisfying the equilibrium equation

$$\frac{\partial \sigma_{ij}}{\partial x_j} = 0, \tag{2}$$

and s is the natural strain tensor.[1] This made it possible to define a Helmholtz free-energy density f as a function of s. However, its Legendre conjugate, known as the Gibbs free energy in the hydrostatic case, still needed to be determined.

Consider the Helmholtz free energy f of a unit mass of a solid as a function of the variable

$$\kappa = \frac{\Delta + E/3}{\rho}, \tag{3}$$

where Δ is the deviatoric (traceless) part of s [1, 2] and ρ is the mass density at a particular point. Since

$$\langle \kappa \rangle = 1/\rho, \tag{3'}$$

and $\rho = \rho_0 e^{-\langle s \rangle}$ [1] (subscript 0 refers to the undeformed state), it is clear that f must be expressed as a function of

$$s \equiv \Delta + \frac{\langle s \rangle}{3} E = \frac{\kappa}{\langle \kappa \rangle} + \frac{E}{3}[\ln(\rho_0(\langle \kappa \rangle)) - 1]. \tag{4}$$

Therefore, the derivative of s with respect to κ is

$$\mathcal{D}^s_\kappa = \rho \mathcal{U}, \tag{5}$$

where

$$\mathcal{U} = \mathcal{I} - \rho\kappa\rangle\langle E + \frac{E\rangle\langle E}{3} = \mathcal{I} - \Delta\rangle\langle E.$$

The following operators are frequently used below:

$$\begin{gathered} \mathcal{U}^{-1} \equiv \mathcal{U}_+ = \mathcal{I} + \Delta\rangle\langle E, \quad \mathcal{U}^t = \mathcal{I} - E\rangle\langle \Delta, \\ \mathcal{U}^{t-1} \equiv \mathcal{U}_+^{\mathrm{T}} = \mathcal{I} + E\rangle\langle \Delta. \end{gathered} \tag{5'}$$

Strictly speaking, the identity operator $\mathcal{I}$ in $\mathbb{E}(3) \otimes \mathbb{E}(3)$ ($\mathcal{I}_{ijkl} = \delta_{ik}\delta_{jl}$) should be replaced here by the operator Π_+ of projection onto the subspace $\mathrm{Sym}\,\mathbb{E}(3) \otimes \mathbb{E}(3)$ (which is equivalent to the identity operator in this subspace),

$$(\Pi_+)_{ijkl} = (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})/2. \tag{5''}$$

However, the latter operator is required only when the symmetry of certain final results with respect to permutation of pairs of indices must be demonstrated (see [5] for details concerning operators in $\mathrm{Sym}\,\mathbb{E}(3) \otimes \mathbb{E}(3)$).

[1] An expression for the work required to create an incremental strain δs that is coaxial with σ in a uniform stress field was originally considered in [4].

By virtue of (5), infinitesimal work (1) per unit mass can be expressed as

$$\frac{1}{\rho}\langle \sigma ds \rangle = \frac{1}{\rho}\langle \sigma \mathcal{D}^s_\kappa d\kappa \rangle = \langle \theta d\kappa \rangle,$$

which implies that

$$df = -\eta dT + \langle \theta d\kappa \rangle. \tag{6}$$

Here, η is the entropy per unit mass and

$$\theta = \frac{\partial f}{\partial \kappa} = \frac{\partial f}{\partial s}\mathcal{D}^s_\kappa = \sigma\mathcal{U} = \sigma - \langle \sigma\Delta \rangle E \equiv \mathcal{U}^t\sigma. \tag{7}$$

Recalling the expression for σ from [1] (see also [2]),

$$\sigma = -pE + \tau, \quad \tau = 2\mu\Delta + \nu\Delta_2, \tag{8}$$

where $\mu = \mu(s)$ is the shear modulus, $\nu = \nu(s)$ is the second shear modulus, and Δ_2 is the deviatoric part of the tensor Δ^2, one obtains

$$\theta = -\pi E + \tau,$$

where

$$\begin{gathered} \pi = p + \langle \sigma\Delta \rangle, \quad \langle \sigma\Delta \rangle = \langle \tau\Delta \rangle = 4\mu k_2 + 3\nu k_3, \\ 2k_2 = \langle \Delta^2 \rangle, \quad 3k_3 = \langle \Delta^3 \rangle. \end{gathered} \tag{7'}$$

2.2. Scalar thermodynamic state variables for an isotropic solid are invariant under rotation, i.e., functions of any set of three functionally independent invariants $\{X_1, X_2, X_3\}$ of the tensor representing strain. If

$$X_1 = I_s = \langle s \rangle, \quad X_2 = k_2, \quad X_3 = k_3, \tag{9}$$

then equation of state (8) can be rewritten in a more general form:

$$\frac{\sigma}{\rho} = \frac{\partial f}{\partial s} = f^a e_a, \tag{10}$$

where

$$f^a = \frac{\partial f}{\partial X_a}, \quad e_a = \frac{\partial X_a}{\partial s},$$

with

$$e_1 = E, \quad e_2 = \Delta, \quad e_3 = \Delta_2 \tag{11}$$

(see (8) and [1, 2]).

Similarly, define the "extensive" deviatoric part and second deviatoric part of the tensor κ,

$$\Omega = \frac{\Delta}{\rho} = \kappa - \frac{\langle\kappa\rangle}{3}E, \quad \Omega_2 = \frac{\Delta_2}{\rho^2} = \Omega^2 - \frac{\langle\Omega^2\rangle}{3}E, \tag{12}$$

and choose the following set of functionally independent invariants of κ

$$\tilde{X}_1 = I_\kappa = \langle\kappa\rangle = \frac{1}{\rho} = v, \quad \tilde{X}_2 = k_{2\kappa} = \frac{1}{2}\langle\Omega^2\rangle, \quad \tilde{X}_3 = k_{3\kappa} = \frac{1}{3}\langle\Omega^3\rangle, \tag{9'}$$

where v is the volume per unit mass. Then, the equation of state becomes

$$\theta = \tilde{f}^a\tilde{e}_a, \tag{10'}$$

where

$$\tilde{f}^a = \frac{\partial f}{\partial\tilde{X}_a}, \quad \tilde{e}_a = \frac{\partial\tilde{X}_a}{\partial\kappa} = \rho^{1-a}e_a.$$

Since this implies that

$$\tilde{e}_1 = E, \quad \tilde{e}_2 = \Omega, \quad \tilde{e}_3 = \Omega_2,$$

it follows that

$$\theta = -\pi E + 2m\Omega + n\Omega_2, \quad \pi = -\frac{\partial f}{\partial I_\kappa} = -\frac{\partial f}{\partial v}, \quad 2m = \frac{\partial f}{\partial k_{2\kappa}}, \quad n = \frac{\partial f}{\partial k_{3\kappa}}. \tag{13}$$

By virtue of Eqs. (8), (11), and (12),

$$m = \rho\mu, \quad n = \rho^2\nu.$$

The following expressions for cross derivatives are straightforward corollaries to (13):

$$\frac{\partial\pi}{\partial k_{2\kappa}} = -2\frac{\partial m}{\partial I_\kappa}, \quad \frac{\partial\pi}{\partial k_{3\kappa}} = -\frac{\partial n}{\partial I_\kappa}, \quad 2\frac{\partial m}{\partial k_{3\kappa}} = \frac{\partial n}{\partial k_{2\kappa}}, \tag{14}$$

which can be written (for covariant $\tilde{X}_i$) as

$$\pi^2 = -2m^1, \quad \pi^3 = -n^1, \quad 2m^3 = n^2.$$

2.3. As an illustration, let us define the elastic modulus of an isotropic solid. Under an arbitrary choice of invariants, the general definition of the isothermal modulus is

$$\mathcal{M} = \frac{1}{\rho}\left(\frac{\partial\theta}{\partial\kappa}\right)_T = \frac{1}{\rho}\frac{\partial^2 f}{\partial\tilde{X}_a\partial\tilde{X}_b}\tilde{\mathcal{P}}_{ab} + \frac{1}{\rho}\frac{\partial f}{\partial\tilde{X}_a}\tilde{\mathcal{D}}_a, \tag{15}$$

where

$$\tilde{\mathcal{P}}_{ab} = \tilde{e}_a\rangle\langle\tilde{e}_b = \rho^{2-a-b}\mathcal{P}_{ab}, \quad \mathcal{P}_{ab} = e_a\rangle\langle e_b, \quad \tilde{\mathcal{D}}_a = \frac{\partial\tilde{e}_a}{\partial\kappa} \tag{16}$$

(see also [5]). When the invariants are defined by (9'), it can be shown directly that

$$\delta\Omega = \left(\Pi_+ - \frac{1}{3}\mathcal{P}_{11}\right)\delta\kappa, \quad \delta\Omega_2 = 2\left[\mathcal{A}_+^\Omega - \frac{1}{3}(\tilde{\mathcal{P}}_{12} + \tilde{\mathcal{P}}_{21})\right]\delta\kappa, \tag{17}$$

where

$$\mathcal{A}_+^\Omega x = \frac{1}{2}(\Omega x + x\Omega) \equiv \frac{1}{2}[(\Omega\otimes E) + (E\otimes\Omega)]x,$$

$$\forall x \in \mathrm{Sym}\,\mathbb{E}(3)\otimes\mathbb{E}(3).$$

Accordingly,

$$\tilde{\mathcal{D}}_2 = \Pi_+ - \frac{1}{3}\mathcal{P}_{11}, \quad \tilde{\mathcal{D}}_3 = 2\mathcal{A}_+^\Omega - \frac{2}{3}(\tilde{\mathcal{P}}_{12} + \tilde{\mathcal{P}}_{21})$$

and the elasticity modulus is

$$\mathcal{M} = M^{ab}\mathcal{P}_{ab} + 2\mu\Pi_+ + 2\nu\mathcal{A}_+^\Delta, \tag{18}$$

where the symmetric matrix M^{ab} has the entries

$$M^{11} = M - \frac{2\mu}{3}, \quad M^{12} = 2v\left(\frac{\partial m}{\partial v} - \frac{n}{3}\right), \quad M^{13} = v\frac{\partial n}{\partial v},$$

$$M^{22} = 2v\frac{\partial m}{\partial k_{2\kappa}}, \quad M^{23} = v\frac{\partial n}{\partial k_{2\kappa}}, \quad M^{33} = v\frac{\partial n}{\partial k_{3\kappa}},$$

and

$$M = -v\frac{\partial\pi}{\partial v}. \tag{18'}$$

2.4. To relate the tensor of thermodynamic moduli to the tensor $\mathcal{H}$ introduced in [3], combine (4) with (7) to obtain

$$\mathcal{M} = \frac{1}{\rho}\frac{\partial\theta}{\partial\kappa} = \frac{1}{\rho}\mathcal{D}_\sigma^\theta\frac{\partial\sigma}{\partial s}\mathcal{D}_\kappa^s = \mathcal{U}^t\mathcal{Y}\mathcal{U} - E\rangle\langle\tau\mathcal{U}, \tag{19}$$

where $\mathcal{Y} = \partial\sigma/\partial s$ and

$$\mathcal{D}_\sigma^\theta = \mathcal{U}^t - E\rangle\langle\tau\mathcal{Y}^{-1}. \tag{20}$$

It was shown in [3] that

$$\mathcal{Y} = \mathcal{H} + \frac{1}{2}(E\rangle\langle\tau - \tau\rangle\langle E).$$

Therefore,

$$\mathcal{M} = \mathcal{U}^t\mathcal{K}\mathcal{U} - \frac{1}{2}(E\rangle\langle\tau\mathcal{U} + \mathcal{U}^t\tau\rangle\langle E),$$
$$\mathcal{K} = \tilde{\mathcal{M}} + \frac{1}{2}(E\rangle\langle\tau + \tau\rangle\langle E), \quad (21)$$

where $\tilde{\mathcal{M}} = \mathcal{U}_+^t\mathcal{M}\mathcal{U}_+$.

The modulus $\mathcal{K}$ was defined in [3] somewhat artificially by using its symmetries with respect to both permutation of indices in the first and last index pairs and permutation of the pairs of indices. Consequently, it has the smallest number of independent components, which is equal to the number of independent components in the linear elastic modulus (i.e., 21 components at most). However, these properties characterize the quantity $\mathcal{M}$, which should, with more reason, be called—and is called in what follows—the tensor of elastic moduli. One additional reason for this is that, according to (5), (5'), and (21), both definitions of the modulus are equivalent on the hydrostatic axis, in which case $\Omega = \Omega_2 = 0$, $k_{2\kappa} = k_{3\kappa} = 0$, $\pi = p$, and $M = K$ in (18'), where K is the conventional bulk modulus [6]. This leads to the expression [3]

$$\mathcal{M} = M^{11}\mathcal{P}_{11} + 2\mu\Pi_+ = \left(K - \frac{2\mu}{3}\right)E\rangle\langle E + 2\mu\Pi_+, \quad (18'')$$

or (in indicial notation, see Section 1 and (5''))

$$M_{ijkl} = K\delta_{ij}\delta_{kl} + \mu\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} - \frac{2}{3}\delta_{ij}\delta_{kl}\right).$$

As in [3], its distinction from the conventional linear approximation (see [7]) lies in the fact that it contains the bulk modulus $K(p)$ and shear modulus $\mu(p)$ determined under pressure.

Formula (21) can be used to calculate $\mathcal{M}$ if the modulus $\mathcal{K}$ is known. The latter modulus is easier to calculate and interpret, because the natural strain tensor s is easier to conceive as compared to κ.

3. GIBBS FREE ENERGY

3.1. According to (6), we can subtract the quantity

$$\langle\theta\kappa\rangle = \langle\sigma\kappa\rangle - \langle\sigma\Delta\rangle\langle\kappa\rangle = \frac{\langle\sigma\rangle}{3\rho} = -\frac{p}{\rho}, \quad (22)$$

from f (i.e., apply the Legendre transform) to define the function

$$\varphi = f - \langle\theta\kappa\rangle = f + \frac{p}{\rho} \quad (23)$$

such that

$$d\varphi = -\eta dT - \langle\kappa d\theta\rangle \quad (24)$$

and therefore

$$\left(\frac{\partial\varphi}{\partial\theta}\right)_T = -\kappa. \quad (24')$$

3.2. By the first law of thermodynamics, the rate of variation of the energy $\mathcal{E}$ of a near-equilibrium solid with rigid boundaries is

$$\dot{E} = \dot{Q} + \dot{R}121$$

where $\dot{Q}$ is the rate of heat transfer to the solid and $\dot{R}$ is the work done on it per unit time ($\dot{R} = 0$ under the rigid-boundary condition). Therefore, the second law of thermodynamics, $\dot{Q} \le T\dot{H}$ (H is the total enthalpy of the solid), dictates that (see [6])

$$\dot{F} \le 0,$$

where $F = \int\rho dVf$ is the total Helmholtz free energy of the solid. This implies that F is minimal in equilibrium. In particular, when κ deviates from an equilibrium value, it holds that

$$\delta F = \int\rho dV\left(\langle\theta\delta\kappa\rangle + \frac{1}{2}\rho\langle\delta\kappa\mathcal{M}_T\delta\kappa\rangle + \ldots\right) \ge 0, \quad (25)$$

where $\mathcal{M}_T$ is the isothermal elastic modulus defined by (15).

By virtue of (1) and (2), the first summand in this integral is

$$\int dV\langle\sigma\delta s\rangle = \int dV\sigma_{ij}\frac{\partial\delta x_i}{\partial x_j} = \oint dS_i\sigma_{ij}\delta x_j - \int dV\frac{\partial\sigma_{ij}}{\partial x_j}\delta x_i = 0$$

since the equilibrium Helmholtz free energy F is defined for $\delta x_i = 0$. Therefore, the quadratic form based on $\mathcal{M}_T$ is positive definite.

Suppose that the state of equilibrium with prescribed surface force density $P_i = \sigma_{ij}n_j$ corresponds to fields σ_e and s_e inside the solid. Using (7) and (3) to find θ_e and κ_e, define

$$J = J(\theta_e, \kappa) = \int\rho dV\langle\theta_e\kappa\rangle \quad (26)$$

with $\theta_e(\mathbf{r}) \equiv \theta(\mathbf{r}_e)$ corresponding to the material point

$$\mathbf{r} = \mathbf{r}_e + \mathbf{u}' = \mathbf{r}_0 + \mathbf{u}$$

that is located at $\mathbf{r}_e$ in equilibrium. Here, $\mathbf{u}$ is the total displacement of the point due to deformation and $\mathbf{u}'$ is

the displacement relative to its equilibrium location. Since $d\tilde{m} = \rho dV$ (where $\tilde{m}$ is mass) is invariant under the change of variables $\mathbf{r}_e \longleftrightarrow \mathbf{r}$, integrals of the form $\int \rho dV j(\mathbf{r})$ can be written as

$$\int_{V_e} \rho_e dV_e j(\mathbf{r}_e + \mathbf{u}')$$

or even

$$\int_{V_0} \rho_0 dV_0 j(\mathbf{r}_0 + \mathbf{u}),$$

where $\mathbf{r}_0$ is the location of the point in the unloaded state of the solid and the subscripts "0" and "e" correspond to the unloaded and equilibrium states of the solid, respectively. Then, (26) is rewritten as

$$J = \int \rho dV \langle \theta_e(\mathbf{r} - \mathbf{u}')\kappa(\mathbf{r})\rangle = \int_{V_e} \rho_e dV_e \langle \theta_e(\mathbf{r}_e)\kappa(\mathbf{r}_e + \mathbf{u}')\rangle$$

and the change from $\mathbf{r}$ to $\mathbf{r}_e$ makes it possible to ignore the condition $\int dV/\langle\kappa\rangle = \tilde{m}$.

Now, a "nonequilibrium" functional Φ (corresponding to an arbitrary κ) can be defined:

$$\begin{aligned}\Phi(T, \{\theta_e\}, \{\kappa\}) &= F(T, \{\kappa\}) - J(\theta_e, \kappa)\\ &= \int_{V_e} \rho_e dV_e (f - \langle\theta_e\kappa\rangle).\end{aligned} \tag{27}$$

By virtue of (25), a local minimum of this functional is reached for a one-phase system with a prescribed surface force density $\mathbf{P}$ at the equilibrium values $\kappa_e(\mathbf{r}_e)$:

$$\begin{aligned}\delta\Phi &= \oint dS_i(\sigma_{ij} - \sigma_{ij}^e)\delta x_j\\ &+ \frac{1}{2}\int \rho^2 dV \langle \delta\kappa \mathcal{M}_T \delta\kappa\rangle + \ldots \geq 0,\end{aligned} \tag{25'}$$

because $\sigma_{ij}n_j = \sigma_{ij}^e n_j = P_i$ at equilibrium.

Consequently, the equilibrium (minimal) value of the functional Φ is

$$\begin{aligned}\Phi(T, \{\mathbf{P}\}) &= F(T, \tilde{m}, \{\kappa_e\}) - J(\tilde{m}, \{\kappa_e\}, \{\theta\})\\ &= \int \rho dV \varphi(T, \theta),\end{aligned} \tag{28}$$

where $\theta \equiv \theta_e$ and κ_e are obtained by solving simultaneously either equation of state (8) or (13) and equilibrium equation (2) subject to the boundary condition $\sigma_{ij}n_j|_S = P_i$, calculating κ from (3) at every point, and using relation (13) to express κ_e in terms of θ; φ is determined from (23); the surface force density is assumed to satisfy the conditions of zero total force and torque applied to the solid:

$$\oint \mathbf{P} dS = 0, \quad \oint (\mathbf{r} \times \mathbf{P}) dS = 0. \tag{29}$$

This functional is called the Gibbs free energy of an elastically deformed solid. There exist an infinite number of functions minimized at equilibrium.[2] They correspond to different bulk moduli and dimensions of the ambient medium. In hydrostatics, a unified description is developed by assuming that the ambient medium (called barostat and/or thermostat) is infinitely large. Since this cannot be done in the case of a solid (normal surface stress cannot be held constant), the force density

$$\mathbf{P} = \hat{\sigma}_e \mathbf{n} = \hat{\theta}_e \mathbf{n} + \langle \theta_e \Delta\rangle \mathbf{n}$$

always fluctuates with both $\mathbf{n}$ and Δ. Expression (28) can be the Gibbs free energy, because $\varphi(T, \theta)$ and $f(T, \kappa)$ are Legendre conjugate quantities. Thus, the Gibbs free energy per unit mass of an arbitrarily deformed solid is given by (23), as in hydrostatics [8].

By virtue of (22), the equilibrium value of J can be expressed as

$$J = -\int p dV = \frac{1}{3}\oint dS_i \sigma_{ij} x_j, \tag{30}$$

where all quantities are taken at equilibrium.

3.3. In accordance with a conventional approach (e.g., see [6, 9]), the solid is divided into macroscopic regions and θ_e is assumed to be constant within each region. Then, the partition function for the T–θ distribution defined by analogy with the T–p distribution [9] is written as follows (with index pair in a tensor defined, for example, by Vogt's rule):

$$\begin{aligned}Y = \int \mathcal{D}\varepsilon(\mathbf{r}) \prod_{a=1}^{6} \mathcal{D}\kappa_a(\mathbf{r})\\ \times \exp\left\{-\beta \int_V \rho dV [\varepsilon(\mathbf{r}) - T\eta(\varepsilon(\mathbf{r}), \kappa(\mathbf{r})) - \langle\theta_e(\mathbf{r})\kappa(\mathbf{r})\rangle]\right\},\end{aligned} \tag{31}$$

where $\beta = 1/T$, the integrals are interpreted as functional ones with respect to ε and κ_a ($a \equiv (ij)$), temperature is measured in energy units so that entropy is a dimensionless quantity, and (see [9])

$$\Phi(T, \{\mathbf{P}\}) = -T \ln Y. \tag{31'}$$

[2] Their number is greater than the number of their analogs in hydrostatics, since the number of degrees of freedom in deviation from equilibrium is greater at every point.

3.4. When the set of independent invariants of θ consists of

$$I_\theta = \langle\theta\rangle = -3\pi, \quad k_{2\theta} = \frac{1}{2}\langle\tau^2\rangle, \quad k_{3\theta} = \frac{1}{3}\langle\tau^3\rangle, \tag{32}$$

then the basis for the decomposition of κ is

$$\tilde{e}'_1 = E, \quad \tilde{e}'_2 = \tau, \quad \tilde{e}'_3 = \tau_2 = \tau^2 - \frac{\langle\tau^2\rangle}{3}E.$$

By analogy with subsections 2.2 and 2.3, (24') yields

$$\kappa = \frac{v}{3}E + c\tau + d\tau_2, \quad v = \frac{1}{\rho} = -3\frac{\partial\varphi}{\partial I_\theta} = \frac{\partial\varphi}{\partial\pi}, \qquad c = -\frac{\partial\varphi}{\partial k_{2\theta}}, \quad d = -\frac{\partial\varphi}{\partial k_{3\theta}}. \tag{32'}$$

Consequently, cross derivatives are expressed as

$$\frac{\partial v}{\partial k_{2\theta}} = -\frac{\partial c}{\partial\pi}, \quad \frac{\partial v}{\partial k_{3\theta}} = -\frac{\partial d}{\partial\pi}, \quad \frac{\partial c}{\partial k_{3\theta}} = \frac{\partial d}{\partial k_{2\theta}},$$

and can be rewritten (in accordance with the ordering of the invariants) as follows:

$$v^2 = 3c^1, \quad v^3 = 3d^1, \quad c^3 = d^2,$$

where

$$x^1 = \frac{\partial x}{\partial I_\theta} = -\frac{1}{3}\frac{\partial x}{\partial\pi}.$$

The derivative of (32') is the compressibility modulus $\mathscr{C} = \mathscr{M}^{-1}$:

$$\mathscr{C} = \rho\frac{\partial\kappa}{\partial\theta} = C^{ab}\tilde{\mathscr{P}}'_{ab} + \rho c\Pi_+ + 2\rho d\tilde{\mathscr{A}}^\tau_+, \tag{33}$$

where $\tilde{\mathscr{P}}'_{ab} = \tilde{e}'_a\rangle\langle\tilde{e}'_b$ and the entries of the symmetric matrix C^{ab} are

$$C^{11} = \frac{\rho}{3}(v^1 - c), \quad C^{12} = \rho\left(c^1 - \frac{2}{3}d\right), \quad C^{13} = \rho d^1,$$

$$C^{22} = \rho c^2, \quad C^{23} = \rho d^1, \quad C^{33} = \rho d^3.$$

On the hydrostatic axis, the compressibility tensor reduces to

$$\mathscr{C} = C^{11}\mathscr{P}_{11} + \rho c\Pi_+.$$

Since $\mathscr{M}$ and $\mathscr{C}$ are mutually inverse, this entails the obvious relation $c = 1/2\mu\rho$ and

$$\mathscr{C} = \frac{1}{2\mu}\Pi_+ + \left(\frac{1}{9K} - \frac{1}{6\mu}\right)E\rangle\langle E. \tag{33'}$$

3.5. Since moduli under pressure are generally measured in ultrasonic experiments and, therefore, can be treated as adiabatic ones, the following identity can be proved by analogy with [6]:

$$\mathscr{M}_T = \mathscr{M}_\eta + \frac{T}{c_\kappa}a\rangle\langle a, \quad a = \left(\frac{\partial\theta}{\partial T}\right)_\kappa, \tag{34}$$

which implies that

$$\mathscr{C}_T = \mathscr{C}_\eta - \frac{T}{c_\kappa}\frac{b\rangle\langle b}{1 + T\langle ab\rangle/c_\kappa}, \quad b \equiv b\rangle = \mathscr{C}_\eta a, \tag{35}$$

and, therefore, $\langle b = \langle a\mathscr{C}_\eta$.

4. FLUCTUATIONS AND THERMODYNAMIC INEQUALITIES

4.1. Let us consider thermodynamic inequalities and fluctuations of state variables. The results presented here are obtained for a solid of unit mass. Suppose that T and θ are held constant, while ε and κ fluctuate. Denote an incremental deviation of any variable from its value at equilibrium by δ and, following [6], replace V with κ and $a\delta V$ (where a is a scalar) with $\langle a\delta\kappa\rangle$ (where a is a tensor):

$$\delta\varphi = \frac{1}{2}[\delta\eta\delta T + \langle\delta\theta\delta\kappa\rangle]. \tag{36}$$

Since the thermodynamic definition of enthalpy, $w = \varepsilon - \langle\theta\kappa\rangle$, implies that $(\partial T/\partial\theta)_\eta = -(\partial\kappa/\partial\eta)_\theta$, the expression above can be rewritten in terms of η and θ as

$$\delta\varphi = \frac{1}{2}\left[\left(\frac{\partial T}{\partial\eta}\right)_\theta(\delta\eta)^2 + \langle\delta\theta\left(\frac{\partial\kappa}{\partial\theta'}\right)_\eta\delta\theta'\rangle\right]. \tag{37}$$

Since identity (6) for Helmholtz free energy implies that $(\partial\theta/\partial T)_\kappa = -(\partial\eta/\partial\kappa)_T$, expression (36) can be rewritten in terms of T and κ as

$$\delta\varphi = \frac{1}{2}\left[\left(\frac{\partial\eta}{\partial T}\right)_\kappa(\delta T)^2 + \langle\delta\kappa\left(\frac{\partial\theta}{\partial\kappa'}\right)_T\delta\kappa'\rangle\right].$$

Following [6], we obtain

$$\overline{(\delta\eta)^2} = c_\theta, \quad \overline{\delta\theta\rangle\langle\delta\theta'} = T\rho\mathscr{M}_\eta, \quad \overline{\delta\eta\delta\theta} = 0, \qquad \overline{(\delta T)^2} = \frac{T^2}{c_\kappa}, \quad \overline{\delta\kappa\rangle\langle\delta\kappa'} = \frac{T}{\rho}\mathscr{C}_T, \quad \overline{\delta T\delta\kappa} = 0, \tag{38}$$

where $\rho\mathscr{M}_\eta = (\partial\theta/\partial\kappa)_\eta$ and c_θ and c_κ denote the specific heats at constant stress and strain, respectively.

The quadratic forms based on the moduli $\mathscr{M}_\eta$ and $\mathscr{M}_T$ are positive definite. The corresponding Sylvester inequalities, which ensure local stability of a physical system, are generally called thermodynamic inequalities.

By virtue of (38), Eq. (5) yields

$$\overline{\delta s\rangle\langle\delta s} = \rho T\mathcal{U}\mathcal{C}_T\mathcal{U}^T, \quad \overline{\delta T\delta s} = 0. \tag{39}$$

The projections of a symmetric rank 2 tensor onto the isotropic subspace of $\mathrm{Sym}\,\mathbb{E}(3) \otimes \mathbb{E}(3)$ and its orthogonal (deviatoric) complement are $\mathcal{P}_{11}/3$ and $\mathcal{P}_\perp = \Pi_+ - \mathcal{P}_{11}/3$, respectively. Therefore, since

$$\mathcal{P}_\perp\mathcal{U} = \Pi_+ - \rho\kappa\rangle\langle E,$$

it follows from (38) and (39), respectively, that

$$\overline{(\delta v)^2} = \frac{T}{\rho}\langle\mathcal{C}_T\rangle \equiv \frac{T}{\rho}(\mathcal{C}_T)_{iijj}, \quad \overline{\delta T\delta v} = 0,$$

and

$$\overline{\delta\Delta\rangle\langle\delta\Delta} = \rho T(\Pi_+ - \rho\kappa\rangle\langle E)\mathcal{C}_T(\Pi_+ - E\rangle\langle\rho\kappa), \quad \overline{\delta T\delta\Delta} = 0. \tag{39'}$$

In particular, the first expression in (39') is written in indicial notation as

$$\begin{aligned}\overline{\delta\Delta_{ij}\rangle\langle\delta\Delta_{kl}} &= \rho T((\mathcal{C}_T)_{ijkl} - \rho\kappa_{ij}(\mathcal{C}_T)_{mmkl} \\ &\quad - \rho(\mathcal{C}_T)_{ijmm}\kappa_{kl} + \rho^2(\mathcal{C}_T)_{mmnn}\kappa_{ij}\kappa_{kl}) \\ &= (\mathcal{C}_T)_{ijkl} - \Delta_{ij}(\mathcal{C}_T)_{mmkl} - \frac{1}{3}\delta_{ij}(\mathcal{C}_T)_{mmkl} \\ &\quad - (\mathcal{C}_T)_{ijmm}\Delta_{kl} - \frac{1}{3}(\mathcal{C}_T)_{ijmm}\delta_{kl} \\ &\quad + (\mathcal{C}_T)_{mmnn}\left(\Delta_{ij}\Delta_{kl} + \frac{1}{3}\delta_{ij}\Delta_{kl} + \frac{1}{3}\Delta_{ij}\delta_{kl} + \frac{1}{9}\delta_{ij}\delta_{kl}\right).\end{aligned}$$

Fluctuations of the stress tensor are calculated by using the identity $\delta\sigma = \mathcal{D}_\theta^\sigma\delta\theta$:

$$\overline{\delta\sigma\rangle\langle\delta\sigma'} = \mathcal{D}_\theta^\sigma\overline{\delta\theta\rangle\langle\delta\theta'}\mathcal{D}_\theta^{\sigma t} = \rho T\mathcal{D}_\theta^\sigma\mathcal{M}_\eta\mathcal{D}_\theta^{\sigma t}. \tag{40}$$

The operator $\mathcal{D}_\theta^\sigma$ can readily be found as the inverse of $\mathcal{D}_\theta^\sigma$ given by (20). Using the expression

$$\mathcal{Y}^{-1} = \tilde{\mathcal{C}} - \frac{\tilde{\mathcal{C}}\rangle\langle\tau\tilde{\mathcal{C}}}{1 + \langle\tau\tilde{\mathcal{C}}\rangle} \tag{41}$$

with $\tilde{\mathcal{C}} = \tilde{\mathcal{M}}^{-1}$, which can be derived from (19), one finds that

$$\mathcal{D}_\theta^\sigma = \mathcal{U}_+^t + E\rangle\langle\tau\mathcal{U}\mathcal{C}.$$

Hence,

$$\begin{aligned}\overline{\delta\sigma\rangle\langle\delta\sigma'} &= \rho(\tilde{\mathcal{M}}_\eta + E\rangle\langle\tau + \tau\rangle\langle E + E\rangle\langle\tau\tilde{\mathcal{C}}_\eta\tau\rangle\langle E) \\ &= \rho\left(\mathcal{K}_\eta + \frac{1}{2}(E\rangle\langle\tau + \tau\rangle\langle E) + E\rangle\langle\tau\tilde{\mathcal{C}}_\eta\tau\rangle\langle E\right).\end{aligned}$$

4.2. By virtue of (18") and (33'), Eqs. (39) and (40) yield the following relations valid on the hydrostatic axis:

$$\overline{(\delta v)^2} = \frac{T}{\rho K_T}, \quad \overline{\delta\Delta\rangle\delta v} = 0,$$

$$\overline{\delta\Delta\rangle\langle\delta\Delta} = \frac{\rho T}{2\mu_T}\left(\Pi_+ - \frac{1}{3}E\rangle\langle E\right),$$

$$\overline{(\delta p)^2} = \rho T K_\eta, \quad \overline{\delta p\tau} = 0,$$

$$\overline{\tau\rangle\langle\tau} = 2T\mu_\eta\left(\Pi_+ - \frac{1}{3}E\rangle\langle E\right).$$

In particular, volume and shear fluctuations (as well as pressure and shear-stress fluctuations) are statistically independent for arbitrary values of strain and stress.

Fluctuations of intensive quantities for a solid of arbitrary mass $\tilde{m}$ in uniform stress field are given by the right-hand sides of these expressions divided by $\tilde{m}$. Fluctuations of extensive quantities are obtained by multiplying similar quotients by $\tilde{m}^2$.

This leads to the well-known formulas (see [6])

$$\begin{aligned}\overline{(\delta V)^2} &= \frac{T\tilde{m}}{\rho K_T} = -T\left(\frac{\partial V}{\partial p}\right)_T, \\ \overline{(\delta p)^2} &= \frac{\rho T}{\tilde{m}}K_\eta = -T\left(\frac{\partial p}{\partial V}\right)_\eta,\end{aligned} \tag{42}$$

where $V = \tilde{m}v$ is the volume of the solid. For self-averaging quantities,

$$\overline{\delta\Delta_V\rangle\langle\delta\Delta_V} = \frac{T}{2\mu_T V}\left(\Pi_+ - \frac{1}{3}E\rangle\langle E\right),$$

$$\overline{\tau_V\rangle\langle\tau_V} = 2\frac{T\mu_\eta}{\rho V}\left(\Pi_+ - \frac{1}{3}E\rangle\langle E\right).$$

The formulas obtained in this section are similar to those used in linear elasticity theory. This is not surprising, since the point corresponding to the unloaded state lies on the hydrostatic axis and the nonlinear theory of natural strain is intentionally constructed to be as similar in form to linear elasticity theory as possible. (However, one should bear in mind the meaning of the quantities contained in these formulas.) The fact that expressions (38)–(40) are valid for arbitrary strain can be used, for example, in analyzing the thermodynamic stability of rubber.

5. BEHAVIOR OF THE GIBBS FREE ENERGY IN THE NEIGHBORHOOD OF THE HYDROSTATIC AXIS AND UNDER LOW-STRESS CONDITIONS

5.1. Denote the value of $\varphi(\theta)$ on the hydrostatic axis (at $\theta = -pE$) by $\varphi(p)$. By virtue of (7'), the first term in the power series expansion of φ as a function of τ (at a constant p) is

$$-\langle \kappa_p(\theta + pE)\rangle = -\frac{1}{3\rho_p}\langle(\theta + pE)\rangle = -\frac{1}{\rho_p}\left(\frac{\langle\theta\rangle}{3} + p\right) = -\frac{1}{\rho_p}(p - \pi) = \frac{1}{\rho_p}\langle\tau\Delta\rangle \equiv \frac{1}{\rho_p}\langle\tau\delta s\rangle. \quad (43)$$

By the definition of $\mathcal{Y}$ in (19), expression (43) can be rewritten as $\langle\tau\mathcal{C}\tau\rangle/\rho_p$ since it follows from (41) that $\mathcal{Y}^{-1} = \mathcal{C}$ on the hydrostatic axis.

By virtue of (18), the second term in the expansion, $-\langle(\theta + pE)(\partial\kappa/\partial\theta')(\theta' + pE')\rangle$, is

$$-\frac{1}{2\rho_p}\langle\tau\mathcal{M}_T^{-1}\tau\rangle = -\frac{1}{2\rho_p}\langle\tau\tilde{\mathcal{C}}_T\tau\rangle \approx -\frac{1}{2\rho_p}\langle\tau\mathcal{C}_T\tau\rangle.$$

Therefore, the final result is

$$\varphi(\theta) = \varphi(p) + \frac{1}{2\rho_p}\langle\tau\mathcal{C}\tau\rangle. \quad (44)$$

According to (33'), this expression reduces to

$$\varphi(\theta) = \varphi(p) + \frac{\langle\tau^2\rangle}{4\mu(p)\rho_p}$$

for an isotropic solid.

5.2. For small p, the power series expansion of $\varphi(p)$ combined with

$$\frac{\partial\varphi(p)}{\partial p} = \frac{1}{\rho}, \quad \frac{\partial I}{\partial p} = -\frac{1}{K},$$

yields

$$\varphi = \varphi_0 + \frac{p}{\rho_0} + \frac{\langle\tau^2\rangle}{4\mu_0\rho_0} - \frac{p^2}{2\rho_0 K_0}.$$

The thermodynamic state variables F and Φ characterizing an elastically deformed solid were defined in [7] per unit volume of the initial (undeformed) solid. The functions f and φ are defined here per unit mass, Φ and F should be compared to $\rho_0\varphi$ and $\rho_0 f$, respectively:

$$\rho_0(\varphi - f) = \rho_0\frac{p}{\rho} = pe^I \approx p + pI_s \approx p - \frac{p^2}{K_0},$$

$$I_s = -\frac{p}{K_0} + O(p^2) + O(\langle\tau^2\rangle)$$

(see [1]). In the present notation, a similar difference considered in [7] is written as

$$\Phi - F = -\langle\sigma u\rangle = -\frac{\langle\tau^2\rangle}{2\mu_0} - \frac{p^2}{K_0},$$

where $u \approx s$ is the strain tensor in linear theory. The power series expansions of $\rho_0 f$ and F in terms of small strain increments are obviously identical to second order inclusive. The inequality of the differences written out above is not unexpected, because a nonzero term proportional to p implies that all higher order terms in the expansion of φ in powers of the strain tensor have different values. Therefore, Φ cannot be interpreted as the Gibbs free energy (in particular, it contains no term linear in p, as noted in [7]). However, it was never claimed in [7] that Φ reaches a minimum at equilibrium (i.e., is the Gibbs free energy). Therefore, the results obtained in studies relying on the assumption that Φ and its modifications include corrections that are linear in the strain tensor should be revised.

5.3. It was postulated in [10] that the chemical potential per unit mass φ_A is a function of σ whose derivative defines an "extensive" tensor V_A:

$$\frac{\partial\varphi_A}{\partial\sigma} = -V_A.$$

Since Φ has a local minimum (see Section 3), it can be hoped that $\varphi_A \equiv \varphi$ by virtue of (4). As a result, V_A can be calculated as

$$V_A = \kappa\mathcal{D}_\sigma^\theta = \frac{1}{\rho}\left(\frac{E}{3} - \mathcal{Y}^{T-1}\tau\right).$$

Thus, it is clear that φ_A cannot be derived from any well-defined thermodynamic variable by a Legendre transform, because the transform cannot depend on any elastic property of a particular material (such as the modulus $\mathcal{Y}$). One would not expect otherwise, since σ can be used as a thermodynamic variable only in the linear approximation and must be replaced with the tensor θ in an exact theory.

6. PHASE EQUILIBRIUM IN A NONUNIFORM STRESS FIELD

A phase transformation caused by varying pressure does not take place simultaneously at every point of a pressure chamber, because the stress distribution over the chamber is nonuniform. Therefore, there can exist an equilibrium boundary between phases in a nonuniform stress field. Let us derive a condition for thermodynamic equilibrium on such a boundary, using the maximum entropy principle for an isolated system consisting of subsystem 0 (called tensothermostat by analogy with the hydrostatic case) and subsystem 1 + 2 comprising the two phases of the material under study. The phases can exchange material particles, as well as

energy and strain. All boundaries between subsystems transfer the total stress, whereas the outer boundary is supposed to be rigid.

In terms of an equilibrium $\eta(\varepsilon, \kappa)$, the first law of thermodynamics is written as

$$\delta\eta(\varepsilon, \kappa) = \frac{1}{T}\delta\varepsilon - \frac{1}{T}\langle\theta\delta\kappa\rangle. \quad (45)$$

In [11], a thermodynamic condition for phase equilibrium based on the maximum entropy principle was considered for an isolated subsystem 1 + 2 with a fixed outer boundary. Reasons for revisiting this problem are discussed at the end of this section.

Suppose that the virtual transition of a small amount of a material from phase 1 to phase 2 results in the displacement of the phase boundary from a locus Γ to a locus Γ', while the outer boundary Γ_0 with the tensostat shifts to a locus Γ_0'. To be specific, assume that the mass of the denser phase 2 increases. In the figure, $\tilde{\Gamma}$ and $\tilde{\Gamma}'$ are the loci of the material boundaries Γ and Γ' after and before the transformation, respectively. In the domains $\tilde{V}_1$ (confined between $\tilde{\Gamma}'$ and Γ_0) and $\tilde{V}_2$ (bounded by $\tilde{\Gamma}$), the states of phases 1 and 2, respectively, undergo purely elastic changes.

The equilibrium entropy of subsystem *i* is expressed in terms of its density as

$$H_i = \int_{V_i} \rho_i dV \eta_i. \quad (46)$$

For the same subsystem subjected to an additional elastic deformation without any change in its mass,

$$H_i' = \int_{V_i'} \rho_i' dV' \eta_i'(\mathbf{r}') = \int_{V_i} \rho_i dV \eta_i'(\mathbf{r} + \mathbf{u}').$$

After a portion of mass changes from phase 1 into phase 2, the corresponding expressions are written for the respective phases as follows:

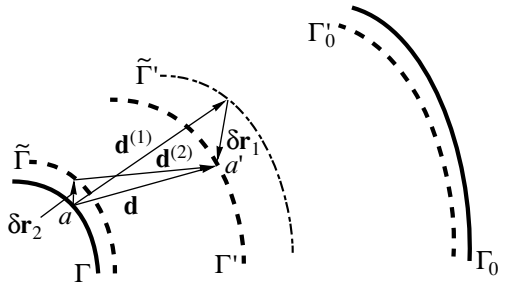
$$H_1 = \int_{V_1} \rho_1 dV \eta_1 = \int_{\tilde{V}_1} \rho_1 dV \eta_1 + \int_{V_1\setminus\tilde{V}_1} \rho_1 dV \eta_1,$$

$$H_1' = \int_{V_1'} \rho_1' dV \eta_1' = \int_{\tilde{V}_1} \rho_1 dV \eta_1',$$

$$H_2 = \int_{V_2} \rho_2 dV \eta_2, \quad (47)$$

$$H_2' = \int_{V_2'} \rho_2' dV' \eta_2' = \int_{\tilde{V}_2} \rho_2' dV' \eta_2'$$

$$+ \int_{V_2'\setminus\tilde{V}_2} \rho_2' dV' \eta_2' = \int_{V_2} \rho_2 dV \eta_2' + \int_{V_2'\setminus\tilde{V}_2} \rho_2' dV' \eta_2'.$$



Changes in phase boundaries due to a phase transition.

Denoting by $\delta\eta$ the corresponding change in entropy per unit mass at a material point, $\eta_1'(\mathbf{r} + \mathbf{u}) - \eta_1(\mathbf{r})$, we have

$$\delta H_1 = H_1' - H_1 = \int_{\tilde{V}_1'} \rho_1 dV \delta\eta_1 - \int_{V_1\setminus\tilde{V}_1'} \rho_1 dV \eta_1, \quad (48)$$

$$\delta H_2 = \int_{V_2} \rho_2 dV \delta\eta_2 + \int_{V_2'\setminus\tilde{V}_2} \rho_2' dV' \eta_2'. \quad (49)$$

By using (45), δH_1 is rewritten as

$$\delta H_1 = \int_{\tilde{V}_1} \rho_1 dV \left(\frac{\delta\varepsilon_1}{T_1} - \frac{\langle\theta_1\delta\kappa_1\rangle}{T_1}\right) - \int_{V_1\setminus\tilde{V}_1} \rho_1 dV \eta_1.$$

Since the movement of the equilibrium boundary between phases is sufficiently slow for temperature to remain equal at all points of subsystem 1, the equation

$$\delta\mathscr{E}_1 = \int_{\tilde{V}_1} \rho_1 dV \delta\varepsilon_1 - \int_{V_1\setminus\tilde{V}_1} \rho_1 dV \varepsilon_1, \quad (50)$$

analogous to (48), can be used to obtain

$$\delta H_1 = \frac{\delta\mathscr{E}_1}{T_1} - \int_{\tilde{V}_1} \rho_1 dV \frac{\langle\theta_1\delta\kappa_1\rangle}{T_1} + \int_{V_1\setminus\tilde{V}_1} \rho_1 dV \left(\frac{\varepsilon_1}{T_1} - \eta_1\right).$$

Similar expressions are obtained for δH_2 and δH_0. Finally,

$$\delta H = \frac{\delta\mathscr{E}_1}{T_1} + \frac{\delta\mathscr{E}_2}{T_2} + \frac{\delta\mathscr{E}_0}{T_0} - \int_{\tilde{V}_1} \rho_1 dV \frac{\langle\theta_1\delta\kappa_1\rangle}{T_1}$$

$$- \int_{V_2} \rho_2 dV \frac{\langle\theta_2\delta\kappa_2\rangle}{T_2} - \int_{V_0} \rho_0 dV \frac{\langle\theta_0\delta\kappa_0\rangle}{T_0} \quad (51)$$

$$+ \int_{V_1\setminus\tilde{V}_1} \rho_1 dV \left(\frac{\varepsilon_1}{T_1} - \eta_1\right) - \int_{V_2'\setminus\tilde{V}_2} \rho_2' dV' \left(\frac{\varepsilon_2'}{T_2} - \eta_2'\right).$$

By substituting the expression for $\delta\mathscr{E}_0$ derived from the

energy balance equation

$$\delta\mathscr{E}_0 + \delta\mathscr{E}_1 + \delta\mathscr{E}_2 = 0,$$

it is verified that $T_0 = T_1 = T_2 = T$. Therefore,

$$T\delta H = -\int_{\tilde{V}_1}\rho_1 dV\langle\theta_1\delta\kappa_1\rangle - \int_{V_2}\rho_2 dV\langle\theta_2\delta\kappa_2\rangle - \int_{V_0}\rho_0 dV\langle\theta_0\delta\kappa_0\rangle + \int_{V_1\setminus\tilde{V}_1}\rho_1 dV f_1 - \int_{V_2'\setminus\tilde{V}_2}\rho_2' dV' f_2', \quad (51')$$

where $f_i = \varepsilon_i - T\eta_i$.

Here, the first three terms can be represented as follows:

$$-\int_{\tilde{V}_1}\rho_1 dV\langle\theta_1\delta\kappa_1\rangle - \int_{V_2}\rho_2 dV\langle\theta_2\delta\kappa_2\rangle - \int_{V_0}\rho_0 dV\langle\theta_0\delta\kappa_0\rangle$$
$$= -\int_{\tilde{V}_1} dV\langle\sigma_1\delta s_1\rangle - \int_{V_2} dV\langle\sigma_2\delta s_2\rangle - \int_{V_0} dV\langle\sigma_0\delta s_0\rangle$$
$$= \oint_{\tilde{\Gamma}_1} dS_i\sigma_{ij}^{(1)}\delta x_j^{(1)} - \oint_{\Gamma_0} dS_i\sigma_{ij}^{(1)}\delta x_j^{(1)} - \oint_{\Gamma} dS_i\sigma_{ij}^{(2)}\delta x_j^{(2)} + \oint_{\Gamma_0} dS_i\sigma_{ij}^{(0)}\delta x_j^{(0)}.$$

Since $\delta x_j^{(1)} = \delta x_j^{(0)}$ on Γ_0, the second and fourth terms cancel out. For a displacement d of the boundary at point 1 in phase 1,

$$\oint_{\tilde{\Gamma}_1} dS_i\sigma_{ij}^{(1)}\delta x_j^{(1)} - \oint_{\Gamma} dS_i\sigma_{ij}^{(2)}\delta x_j^{(2)} \approx \oint_{\Gamma} dS_i\sigma_{ij}^{(1)}(\delta x_j^{(1)} - \delta x_j^{(2)})$$

to the first order in d.

Let the vector $\mathbf{d}$ be pointing from a point a on Γ to an arbitrary point a' on Γ'. Consider adjacent points 1 and 2 in the vicinity of a in phases 1 and 2, respectively, and analogous points 1' and 2' in the vicinity of a'. For a coherent nucleus (the only case amenable to thermodynamic analysis), two adjacent material points belonging to the same phase remain adjacent after they are separated by a phase boundary. Suppose that points 1' and 2 are the locations of material points $\tilde{1}$ and $\tilde{2}$ after their respective elastic displacements by $\delta\mathbf{r}_1$ and $\delta\mathbf{r}_2$ due to an increase in the mass of phase 2. Denoting the respective inelastic displacements by $\mathbf{d}_1$ and $\mathbf{d}_2$, we have

$$\mathbf{d} = \mathbf{d}^{(1)} + \delta\mathbf{r}_1 = \mathbf{d}^{(2)} + \delta\mathbf{r}_2. \quad (52)$$

If the boundary can be represented by an analytic function, then

$$\int_{V_1\setminus\tilde{V}_1}\rho_1 dV f_1 \approx \oint_{\Gamma} dS_i d_i^{(1)}\rho_1 f_1$$

and

$$\int_{V_2'\setminus\tilde{V}_2}\rho_2' dV' f_2' \approx \oint_{\Gamma} dS_i d_i^{(2)}\rho_2 f_2.$$

Since $\sigma_{ij}^{(1)}n_i = \sigma_{ij}^{(0)}n_i$ on Γ_0, expression (51') can be rewritten as

$$T\delta H = \oint_{\Gamma} dS_i(\sigma_{ij}^{(1)}\delta x_j^{(1)} - \sigma_{ij}^{(2)}\delta x_j^{(2)} + d_i^{(1)}\rho_1 f_1 - d_i^{(2)}\rho_2 f_2). \quad (51'')$$

Under the no-slip condition on the boundary, the invariance of the area dS_i in the volume element $dV = dS_i d_i$ and the mass-conservation condition for a mass element undergoing the transition imply that

$$\rho_1\mathbf{n}\cdot\mathbf{d}^{(1)} = \rho_2\mathbf{n}\cdot\mathbf{d}^{(2)}, \quad \mathbf{d}_\parallel^{(1)} = \mathbf{d}_\parallel^{(2)}. \quad (52')$$

Since mechanical equilibrium on the phase boundary implies that $\sigma_{ij}^{(1)}n_i = \sigma_{ij}^{(2)}n_i = P_j$, relations (52) and (52') can be used to rewrite (51'') as

$$T\delta H = \oint_{\Gamma} dS_i d_i^{(2)}\rho_2\left\{f_1 - \frac{P_n}{\rho_1} - \left(f_2 - \frac{P_n}{\rho_2}\right)\right\}.$$

The inelastic component $\mathbf{d}^{(2)}$ of the total displacement $\mathbf{d}$ in phase 2 is arbitrary, and the equilibrium condition $\delta H = 0$ implies that the integrand is zero, i.e.,

$$f_1 - \frac{P_n}{\rho_1} = f_2 - \frac{P_n}{\rho_2}, \quad (53)$$

where $P_n = \mathbf{n}\cdot\mathbf{P} = n_i\sigma_{ij}^{(1)}n_j = n_i\sigma_{ij}^{(2)}n_j$. Since the difference of f_1 of f_2 can be large under high pressure, it is reasonable to single out the isotropic part in σ_i and rewrite (53) as

$$\varphi_1 - \frac{n_i\tau_{ij}^{(1)}n_j}{\rho_1} = \varphi_2 - \frac{n_i\tau_{ij}^{(2)}n_i}{\rho_2}. \quad (53')$$

A similar thermodynamic equilibrium boundary condition was obtained in [11]. However, the results presented in [3, 5] can be used to show that Eq. (10a) in [11] is not equivalent to (53'). This may be explained by the insufficiently rigorous definition of the chemical

potential of an elastically deformed solid in [11]. Moreover, distortion was used as a measure of deformation in [11]. For this reason, all results obtained in that study are not symmetric and are therefore difficult to apply, in particular, in deriving formulas corresponding to linear approximation.

The meanings of the quantities used in this paper imply that the Gibbs free energies of both phases are measured relative to the Helmholtz free energy of a deformed solid. The Helmholtz free energy is measured relative to the unloaded state of the solid (which may not be observable), and the change in strain is determined by the equations of state for the phases subject to matching conditions on the boundary (Eqs. (52') for virtual displacements) and continuity conditions for P_i.

According to Eq. (53'), the Gibbs free energy $\Phi = \Phi_1 + \Phi_2$ does not remain constant in the course of a phase transition, in contrast to the hydrostatic case. More precisely, the equilibrium value of Φ changes by the following increment as the boundary shifts by a vector **d**:

$$\delta\Phi = \int\limits_{\tilde V_1} \rho_1 dV \delta\varphi_1 + \int\limits_{V_2} \rho_2 dV \delta\varphi_2$$

$$+ \oint\limits_\Gamma dS_i d_i^{(1)} \rho_1 (\varphi_2 - \varphi_1) = \int\limits_{\tilde V_1} \rho_1 dV \delta\varphi_1 \qquad (54)$$

$$+ \int\limits_{V_2} \rho_2 dV \delta\varphi_2 + \oint\limits_\Gamma dS_i [d_i^{(2)} n_i \tau_{ij}^{(2)} n_j - d_i^{(1)} n_i \tau_{ij}^{(1)} n_j].$$

Let us rearrange this expression to make it more intelligible. The first two summands are associated with the change in the elastic strain field due to displacement of the phase boundary. The first summand can be rearranged as follows:

$$\delta\Phi_{1\mathrm{el}} = \int\limits_{\tilde V_1} \rho_1 dV \delta\varphi_1 = \int\limits_{\tilde V_1} \rho_1 dV \left[\delta f_1 + \delta\left(\frac{p_1}{\rho_1}\right)\right]$$

$$= \int\limits_{\tilde V_1} dV (\langle \sigma^{(1)} \delta s_1 \rangle + p_1 \delta \langle s_1 \rangle + \delta p_1)$$

$$= \int\limits_{\tilde V_1} dV (\langle \tau^{(1)} \delta s_1 \rangle + \delta p_1) = \oint\limits_{\Gamma_0} dS_i \tau_{ij}^{(1)} \delta x_j$$

$$- \oint\limits_\Gamma dS_i \tau_{ij}^{(1)} \delta x_j + \int\limits_{\tilde V_1} dV \left(\delta p_1 - \frac{\partial p_1}{\partial x_j} \delta x_j\right).$$

Adding an analogous expression for $\delta\Phi_{2\mathrm{el}}$ and using the relations $(\tau^{(2)} - \tau^{(1)})\mathbf{n} = (\sigma^{(2)} - \sigma^{(1)})\mathbf{n} + (p_2 - p_1)\mathbf{n}$, $\mathbf{n} \cdot \mathbf{d}_1 + \mathbf{n} \cdot \delta\mathbf{r}_1 = \mathbf{n} \cdot \mathbf{d}_2 + \mathbf{n} \cdot \delta\mathbf{r}_2$, and $\delta\mathbf{r}_\parallel^{(1)} = \delta\mathbf{r}_\parallel^{(2)}$ (see (52)) leads to

$$\delta\Phi = \oint\limits_\Gamma dS_i d_i (p_2 - p_1) + \oint\limits_{\Gamma_0} dS_i \tau_{ij}^{(1)} \delta x_j + \int\limits_{\tilde V} dV \bar\delta p$$

$$= \oint\limits_\Gamma dS_i d_i (p_2 - p_1) + \delta R + \oint\limits_{\Gamma_0} dS_i p_1 \delta x_i + \int\limits_{\tilde V} dV \bar\delta p, \qquad (54')$$

where δR is the work done by external forces, the third summand is minus the work done by external pressure (note that p_1 is not equal to external pressure), and $\bar\delta p = \delta p - (\partial p/\partial x_j)\delta x_j$ is the variation of p at a certain point in space (not material point).

If the temperature of the entire system is assumed to be constant from the outset, then it can easily be shown that the resulting equilibrium condition can be derived from the following relation (see (25')):

$$\delta F - \oint\limits_\Gamma dS_i \sigma_{ij}^{\mathrm{e}} \delta x_j = 0.$$

7. CONCLUSIONS

1. Legendre conjugate measures of deformation and stress are introduced and used to define a thermodynamic potential of an elastically deformed solid. Since the thermodynamic potential is minimized in a homogeneous state at constant values of temperature and forces applied to a solid in equilibrium, it is interpreted as the Gibbs free energy. It is shown that the quadratic terms of its expansion in terms of the stress tensor known from the linear theory change when the potential contains a term proportional to p.

2. In contrast to the hydrostatic case, the Gibbs free energy of a multiphase system changes with its phase state, and the change is not entirely due to change in the elastic equilibrium field.

3. The most remarkable result is the similarity of the expression for the Gibbs free energy density corresponding to an arbitrary elastic strain to its counterpart in hydrostatics.

4. Correlation functions are calculated for strain and stress fluctuations in an arbitrarily deformed solid.

5. A thermodynamic equilibrium boundary condition is found for coexisting phases.

REFERENCES

1. A. P. Kochkin, Indian J. Pure Appl. Math. **17**, 564 (1986).
2. A. P. Kochkin, Zh. Éksp. Teor. Fiz. **109**, 1823 (1996) [JETP **82**, 983 (1996)].

3. A. P. Kochkin, Zh. Éksp. Teor. Fiz. **117**, 723 (2000) [JETP **90**, 628 (2000)].

4. H. Hencky, Z. Tech. Phys. **9**, 215 (1928).

5. A. P. Kochkin, Available from VINITI, No. 2120-V99 (1999).

6. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics*, 3rd ed. (Nauka, Moscow, 1976; Pergamon, Oxford, 1980), Part 1.

7. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 7: *Theory of Elasticity*, 4th ed. (Pergamon, Oxford, 1986; Nauka, Moscow, 1987).

8. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics*, 3rd ed. (Nauka, Moscow, 1986; Pergamon, Oxford, 1987).

9. R. Kubo, *Statistical Mechanics: An Advanced Course with Problems and Solutions* (North-Holland, Amsterdam, 1965; Mir, Moscow, 1967).

10. R. G. Arkhipov, Zh. Éksp. Teor. Fiz. **92**, 1021 (1987) [Sov. Phys. JETP **65**, 579 (1987)].

11. M. A. Grinfel'd, Dokl. Akad. Nauk SSSR **251**, 824 (1980).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 133–146.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 155–169.

SOLIDS
Structure

Generation of Acoustic Solitons by a Single-Mode Electromagnetic Field

A. A. Zabolotskii

Institute of Automation and Electrometry, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia

e-mail: zabolotskii@iae.nsk.su

Received November 17, 2003

Abstract—Mechanisms of acoustic pulse generation by a single-mode electromagnetic field propagating in a photoelastic material are analyzed. The anisotropy induced by acoustic excitations in an isotropic medium leads to nonlinear coupling between the polarization components of a single-mode electromagnetic field. For different conditions, it is shown that the acoustic–electromagnetic wave interaction due to mixing of the polarization components of light and acoustic waves can give rise to soliton-like coherent acoustic excitations in a thin crystal plate. When spatial dispersion is ignored, the governing system of equations for unidirectional acoustic solitons can be reduced to an integrable model. It is shown that qualitatively different scenarios of formation of acoustic solitons are possible, depending on the directions of deformation and field polarization. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

When a single-mode electromagnetic field propagates in an isotropic medium, coherent soliton-like and other excitations can develop as a result of balance between group-velocity dispersion and Kerr nonlinearity or parametric interaction between several field modes with different carrier frequencies (e.g., see [1, 2]). Of particular interest are systems in which dynamic interaction regimes do not lead to any steady state under time-independent external conditions. The interactions associated with evolution of quasi-monochromatic optical pulses have been analyzed in detailed studies, which in many cases relied on specific choices of time scales and resonance conditions [3]. This has led to good understanding of the physical patterns of fundamental nonlinear interactions between solitary waves of different nature. Analogous analytical results concerning single-phase solutions were obtained in studies of acoustic–electromagnetic wave interactions governed by systems of equations related to some of those employed in nonlinear optics [4–6].

Propagation of waves of moderate amplitude is dominated by quadratic nonlinearity, which leads to generation of solitons via either parametric interaction of two fundamental waves with a wave with difference carrier frequency or (in the degenerate case) second harmonic generation [2]. Resonant three-wave mixing in solids have been analyzed in detailed studies [2]. If effects due to acoustic–electromagnetic wave interactions in an elastic crystalline material are strong, then the reverse effect of elastic strain on the dielectric constant must be taken into account. Dependence of the dielectric constant on strain leads to a nonlinear coupling between polarization components described by equations analogous to those of the slowly varying envelope approximation for three- or four-wave mixing of optical pulses. Nonlinear acoustic–electromagnetic wave interactions of this kind in solids have been analyzed in numerous studies (e.g., see [4–6]). It was shown in [4, 5] that the interaction between orthogonally polarized components of the ordinary and extraordinary waves in a uniaxial crystal can result in the generation of a nonpiezoactive longitudinal acoustic wave via resonant parametric interaction. The simplest one-soliton solutions were found in those studies by solving appropriate evolution equations derived in the slowly varying envelope approximation.

Currently, conditions for subnanosecond and picosecond pulse generation and evolution are the subject of special interest. The length of a pulse of duration on the order of 1 ps propagating in a solid with a velocity of about 5×10^5 cm/s is only a few interatomic distances. Therefore, pulses of this kind can be used to explore the structure of various crystalline materials, complex molecular structures [7], and quantum wells in semiconductor films [8]. Acoustic pulse evolution was analyzed by Sazonov with coauthors (e.g., in [9–11]) and the present author (e.g., in [12, 13]) under light–matter interaction conditions different from those considered in this study without applying the slowly varying envelope approximation. Studies of such pulses are of great current interest, as well as the ongoing studies of their optical counterparts, femtosecond pulses and micropulses (see [14]). Even though the models that describe the evolution of acoustic and optical pulses without using the slowly varying envelope approximation bear some resemblance, they are substantially dif-

ferent in certain respects (cf. the models developed in [14] and [12, 13]).

Intensive studies of acoustic–electromagnetic wave interactions involving nanosecond acoustic pulse generation and polarization effects are being conducted for anisotropic crystals characterized by nonlinear moduli strongly depending on the directions and polarizations of the interacting waves. As an example, consider the interaction between parallel acoustic and electromagnetic waves in the optically anisotropic $LiNbO_3$ crystal. Suppose that the ordinary and extraordinary waves propagating along the x axis interact with an acoustic wave whose frequency is nearly equal to the difference of the electromagnetic-wave frequencies. This interaction (due to linear photoelasticity) can lead to propagation of solitons [15]. Localization of the bound optical component of acoustic–electromagnetic solitons was observed experimentally in [16]. In the theoretical analyses of soliton solutions to parametric-interaction problems presented in [4–6], both electromagnetic and acoustic fields were described in quasi-monochromatic approximation. However, the frequency difference between the extraordinary and ordinary waves imposes a restrictive condition on the duration of the generated acoustic pulses. For example, the difference frequency is about 10^8 s^{-1} for $BaTiO_3$, which is characterized by $3m$ symmetry at temperatures below –90°C [5]. A higher difference frequency, about 4×10^9 s^{-1}, is attained for $LiNbO_3$. This implies that the durations of quasi-monochromatic solitons must be at least 10^{-7} and 10^{-8} s, respectively. Generation of acoustic pulses with duration comparable to (or even shorter than) that corresponding to the difference frequency would make it possible to implement nanosecond (or even shorter) pulses with this mechanism of acoustic–electromagnetic wave interaction. Furthermore, the attenuation factor for an acoustic pulse is proportional to the squared frequency in the range between 10^8 and 10^{10} s^{-1} [5]. Therefore, the attenuation factor is more than two orders of magnitude larger for quasi-monochromatic pulses as compared to pulses with durations of about Ω^{-1}, which may include picosecond pulses. Generation of picosecond acoustic solitons in a crystal plate was observed in recent experiments [17]. Solitons formed over a distance of about one millimeter as a result of balance between dispersion due to location of atoms in the crystalline lattice and nonlinearity caused by anharmonicity of the interatomic interaction potential. Analysis of pulse generation caused by acoustic–electromagnetic wave interaction offers additional opportunities and mechanisms for using electromagnetic field to control acoustic-pulse parameters.

The theoretical results obtained in the studies of acoustic–electromagnetic wave interactions mentioned above cannot be used to analyze propagation of subnanosecond—and, particularly, picosecond—acoustic pulses in crystals, because the slowly varying envelope approximation is not applicable to such pulses. Therefore, the development of a theory of acoustic–electromagnetic wave interactions (in particular, at the difference frequency) that does not rely on this approximation is of great interest and practical importance. Theories describing the evolution of pulses with durations comparable to $\pi\Omega_{12}^{-1}$ do not make use of frequency resonance conditions. Therefore, new effective mechanisms and conditions of acoustic soliton generation can be suggested in these theories. Note also that the use of short light pulses for generating acoustic solitons in multimode parametric interactions is impeded by technical difficulties due, in particular, to the requirement of pulse timing [6].

In this paper, new mechanisms of acoustic pulse (soliton) generation in photoelastic materials are analyzed without invoking the slowly varying envelope approximation for acoustic pulses. The present analysis is restricted to the case of a single-mode field and a homogeneous medium.

The paper is organized as follows. In the next section, the system of evolution equations is derived for an electromagnetic wave and acoustic excitations. In Section 3, single-phase solutions are presented. In Section 4, an integrable reduced system is derived, the corresponding Lax representations are found and a self-similar nonsoliton solution is obtained. The results obtained in this study are discussed in Section 5. In the Appendix, N-soliton solutions to the integrable reduced system are found.

2. EVOLUTION OF A SINGLE-MODE FIELD IN A TWO-DIMENSIONAL MEDIUM

Consider an isotropic crystal that has large dimensions along the x and y axes and is optically thin along the z axis. Weak deformation of an initially undeformed crystal is described in linear approximation with respect to the strain tensor

$$u_{jk} = \frac{1}{2}\left(\frac{\partial U_j}{\partial x_k} + \frac{\partial U_k}{\partial x_j}\right), \quad (1)$$

where $\mathbf{U}(x, y, z)$ is the vector of displacement of a material point. Accordingly, the dielectric tensor has the form (e.g., see [18])

$$\varepsilon_{jk} = \varepsilon_0\delta_{jk} + a_1u_{jk} + a_2u_{jj}\delta_{jk}, \quad (2)$$

where δ_{jk} is Kronecker's delta and a_1 and a_2 are real constants. When nonlinear effects are neglected, the dynamics of acoustic field are described by the Hamiltonian

$$H_a = \int\left(\frac{1}{2n_0}\sum_j p_j^2 + \frac{1}{2}\sum_{j,k,l,m}\lambda_{jklm}\frac{\partial U_j}{\partial x_k}\frac{\partial U_l}{\partial x_m}\right)d\mathbf{r}, \quad (3)$$

where n_0 is the mean density of the crystal, p_j is a component of the momentum density associated with time-dependent displacement ($j = x, y, z$), and λ_{jklm} is the tensor of elastic moduli [18]. The integral in (3) is over the crystal volume. It is assumed that both photon and phonon concentrations are large and classical models of both acoustic and electromagnetic fields are valid.

Suppose that the electromagnetic wave vector $\mathbf{k} = (k_x, k_y, k_z)$ makes an angle θ_e with the z axis (see below). Since the crystal plate is optically thin along the z axis, both the derivative of the electric-field envelope and the acoustic waves associated with U_z are negligible. The coordinate system (X, Y, Z) associated with light propagation is defined by the Euler angle θ_e between the z and Z axes and the Euler angle ϕ_e between the x axis and the projection of the Z axis on the xy plane [19, 20]. The corresponding dielectric tensor $\hat{\varepsilon}'$ is related to the dielectric tensor $\hat{\varepsilon}$ in the case of $\theta_e = \phi_e = 0$ as follows:

$$\hat{\varepsilon}' = \hat{D}^{-1}\hat{\varepsilon}\hat{D},$$

where $\hat{D}(\theta_e, \phi_e)$ is a rotation matrix.

Electromagnetic field in the crystal is governed by the Maxwell equations

$$\nabla \times (\nabla \times \tilde{\mathbf{E}}) = -\frac{1}{c^2}\frac{\partial^2 \mathbf{D}}{\partial t^2}, \tag{4}$$

$$(\nabla \cdot \mathbf{D}) = 0, \tag{5}$$

where $\tilde{\mathbf{E}} = (\tilde{E}_x, \tilde{E}_y, \tilde{E}_z)$, $\mathbf{D} = \hat{\varepsilon}\tilde{\mathbf{E}}$ is the polarization vector, and c is the speed of light in the crystal. In the absence of acoustic excitations, $\hat{\varepsilon}$ is a diagonal tensor. Nonlinear dependence of $\hat{\varepsilon}$ on $\tilde{\mathbf{E}}$ is not taken into account.

If

$$\tilde{\mathbf{E}}(x, y, z, t) = \exp(iKz)\mathbf{E}(x, y, t),$$

then E_z can be eliminated by combining (4) with the identity

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - (\Delta_\perp - K^2)\mathbf{E}, \tag{6}$$

where

$$\Delta_\perp = \nabla_\perp \cdot \nabla_\perp = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}, \quad \nabla_\perp = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right).$$

The resulting equation for the transverse field $\mathbf{E}_\perp = (E_x, E_y)$ is

$$\nabla_\perp^2 \mathbf{E}_\perp - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}(\hat{\varepsilon}\mathbf{E})_\perp - K^2\mathbf{E}_\perp - \nabla_\perp(\nabla_\perp \mathbf{E}_\perp + iKE_z) = 0. \tag{7}$$

The longitudinal field component is expressed in terms of the transverse field by using Eq. (5):

$$iK(\hat{\varepsilon}\mathbf{E})_z + \nabla_\perp \cdot (\hat{\varepsilon}\mathbf{E})_\perp = 0. \tag{8}$$

Since the unperturbed medium is assumed to be isotropic, it follows from (2) that the differences between the components of $\hat{\varepsilon}$ for a perturbed medium are on the order of $a_n u_{jk}$ ($n = 1, 2$; $j = x, y, z$). The angle between $\mathbf{k}$ and the z axis is assumed to be small, but sufficiently large for conditions (28) to be satisfied (see below).

By analogy with [19], the longitudinal electric field component is estimated as follows. In the zeroth approximation, the expansion of (8) in terms of $1/Kl_s$, where l_s is the optical-soliton length, yields

$$E_z = -\varepsilon_0^{-1} a_1 u_{xx} \sin\theta_0 \cos\phi_0 E_x \approx -\varepsilon_0^{-1} a_1 u_{xx} \theta_0 E_x$$

if the off-diagonal components of $\hat{\varepsilon}$ are ignored. It is assumed that $\theta_0, \phi_0 \sim \epsilon$ and $a_1 u_{xx} \sim \epsilon$ (ϵ is a small quantity). Henceforth, the contribution of the longitudinal electric-field component is of order ϵ^2 and other contributions of similar order (including the off-diagonal components of $\hat{\varepsilon}$) are neglected.

Since the processes analyzed in this study develop on a (sub)picosecond time scale, the slowly varying envelope approximation is applied to the electromagnetic field polarization components:

$$\mathbf{E}_\perp = \exp(ik_x x + ik_y y - i\omega t)\mathcal{E}(x, y, t),$$

where $\partial_t \mathcal{E} \ll w\mathcal{E}$. Similarly, the time scale of acoustic perturbations u_{xx}, u_{yy}, and u_{xy} is much larger than ω^{-1} for the optical frequency range.

The assumptions introduced above are used to derive the following system of equations from (7):

$$\left[2i(\mathbf{k}_\perp \cdot \nabla_\perp) + \Delta_\perp + i\frac{2\omega}{c^2}\frac{\partial}{\partial t}\right]\mathcal{E}_x = \frac{\omega^2}{c^2}[(q_x + a_2 u_{xx})\mathcal{E}_x + a_1 u_{xy}\mathcal{E}_y], \tag{9}$$

$$\left[2i(\mathbf{k}_\perp \cdot \nabla_\perp) + \Delta_\perp + i\frac{2\omega}{c^2}\frac{\partial}{\partial t}\right]\mathcal{E}_y = \frac{\omega^2}{c^2}[(q_y + a_2 u_{yy})\mathcal{E}_y + a_1 u_{xy}\mathcal{E}_x], \tag{10}$$

where $\mathbf{k}_\perp = (k_x, k_y)$ and

$$q_x = q_y = (K^2 + k_x^2 + k_y^2)c^2\omega^{-2} - \varepsilon_0 = 0$$

for a single-mode field. However, the case of $q_x \neq q_y$ is also of physical interest. This condition is satisfied, for example, by the envelopes E_x and E_y of orthogonally polarized waves with different carrier frequencies ω_x and ω_y, in which case the quantity

$$q_0 = q_x - q_y \propto (\omega_x - \omega_y)$$

can be both positive and negative. This case is discussed in Section 5.

The Stokes vector is defined as

$$\mathbf{S} = (S_x, S_y, S_z)$$
$$\equiv \left(\frac{\mathscr{E}_x\mathscr{E}_y^* + \mathscr{E}_x^*\mathscr{E}_y}{2}, i\frac{\mathscr{E}_x\mathscr{E}_y^* - \mathscr{E}_x^*\mathscr{E}_y}{2}, \frac{\mathscr{E}_x\mathscr{E}_x^* - \mathscr{E}_y\mathscr{E}_y^*}{2}\right).$$

Its magnitude S_0 is

$$S_0^2 = S_x^2 + S_y^2 + S_z^2 = [(|\mathscr{E}_x|^2 + |\mathscr{E}_y|^2)/2]^2. \quad (11)$$

The right-hand sides of (9) and (10) can be used to define the Hamiltonian of interaction between electric and acoustic fields as

$$H_{\text{int}} = A_1\int[g(u_{xx} - u_{yy})S_z + g(u_{xx} + u_{yy})S_0 + u_{xy}S_x]d\mathbf{r}, \quad (12)$$

where

$$A_1 = 2a_1\omega^2/c^2, \quad g = a_2/2a_1.$$

Acoustic waves are governed by the equations

$$\frac{\partial\mathbf{U}}{\partial t} = \frac{\partial H}{\partial\mathbf{P}}, \quad \frac{\partial\mathbf{P}}{\partial t} = -\frac{\partial H}{\partial\mathbf{U}}, \quad (13)$$

where $\mathbf{U} = (U_x, U_y)$ and $\mathbf{P} = (P_x, P_y)$. These equations are used to derive equations

$$\left[\nabla_\perp - \frac{1}{v_x^2}\frac{\partial^2}{\partial t^2}\right]U_x = A_1\left(\frac{\partial S_x}{\partial y} + g\frac{\partial\mathscr{E}_x\mathscr{E}_x^*}{\partial x}\right), \quad (14)$$

$$\left[\nabla_\perp - \frac{1}{v_y^2}\frac{\partial^2}{\partial t^2}\right]U_y = A_1\left(\frac{\partial S_x}{\partial x} + g\frac{\partial\mathscr{E}_y\mathscr{E}_y^*}{\partial y}\right), \quad (15)$$

for elastic-strain waves, where

$$v_x = v_y = v_0 = \sqrt{\lambda_{11}n_0^{-1}}$$

is the phase velocity of acoustic waves, which entail

$$\left[\nabla_\perp - \frac{1}{v_0^2}\frac{\partial^2}{\partial t^2}\right]u_{xx} = A_1\left(\frac{\partial^2 S_x}{\partial x\partial y} + g\frac{\partial^2\mathscr{E}_x\mathscr{E}_x^*}{\partial x^2}\right), \quad (16)$$

$$\left[\nabla_\perp - \frac{1}{v_0^2}\frac{\partial^2}{\partial t^2}\right]u_{yy} = A_1\left(\frac{\partial^2 S_x}{\partial x\partial y} + g\frac{\partial^2\mathscr{E}_y\mathscr{E}_y^*}{\partial y^2}\right), \quad (17)$$

$$\left[\nabla_\perp - \frac{1}{v_0^2}\frac{\partial^2}{\partial t^2}\right]u_{xy} = \frac{A_1}{2}\left(\nabla_\perp S_x + g\frac{\partial^2 S_0}{\partial x\partial y}\right). \quad (18)$$

When the Euler angles are small, diffraction (i.e., terms proportional to $\nabla_\perp\mathscr{E}_l$ with $l = x, y$ in Eqs. (9) and (10)) cannot be neglected. The resulting system of equations is a generalization of the two-dimensional nonlinear Schrödinger equation, which does not have stable soliton solutions (e.g., see review in [21]). Instabilities due to diffraction effects can lead to collapse, i.e., singular behavior of localized solutions. This suggests that the system of Eqs. (9), (10), and (16)–(18) does not describe stable soliton propagation.

When the Euler angles are relatively large, the terms proportional to $\nabla_\perp\mathscr{E}_l$ ($l = x, y$) in Eqs. (9) and (10) may be negligible. A numerical analysis of the resulting system (9), (10), (16)–(18) without diffraction-related terms revealed that a Gaussian-shaped acoustic pulse

$$u_{xy}(x, y, 0) = u_0\exp[-(x^2 + y^2)/x_0^2],$$

$$u_{xx}(x, y, 0) = u_{xx}(x, y, 0) = 0$$

evolves into a ring-shaped pulse. The ring diameter grows with decreasing speed, and the pulse profile becomes irregular. Figure 1 illustrates the numerically predicted evolution of an initially bell-shaped two-dimensional acoustic pulse.

3. SINGLE-PHASE SOLUTION

Since soliton-like solutions are sought in this study, let us consider physical situations in which solitons can be generated. One can find conditions such that system (9), (10), (16)–(18) reduces to a version of the two-component Schrödinger equation and admits soliton and soliton-like periodic solutions. They are obtained in the case of quasi-one-dimensional dynamics, when the second derivative along one coordinate is negligible in a corresponding reference frame, for example, if the projection of the electric-field vector on the y axis is such that

$$k_y\partial_y\mathscr{E}_\perp \gg \partial_y^2\mathscr{E}_\perp,$$

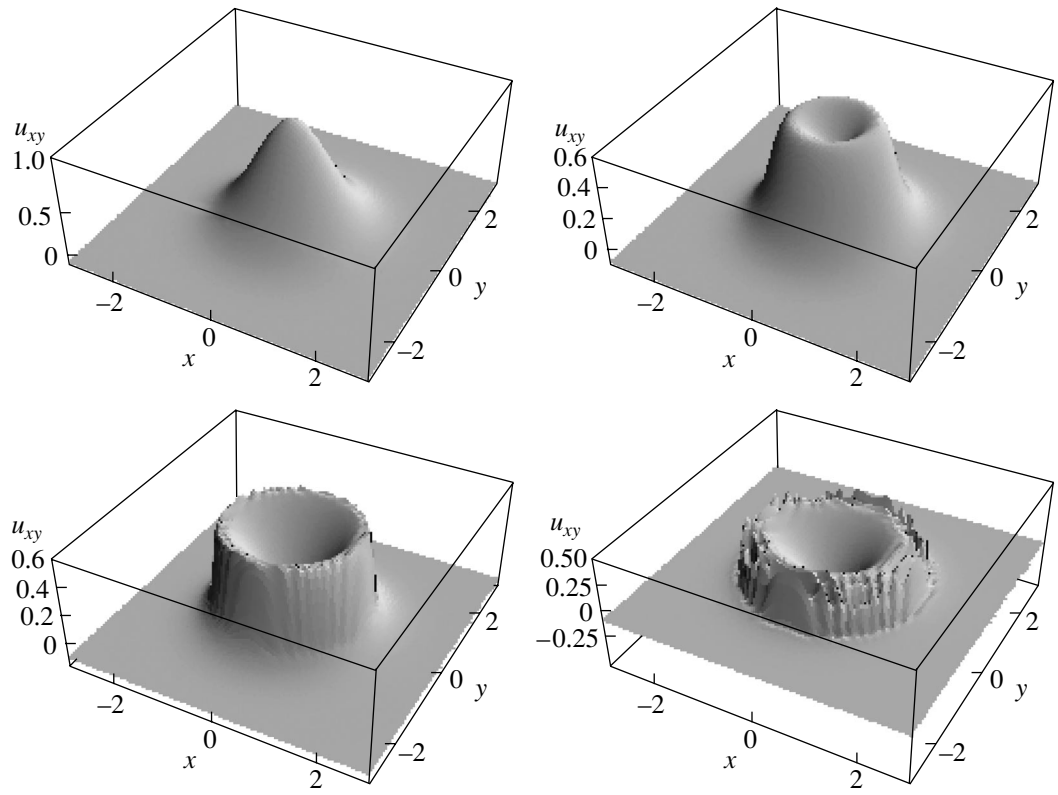


Fig. 1. Profiles of u_{xy} versus x, y for an initial pulse with $u_{xy} = 1.2\exp[-(x^2 + y^2)]$, $u_{xx} = u_{yy} = 0$, $E_x(-5, 0) = 1$, and $E_y(-5, 0) = 0$ at instants separated by equal time intervals.

whereas

$$k_x\partial_x\mathscr{E}_\perp \ll \partial_x^2\mathscr{E}_\perp.$$

Then, $\Delta = \cos^2\gamma_0\partial_x^2$ to the required accuracy in Eqs. (9) and (10). This is the case, for example, when the elliptical cross section of the light beam incident on the crystal is oblate in the x direction and stretched along the y axis. Steady-state periodic and one-soliton solutions to system (9), (10), (16)–(18) can be sought as functions of the single phase variable

$$\theta = \cos\gamma_0 x + \sin\gamma_0 y - Vt,$$

where γ_0 is the angle between the soliton trajectory in the xy plane and the x axis. System (16)–(18) yields

$$\begin{aligned} u_{xx} &= \Lambda(\cos\gamma_0\sin\gamma_0 S_x + g\cos^2\gamma_0\mathscr{E}_x\mathscr{E}_x^*), \\ u_{yy} &= \Lambda(\cos\gamma_0\sin\gamma_0 S_x + g\sin^2\gamma_0\mathscr{E}_y\mathscr{E}_y^*), \\ 2u_{xy} &= \Lambda(S_x + g\cos\gamma_0\sin\gamma_0 S_0), \end{aligned} \tag{19}$$

where

$$\Lambda = \frac{A_1 v_0^2}{v_0^2 - V^2}.$$

Here, the constants of integration are set to zero and $u_{jl} = 0$ ($j, l = x, y$) if $\mathscr{E}_x = \mathscr{E}_y = 0$.

By substituting these results into (9) and (10), the model is reduced to the translation-invariant nonlinear Schrödinger equations

$$\left[2i\left(k_y - \frac{V\omega}{c^2}\right)\frac{\partial}{\partial\theta} + b_0^2\frac{\partial^2}{\partial\theta^2}\right]\mathscr{E}_x = (b_1\mathscr{E}_x + b_2\mathscr{E}_y)S_x + b_3\mathscr{E}_y S_0 + b_x|\mathscr{E}_x|^2\mathscr{E}_x, \tag{20}$$

$$\left[2i\left(k_y - \frac{V\omega}{c^2}\right)\frac{\partial}{\partial\theta} + b_0^2\frac{\partial^2}{\partial\theta^2}\right]\mathscr{E}_y = (b_1\mathscr{E}_y + b_2\mathscr{E}_x)S_x + b_3\mathscr{E}_x S_0 + b_y|\mathscr{E}_y|^2\mathscr{E}_y, \tag{21}$$

where

$$b_0 = \cos\gamma_0, \quad b_1 = a_2\Lambda\cos\gamma_0\sin\gamma_0\omega^2/c^2,$$

$$b_2 = a_1\Lambda\omega^2/2c^2, \quad b_3 = a_1 g\Lambda\cos\gamma_0\sin\gamma_0\omega^2/2c^2,$$

$$b_x = a_2 g\Lambda\cos^2\gamma_0\omega^2/c^2, \quad b_y = a_2 g\Lambda\sin^2\gamma_0\omega^2/c^2.$$

The simplest single-phase solution to system (20), (21) is sought in the form

$$\mathcal{E}_x = \Lambda_0 F_0 G(s_0\theta)\exp(i f_0\theta), \quad (22)$$

$$\mathcal{E}_y = \Lambda_0^{-1} F_0 G(s_0\theta)\exp(i f_0\theta), \quad (23)$$

where Λ_0, F_0, f_0, and s_0 are real constants and $G(\theta)$ is a real-valued function. Substituting (22) and (23) into (19)–(21), we find that Λ_0 is a solution to the algebraic equation

$$\begin{aligned}\Lambda_0^2 1 - g\cos^2\gamma_0 - \Lambda_0^{-2}(1 - g\sin^2\gamma_0)\\ = g\sin\gamma_0\cos\gamma_0(\Lambda_0^{-4} - \Lambda_0^4)/2,\end{aligned} \quad (24)$$

where

$$f_0 = 2(V\omega/c^2 - k_y), \quad F_0^2 = 4\kappa_0|B_0|,$$

$$B_0^{-1} = (2b_1 + b_3) + 2b_2\Lambda_0^{-2} + b_3\Lambda_0^{-4} + 2b_x\Lambda_0^2.$$

The function G solves the equation

$$\partial_\theta^2 G(s_0\theta) = \kappa_0 2G(s_0\theta)^3 - f_0^2 G(s_0\theta), \quad (25)$$

where

$$s_0 = 1/\cos\gamma_0, \quad \kappa_0 = \operatorname{sgn} B_0.$$

One of the simplest solutions to (25) corresponds to $\kappa_0 = -1$ and has the form

$$G(s_0\theta) = \eta_0 \mathrm{dn}(s_0\eta_0\theta, m), \quad (26)$$

where dn is the Jacobi ellipsoidal function with

$$m^2 = \frac{f_1^2 - \sqrt{f_1^4 - C}}{f_1^2 + \sqrt{f_1^4 - C}}, \quad \eta_0 = f_1^2 + \sqrt{f_1^4 - C},$$

$$f_1 = f_0 G_0,$$

and arbitrary real constants. When $C = f_1^4$, solution (26) reduces to the one-soliton solution

$$G(s_0\theta) = f_1^2 \cosh(f_1^2 s_0\theta). \quad (27)$$

Solutions (26) and (27) describe a long-wavelength wave packet or an electromagnetic solitary wave propagating across the plate. The pulse profiles are constant along the direction perpendicular to the trajectory of soliton propagation.

4. INTEGRABLE REDUCED MODEL

4.1. Derivation of the Model

Generally, only single-phase solutions can be found to nonintegrable equations, such as (9), (10), and (16)–(18). A more detailed analysis and control of nonlinear pulse dynamics should rely on solution of an initial–boundary value problem. This can be done only by applying the inverse scattering method (ISM) to nonlinear evolution equations. Therefore, integrable reduced models amenable to analysis by ISM should be formulated. Furthermore, soliton-like solutions and other coherent structures should be sought as a result of balance between dispersion, cross modulation, nonlinear mixing, and other physical effects modeled by corresponding terms in the governing equations. Conditions for the formation of a particular structure should be determined by finding appropriate combinations of the scales that characterize the amplitudes of fields. Soliton-like and other solutions corresponding to integrable models can also be used as zeroth-order approximations in perturbative analyses of near-integrable models.

An integrable reduced model is derived from system (9), (10), (16)–(18) by applying the slowly varying envelope approximation with respect to the transverse coordinates:

$$\partial_j \mathcal{E} \ll k_j \mathcal{E}, \quad j = x, y. \quad (28)$$

Under these conditions, the terms proportional to $\Delta_\perp$ in Eqs. (9) and (10) can be neglected. These conditions can be combined with the assumption of small Euler angles introduced above in analyzing picosecond pulse propagation for a wide range of pulse duration.

Again, let us consider one-dimensional pulse propagation in the xy plane at an angle γ_0 to the x axis. Suppose that the elliptical cross section of the light beam incident on the crystal is oblate in the direction of the vector $(\cos\gamma_0, \sin\gamma_0, 0)$ and elongated in the orthogonal direction in the xy plane. A numerical analysis of system (9), (10), (16)–(18) without diffraction-related terms revealed that the propagation of such a beam can be described in a quasi-one-dimensional approximation. A highly elliptic, large-amplitude, Gaussian-shaped pulse splits into two independent pulses propagating in opposite directions. The rightward-propagating pulse is similar in shape to the one-soliton solution obtained below in the one-dimensional approximation for unidirectional acoustic pulse propagation. Note that no splitting of this kind was observed in numerical simulations if the initial Gaussian-shaped pulse had a nearly symmetric profile. It can also be shown analytically that quasi-one-dimensional dynamics is observed when the incident pulse has a uniform field distribution along the coordinate perpendicular to the propagation direction.

When diffraction-related terms in Eqs. (9) and (10) are neglected in analyzing one-dimensional pulse prop-

agation, the Stokes vector satisfies the following corollary to system (9), (10):

$$\begin{aligned} \partial_{\tilde{\chi}} S_x &= -g(u_{xx} - u_{yy})S_y, \\ \partial_{\tilde{\chi}} S_y &= g(u_{xx} - u_{yy})S_x - 2u_{xy}S_z, \\ \partial_{\tilde{\chi}} S_z &= 2u_{xy}S_y, \end{aligned} \tag{29}$$

where

$$\partial_{\tilde{\chi}} = 2A_1^{-1}(k_x\partial_x + k_y\partial_y + \omega c^{-2}\partial_t).$$

However, system (14), (15), (29) is too difficult to analyze. The widespread slowly varying envelope approximation cannot be applied to derive an integrable reduced model, because the Stokes vector does not contain fast-oscillating terms. Without using this approximation, Eqs. (14), (15), and (29) can be simplified if $A_1 S_0/G_0 \ll 1$ (i.e., if acoustic–electromagnetic wave interaction is sufficiently weak; see Eq. (33) below). This condition is widely used in fluid dynamics and is analogous to the unidirectional approximation used in [22, 23] to derive reduced Maxwell–Bloch equations for a two-level laser medium.

In the unidirectional approximation for acoustic pulse propagation used here to simplify analysis of system (16)–(18),

$$\partial_x = \cos\gamma_0\partial_{\tilde{\chi}}, \quad \partial_y = \sin\gamma_0\partial_{\tilde{\chi}},$$

where

$$\tilde{\chi} = \cos\gamma_0 x + \sin\gamma_0 y,$$

and it formally holds that

$$\partial_{\tilde{\chi}} \approx -v_0^{-1}\partial_t + \mathbb{O}(\epsilon).$$

The factor multiplying the right-hand sides of (14) and (15) is similar in order of magnitude to the derivatives

$$\partial_{\tilde{\chi}} = \partial_{\tilde{\chi}} + v_0^{-1}\partial_t$$

of acoustic field amplitudes. Under this condition, acoustic pulses propagate with a velocity close to the phase velocity v_0, and the derivatives with respect to x and y on the right-hand sides of (16)–(18) can be replaced with $\cos\gamma_0\partial_{\tilde{\chi}}$ and $\sin\gamma_0\partial_{\tilde{\chi}}$, respectively, up to $\mathbb{O}(\epsilon^2)$ terms. As a result, unidirectional one-dimensional pulse propagation is obtained for arbitrary initial and boundary conditions. If the initial pulse is a plane wave, then unidirectional pulse propagation can be obtained when the factors on the right-hand sides of (16)–(18) are of order unity.

The derivatives of electromagnetic field components become

$$2A_1^{-1}(k_x\cos\gamma_0 + k_y\sin\gamma_0)\partial_{\tilde{\chi}}$$

since the ratio v_0/c is about 10^{-5} for solids, and the time derivative in $\partial_{\tilde{\chi}}$ can be dropped. Furthermore, system (29) entails the invariance of the Stokes-vector magnitude:

$$\partial_{\tilde{\chi}} S_0 = 0. \tag{30}$$

To simplify analysis, assume that $\gamma_0 \ll 1$; i.e., the terms proportional to $\sin^2\gamma_0$ on the right-hand sides of (16)–(18) can be neglected. As a result, system (16)–(18) is reduced to

$$\frac{\partial\mathcal{H}}{\partial\tilde{\chi}} + \frac{1}{v_0}\frac{\partial\mathcal{H}}{\partial t} = -\frac{gA_1}{2}\frac{\partial S_z}{\partial\tilde{\chi}}, \tag{31}$$

$$\frac{\partial\mathcal{G}}{\partial\tilde{\chi}} + \frac{1}{v_0}\frac{\partial\mathcal{G}}{\partial t} = -\frac{A_1}{2}\frac{\partial S_x}{\partial\tilde{\chi}}, \tag{32}$$

where

$$\mathcal{H} = u_{xx} - u_{yy}, \quad \mathcal{G} = 2u_{xy}.$$

By virtue of (29), (31), and (32), the acoustic field components are related as follows:

$$\mathcal{G}^2 + \mathcal{H}^2 = G_0^2. \tag{33}$$

The real function $G_0(\tilde{\chi}, t)$ is determined by the equation

$$\partial_{\tilde{\chi}} G_0 = -v_0^{-1}\partial_t G_0.$$

In view of the essentially nonresonant nature of the acoustic–electromagnetic wave interaction analyzed here, a multimode electromagnetic wave packet should be considered in the general case. The present model is analogous to that describing interaction between electromagnetic fields and ion waves in plasmas. In both models, the contributions due to different electromagnetic field modes are taken into account by performing weighted averaging of the right-hand sides of Eqs. (31) and (32).

Without loss of generality, suppose that

$$G_0 = \text{const}, \quad S_0 = \text{const}.$$

By virtue of (33), the integrable system of Eqs. (29), (31), and (32) becomes

$$\partial_\tau G = \langle\partial_\chi R_x\rangle, \tag{34}$$

$$\partial_\tau H = g\langle\partial_\chi R_z\rangle, \tag{35}$$

$$\partial_\chi R_y = gHR_x - GR_z, \tag{36}$$

$$\partial_\chi R_x = -gHR_y, \tag{37}$$

$$\partial_\chi R_z = GR_y, \tag{38}$$

where

$$G = \frac{\mathcal{G}}{G_0}, \quad H = \frac{\mathcal{H}}{G_0}, \quad R_\gamma = -\frac{S_\gamma}{S_0}, \quad \gamma = x, y, z,$$

$$\tau = \frac{(\tilde{\chi} + v_0 t)A_1 S_0}{2(k_x \cos\gamma_0 + k_y \sin\gamma_0)},$$

$$\chi = \frac{A_1 G_0 \tilde{\chi}}{2(k_x \cos\gamma_0 + k_y \sin\gamma_0)}.$$

The angle brackets in (34) and (35) denote averaging over the electric-field carrier frequencies:

$$\langle f(\omega)\rangle = \int_{-\infty}^{\infty} f(\omega')\mathcal{D}(\omega)d\omega', \tag{39}$$

where $\mathcal{D}(\omega)$ is a frequency distribution function.

4.2. Inverse Scattering Method

To obtain a solution defined on the entire χ axis, assume that the acousto-optic system is in the stable ground state corresponding to an energy minimum at $\chi \longrightarrow \pm\infty$. Suppose that the acoustic pulse with $H(\chi, 0)$ and $G(\chi, 0)$ introduced into the system gives rise to acoustic- and electromagnetic-field solitons. Details of the required initial pulse shape are discussed below.

System (34)–(38) is the compatibility condition for the following linear systems:

$$\partial_\chi \mathbf{\Phi} = \hat{L}\mathbf{\Phi}, \tag{40}$$

$$\partial_\tau \mathbf{\Phi} = \hat{A}\mathbf{\Phi}, \tag{41}$$

where

$$\hat{L} = \begin{pmatrix} -i\lambda_- H & \lambda_+ G \\ -\lambda_+ G & i\lambda_- H \end{pmatrix}, \tag{42}$$

$$\hat{A} = \frac{g}{g^2 - 4\lambda_-^2} \times \left\langle \begin{matrix} -i\lambda_- R_z & \lambda_+(gR_x - 2i\lambda_- R_y) \\ -\lambda_+(gR_x + 2i\lambda_- R_y) & i\lambda_- R_z \end{matrix} \right\rangle, \tag{43}$$

$$\lambda_\pm = \xi \pm r^2 \xi^{-1},$$

with ξ denoting a spectral parameter,

$$G^2 + E^2 = 1, \quad r = \frac{1}{4}\sqrt{g^2 - 1}.$$

Since both (40) and (41) are independent of ω, it can be assumed that $\mathcal{D}(\omega) = \delta(\omega)$.

Physically meaningful solutions must describe soliton propagation against the background corresponding to a stable vacuum solution. Let us analyze the linear stability of the homogeneous vacuum solution $\{R_\gamma^{(0)}, G^{(0)}, H^{(0)}\}$ to system (34)–(38). The solution satisfies the equations

$$gR_x^{(0)}H^{(0)} = R_z^{(0)}G^{(0)}, \quad R_y^{(0)} = 0. \tag{44}$$

Suppose that

$$R_\gamma = R_\gamma^{(0)} + \tilde{R}_\gamma \exp(i\nu t + iq\tilde{\chi}),$$

$$H = H^{(0)} + \tilde{H}\exp(i\nu t + iq\tilde{\chi}),$$

$$G = G^{(0)} + \tilde{G}\exp(i\nu t + iq\tilde{\chi}),$$

where the tilde denotes small perturbation amplitudes. In the long-wavelength limit ($q \longrightarrow 0$), the dispersion relation reduces to

$$\nu^2 = \nu_0^2\sqrt{1 + (g^2 - 1)H^{(0)2}} \times [\sqrt{1 + (g^2 - 1)H^{(0)2}} - gR_z^{(0)}\mathrm{sgn}(H^{(0)})], \tag{45}$$

where ν_0 is a real constant. According to (45), the solution to (44) is stable if $\mathrm{sgn}(gR_z^{(0)}H^{(0)}) < 0$. If $\mathrm{sgn}(gR_z^{(0)}H^{(0)}) > 0$, then stable and unstable solutions correspond to $g^2 < 1$ and $g^2 > 1$, respectively. If $g = 1$ and $\mathrm{sgn}(R_z^{(0)}H^{(0)}) > 0$, then the solution to (44) corresponds to neutral stability. The solutions obtained below demonstrate that stability of the background solution corresponds to soliton propagation with a velocity lower than the phase velocity of sound propagating in the medium. Formally, there exist solutions corresponding to solitons that propagate against an unstable background solution to (44) with a velocity higher than the phase velocity of sound propagating in the medium. However, they are physically meaningless because of instability, and the actual solution will be a quasi-self-similar one describing the decay of an unstable initial state. It can be shown that the group velocity in this solution is lower than the phase velocity of sound propagating in the medium.

The spectral parameter in (40) and (41) differs from that employed in [13]. It is chosen so that it may not be necessary to introduce Riemann surfaces when $g^2 \leqslant 1$ in certain cases. One of these cases corresponds to the conditions

$$R_x(0, \tau) = R_y(0, \tau) = 0, \quad R_z(0, \tau) = -1, \quad G(\chi, 0) = 0, \quad H(\chi, 0) = 1. \tag{46}$$

The ISM apparatus corresponding to the parameterization used here and to the background solution defined by (46) is simpler than that developed in [13] for a different problem.

Solutions to (40) have the symmetry

$$\mathbf{\Phi}(\xi) = \hat{M}\mathbf{\Phi}(\lambda^*)^*\hat{M}^{-1}, \tag{47}$$

where

$$\hat{M} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{48}$$

The standard Jost functions $\mathbf{\Psi}^{\pm}$ are defined as solutions to (40) having the asymptotic form

$$\mathbf{\Psi}^{\pm} = \hat{\mathbf{\Psi}}_0 \exp(-i\lambda_-\hat{\sigma}_3\tau), \quad \tau \longrightarrow \pm\infty, \tag{49}$$

where $\hat{\sigma}_3$ is a Pauli matrix and

$$\hat{\mathbf{\Psi}}_0 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{50}$$

These solutions are related by a scattering matrix $\hat{T}$:

$$\mathbf{\Psi}^- = \mathbf{\Psi}^+\hat{T}, \quad \mathbf{\Psi}^{\pm} = \begin{pmatrix} \psi_1^{\pm} & -\psi_1^{\pm *} \\ \psi_2^{\pm} & \psi_2^{\pm *} \end{pmatrix}. \tag{51}$$

Symmetry (47), (48) entails the following form of the scattering matrix:

$$\hat{T} = \begin{pmatrix} a(\xi) & -b^*(\xi) \\ b(\xi) & a^*(\xi) \end{pmatrix}. \tag{52}$$

The functions $a(\xi)$ and $b(\xi)$, as well as the Jost functions, have standard analytic properties: $a(\xi)$, $\psi^+(\xi)$, and ψ^- admit analytic continuation to the upper half-plane; $a^*(\xi)$, $\psi^{+*}(\xi)$, and $\psi^{-*}(\xi)$, to the lower half-plane.

The following symmetries are valid:

$$a(-r^2/\xi) = a(\xi), \quad b(-r^2/\xi) = b(\xi), \quad \xi \in \mathbb{R}. \tag{53}$$

Let ξ_m and ξ_m^* ($m = 1, 2, \ldots, 2N$) be the poles in the complex plane such that

$$\begin{gathered} b_n(\xi_n) = -b_n^*(\xi_n^*), \quad b_{N+n} = -b_n, \\ n = 1, 2, \ldots, N, \\ c_{N+n} = -r^2\xi_n^{-2}c_n, \quad \xi_{N+n} = -r^2/\xi_n, \\ \lambda_{\pm, N+n} = \pm\lambda_{\pm, n}, \end{gathered} \tag{54}$$

where

$$c_n = b(\xi_n)/(ida/d\xi|_{\xi=\xi_n}), \quad \lambda_{\pm, m} = \xi_m \pm r^2\xi_m^{-1}.$$

By substituting the components of the Jost functions represented as

$$\begin{aligned} \mathbf{\Psi}(\chi, \xi) &= \boldsymbol{\psi}^+(\chi, \xi) \\ &+ \int_\chi^\infty \begin{pmatrix} \lambda_- K_1(\chi, s) & \lambda_+ K_2(\chi, s) \\ -\lambda_+ K_2^*(\chi, s) & \lambda_- K_1^*(\chi, s) \end{pmatrix} \boldsymbol{\psi}^+(s, \xi) ds \end{aligned} \tag{55}$$

into (51); integrating the resulting equations

$$\psi_1^{+*} = \frac{\psi_1^{-*}}{a} - \frac{b}{a}\psi_1^+, \tag{56}$$

$$\psi_2^{+*} = \frac{\psi_2^{-*}}{a} - \frac{b}{a}\psi_2^+ \tag{57}$$

with the weighting factors

$$\int_{-\infty}^{\infty} \exp(-i\lambda_{\pm}y)/2\pi$$

and using the relations

$$\int_{-\infty}^{\infty} \exp(i\lambda_-\chi)d\xi = 2\pi\delta(\chi), \tag{58}$$

$$\begin{gathered} \int_{-\infty}^{\infty} \Omega(\xi)\xi^{-1}\exp(i\lambda_-\chi)d\xi = 0, \\ \Omega(\xi) = \Omega(-r^2/\xi), \end{gathered} \tag{59}$$

one obtains the Marchenko equations for $y \geq \chi$:

$$K_2^*(\chi, y) - F_1(\chi, y) = \int_\chi^\infty K_1(\chi, s)F_2(s + y)^* ds, \tag{60}$$

$$K_1^*(\chi, y) = -\int_\chi^\infty K_2(\chi, s)F_3(s + y)ds, \tag{61}$$

where the kernels F_k are

$$F_1(\chi) = \sum_{n=1}^{4N} \frac{c_n}{\lambda_{-,n}} \exp(i\lambda_{-,n}\chi) + \frac{1}{2\pi}\int \frac{\rho}{\lambda_-} \exp(i\lambda_-\chi)d\xi, \quad (62)$$

$$F_2(\chi) = \sum_{n=1}^{4N} c_n \exp(i\lambda_{-,n}\chi) + \frac{1}{2\pi}\int \rho \exp(i\lambda_-\chi)d\xi, \quad (63)$$

$$F_3(\chi) = \sum_{n=1}^{4N} \frac{c_n\lambda_{+,n}}{\lambda_{-,n}} \exp(i\lambda_{-,n}\chi) + \frac{1}{2\pi}\int \frac{\rho\lambda_+}{\lambda_-} \exp(i\lambda_-\chi)d\xi, \quad (64)$$

with

$$\rho(\xi) = b(\xi)/a(\xi), \quad \xi \in \mathbb{R}.$$

The symmetries formulated above and real-valuedness conditions for G and H imply that an N-soliton solution is associated with a set of $4N$ poles, among which the $2N$ poles ξ_n and ξ_{n+N} ($n = 1, 2, \ldots, N$) are related as follows:

$$\xi_{N+n} = -r^2/\xi_n.$$

Each pole is associated with an additional pole $\tilde{\xi}_m = -\xi_m^*$ ($m = 1, 2, \ldots, 2N$) if $\mathrm{Re}\,\xi_m \neq 0$. The existence of such pairs of poles combined with the symmetry condition

$$c_n(\xi) = -c_n^*(-\xi^*) \quad (65)$$

ensures that G and H are real-valued.

The kernels F_1, F_2, and F_3 corresponding to the radiationless soliton spectrum and preserving symmetries (54) and (65) are

$$F_1(\chi) = \sum_{n=1}^{N} (C_n \exp(i\lambda_{-,n}\chi)) + C_n^* \exp(-i\lambda_{-,n}^*\chi), \quad (66)$$

$$F_2(\chi) = \sum_{n=1}^{N} (C_n\lambda_{-,n} \exp(i\lambda_{-,n}\chi) - C_n^*\lambda_{-,n}^* \exp(-i\lambda_{-,n}^*\chi)), \quad (67)$$

$$F_3(\chi) = \sum_{n=1}^{N} \left(\frac{C_n\lambda_{+,n}^2}{\lambda_{-,n}} \exp(i\lambda_-\chi) - \frac{C_n^*\lambda_{+,n}^{2*}}{\lambda_{-,n}^*} \exp(-i\lambda_-^*\chi) \right), \quad (68)$$

where

$$C_n = c_n/\xi_n, \quad \lambda_{\pm,n} = \xi_n \pm r^2/\xi_n, \quad \lambda_{\pm,n}^* = \xi_n^* \pm r^2/\xi_n^*.$$

Substituting the expression for $\mathbf{\Psi}^+$ into (40) and collecting the terms proportional to powers of λ, we obtain

$$K_1(\chi, \chi)[1 + H(\chi)] = G(\chi)[1 - iK_1(\chi, \chi)]. \quad (69)$$

This relation is combined with $G^2 = 1 - H^2$ to express G and H in terms of $K_{1,2}$ as

$$H(\chi) = \frac{[1 - iK_1(\chi, \chi)]^2 - K_2^2(\chi, \chi)}{[1 - iK_1(\chi, \chi)]^2 + K_2^2(\chi, \chi)}, \quad (70)$$

$$G(\chi) = \frac{2[1 - iK_1(\chi, \chi)]K_2(\chi, \chi)}{[1 - iK_1(\chi, \chi)]^2 + K_2^2(\chi, \chi)}. \quad (71)$$

The N-soliton solutions corresponding to the sets

$$\{\xi_n, -\xi_n^*, -r^2/\xi_n, r^2/\xi_n^*\}, \quad n = 1, 2, \ldots, N$$

can be obtained by a standard method. The matrix form of the solutions is given in the Appendix.

The function $c_n(\tau)$ is determined from system (41) corresponding to $R_z(\pm\infty, \tau) = 1$ by using the equation

$$\partial_\tau \hat{T} = \hat{T}\hat{A}'(\chi, \tau)\big|_{\chi\to\infty} - \hat{A}'(\chi, \tau)\big|_{\chi\to-\infty}\hat{T}, \quad (72)$$

where

$$\hat{A}'(\chi, \tau) = \mathbf{\Psi}^+(\chi)\hat{A}(\chi, \tau)[\mathbf{\Psi}^+(\chi)]^{-1}.$$

For the initial and boundary conditions chosen here to describe soliton dynamics,

$$c_n(\tau) = c_n(0)\exp\left(\frac{2ig\lambda_{-,n}\tau}{g^2 - 4\lambda_{-,n}^2}\right). \quad (73)$$

If

$$C_1 = |C_1|\exp(i\phi_1), \quad \xi = r\exp(i\alpha),$$

then the one-soliton particular solution, which corresponds to an eigenvalue $\lambda_{-,1}$ lying on the imaginary axis, has the form

$$H = 1 - \frac{2}{f^2(\chi, \tau) + 1}, \quad (74)$$

$$G = \frac{2f(\chi, \tau)}{f^2(\chi, \tau) + 1}, \quad (75)$$

where

$$f(\chi, \tau) = \frac{2|C_1|\exp[-4r\theta\sin\alpha]\cos\phi_1}{1 + 8r\left|C_1^2\cos^2\alpha\right||\sin\alpha|^{-1}\cos\phi_1\exp[-8r\theta\sin\alpha]}, \quad (76)$$

with

$$\theta = \chi - \tilde{V}_0\tau, \quad \tilde{V}_0 = g[g^2 + (g^2 - 1)\sin^2\alpha]^{-1}.$$

Figure 2 demonstrates that the solution profile depends on the sign of $\cos\phi_1$. The expressions for K' and K_2 obtained in the Appendix imply that the soliton corresponding to the more general case of $\xi = r_1\exp(i\alpha)$ ($r_1 \neq r$) has an additional degree of freedom analogous to the oscillatory mode of the breather described by the sine-Gordon equation.

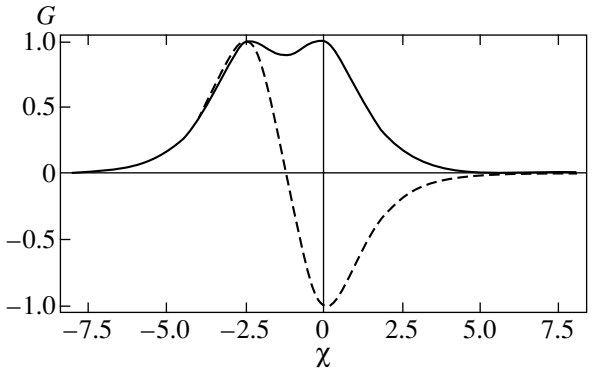


Fig. 2. Soliton profile $G(\chi)$ given by (75) for $r = 0.25$, $|c_1| = 1$, and $\alpha = 2\pi/5$: solid and dashed curves correspond to $\phi_1 = \pi/3$ and $\phi_1 = 2\pi/3$, respectively.

5. DISCUSSION

It is shown in this study that solitons can be generated as a result of acoustic–electromagnetic wave interaction when a single-mode field propagates in an isotropic photoelastic material. The evolution equations describing nonlinear coupling between polarization components admit single-phase solutions and reduce (under realistic assumptions) to integrable models. These models describe (multi)soliton solutions and solutions of other types, which arise as a result of parametric mixing due to off-resonance interaction between light and acoustic waves.

The mechanism of nonlinear interaction between acoustic and electromagnetic fields described in this paper provides a basis for using electric field to control generation and evolution of picosecond acoustic pulses. Acoustic pulse propagation can be analyzed in one-dimensional approximation if the initial pulse has the form of a wave or a highly elliptic Gaussian-shaped pulse. According to a numerical analysis, the initial pulse splits into two independent packets of pulses propagating in opposite directions. Thus, it is shown that an initial pulse having the form of a wave can be described in the unidirectional approximation if the factors multiplying the right-hand sides of Eqs. (31) and (32) are of order unity.

Picosecond acoustic solitons can propagate against the background corresponding to a stable vacuum solution if electric field of required form is applied to the crystal. The stable background state described by model (34)–(38) may correspond to the conditions

$$u_{xx} > 0, \quad u_{yy} = u_{xy} = 0, \quad E_x \neq 0, \quad E_y = 0,$$

when the plate is deformed in the direction of initial incident-field polarization (compressed along the x axis). In the case of a two-mode field, when $q_0 \neq 0$ (see (9) and (10)), there exists an additional possibility of controlling the generation of solitons. When $u_{xx} = 0$ and $u_{yy} = u_{xy} = 0$, the role of an initial condition is played by the detuning q_0. Depending on its sign (instead of the sign of $H(\chi, 0)$), either stable or unstable pulse propagation regime is obtained; i.e., detuning can be used to control acoustic wave generation.

According to the solution to an initial–value problem for integrable system (34)–(38) with coupling via the Stokes vector, when the incident electric field is concentrated in a high-power pulse, the acoustic field u_{xy} is a packet of solitons. Strictly speaking, the problem of acoustic pulse generation is equivalent to a spectral problem in which the potential is expressed in terms of $E_{x,y}$ or the Stokes vector **S**. However, the spectral-parameter dependence in the spectral problem corresponding to system (34)–(38) is much more complicated than in (40), and the corresponding analysis requires further development of the ISM apparatus.

In the "initial value" problem (40) considered in this paper, solitons are generated by the initial deformation represented by $G(\chi, 0)$. A criterion for soliton generation can be obtained by using a gauge transformation analogous to that applied in [24] to transform spectral problem (40) into the Zakharov–Shabat problem with the potential $q(\chi)$ defined in terms of G by the relation

$$q(\chi) = i(\sqrt{1 - g^2}G + \partial_\chi G/\sqrt{1 - G^2}). \quad (77)$$

According to the Zakharov–Shabat theory, solitons can be generated if the integral

$$I_0 = \left|\int q(s)ds\right|$$

is sufficiently large (approximately equal to 3) (e.g., see [25]). This condition requires that either the integral of G is sufficiently large or G is a sufficiently fast-varying monotonic function of χ. Thus, acoustic solitons can be generated by creating either an almost uniform strain $u_{xy} \neq 0$ in the crystal or a sufficiently intense

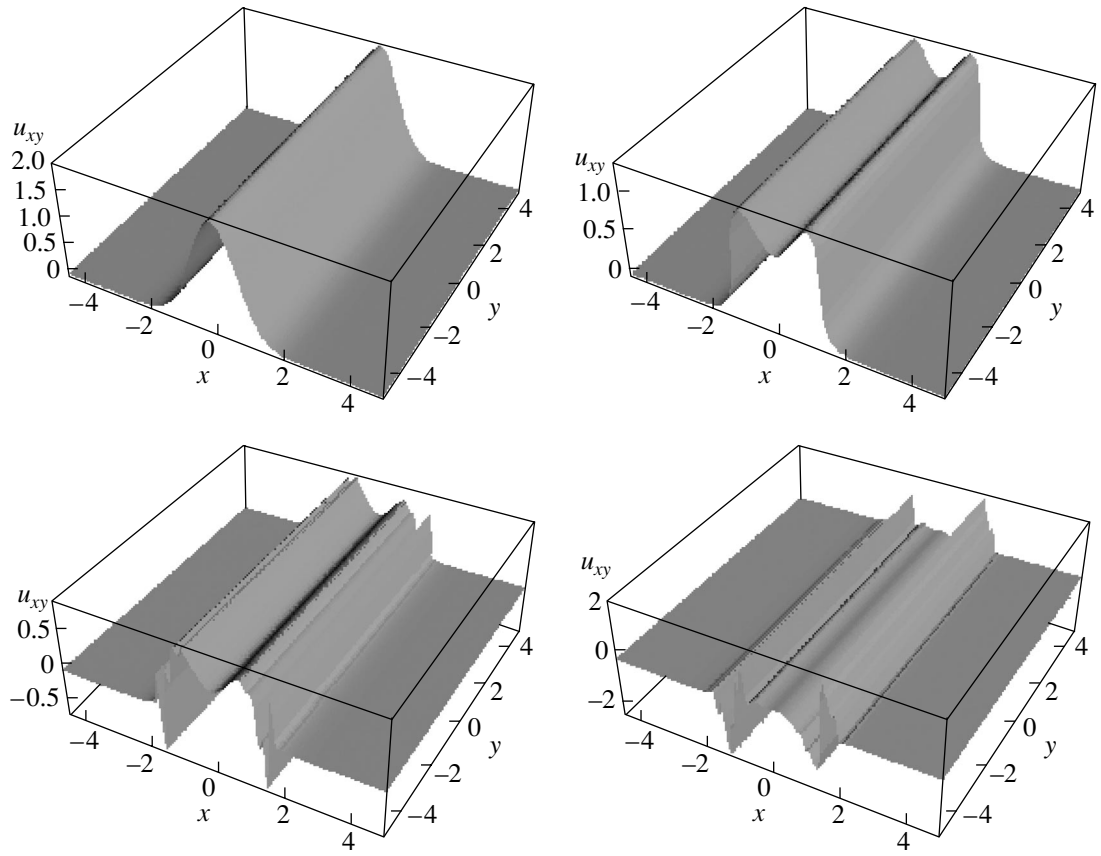


Fig. 3. Transformation of an initial pulse with $u_{xy} = 2\exp[-(x^2 + 0.00001y^2)]$, $u_{xx} = u_{yy} = 0$, $E_x(-5, \tau) = 1$, and $E_y(-5, \tau) = 0$ into solitons. Profiles of u_{xy} versus x, y are shown at instants separated by equal time intervals.

acoustic pulse. Figure 2 shows that the resulting solution is either two-hump or soliton–antisoliton bound state, depending on the value of C_1. Note that a rightward-propagating soliton–antisoliton pair was obtained numerically for a Gaussian shaped initial pulse with $\max u_{xy}(x, y, 0, 0) = 2$ (see Fig. 3).

A different pattern will be observed if an unstable steady state is used as an initial–boundary condition, such as

$$u_{yy} > 0, \quad u_{xx} = u_{xy} = 0,$$

$$E_x \neq 0, \quad E_y = 0, \quad g > 1,$$

in which case the initial strain is perpendicular to the initial polarization. If the initial perturbation is a picosecond electromagnetic pulse such that $S_x(0, \tau) \neq 0$, then packets of acoustic and electromagnetic pulses of similar duration will be generated as a result of decay of the unstable state. Detailed analysis of this process lies outside the scope of the present study. It should only be noted that the ensuing dynamics are described by a nonsoliton solution that can be expressed, under a certain choice of initial conditions, in terms of a Painlevé transcendental function after applying a gauge transformation to system (40), (41). This solution is analogous to that describing Raman scattering or a phonon avalanche in a paramagnetic crystal with spin-1/2 impurities [13].

It is clear that additional mechanisms can be used to generate acoustic and electromagnetic pulses and control their parameters in the case of finite Euler angles (see Section 2) and anisotropic unperturbed crystal. However, acoustic–electromagnetic wave interaction in such systems are described by substantially more complicated equations, which require separate analysis. It would be interesting to extend the results of this study to the case of a medium with a periodically modulated density or dielectric constant.

ACKNOWLEDGMENTS

I thank A.I. Plekhanov and F.Kh. Gel'mukhanov for discussion of the physical aspects of the problem. This work was supported, in part, by the Russian Foundation for Basic Research, project no. 03-02-16297; under the Interdisciplinary Integration for Basic Studies of the Siberian Division of the Russian Academy of Sciences, project no. 84; and under the Program of Basic

Research of the Presidium of the Russian Academy of Sciences, grant no. 8-2.

APPENDIX

By representing the kernels as

$$\begin{aligned} F_1(\chi + y) &= D_1(\chi)D_0(y)^T, \\ F_2(\chi + y) &= D_2(\chi)D_0(y)^T, \\ F_3(\chi + y) &= D_3(\chi)D_0(y)^T, \end{aligned} \tag{78}$$

with

$$\begin{aligned} D_0(\chi)_n &= \exp(i\lambda_{-,n}\chi), \\ D_0(\chi)_{n+N} &= \exp(-i\lambda^*_{-,n}\chi), \\ D_1(\chi)_n &= c_n\xi_n^{-1}D_0(\chi)_n, \\ D_1(\chi)_{n+N} &= c_n^*\xi_n^{-1*}\exp(-i\lambda^*_{-,n}\chi), \\ D_2(\chi)_n &= c_n\lambda_{-,n}\xi_n^{-1}\exp(i\lambda_{-,n}\chi), \\ D_2(\chi)_{n+N} &= -c_n^*\lambda^*_{-,n}\xi_n^{-1*}\exp(-i\lambda^*_{-,n}\chi), \\ D_3(\chi)_n &= c_n\lambda_{+,n}^2\lambda_{-,n}^{-1}\xi_n^{-1}\exp(i\lambda_{-,n}\chi), \\ D_3(\chi)_{n+N} &= -c_n^*\lambda_{+,n}^{2*}(\lambda_{-,n}\xi_n)^{-1*}\exp(-i\lambda^*_{-,n}\chi), \\ n &\le N, \end{aligned} \tag{79}$$

and using the analogous representation

$$\begin{aligned} K_1(\chi, y) &= Q_1(\chi)D_0(y)^T, \\ K_2(\chi, y) &= Q_2(\chi)D_0(y)^T, \end{aligned} \tag{80}$$

the Marchenko equations (60) and (61) can be rewritten in matrix form as

$$Q_1(\chi) = -Q_2(\chi)P_1(\chi), \tag{81}$$

$$Q_2(\chi) = D_1^*(\chi) + Q_1(\chi)P_2(\chi), \tag{82}$$

where P_1 and P_2 are $2N$-by-$2N$ matrices:

$$P_1(\chi)_{nm} = \frac{D_0^*(\chi)_n D_3(\chi)_m}{i(\lambda^*_{-,n} - \lambda_{-,m})}, \tag{83}$$

$$P_2(\chi)_{nm} = \frac{D_0(\chi)_n D_2^*(\chi)_m}{i(\lambda^*_{-,m} - \lambda_{-,n})}. \tag{84}$$

Their formal solution leads to expressions for $K' = 1 - iK_1$ and K_2:

$$\begin{aligned} K'(\chi, \chi) &= 1 + \mathrm{Tr}[iP_1D_0^{T*}D_1^*(I + P_1P_2)] \\ &= \frac{\det(I + P_1P_2 + iP_1D_0^TD_1^*)}{\det(I + P_1P_2)}, \end{aligned} \tag{85}$$

$$\begin{aligned} K_2(\chi, \chi) &= \mathrm{Tr}[D_0^{T*}D_1^*(I + P_1P_2)] \\ &= \frac{\det(I + P_1P_2 + D_0^{T*}D_1^*)}{\det(I + P_1P_2)} - 1. \end{aligned} \tag{86}$$

If $N = 1$, then

$$\begin{aligned} K'(\chi, \chi; \tau) &= 1 + \frac{f(\chi, \tau)}{D(\chi, \tau)} \\ &\times \left\{ 2\mathrm{Re}[\exp(i\theta_1 + i\phi_1)\sin(\theta_1 + i\gamma/2)] \right. \\ &\left. + \frac{f(\chi, \tau)}{2\zeta}\mathrm{Im}[\sin(\theta_1 + i\gamma/2)S(\chi, \tau)] \right\}, \end{aligned} \tag{87}$$

$$\begin{aligned} K_2(\chi, \chi; \tau) &= \frac{|C_1|\exp(-2\zeta(\chi - V_0\tau))}{D(\chi, \tau)} \\ &\times \left\{ \cos(2\eta(\chi - V_1\tau) + \phi_1) + \frac{f(\chi, \tau)}{2\zeta}\mathrm{Im}[e^{\beta/2}S(\chi, \tau)] \right\}, \end{aligned} \tag{88}$$

where

$$\begin{aligned} D(\chi, \tau) &= 1 + \sinh^2(\gamma)|f(\chi, \tau)|^2(2\zeta)^{-2} \\ &- f(\chi, \tau)\zeta^{-1}\mathrm{Re}[\exp(i\phi_1 + i\beta + i\phi_2/2)\sin(\theta_0 + i\gamma)], \\ f(\chi, \tau) &= \frac{|C_1\lambda_{+,1}|^2\exp(-4\zeta(\chi - V_0\tau))}{|\lambda_{-,1}|}, \\ S(\chi, \tau) &= \exp(i\psi_1)\sin(\theta_0) + \exp(-i\psi_1)\sin(\theta_0 - i\gamma), \\ \theta_0 &= 2\eta(\chi - V_1\tau) - \beta/2 + \phi_2/2, \\ \theta_1 &= 2\eta(\chi - V_1\tau) - \beta/2, \\ \psi_1 &= \beta/2 + \pi/4 + \phi_2, \quad \gamma = \ln|\lambda_{-,1}/\zeta|, \\ \lambda_{-,1} &= \eta + i\zeta = i|\lambda_{-,1}|\exp(i\beta), \\ C_1 &= |c_1/\zeta_1|\exp(i\phi_1), \\ \phi_2 &= \phi_1 + 2\arg(\lambda_{+,1}) - \beta - \pi/2, \\ V_0 &= \frac{g^2 + 4g(\zeta + \eta)}{[g^2 + 4(\zeta^2 - \eta^2)]^2 + (8\eta\zeta)^2}, \\ V_1 &= \frac{-g^2 + 4g(\zeta + \eta)}{[g^2 + 4(\zeta^2 - \eta^2)]^2 + (8\eta\zeta)^2}. \end{aligned}$$

REFERENCES

1. Yu. N. Karamzin and A. P. Sukhorukov, Pis'ma Zh. Éksp. Teor. Fiz. **20**, 734 (1974) [JETP Lett. **20**, 339 (1974)]; Zh. Éksp. Teor. Fiz. **68**, 834 (1975) [Sov. Phys. JETP **41**, 414 (1975)].
2. A. P. Sukhorukov, *Nonlinear Wave Interactions in Optics and Radiophysics* (Nauka, Moscow, 1988) [in Russian].
3. A. I. Maimistov and A. M. Basharov, *Nonlinear Optical Waves* (Kluwer Academic, Dortrecht, 1999).
4. G. N. Burlak, V. V. Grimal'skiĭ, and N. Ya. Kotsarenko, Fiz. Tverd. Tela (Leningrad) **27**, 631 (1985) [Sov. Phys. Solid State **27**, 392 (1985)].
5. G. N. Burlak, Zh. Éksp. Teor. Fiz. **96**, 1607 (1990) [Sov. Phys. JETP **70**, 908 (1990)].
6. G. N. Burlak, V. V. Grimal'skiĭ, and K. Ishkabulov, Fiz. Tverd. Tela (St. Petersburg) **39**, 1101 (1997) [Phys. Solid State **39**, 987 (1997)].
7. D. I. Abarbanel, R. Brown, J. J. Sidorovich, and L. S. Tsimring, Rev. Mod. Phys. **65**, 1331 (1993).
8. M. Schall, M. Walther, and P. Uhd Jepsen, Phys. Rev. B **64**, 094301 (2001).
9. S. V. Sazonov, J. Phys.: Condens. Matter **4**, 6485 (1992).
10. S. V. Voronkov and S. V. Sazonov, Fiz. Tverd. Tela (St. Petersburg) **43**, 1969 (2001) [Phys. Solid State **43**, 2051 (2001)].
11. S. V. Voronkov and S. V. Sazonov, Zh. Éksp. Teor. Fiz. **120**, 269 (2001) [JETP **93**, 236 (2001)].
12. A. A. Zabolotskii, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 709 (2002) [JETP Lett. **76**, 607 (2002)].
13. A. A. Zabolotskii, Phys. Rev. E **67**, 066606 (2003).
14. A. I. Maimistov, Kvantovaya Élektron. (Moscow) **30**, 287 (2000).
15. G. N. Burlak, V. V. Grimal'skiĭ, and N. Ya. Kotsarenko, Zh. Éksp. Teor. Fiz. **90**, 1487 (1986) [Sov. Phys. JETP **63**, 872 (1986)].
16. A. S. Shcherbakov and I. B. Pozdnov, Pis'ma Zh. Tekh. Fiz. **20** (9), 12 (1994) [Tech. Phys. Lett. **20**, 687 (1994)].
17. H.-Y. Hao and H. J. Maris, Phys. Rev. B **64**, 4302 (2001).
18. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 3rd ed. (Pergamon, New York, 1984; Nauka, Moscow, 1992).
19. N. N. Rosanov and S. V. Fedorov, Phys. Rev. E **63**, 066601 (2001).
20. A. Bourgeade and E. Freysz, J. Opt. Soc. Am. B **17**, 226 (2000).
21. L. Bergé, Phys. Rep. **303**, 259 (1998).
22. P. J. Coudrey, J. C. Eilbeck, J. D. Gibbon, and R. K. Bullough, J. Phys. A: Math. Gen. **6**, L53 (1973).
23. J. C. Eilbeck, J. D. Gibbon, P. J. Caudrey, and R. K. Bullough, J. Phys. A: Math. Gen. **6**, 1337 (1973).
24. A. A. Zabolotskii, Zh. Éksp. Teor. Fiz. **115**, 1168 (1999) [JETP **88**, 642 (1999)].
25. V. E. Zakharov, S. V. Manakov, S. P. Novikov, and L. P. Pitaevskiĭ, *Theory of Solitons: the Inverse Scattering Method* (Nauka, Moscow, 1980; Consultants Bureau, New York, 1984).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 147–156.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 170–180.

SOLIDS
Electronic Properties

Resonant Tunneling through a Double-Barrier Quantum Well in a Transverse Magnetic Field

A. Yu. Serov and G. G. Zegrya

Ioffe Physicotechnical Institute, Russian Academy of Sciences,
ul. Politekhnicheskaya 26, St. Petersburg, 194021 Russia
e-mail: andserov@as11585.spb.edu
e-mail: zegrya@theory.ioffe.rssi.ru
Received October 8, 2003

Abstract—We theoretically analyze the tunneling of electrons through a heterostructure with two barriers and a quantum well between them in a magnetic field perpendicular to the current. We take into account the contribution from electrons with various positions of the magnetic oscillator center to the current. The region of the *Z*-shaped current–voltage characteristic for the heterostructure is shown to narrow as the magnetic field strengthens. Our analysis reveals a critical magnetic field strength at which the *Z*-shaped current–voltage characteristic transforms into an *N*-shaped one. We compare our results with experimental data. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Semiconductor systems with highly nonlinear current–voltage characteristics have gained widespread acceptance in electronics. They are used as high-frequency oscillators, logic gates, and switches [1]. A heterostructure with two barriers and a quantum well between them is an example of a system with a highly nonlinear current–voltage characteristic. This structure is called a resonant tunnel diode and was first studied by Esaki and Chang [2]. One or more energy levels can exist in the quantum well. The region on the left acts as a reservoir of conduction electrons, while the region on the right acts as a collector. Contacts are adjacent to these regions. The following peculiarity of the double barrier is used in a resonant tunnel diode: its tunnel transparency has a distinct resonant pattern. This implies that the current strongly depends on the location of the energy level in the quantum well with respect to the Fermi energy of the electrons in the emitter. The current reaches its maximum at voltages at which the energy of the electrons in the emitter is equal to the energy of the discrete level in the quantum well. At higher voltages, the energy of the incident electrons is larger than the energy of the discrete level in the quantum well, and the tunnel transparency of the barrier decreases.

Chang *et al.* [2] studied the *N*-shaped current–voltage characteristic of a resonant tunnel diode. Ricco and Azbel [3] suggested that the charge carriers in the quantum well affect the potential of the system and the shape of the current–voltage characteristic. Subsequently, Goldman *et al.* [4] established that such a system could have a *Z*-shaped current–voltage characteristic. These authors explained the existence of a *Z*-shaped segment in the current–voltage characteristic by the influence of the electric field produced by the charge carriers accumulated in the quantum well on the potential of the structure; many authors subsequently used this model in their studies [5–8]. However, Sollner [9] offered an alternative explanation for this effect where the high-frequency oscillations of the current in the external circuit in the range of negative differential resistance are responsible for the *Z*-shaped pattern. Based on a dynamical model, Buot [10] showed that not only the *Z*-shaped pattern of the current–voltage characteristic, but also the high-frequency oscillations of the current that were presumably observed by the authors of [9] are the result of the accumulation of electrons in the quantum well and the screening of the potential by these electrons. Below, we do not consider these oscillations.

For the *Z*-shaped current–voltage characteristic, two output currents are possible at one voltage. Below, this possibility is called the bistability of the current–voltage characteristic. The possibility of several states being realized when a load is connected should not be meant by the bistability.

The effect of a magnetic field was analyzed to further study the transport properties of a resonant tunnel diode. A magnetic field parallel to the current was found to have no effect on such parameters of the current–voltage characteristic as the width of the bistability region and the peak current. This is because a longitudinal magnetic field does not affect the motion of charge carriers along the current, but changes the density of the states. Experiment and theory [11–13] show that a longitudinal magnetic field gives rise to steps in the current–voltage characteristic the number of which is related to the number of Landau levels.

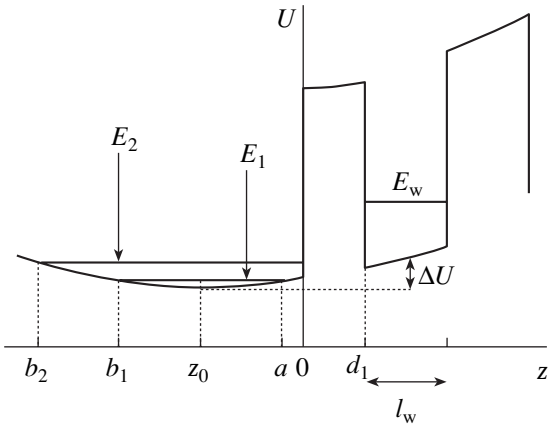


Fig. 1. The potential of the system in a magnetic field: z_0 is the position of the center of the magnetic oscillator; E_1 and E_2 are the quantized energy levels in the emitter; b_1 and b_2 are the left turning points for the electrons with energies E_1 and E_2, respectively, a is the right turning point for the electron with energy E_1; ΔU is the shift of the bottom of the quantum well relative to the emitter.

The tunneling through a resonant tunnel diode in a magnetic field perpendicular to the current is a more complex problem, because the magnetic oscillator component of the potential is superimposed on the heterobarrier potential. The single-barrier tunneling in a transverse magnetic field was investigated in [14]. The coordinates of the center of the magnetic component of the potential can take on various values on which the tunneling probability strongly depends. For the current through a heterostructure to be calculated, the contribution from electrons with various positions of the magnetic oscillator center must be taken into account.

It was experimentally found [15] that the region of the *Z*-shaped current–voltage characteristic narrows and the peak current decreases as the transverse magnetic field increases in strength. Zaslavsky *et al.* [15] showed the existence of a critical magnetic field strength above which the current–voltage characteristic degenerates into an *N*-shaped one. However, their paper contains no quantitative estimates and calculations that would confirm the disappearance of the *Z*-shaped pattern. The tunneling through a resonant tunnel diode in a transverse magnetic field was theoretically calculated in [16, 17]. However, the authors of these papers did not investigate the influence of the charge accumulated in the quantum well on the potential and, hence, the bistability effect of the current–voltage characteristic.

In this paper, we use the formalism of the tunnel matrix element [18] to calculate the current–voltage characteristic of a resonant tunnel diode in transverse electric and magnetic fields by taking into account the screening of the potential by the charge carriers accumulated in the quantum well. We show that a transverse magnetic field causes the bistability region in the current–voltage characteristic to narrow, and that the *Z*-shaped characteristic degenerates into an *N*-shaped one at a certain critical field strength. The narrowing of the bistability region is accompanied by a decrease in the peak resonant current. We derive an expression for determining the critical magnetic field at which the *Z*-shaped pattern of the current–voltage characteristic disappears. Our results are in agreement with experimental data [15].

2. THE HAMILTONIAN OF THE SYSTEM, FORMULATION OF THE PROBLEM

Let us consider a resonant tunnel diode in a magnetic field perpendicular to the current. We direct the magnetic field vector along the *y* axis, while the current flows along the *z* axis. We choose a calibration of the vector potential of the magnetic field that allows the problem to be reduced to one-dimensional:

$$\mathbf{A} = (Hz, 0, 0), \tag{1}$$

where H is the magnetic field strength.

In this case, the Hamiltonian of the system takes the form

$$\hat{H} = \frac{1}{2m}\left(\hat{p}_x - \frac{e}{c}Hz\right)^2 + \frac{\hat{p}_y^2}{2m} + \frac{\hat{p}_z^2}{2m} + U_b(z), \tag{2}$$

where U_b is the potential of the two barriers, and e is the electron charge. We see from (2) that a magnetic component quadratic in z coordinate is added to U_b (Fig. 1). Note that the position of the center of the parabola is specified by the electron momentum p_x and, hence, can differ for different electrons. The potential energy for the electrons is

$$\begin{aligned} U(z) &= U_b + \frac{m\omega_c^2(z - z_0)^2}{2}, \\ z_0 &= \frac{cp_x}{eH}, \quad \omega_c = \frac{eH}{mc}, \end{aligned} \tag{3}$$

where ω_c is the cyclotron frequency in the magnetic field, and z_0 is the center of the magnetic oscillator.

Note that the energy level of an electron in the quantum well in a transverse magnetic field shifts upward relative to the emitter by (see Fig. 1)

$$\Delta U(z_0) = \frac{m\omega_c^2(|z_0| + d_1 + l_w/2)^2}{2}, \tag{4}$$

where d_1 is the width of the heterobarrier, and l_w is the width of the quantum well.

The potential difference applied to the structure shifts the bottom of the quantum well downward while also shifting the energy levels relative to the emitter. Apart from the external voltage, the location of the

electron energy level in the quantum well is affected by the electric field produced by the electrons in it. It is this field that leads to the bistability of the current–voltage characteristic. Let us determine the total shift of the energy level of an electron, ΔE, in the quantum well as a function of the external voltage and the charge carrier density. Assuming the electric field produced by the charges in the quantum well to the field of a charged plane, we derive an expression for the energy level shift in the quantum well [4]:

$$\Delta E = e\phi - \frac{eU}{2} = \frac{2\pi e^2 (d_1 + l_w/2)n}{\epsilon} - \frac{eU}{2}. \quad (5)$$

Here, n is the two-dimensional electron density in the quantum well, ϵ is the permittivity, and U is the applied voltage.

Figure 2 shows the energy band structure without the field of the electrons in the quantum well (solid line) and with its influence on potential (5) (dashed line). It follows from Fig. 2 that the bottom of the quantum well shifts upward in the field produced by the electrons in the quantum well. The higher the density of the electrons in the well, the stronger the field produced by them and the higher the shift of the well bottom relative to the emitter, along with the location of the energy level. It thus follows that bistability can emerge when, depending on the electron density, the energy level in the quantum well is below or above the bottom of the emitter conduction band. In the former case, the current through the structure will be weak; in the latter case, a resonant current will be observed.

To calculate the current through the structure, we use a model in which the current of coherent electrons (passing through the heterostructure without reflection) is assumed to be weak due to the scattering processes in the quantum well. Experimental circumstantial evidence for this fact was given by Zaslavsky *et al.* [13], who studied the tunneling through an asymmetric resonant tunnel heterostructure. These authors considered the forward (the first barrier of the heterostructure is weaker than the second barrier) and backward (the second barrier of the heterostructure is weaker than the first barrier) shifts under the applied potential difference; the widths of the heterostructure barriers were equal. The coherent tunneling model (the Fabry–Perot resonator model) must yield equal values for the coherent contribution to the current in both cases. However, for the forward shift (a *Z*-shaped current–voltage characteristic), the peak current is an order of magnitude larger than that for the backward shift. Even if we assume that the entire current for the backward shift is determined by the coherent tunneling, this current may be ignored for the forward shift.

Thus, electrons from the emitter tunnel through the first barrier into the quantum well from which they tunnel into the collector after multiple reflections from the barrier walls. Since we assume the states in the emitter

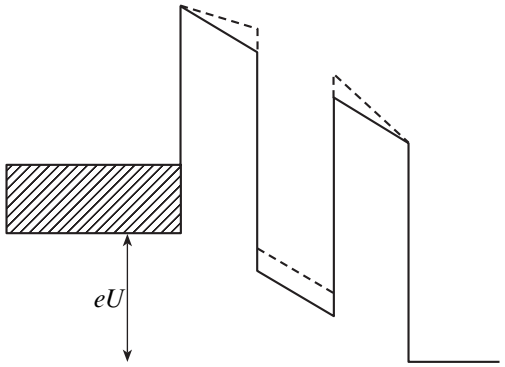


Fig. 2. The potential of the system in the absence of a magnetic field with (solid line) and without (dashed line) electron screening.

to be constantly occupied, the electrons from the quantum well cannot return to the emitter because of the Pauli exclusion principle. The current from the emitter into the quantum well is determined by the potential of the system (the location of the energy level in the quantum well relative to the emitter) and, hence, depends on the applied voltage and the charge carrier density in the quantum well. In turn, the current from the quantum well into the collector does not depend on the voltage and is proportional to the charge carrier density in the quantum well. To obtain the current–voltage characteristic, we must seek for the intersection of the dependences of the currents into and out of the quantum well on the charge carrier density in the well at fixed voltages.

3. THE TUNNEL MATRIX ELEMENT

To calculate the current from the emitter into the quantum well, we use the formalism of the tunnel matrix element suggested by Bardeen [18]. This issue is studied in more detail in [19, 20]. Bardeen applied the following approach to this problem. Rather than introducing the states that would be exact solutions of some approximate Hamiltonian, he introduced approximate solutions of the exact Hamiltonian. Following Bardeen, we choose these states of the charge carriers in the subbarrier region:

$$\psi_w(z) = b_w e^{\kappa_2 z}, \quad z \le d_1, \quad (6)$$

$$\psi_e(e) = b_e e^{-\kappa_2 z}, \quad z \ge 0, \quad (7)$$

where ψ_w is the wave function of the electron to the right of the barrier, which decreases at $z \le 0$ rather than satisfying the Schrödinger equation in this range. Analogously, ψ_e, the wave function of the electron in the emitter, continues to decrease at $z \ge d_1$. In this case, ψ_w is the exact solution of the problem with the Hamilto-

nian $\hat{H}$ for $z \geq 0$, and ψ_e is the exact solution at $z \leq d_1$. The tunnel matrix element is [18, 19]

$$T_{we} = \int \psi_w^*(\hat{H} - E_e)\psi_e dz. \quad (8)$$

Using functions (6), we obtain

$$T_{we} = \frac{\hbar^2 \kappa_2}{m} b_w^* b_e. \quad (9)$$

Using the boundary conditions for the wave function and its derivative at the boundary of the barrier at $z = d_1$, we can derive the following expression for the coefficient b_w:

$$b_w = \frac{2ik_3}{\kappa_2 + ik_3}\sqrt{\frac{1}{l_w}}e^{-\kappa_2 d_1}, \quad (10)$$

where $k_3 = \hbar^{-1}\sqrt{2mE_w}$.

To calculate the coefficient b_e, we use the quasi-classical wave functions in the emitter and the subbarrier region

$$\psi_e = \sqrt{\frac{2m\omega}{\pi p}}\cos\left(\frac{i}{\hbar}\int_a^z p dz + \frac{\pi}{4}\right), \quad z < a, \quad (11)$$

$$\psi_e = \frac{1}{2}\sqrt{\frac{2m\omega}{\pi p}}\exp\left(-\frac{1}{\hbar}\int_a^z |p| dz\right), \quad z > a, \quad (12)$$

where a is the turning point (Fig. 1) and $w = 2\pi\nu$ is the electron–barrier collision frequency in the emitter. Thus, b_e can be written as

$$b_e = \frac{1}{2}\sqrt{\frac{2m\omega}{\pi|p|}}\exp\left(-\frac{1}{\hbar}\int_a^0 |p| dz\right). \quad (13)$$

To simplify the latter expression, we introduce $l_e = |z_0 - a|$. Given that

$$|p| = \sqrt{2m(U-E)} = m\omega_c\sqrt{(z-z_0)^2 - l_e^2},$$

and integrating in (13), we obtain

$$b_e = \frac{1}{2}\sqrt{\frac{2m\omega}{\pi|p|}}\exp\left(-\frac{m\omega_c}{2\hbar}[|z_0|\sqrt{z_0^2 - l_e^2} - l_e^2\ln(|z_0| + \sqrt{z_0^2 - l_e^2}) + l_e^2\ln l_e]\right). \quad (14)$$

Substituting (14) and (10) into (9) yields the final expression for the tunnel matrix element

$$|T_{we}|^2 = \frac{4k_3^2}{k_3^2 + \kappa_2^2}\frac{\hbar^3\kappa_2\nu}{ml_w} \times \exp\left(-2\kappa_2 d_1 - \frac{m\omega_c}{\hbar}[|z_0|\sqrt{z_0^2 - l_e^2} - l_e^2\ln(|z_0| + \sqrt{z_0^2 - l_e^2}) + l_e^2\ln l_e]\right). \quad (15)$$

The electrons in the emitter for which the relation $|b - z_0| > |z_0|$ holds (see the level E_1 in Fig. 1) do not reach the barrier; the quasi-classical frequency of their collisions with the walls of the potential well is

$$\nu = \omega_c/2\pi. \quad (16)$$

For the remaining electrons in the emitter (see the level E_2 in Fig. 1), the electron–barrier collision frequency can be determined by using the quasi-classical relations

$$T = 2\int_{b_2}^0 \frac{1}{v(z)}dz,$$
$$v(z) = \frac{1}{m}\sqrt{2m(E-U)} = \omega_c\sqrt{(b_2 - z_0)^2 - (z - z_0)^2}, \quad (17)$$
$$E = \frac{m\omega_c^2(b_2 - z_0)^2}{2}, \quad U = \frac{m\omega_c^2(z - z_0)^2}{2}.$$

Here, T is the oscillation period of the electron in the emitter, and E is one of the dimensional quantization energy levels in the emitter, which can be found by using the quasi-classical approximation.

Integrating (17) yields

$$T = \frac{\pi}{\omega_c} + \frac{2}{\omega_c}\arcsin\left(\frac{z_0}{b_2 - z_0}\right). \quad (18)$$

Consequently, the electron–barrier collision frequency is

$$\nu = \frac{1}{T} = \frac{\omega_c}{\pi + 2\arcsin\left(\frac{z_0}{b_2 - z_0}\right)}. \quad (19)$$

For $b_2 - z_0 = z_0$, the barrier no longer affects the quantization conditions and ν takes a simpler form that agrees with (16):

$$\nu = \omega_c/2\pi. \quad (20)$$

For $b = z_0$, the collision frequency doubles:

$$\nu = \omega_c/\pi. \tag{21}$$

4. CALCULATING THE CURRENT FROM THE EMITTER INTO THE QUANTUM WELL

Knowing the tunnel matrix element, we can now calculate the current through the heterostructure. According to the golden rule of quantum mechanics, the probability of the transition from state a to state b per unit time can be expressed in terms of the matrix element as

$$P_{ba} = \frac{2\pi}{\hbar}|T_{ba}|^2 A_b(E_b). \tag{22}$$

Here, T_{ba} is the tunnel matrix element of the transition from state a to state b, and A_b describes the density of the states in b. Taking into account the number of electrons in the emitter and the Pauli exclusion principle, we can express the density of the current from the emitter into the quantum well as [7]

$$J_{e\mathrm{w}} = \frac{2e}{L_x L_y} \times \sum \frac{2\pi}{\hbar}|T_{\mathrm{w}e}|^2 A_{\mathrm{w}}(E_z) f(E)(1 - f_{\mathrm{w}}(E)). \tag{23}$$

Here, the summation is over the electron quantum states in the emitter, $f(E)$ is the Fermi–Dirac distribution function for the charge carriers in the emitter, f_{w} is the Fermi–Dirac distribution function in the quantum well, T_{we} is the tunnel matrix element, L_x and L_y are the cross-sectional dimensions of the structure, and A_{w} is described by the Lorentz function:

$$A_{\mathrm{w}} = \frac{1}{2\pi}\frac{\Gamma}{(E - E_{\mathrm{w}})^2 + \Gamma^2/4}. \tag{24}$$

Here, $\Gamma = \Gamma_l + \Gamma_r + \Gamma_s$ characterizes the smearing of the dimensional quantization energy level in the quantum well; Γ_l and Γ_r describe the reciprocals of the tunneling times from the quantum well through the first and second barriers, respectively; Γ_s characterizes the reciprocal of the scattering time in the quantum well; and E_{w} is the particle energy in the quantum well measured from its bottom. The reciprocals of the tunneling times from the well into the emitter and the collector can be estimated by using the formula

$$\Gamma_r = \hbar/\tau = \hbar\nu_{\mathrm{w}} D. \tag{25}$$

Here, ν_{w}, the classical electron–barrier collision frequency, is expressed in terms of the energy of the dimensional quantization level in the well and the width of the quantum well, and D is the barrier penetration coefficient.

Given the screening of the potential by the electrons (5) accumulated in the quantum well and the influence of the magnetic field (4), A_{w} takes the form

$$A_{\mathrm{w}} = \frac{1}{2\pi}\Gamma\left[\left(E_z - E_{\mathrm{w}} - \frac{2\pi e^2(d_1 + l_{\mathrm{w}}/2)n}{\epsilon} + \frac{eU}{2} - \Delta U(z_0)\right)^2 + \frac{\Gamma^2}{4}\right]^{-1}. \tag{26}$$

To calculate the current from the emitter into the quantum well, we must add up the contributions from all of the electron oscillator centers. The problem is simplified by the fact that, the farther the position of the oscillator center from the barrier, the smaller its contribution to the current, because distant oscillators need to overcome not only the potential barrier of the heterostructure, but also the region under the parabolic potential to fall into the quantum well (see E_1 in Fig. 1). For this reason, their tunneling probability decreases sharply. Let us calculate the energy levels for each oscillator and add up their contributions to the current.

The quasi-classical approximation, which is known to provide a good accuracy for the parabolic potential, is used to calculate the energy levels of the oscillators in the emitter. To determine the energy of the electrons with a turning point at the heteroboundary ($|b - z_2| > |z_0|$, the energy level E_2 in Fig. 1), we use the quasi-classical quantization condition

$$\pi\hbar\left(n_l + \frac{1}{2}\right) = \int_{b_2}^{0} dz\, p_z = \frac{1}{4}m\omega_{\mathrm{c}}|b_2 - z_0| \times \left[\pi z_0 - \pi b_2 + 2\arcsin\left(\frac{z_0}{b_2 - z_0}\right)|z_0 - b_2| - 2z_0\sqrt{1 - \frac{z_0^2}{(b_2 - z_0)^2}}\right], \tag{27}$$

where

$$p_z = \sqrt{2m(E_z - U)} = m\omega_{\mathrm{c}}\sqrt{(b_2 - z_0)^2 - (z - z_0)^2}. \tag{28}$$

Solving Eq. (27) for b_2 at various values of n_l, we can find the energy levels

$$E_{nl} = \frac{m\omega_{\mathrm{c}}^2(b_2 - z_0)^2}{2}. \tag{29}$$

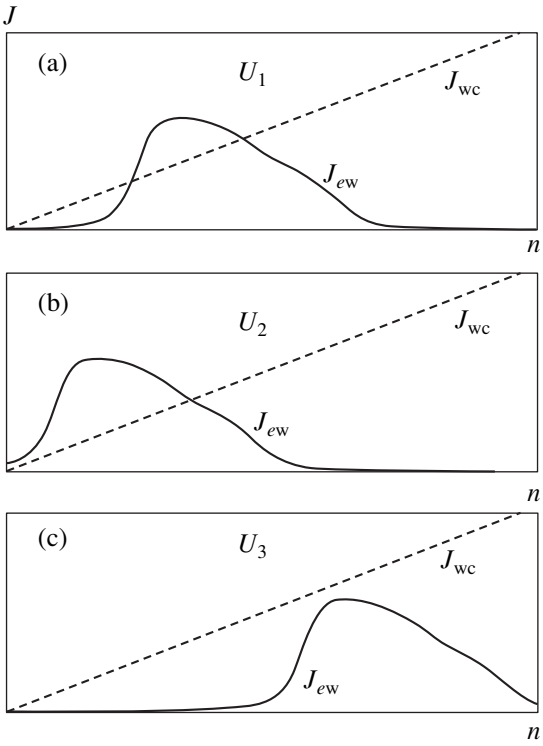


Fig. 3. Currents from the emitter into the quantum well (solid line) and from the well into the collector (dashed line) versus electron density in the quantum well for voltages $U_2 < U_1 < U_3$. The intersections of the curves specify the points of the current–voltage characteristic.

When the condition $|b - z_0| > |z_0|$ is satisfied for the electrons in the emitter (the energy level E_1 in Fig. 1), the barrier does not affect their spectrum:

$$E_{nl} = \hbar\omega_c(n_l + 1/2). \tag{30}$$

We can now calculate the density of the current from the emitter into the quantum well by adding up the contributions from all oscillators:

$$J_{ew} = \frac{4e}{L_x\hbar^2}\sum_{z_0,\,n_l}^{E_{nl(z_0)}<E_f}\sqrt{2m(E_f - E_{nl(z_0)})}\,|T_{we}|^2\frac{1}{2\pi}$$

$$\times\Gamma\left[\left(E_z - E_w - \frac{2\pi e^2(d_1 + l_w/2)n}{\epsilon}\right.\right. \tag{31}$$

$$\left.\left. + \frac{eU}{2} - \Delta U(z_0)\right)^2 + \frac{\Gamma^2}{4}\right]^{-1}.$$

The current from the quantum well into the collector is

$$J_{wc} = \frac{e}{\hbar}\Gamma_r n. \tag{32}$$

5. CONSTRUCTING A STATIONARY CURRENT–VOLTAGE CHARACTERISTIC

As has been noted above, the current–voltage characteristic is stationary when the current from the emitter into the quantum well is equal to the current from the well into the collector, $J_{ew} = J_{wc}$. Figure 3 presents the $J_{ew}(n)$ and $J_{wc}(n)$ curves for a fixed voltage in the absence of a magnetic field. Figure 3a shows such a potential difference across the structure that the energy level in the quantum well at $n = 0$ is below the bottom of the emitter conduction band. We see from the figure that the current from the emitter into the quantum well increases as the density increases, i.e., as the energy level in the quantum well shifts upward. This is because the energy level in the well rises above the bottom of the emitter conduction band. Consequently, several intersections of the J_{ew} and J_{wc} curves are possible: one intersection with a low electron density in the quantum well (see Fig. 3c), three intersections (one with a low density and two with a high density) (see Fig. 3a), and one intersection with high densities (see Fig. 3b).

The case of three intersections must yield three values in the current–voltage characteristic, but the middle value of the current is unstable against electron density fluctuations in the quantum well. The reason is as follows: at the middle point of intersection, $J_{ew}(n)$ is sharper than $J_{wc}(n)$, and the current from the emitter into the well increases faster than does the current from the well into the collector for a small positive density variation, which leads to the accumulation of electrons in it until the currents become equal. Thus, a positive electron density fluctuation in the quantum well will cause a switch to the third point of intersection of the $J_{ew}(n)$ and $J_{wc}(n)$ curves. In contrast, for the first and third points of intersection, the electron density fluctuation in the quantum well will relax, because $J_{wc}(n)$ varies faster than $J_{ew}(n)$; thus, these points of intersection are stable. A *Z*-shaped current–voltage characteristic is observed for the case illustrated by Fig. 3a.

Solving the equation $J_{ew}(n) = J_{wc}(n)$ for the electron density in the quantum well for various voltages yields the current–voltage characteristic.

For our calculations, we use parameters of the heterostructure for which a *Z*-shaped current–voltage characteristic can be observed [15]: a GaAs quantum well 56 Å in width enclosed between two $Al_xGa_{1-x}As$ barriers 85 Å in width with $x = 0.4$ and 0.5, $E_f = 40$ meV, and $E_w = 93$ meV.

6. ANALYSIS OF THE CURRENT–VOLTAGE CHARACTERISTIC

The computed current–voltage characteristics are shown in Fig. 4. We see that the peak current decreases with increasing magnetic field strength, which is accompanied by a narrowing of the bistability region. There is a critical magnetic field strength at which the

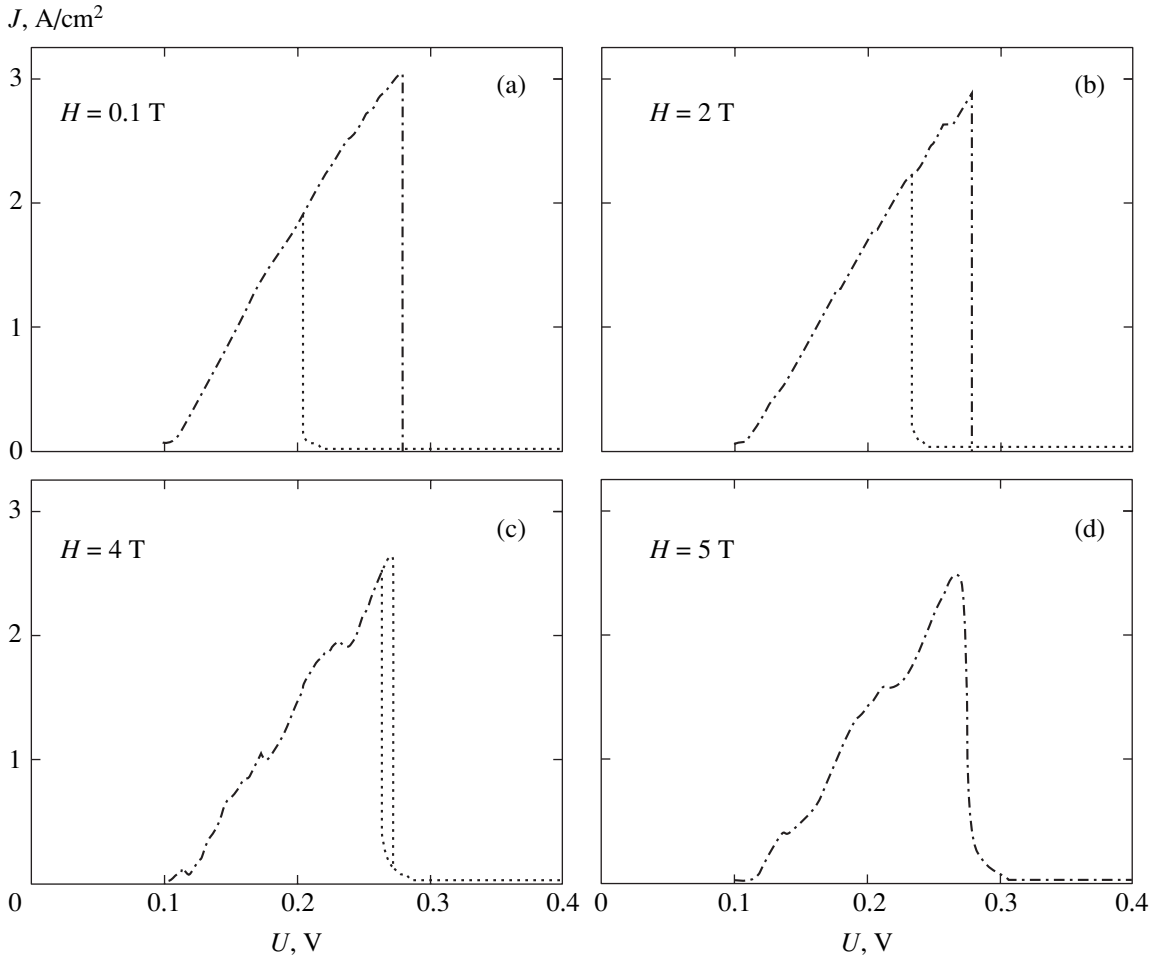


Fig. 4. Computed current–voltage characteristics in a magnetic field of strength H transverse to the current.

Z-shaped current–voltage characteristic transforms into an N-shaped one.

The current–voltage characteristic exhibits a Z-shaped pattern in the range of voltages in which the energy level in the quantum well is near the bottom of the emitter conduction band. If a sufficient number of electrons is accumulated in the quantum well at such voltages, then its bottom shifts upward—a resonant current is observed; if the number of electrons in the quantum well is small, then the energy level in it is below the bottom of the conduction band—the current is weak.

For the bistability to exist, the currents from the emitter into the quantum well and from the well into the collector must be equal at two different electron densities. This requires that $J_{ew}(n)$ be sharp enough in the range of densities after the first intersection with $J_{wc}(n)$. The width of the region of the Z-shaped pattern is characterized by the peak $J_{ew}(n)$ (see Fig. 6); the larger the peak $J_{ew}(n)$, the wider the region of the Z-shaped pattern in the current–voltage characteristic. If, however, the slope and peak value of the current from the emitter into the quantum well decrease, then the bistability region narrows.

The peak current decreases, because the location of the energy level in the quantum well shifts by different values (4) for different oscillators. Therefore, not all of the oscillators contribute to the current at the same voltage; the dependence of J_{ew} on the charge carrier density in the quantum well is smeared, which is accompanied by a decrease in the peak current. In Fig. 5, the currents from the emitter into the quantum well and from the well into the collector are plotted against electron density in the presence and absence of a magnetic field. It follows from Fig. 5 that, in the presence of a strong magnetic field, the $J_{ew}(n)$ curve is smoother and cannot have three points of intersection with the $J_{wc}(n)$ curve.

7. DEPENDENCE OF THE WIDTH OF THE Z-SHAPED PATTERN ON TRANSVERSE MAGNETIC FIELD STRENGTH

Figure 5 shows that a magnetic field causes the dependence of the current from the emitter into the quantum well on the electron density in the well, $J_{ew}(n)$, to be smeared. At a critical magnetic field, this smear-

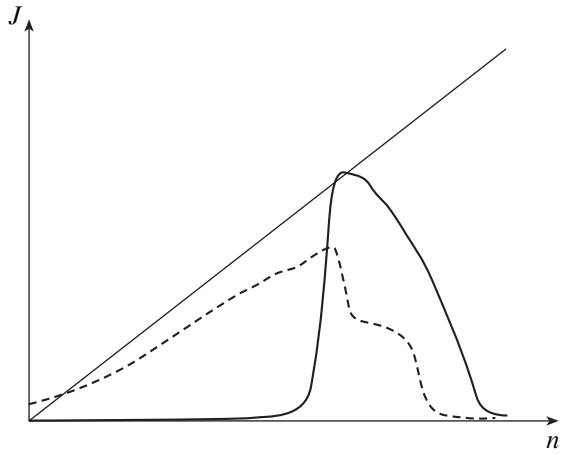


Fig. 5. The currents from the emitter into the quantum well in the absence of a magnetic field (solid line) and in a transverse magnetic field (dashed line); the current from the quantum well into the collector (straight line).

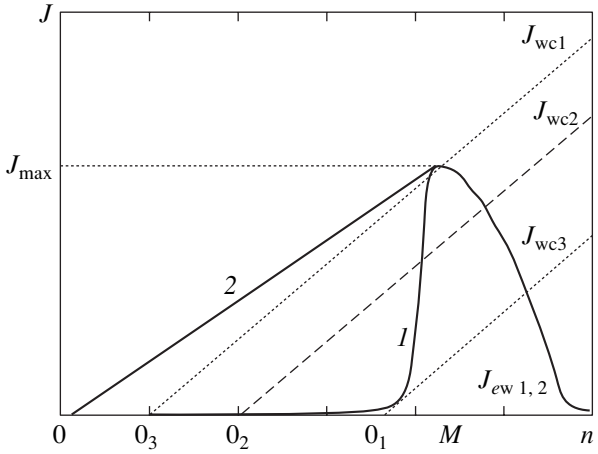


Fig. 6. Current versus density at various voltages: J_{wc1}, J_{wc2}, J_{wc3} are the currents from the quantum well into the collector at various voltages; 0_1, 0_2, 0_3 ate the points of intersection of the straight lines J_{wc} with the x axis that correspond to a zero electron density in the quantum well; J_{ew} is the current from the emitter into the quantum well in the absence of a magnetic field; *1* is the current from the emitter into the quantum well in a transverse magnetic field weaker than the critical one; *2* is the current from the emitter into the quantum well in a transverse magnetic field stronger than the critical one.

ing is so significant that $J_{ew}(n)$ and $J_{wc}(n)$ cannot have two intersections, and, hence, the current–voltage characteristic cannot have bistability.

To analyze the smearing of the $J_{ew}(n)$ curve, let us examine Fig. 6 in which the current through the heterostructure is plotted against the electron density in the quantum well. It follows from (5) that a change in the potential difference applied to the structure causes the coordinate origin for $J_{ew}(n)$ to shift. Thus, a zero electron density in the quantum well for the voltage U_i applied to the structure corresponds to the point of intersection of the x axis of the straight line $J_{wci}(n)$ (0_i) (where $i = 1, 2, 3$). The higher the applied voltage, the larger the leftward shift of the coordinate origin; i.e., the voltage U_1 corresponding to 0_1 is higher than the voltage U_2 corresponding to 0_2 (see Fig. 6). In other words, the coordinate origin in Fig. 6 is different for the J_{wc1}, J_{wc2}, and J_{wc3} curves, because the energy level in the quantum well sinks in energy under the voltage applied to the heterostructure, and a large number of electrons must be accumulated in the quantum well to return this level to its original location. It follows from Fig. 6 that three points of intersection of the $J_{ew}(n)$ and $J_{wc}(n)$ curves are possible in the range between the voltages corresponding to the coordinate origins 0_1 and 0_3.

The shift of the coordinate origin along the x axis, $|0_1 0_2|$, for the change in voltage $\Delta U = U_1 - U_2$ is given by the relation (see (5))

$$\frac{2\pi e^2 (d_1 + l_w/2)|0_1 0_2|}{\epsilon} - \frac{e|U_1 - U_2|}{2} = 0. \qquad (33)$$

Denote the point on the x axis that corresponds to the peak of the $J_{ew}(n)$ curve by M. The width of the bistability region in the current–voltage characteristic is proportional to the length of the segment $|0_3 0_1|$:

$$|0_3 0_1| = |0_3 M| - |0_1 M|. \qquad (34)$$

The length of the segment $|0_3 M|$ can be determined from relation (32)

$$|0_3 M| = \hbar J_{max}/e\Gamma_r. \qquad (35)$$

To determine the length of the segment $|0_1 M|$, i.e., the smearing of the $J_{ew}(n)$ curve, we must determine the location of the energy level in the quantum well relative to the emitter at which the current from the emitter into the well reaches its maximum (corresponds to point M in Fig. 6) and the location at which the current decreases to zero (corresponds to point 0_1 in Fig. 6).

We estimate the shift of the energy level in the quantum well relative to the emitter at which the current through the structure decreases appreciably from the condition that the matrix element as a function of the coordinate of the magnetic oscillator center z_0 be small:

$$B|T_{we}(z_{cr})|^2 = |T_{we}(0)|^2, \qquad (36)$$

where B varies within the range 10–100.

Hence, the expression for z_{cr} (15) is

$$z_{cr} = \sqrt{2\hbar \ln B/m\omega_c}. \qquad (37)$$

The shift of the energy level in the quantum well rela-

tive to the emitter for the oscillator with $z_0 = z_{cr}$ is

$$\Delta_1 = E_w + \frac{m\omega_c^2}{2}\left(z_{cr} + d_1 + \frac{l_w}{2}\right)^2. \quad (38)$$

The current reaches its maximum when the energy level in the quantum well coincides with the lower energy level of the oscillator whose center lies at the left boundary of the first heterobarrier ($z_0 = 0$). In this case, the shift of the energy level in the quantum well relative to the emitter is given by

$$\Delta_2 = E_w + \frac{m\omega_c^2}{2}\left(d_1 + \frac{l_w}{2}\right)^2. \quad (39)$$

Thus (see (5)),

$$|0_1 M| = \frac{\epsilon(\Delta_1 - \Delta_2)}{2\pi e^2(d_1 + l_w/2)}. \quad (40)$$

Substituting the latter expression into (34) yields

$$|0_1 0_3| = \frac{\hbar J_{max}}{e\Gamma_r} - \frac{\epsilon}{2\pi e^2(d_1 + l_w/2)} \times \frac{m\omega_c^2}{2}(2z_{cr}d_1 + z_{cr}l_w + z_{cr}^2). \quad (41)$$

The width of the bistability region in the current–voltage characteristic takes the form

$$\Delta U = \frac{4\pi(d_1 + l_w/2)}{\epsilon}\frac{\hbar J_{max}}{\Gamma_r} - \frac{m\omega_c^2}{e}(2z_{cr}d_1 + z_{cr}l_w + z_{cr}^2). \quad (42)$$

Taking into account (37), we obtain

$$\Delta U = \frac{4\pi(d_1 + l_w/2)}{\epsilon}\frac{\hbar J_{max}}{\Gamma_r} - \frac{eH^2}{mc^2}\sqrt{\frac{2c\hbar \ln B}{eH}}\left(2d_1 + l_w + \sqrt{\frac{2c\hbar \ln B}{eH}}\right). \quad (43)$$

The critical magnetic field strength can be calculated from (43). As follows from Fig. 6, point 0_1 shifts leftward as the magnetic field strength increases; thus, the region of the *Z*-shaped pattern in the current–voltage characteristic narrows. At the critical magnetic field strength at which $|0_1 M| = |0_3 M|$, the *Z*-shaped pattern disappears; i.e., the region of the *Z*-shaped pattern has a zero width. For example, we see from Fig. 6 that the $J_{wc}(n)$ curve cannot intersect curve *2* at three points. The width of the bistability region is plotted against the magnetic field strength for $B = 30$ in Fig. 7. To simplify our calculation, we disregarded the influence of the magnetic field on the peak current J_{max}.

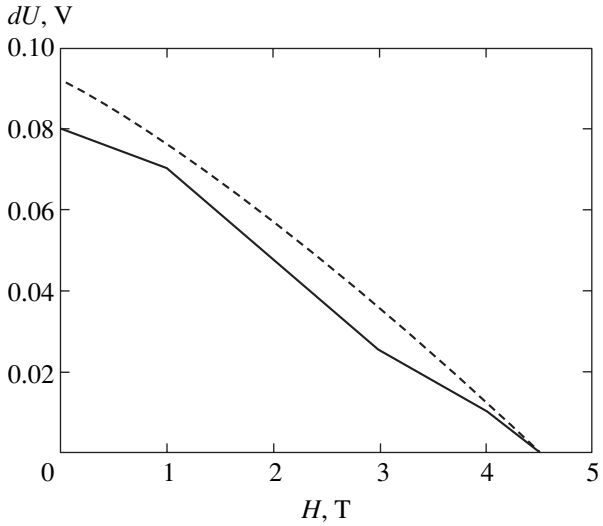


Fig. 7. Width of the bistability region versus magnetic field strength: the solid and dashed lines represent the numerical and analytical calculations, respectively.

It follows from Eq. (43) and Fig. 7 that there is a critical magnetic field strength at which the width of the region of the *Z*-shaped current–voltage characteristic, ΔU, becomes zero. As the transverse magnetic field strength increases further, the current–voltage characteristic transforms into an *N*-shaped one.

Our results are in agreement with the experimental data obtained by Zaslavsky *et al.* [15], who investigated a heterostructure with a GaAs quantum well 56 Å in width between two $Al_xGa_{1-x}As$ barriers 85 Å in width with $x = 0.4$ and 0.5. The authors observed a *Z*-shaped current–voltage characteristic in the case of a positive potential difference. According to experimental data, the bistability region disappears when the magnetic field strength is equal to 5 T. This result is consistent with our calculation.

8. CONCLUSIONS

Using the formalism of the tunnel matrix element, we calculated the current–voltage characteristic of a resonant tunnel diode in a magnetic field perpendicular to the current. We took into account the contribution from electrons with various positions of the magnetic oscillator center to the current. Our analysis shows that, in contrast to a longitudinal magnetic field, a transverse magnetic field strongly affects the shape of the current–voltage characteristic. It follows from our data that the region of the *Z*-shaped current–voltage characteristic narrows as the magnetic field strength increases. There is a critical magnetic field strength at which the bistability region disappears and the current–voltage characteristic transforms into an *N*-shaped one. In addition, the peak current slightly decreases as the magnetic field

strengthens. The reason is that for electrons with different positions of the magnetic oscillator center, the energy level in the quantum well shifts by different values relative to the emitter. As a result, the current through the heterostructure decreases, the potential screening effect is smoothed out, and, hence, the region of the *Z*-shaped current–voltage characteristic narrows.

ACKNOWLEDGMENTS

This study was supported in part by the INTAS (grant no. 0175wp), the Russian Foundation for Basic Research (project nos. 04-07-90148b, 04-02-16786, and 02-02-17610), and the Program of the Russian Academy of Sciences.

REFERENCES

1. T. Waho, K. J. Chen, and M. Yamamoto, IEEE J. Solid-State Circuits **33**, 268 (1998).
2. L. L. Chang, L. Esaki, and R. Tsu, Appl. Phys. Lett. **24**, 593 (1974).
3. B. Ricco and M. Ya. Azbel, Phys. Rev. B **29**, 1970 (1984).
4. V. J. Goldman, D. C. Tsui, and J. E. Cunningham, Phys. Rev. Lett. **58**, 1256 (1987).
5. J. Zang and J. L. Birman, Phys. Rev. B **46**, 5020 (1992).
6. N. C. Kluksdahl, A. M. Kriman, D. K. Ferry, and C. Ringhofer, Phys. Rev. B **39**, 7720 (1989).
7. M. Meixner, P. Rodin, E. Schöll, and A. Wacker, Eur. Phys. J. B **13** (1), 157 (2000).
8. J. O. Sofo and C. A. Balseiro, Phys. Rev. B **42**, 7292 (1990).
9. T. C. L. G. Sollner, Phys. Rev. Lett. **59**, 1622 (1987).
10. F. A. Buot, Phys. Rep. **234**, 73 (1993).
11. W. Pötz, Phys. Rev. B **41**, 12111 (1990).
12. P. Orellana, F. Claro, E. Anda, and S. Makler, Phys. Rev. B **53**, 12967 (1996).
13. A. Zaslavsky, D. C. Tsui, M. Santos, and M. Shayegan, Phys. Rev. B **40**, 9829 (1989).
14. M. I. D'yakonov and M. É. Raĭkh, Zh. Éksp. Teor. Fiz. **88**, 1898 (1985) [Sov. Phys. JETP **61**, 1125 (1985)].
15. A. Zaslavsky, Y. P. Li, D. C. Tsui, *et al.*, Phys. Rev. B **42**, 1374 (1990).
16. V. M. Genkin and A. S. Mel'nikov, Fiz. Tverd. Tela (Leningrad) **31** (8), 249 (1989) [Sov. Phys. Solid State **31**, 1418 (1989)].
17. G. Yong, Y. Li, X. Kong, and C. Wei, Phys. Rev. B **50**, 17249 (1994).
18. J. Bardeen, Phys. Rev. Lett. **6**, 59 (1961).
19. E. O. Kane, in *Tunneling Phenomena in Solids*, Ed. by E. Burstein and S. Lundqvist (Plenum, New York, 1969; Mir, Moscow, 1973).
20. W. A. Harrison, Phys. Rev. **123**, 85 (1961).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 157–168.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 181–193.

SOLIDS
Electronic Properties

Transition in a Luttinger Liquid with Two-Level Impurities to the Fermi-Liquid State

L. A. Manakova
Russian Research Centre Kurchatov Institute, Moscow, 123182 Russia
e-mail: manakova@kurm.polyn.kiae.su
Received November 4, 2003

Abstract—The mechanisms leading to instability of the non-Fermi-liquid state of a Luttinger liquid with two-level impurities are proposed. Since exchange scattering in 1D systems is two-channel scattering in a certain range of parameters, several types of non-Fermi-liquid excitations with different quantum numbers exist in the vicinity of the Fermi level. These excitations include, first, charge density fluctuations in the Luttinger liquid and, second, many-particle excitations due to two-channel exchange interaction, which are associated with band-type as well as impurity fermion states. It is shown that mutual scattering of many-particle excitations of various types leads to the emergence of an additional Fermi-liquid singularity in the vicinity of the Fermi level. The conditions under which the Fermi-liquid state with a new energy scale (which is much smaller than the Kondo temperature) is the ground state of the system are formulated. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The interest in low-dimensional, in particular, one-dimensional (1D), systems has been revived in recent years in connection with the obtaining of degenerate quantum gases in quasi-one-dimensional magnetic and optical traps as well as in 1D optical gratings. The previous peak in interest in 1D systems about ten years ago (see references cited in [1, 2] and in this work) was due to the development of so-called quantum wires, viz., 1D electronic systems in inversion layers of GaAs [3]. Degenerate atomic Fermi gases have already been obtained in traps [4–7]. It should be noted that Fermi gases with practically any number of particles and any intensity of the interaction between them can be produced in traps. This can be done by using the Feschbach resonance [7]. Thus, we can expect that various remarkable properties predicted theoretically for 1D systems both in the framework of the low-energy Luttinger model [8], as well as using exactly solvable models (see, for example, [9]), can be experimentally observed in such systems with a quasi-one-dimensional geometry.

The properties of 1D metals in a fairly wide range of parameters can be described using the Luttinger model (see, for example, recent reviews [1, 2] and the literature cited therein). Accordingly, such materials are referred to as Luttinger liquids (in contrast to the Fermi liquid in the 3D case). In such systems, even a weak interaction leads to a qualitative rearrangement of the excitation spectrum at low energies. Namely, 1D metals have no well-defined one-particle excitations. The only stable excitations in the vicinity of the Fermi level are collective charge and spin density fluctuations (acoustic modes). These excitations are dynamically independent, which corresponds to complete separation of the spin and charge degrees of freedom. The interactions also lead to a power decay of all correlation functions over large distances and times. The response to local perturbations is an important problem for real systems. In this case, the behavior of 1D systems also differs qualitatively from the situation in 3D metals. It was shown in the famous publication by Kane and Fisher [10] that the potential scattering of right fermions by left ones (so-called backward scattering) in a Luttinger liquid with a repulsive interaction leads to complete reflection of excitations from the potential at low energies so that only a weak tunneling through the barrier takes place. The X-ray response in a Luttinger liquid has been studied extensively [11–14] (including the situation with backward scattering [14]). The exchange interaction with a spin impurity is one of the central problems for strongly correlated systems [15]. It should be noted that if the number of electron channels participating in the exchange interaction exceeds double the impurity spin, the system has a non-Fermi-liquid fixed point, exhibiting an abnormal behavior of heat capacity and susceptibility [16, 17]. The Kondo effect in Luttinger liquids was studied in [18–20]. It was shown that, as in the 3D case, the problem could be renormalized to the strong-coupling limit. However, two distinguishing features are observed. First, the Kondo effect in a Luttinger liquid exists both for the antiferromagnetic and ferromagnetic interactions. Second, the system has three fixed points, two of which correspond to the one-channel Kondo behavior and one exhibits the two-channel (i.e., non-Fermi-liquid) behavior (we consider the impurity spin 1/2). The conditions for a stable two-channel Kondo behavior relative to the exchange backward scattering were obtained in [20].

It is important for subsequent analysis to emphasize that, in the absence of interaction between fermions, the non-Fermi-liquid (NFL) state associated with a multichannel (two-channel) exchange interaction is unstable to any mechanism removing the degeneracy of the channel participating in exchange scattering. In particular, the instability of the NFL state in the two-channel Kondo model upon the introduction of anisotropy of exchange constants in different channels was considered in [21]. The instability of the same state to potential scattering of many-particle excitations with different quantum numbers in quantum-dimensional structures and in metals containing impurities of *d*- or *f*-elements was demonstrated in [22, 23].

In this study, the mechanisms leading to instability of the NFL state of a Luttinger liquid with two-level (pseudospin) impurities are proposed. Since exchange scattering in 1D systems is of the two-channel type in a certain range of parameters, several types of NFL excitations exist in the vicinity of the Fermi level: density fluctuations of the Luttinger liquid in the charge channel and many-particle excitations generated by the two-channel exchange interaction in the pseudospin channel. It will be shown below that allowance for resonant scattering of 1D fermions (along with their potential backward scattering) from many-particle excitations generated by the two-channel exchange interaction leads in our case to the emergence of additional narrow Fermi-liquid resonances in the vicinity of the Fermi level (even for very weak backward scattering) and to the instability of the NFL state in this sense. A transition from the NFL to the FL state is accompanied by the emergence of a new small energy scale and, as a consequence, by an anomalous increase in the density of states at low energies.

2. HAMILTONIAN AND SCATTERING PROBLEM

1. Let us consider the situation when a 1D lattice with a number of fermion per site smaller than unity (metallic state) contains localized fermions with two internal degrees of freedom. The energy of a localized state is much lower than the Fermi level. Localized fermions do not interact with one another. We assume, however, that the wave functions of localized and band fermions may overlap; for this reason, a localized state interacts with band fermions. Since a two-level atom can be described by a pseudospin variable with two values of the *z* component, we will refer to this situation as a pseudospin impurity in a metallic 1D lattice. We assume that the repulsion of fermions at an impurity site is so strong that only one fermion can occupy this site. We will consider below a 1D system with periodic boundary conditions. It should be emphasized that, in the case of an atomic Fermi gas, the pseudospin degrees of freedom in the optical lattice correspond to two intrinsic atomic levels of the hyperfine structure.

The Hamilton of such a system will be written in the form

$$H = \mathcal{H} + H_{sc} \equiv H_{\mathcal{L}} + H_{0d} + H_h + H_{\text{int}} + H_{sc};$$

$$H_{0d} = \sum_{\sigma} E_d n_{d\sigma}; \tag{1}$$

$$H_h = \sum_{k\alpha\sigma} (V_{kd}^{\alpha} a_{k\alpha\sigma}^{+} d_{\sigma} + \text{H.c.}),$$

where $H_{\mathcal{L}}$ is the Hamiltonian of 1D fermions, which coincides with the Hamiltonian in the Luttinger model in the continuous limit; H_{0d} is the Hamiltonian of a localized level with energy E_d; $n_{d\sigma} = d_{\sigma}^{+} d_{\sigma}$; H_h is the hybridization between band and localized fermions; operators $a_{k\alpha\sigma}^{+}$, $a_{k\alpha\sigma}$, $\alpha = R, L \equiv \pm$, describe the right and left 1D fermions: $\psi_{\alpha\sigma}(x) = (1/\sqrt{L}) \sum_k e^{ikx} a_{k\alpha\sigma}$, L being the chain length (it should be recalled here that the relations $N \longrightarrow \infty$ and $L \longrightarrow \infty$ hold in the continuous limit); and $(2\pi N/L) = k_F$, N being the number of fermions and k_F the Fermi momentum. The term H_{sc} describes the potential scattering of band fermions with different quantum numbers (in other words, backward scattering); namely,

$$H_{sc} = \sum_{kk'} \sum_{\alpha \neq a', \sigma} (T_{kk'}^{\alpha\alpha'} a_{k\alpha\sigma}^{+} a_{k'\alpha'\sigma} + \text{H.c.}). \tag{2}$$

The interaction between band fermions and a localized state is generated, apart from hybridization, by the repulsion between fermions at a deep level. In H_{int} from Eq. (1), this interaction correspond to the term

$$H_{\text{int}} = \frac{1}{2} U \sum_{\sigma} n_{d\sigma} n_{d-\sigma}.$$

2. Excitations in a system with Hamiltonian $H = \mathcal{H} + H_{sc}$ are completely determined by the Green function of complex argument z, $\mathbf{G}_d(z) = \langle d|(z - \hat{H})^{-1}|d\rangle$, which can be evaluated using the equations of motion.

The system of equations of motion for $\mathbf{G}_d(z)$ has the form

$$\mathcal{G}_d^{(0)-1}(z)\mathbf{G}_d(z) - \sum_{p\alpha = L, R} V_{dp}^{\alpha} \mathbf{G}_{d\alpha}(p; z) = \hat{1},$$

$$\mathbf{G}_L^{-1}(p; z)\mathbf{G}_{dL}(p; z)$$

$$- \sum_{p'} T_{pp'}^{LR} \mathbf{G}_{dR}(p'; z) - V_{pd}^{L} \mathbf{G}_d(z) = 0, \tag{3}$$

$$\mathbf{G}_R^{-1}(p; z)\mathbf{G}_{dR}(p; z)$$

$$- \sum_{p'} T_{pp'}^{LR} \mathbf{G}_{dL}(p'; z) - V_{pd}^{R} \mathbf{G}_d(z) = 0.$$

Here, $\mathcal{G}_d^{(0)}(z) = \langle d|(z - (H_{0d} + H_{\text{int}})^{-1}|d\rangle$, $\mathbf{G}_\alpha(p; z) = \langle a_{p\alpha}|(z - \hat{H})^{-1}|a_{p\alpha}\rangle$, and $\mathbf{G}_{d\alpha}(k; z) = \langle d|(z - \hat{H})^{-1}|a_{k\alpha}\rangle$ are the Green functions taking into account all interactions, but disregarding scattering. In expressions (3) and below, the spin indices are temporarily omitted.

In the case of separable matrix elements and/or those exhibiting a weak energy dependence and describing potential scattering [in particular, for $T_{kk'}^{\alpha\alpha'} = T_k^\alpha T_{k'}^{*\alpha'}$ $(\nu \neq \nu')$], $\mathbf{G}_d(z)$ can be evaluated exactly.

The solution for $\mathbf{G}_d(z)$ has the form

$$\mathbf{G}_d(z) = \frac{\mathcal{G}_d(z)}{1 - \Sigma_{sc}(z)\mathcal{G}_d(z)}, \quad \mathcal{G}_d^{-1}(z) = \mathcal{G}_d^{(0)-1}(z) - \sum_\alpha \Sigma_\alpha^{VV}(z),$$

$$\Sigma_{\text{sc}}(z) = \frac{\Sigma_L^{VT}(z)[\Sigma_R^{TV}(z) + \Sigma_R^{TT}(z)\Sigma_L^{TV}(z)] + \Sigma_R^{VT}(z)[\Sigma_L^{TV}(z) + \Sigma_L^{TT}(z)\Sigma_R^{TV}(z)]}{1 - \Sigma_L^{TT}(z)\Sigma_R^{TT}(z)}, \tag{4}$$

$$\Sigma_\alpha^{ab}(z) = \sum_p W_\alpha^{ab}(p)\mathbf{G}_\alpha(p; z).$$

Here, $\mathcal{G}_d(z)$ is the Green function of impurity degrees of freedom disregarding scattering processes associated with Hamiltonian H_{sc}. However, by definition, this function includes all interactions between band and localized fermions, which are generated by the repulsion H_{int} of fermions at an impurity site. The self-energy functions $\Sigma_\alpha^{ab}(z)$ are written in the form of spectral expansions of many-particle Green functions for 1D fermions; W_α^{ab} are the products of matrix elements, which are defined by the superscripts of the self-energy functions; and $\rho_\alpha(\varepsilon)$ is the density of states corresponding to the many-particle excitation spectrum taking into account the interactions. This spectrum is determined by the Green function $\mathbf{G}_\alpha(p; z)$. The complete Green function in relations (4) has singularities of two types. Functions $\mathcal{G}_d(z)$ contains singularities generated by the interaction between band fermions and a localized state. The denominators in relations (4) appear due to scattering of various types of many-particle excitations from one another. It follows from the expression for $\Sigma_{\text{sc}}(z)$ that the system experiences, first, potential scattering of right fermions from left ones (the term containing $\Sigma_L^{TT}(z)\Sigma_R^{TT}(z)$), but taking into account the modification of the density of states due to the exchange interaction and, second, the resonance scattering of 1D fermions from many-particle excitations, which are described by the Green function $\mathcal{G}_d(z)$. It will be shown below that in both cases scattering may lead to the emergence of additional singularities (namely, simple poles) in the Green function in the vicinity of the Fermi level. The poles correspond to Fermi-liquid excitations. The position of poles $z_r = \varepsilon_r + i\gamma_r$ is determined from the solution of the equation

$$1 - \Sigma_{\text{sc}}(z_r)\mathcal{G}_d(z_r) = 0. \tag{5}$$

In turn, the self-energy functions can be written in the form of spectral expansions of many-particle Green functions for 1D fermions,

$$\Sigma_\alpha^{ab}(z) = \sum_p \frac{W_\alpha^{ab}(p)}{z - \varepsilon_\alpha(p)} = W_\alpha^{ab}(p_F)\int d\varepsilon \frac{\rho_\alpha(\varepsilon)}{z - \varepsilon} \equiv W_\alpha^{ab}(p_F)\Sigma_\alpha(z), \tag{6}$$

where W_α^{ab} are the matrix elements defined in relations (4), $\varepsilon_\alpha(p)$ is the low-energy excitation spectrum, and $\rho_\alpha(\varepsilon)$ is the density of states corresponding to this spectrum.

Thus, to solve the scattering problem, we must know the many-particle density of states for 1D fermions in the energy range of importance to us (disregarding scattering) as well as the Green function $\mathcal{G}_d(z)$ for the resonance level.

If U is the maximal parameter of the problem, it is natural to solve first the problem with Hamiltonian $\mathcal{H}$ and to determine $\mathcal{G}_d(z)$ with the help of the methods used in the two-channel Kondo problem; the scattering problem can then be solved by substituting the expressions obtained into relations (4). In other words, we first determine low-energy many-particle excitations, assuming that the interactions of fermions with one another and with the impurity are strong, and then take into account the scattering of fermions from one another.

3. TWO-CHANNEL EXCHANGE INTERACTION

1. In the case of noninteracting band fermions, the interaction between these particles and a localized fermion with a spin 1/2 for large values of U can be derived using the method of projection operators or the

Schrieffer–Wolf transformation to Hamiltonian $\mathcal{H}$ with $H_{\mathcal{L}} \longrightarrow H_{00}$ (see, for example, [15]).

The exchange interaction for a 1D system with Hamiltonian $\mathcal{H}$ can also be obtained analogously in the limit of large values of U ensuring a single filling of the localized state. Namely, a transformation analogous to the Schrieffer–Wolf transformation in the lowest order in V_{dk} gives the Hamiltonian

$$H_{\text{eff}} = H_{\mathcal{L}} + H_{\text{ex}} + H_{\text{ex}}^{(B)} + H'_{\text{sc}} + H',$$

$$H_{\text{ex}} = \sum_{\alpha} \sum_{i=x,y,z} \sum_{\sigma \neq \sigma'} J^i \psi^+_{\alpha\sigma}(0) \hat{\tau}^i_{\sigma,\sigma'} \psi_{\alpha\sigma'}(0) \mathbf{S}^i, \quad (7)$$

$$H_{\text{ex}}^{(B)} = \sum_{\alpha \neq \alpha'} \sum_{i=x,y,z} \sum_{\sigma \neq \sigma'} J^i_B \psi^+_{\alpha\sigma}(0) \hat{\tau}^i_{\sigma,\sigma'} \psi_{\alpha'\sigma'}(0) \mathbf{S}^i.$$

Here, $J^i \sim |V_{dk_F}|^2/E_d$ are the exchange interaction constants, $\mathbf{S}$ is the impurity pseudospin, and τ^i are the Pauli matrices. Hamiltonian $H_{\text{ex}}^{(B)}$ corresponds to the backward exchange scattering emerging in the presence of anisotropy of scattering channels. Term H'_{sc} describes the total potential scattering of fermions taking into account the additional contribution associated with the exchange interaction. Henceforth, we will omit the prime on H_{sc}. Hamiltonian H' includes the terms that are insignificant for low energies.

The interaction in the charge channel of a Luttinger liquid is characterized by parameter K_c (see expression (10) below). It was proved in [20] that, in the case of repulsive interactions ($K_c < 1$), the small anisotropy (J_B) of the exchange scattering channels in the vicinity of the fixed point of strong interaction in J is relevant only for $1/2 < K_c < 1$. In this case, the results obtained in [19] are valid (it was found in [19] that the two-channel Kondo model is absolutely unstable). However, for $K_c < 1/2$, the two-channel Kondo behavior is stable to the exchange anisotropy of the channels. In this case, however, one more mechanism violating the channel degeneracy (weak resonant and potential scattering of many-particle excitations) in the two-channel exchange interaction was discarded. We will consider this mechanism here.

Using the fact that the exchange anisotropy of the channels is irrelevant for $K_c < 1/2$, we consider the situation when the impurity pseudospin has a symmetric coupling with adjoining lattice sites so that $J^i_B = 0$. It will be shown below that Hamiltonian (7) in this case can be reduced to the model of the resonance level (this model was constructed in [24] for the two-channel exchange interaction).

In determining the mechanism of formation of an NFL state in the case of interaction with a pseudospin impurity, the splitting Δ of the two-level impurity plays a significant role. Splitting may be due to the following factors: (i) local distortions in the region of an impurity site; (ii) transformation to the exchange Hamiltonian, and (iii) hyperfine intraatomic interactions. In the presence of splitting, the Hamiltonian acquires the term

$$H_\Delta = \Delta S^z.$$

In order to represent the two-channel exchange interaction in the form of the model of a resonance level, we introduce the boson representation of fermion fields $\psi_{\alpha\sigma}(0)$. In the Luttinger model, these fields are defined as (see, for example, [2] as well as [8])

$$\psi_{\alpha\sigma}(x) = \frac{\hat{\eta}_{\alpha\sigma}}{\sqrt{2\pi a}} \exp[-i\sqrt{4\pi}\Phi_{\alpha\sigma}(x)],$$

$$\Phi_{\alpha\sigma}(x) \quad (8)$$

$$= \frac{1}{\sqrt{2}}[\alpha(\phi_c(x) + \sigma\phi_s(x)) + (\theta_c(x) + \sigma\theta_s(x))],$$

where $\eta_{\alpha\sigma}$ are the so-called Klein factors, $\alpha = \pm$. Fields ϕ_ν and $\Pi_\nu \equiv \partial_x\theta_\nu$ are canonically conjugate: $[\phi_\nu(x), \phi_{\nu'}(x')] = 0$, $[\Pi_\nu(x), \Pi_{\nu'}(x')] = 0$, and $[\phi_\nu(x), \Pi_{\nu'}(x')] = i\delta_{\nu,\nu'}\delta(x - x')$. Boson fields $\phi_{c,s}$ are determined by change and spin density fluctuations, respectively: $\rho_c = \partial_x\phi_c/\sqrt{\pi}$, $\rho_s = \partial_x\phi_s/\sqrt{\pi}$. In turn, the total fermion density can be written with the help of the boson fields in the form $\rho_{\alpha\sigma} = (1/\sqrt{8\pi})(\partial_x\phi_c - \alpha\Pi_c + \sigma(\partial_x\phi_s - \alpha\Pi_s))$. The Hamiltonian of a purely Luttinger liquid has the form

$$H_{\mathcal{L}} = H_c + H_s,$$

$$H_\nu = \int dx \frac{u_\nu}{2}\left(K_\nu\Pi_\nu^2 + \frac{1}{K_\nu}(\partial_x\phi_\nu)^2\right), \quad \nu = c, s,$$

$$u_\nu = \sqrt{\left(v_F + \frac{g_{4\nu}}{\pi}\right)^2 - \left(\frac{g_\nu}{2\pi}\right)^2}, \quad (9)$$

$$K_\nu = \sqrt{\frac{2\pi v_F + 2g_{4\nu} + g_\nu}{2\pi v_F + 2g_{4\nu} - g_\nu}},$$

where $g_c = g_1 - 2g_2$, $g_s = g_1$, $g_{4c} = g_4$, $g_{4s} = 0$, $g_{4\nu}$, g_ν are interactions in a Luttinger liquid. Here, we consider only the case when interactions in a Luttinger liquid are independent of pseudospins. It is well known [1] that g_1 is identically equal to zero in this case. Accordingly, $K_s = 1$ irrespective of the presence of splitting Δ. In the charge channel, the case $g_1 = 0$ corresponds to the parameters

$$u_c = \sqrt{\left(v_F + \frac{g_4}{\pi}\right)^2 - \left(\frac{g_2}{\pi}\right)^2},$$

$$K_c^2 = \frac{\pi v_F + g_4 - g_2}{\pi v_F + g_4 + g_2}. \quad (10)$$

Hamiltonians H_ν lead to the canonical (for bosons) form

$$H'_\nu = \frac{u_\nu}{2}\int dx(\Pi'^2_\nu + (\partial_x\phi'_\nu)^2)$$

with the help of the following redefinition of boson fields: $\phi_\nu = \sqrt{K_\nu}\phi'_\nu$ and $\Pi_\nu = (1/\sqrt{K_\nu})\Pi'_\nu$. In this case, phases $\Phi_{\alpha s}$ in relation (8) are transformed to

$$\Phi'_{\alpha\sigma} = \frac{1}{\sqrt{2}}\left[\pm(\phi'_c\sqrt{K_c} + \sigma\phi'_s\sqrt{K_s}) + \left(\frac{\theta'_c}{\sqrt{K_c}} + \sigma\frac{\theta'_s}{\sqrt{K_s}}\right)\right]. \quad (11)$$

Passing in relations (7) for fields ϕ'_ν and θ'_ν and taking into account relation (11), we obtain

$$\sum_{\sigma\neq\sigma'} J_\perp \psi'^+_{\alpha\sigma}(0)\hat{\tau}^i_{\sigma,\sigma'}\psi'_{\alpha\sigma'}(0)S^+ = \frac{J_\perp}{2\pi a}S^+\left[\exp\left[i\left(\phi'_s\sqrt{4\pi K_s} - \sqrt{\frac{4\pi}{K_s}}\theta'_s\right)\right] + \exp\left[-i\left(\phi'_s\sqrt{4\pi K_s} + \sqrt{\frac{4\pi}{K_s}}\theta'_s\right)\right]\right]_{x=0},$$

$$J_z S^z\sum_\alpha(\rho_{\alpha\uparrow} - \rho_{\alpha\downarrow}) = J_z S^z\sqrt{\frac{2}{\pi K_s}}(\partial_x\theta'_s)_{x=0}.$$

Thus, in the general case, the exchange interaction in a Luttinger liquid has the form

$$H_{ex} = \frac{J_\perp}{\pi a}[S^+\cos(\sqrt{4\pi K_s}\phi'_s(0)) \times \exp(-i\sqrt{4\pi/K_s}\theta'_s(0)) + \text{H.c.}] + J_z S^z\sqrt{\frac{2}{\pi K_s}}(\partial_x\theta'_s)_{x=0} + \Delta S^z. \quad (12)$$

It should be emphasized that the expression for the exchange interaction contains only pseudospin fields.

To reduce expression (12) to the Hamiltonian in the model of a resonant level, we carry out the following transformations.

(1) We introduce fields $\Phi_{L,R}$ instead of ϕ'_s, θ'_s:

$$\phi'_s = \frac{\Phi_L + \Phi_R}{\sqrt{4\pi K_s}}, \quad \theta'_s = (\Phi_L + \Phi_R)\sqrt{\frac{K_s}{4\pi}}.$$

(2) For convenience, we replace the right field Φ_R by the left field Φ'_L,

$$\Phi'_L(x) = -\Phi_R(-x).$$

(3) We introduce the symmetric and antisymmetric fields $\Phi_{S,A}$,

$$\Phi_L = \frac{\Phi_S + \Phi_A}{\sqrt{2}}, \quad \Phi'_L = \frac{\Phi_S - \Phi_A}{\sqrt{2}}.$$

As a result of these transformations, Hamiltonian (12) assumes the form

$$H_{ex} \equiv H^{(s)}_{ex} = \frac{J_\perp}{2\pi a} \times [S^+(e^{-2i(\Phi_S+\Phi_A)} + e^{-2i(\Phi_S-\Phi_A)}) + \text{H.c.}] + \frac{J_z}{2\pi}S^z(\partial_x\Phi_S)_{x=0} + \Delta S^z. \quad (13)$$

Field Φ_S from the transverse part of the interaction is eliminated by rotation about the S^z axis: $U = e^{2i\Phi_S S^z}$. For $g_1 = 0$ (or, which is the same, for $K_s = 1$), the kinetic energy in the spin channel can be reduced to the form corresponding to free boson fields so that

$$H_s = \frac{v_F}{\pi}\int dx([\nabla\Phi_S(x)]^2 + [\nabla\Phi_A(x)]^2). \quad (14)$$

Since $UH_sU^{-1} = H_s - 4(v_F/\pi)(\partial_x\Phi_S)S^z$, we obtain, using the transformation $U(H_s + H^{(s)}_{ex})U^{-1}$, the Hamiltonian $H_s = H_s + H^{(s)}_{ex} = H_\Delta$ in the pseudospin channel in the model of a resonant level; here,

$$H^{(s)}_{ex} = \frac{J_\perp}{\sqrt{2\pi a}}[\psi^+_A(0) + \psi_A(0)](d^+ - d) + \frac{1}{2\pi}(J_z - 8v_F)\psi^+_S(0)\psi_S(0)\left(d^+d - \frac{1}{2}\right),$$
$$\psi^+_{S,A}(x) = \hat{\eta}\frac{e^{-2i\Phi_{S,A}(x)}}{\sqrt{2\pi a}}, \quad S^+ = d^+\hat{\eta}, \quad S^z = \left(d^+d - \frac{1}{2}\right), \quad \hat{\eta}^2 = 1. \quad (15)$$

It should be noted that phases $\Phi_{S,A}(x)$ describe the symmetric and antisymmetric fluctuations of pseudospin density.

The total Hamiltonian of a Luttinger liquid with a two-level impurity has the form

$$H = H_c + H_s + H^{(s)}_{ex} + H'_{sc} \equiv H_0 + H^{(s)}_{ex} + H'_{sc}. \quad (16)$$

If we disregard H_{sc}, Hamiltonian (16) gives excitations of the Luttinger liquid with velocity u_c in the charge channel. In the pseudospin channel, excitations

$\psi^{+}_{S,A}$ are determined by the two-channel exchange interaction, generating, among other things, a resonant many-particle level for low energies.

2. n the absence of scattering H_{sc}, many-particle excitations at the Fermi level are described by the Green function

$$\mathcal{G}_d(z) = \langle d|(z - \mathcal{H})^{-1}|d\rangle \approx \langle d|[z - (H_0 + H^{(s)}_{ex})]^{-1}|d\rangle.$$

It is well known [25, 26] that the behavior of physical quantities in the two-channel model (7), (15) depends on the relation between T_K and Δ, where T_K is the Kondo temperature.

For $T_K \gg \Delta$, the physical properties of the system are determined by the Kondo effect. We will henceforth refer to this region as the Kondo regime. In this case, model (15) can be renormalized to the limit of strong coupling [25, 26]; in this limit, $\Delta \longrightarrow (\Delta^2/T_K) \ll T_K$. The fixed point lies on the line $\tilde{J}_z = (J_z - 8v_F) = 0$ [21] (the Emery–Kivelson (EK) line). On this line, the impurity degrees of freedom are hybridized only with the field associated with antisymmetric pseudospin fluctuations. The Kondo temperature T_K is defined on the EK line so that parameters T_K and Δ are independent. For $T = 0$, the Green function in the vicinity of the Fermi level has the form

$$\mathcal{G}_d(\varepsilon) = \frac{1}{2}\left[\frac{\hat{\tau}_0 - \hat{\tau}_x}{\varepsilon + i\Gamma_K \operatorname{sgn}\varepsilon} + \frac{\hat{\tau}_0 + \hat{\tau}_x}{\varepsilon + i\delta \operatorname{sgn}\varepsilon}\right]. \qquad (17)$$

In the strong coupling limit, $T_K \sim \Gamma_K$. The second term in relation (17) emerges due to the fact that half the impurity degrees of freedom in the two-channel Kondo model are not hybridized with collective variables of band fermions. Since Hamiltonian (15) does not conserve the number of fermions, $\mathcal{G}_d(z)$ has nonzero anomalous matrix elements proportional to $\langle dd\rangle$ and $\langle d^+d^+\rangle$.

On the other hand, if the conditions $T_K \ll \Delta$ and $\tilde{J}_z \ll J_\perp$ are satisfied, model (15) cannot be renormalized to the limit of strong coupling for low temperatures since Δ is renormalized only slightly [25]. In this case, the NFL state is generated by the screening interaction in Hamiltonian (15). Below, this mechanism will be referred to the X-ray-edge regime. For $J_\perp = 0$, the Green function is defined by the well-known expression [27]

$$\mathcal{G}_d(\varepsilon) = A\Gamma(1 - \alpha_S)e^{-i\pi\alpha_S}\left(\frac{W}{\varepsilon - \tilde{\Delta}}\right)^{1 - a_S}, \quad \varepsilon - \tilde{\Delta} > 0. \qquad (18)$$

Here, $A \sim W^{-1}$, $\Gamma(x)$ is the gamma function, $W \sim \varepsilon_F$ is the truncation parameter, $\alpha_S = (\delta_S/\pi)^2$ and $\delta_S \sim \tilde{J}_z^2$ is the phase shift due to the screening interaction in the S channel, and $\tilde{\Delta} = \Delta - \varepsilon_J$, $\varepsilon_J \sim \tilde{J}_z^2/\varepsilon_F$ being the polaron shift induced by the same interaction.

4. CALCULATION OF DENSITY OF STATES AND SELF-ENERGY FUNCTIONS

1. To determine the many-particle density of states for low energies, the following considerations can be used. For an impurity localized in space, the Green function of band fermions has the form

$$\mathbf{G}_{\alpha\alpha\sigma}(k, k'; z) = \delta(k - k')\mathcal{G}_{\alpha\sigma}(k; z) + \mathcal{G}_{\alpha\sigma}(k; z)\mathcal{T}_{\alpha\alpha}(k, k'; z)\mathcal{G}_{\alpha\sigma}(k'; z), \qquad (19)$$

where $\mathcal{G}_\alpha(k; z)$ is the Green function of a pure Luttinger liquid and $\mathcal{T}_{\alpha\alpha}(k, k'; z)$ is the matrix of scattering from a pseudospin impurity. If we assume that the effect of the exchange interaction in the important energy range mainly consists in the formation of the resonant level, the approximate expression for the scattering matrix can be obtained by the method of the equations of motion in the form

$$\mathcal{T}_{\alpha\alpha}(k, k'; z) = \Gamma^\alpha_{kd}\mathcal{G}_d(z)\Gamma^\alpha_{dk'}, \quad \mathcal{G}_d(z) = \langle d|(z - \mathcal{H})^{-1}|d\rangle,$$

where Γ^α_{kd} are the vortex parts weakly depending on energy and $\mathcal{G}_d(z)$ is the Green function of a many-particle resonant level, which is defined by formula (17). Henceforth, we will assume that $\Gamma^\alpha_{kd} = V^\alpha_{kd}$. In this case, the density of states is defined as

$$\rho_{\alpha\sigma}(\varepsilon) = \mp\pi^{-1}\operatorname{Im}[\operatorname{Tr}\mathbf{G}_{\alpha\alpha\sigma}(k, k'; \varepsilon)] = \rho_{0\alpha\sigma}(\varepsilon) \mp \pi^{-1}\operatorname{Im}\left[\operatorname{Tr}\mathcal{G}_d(\varepsilon)\sum_k \left|V^\alpha_{kd}\right|^2\mathcal{G}^2_{\alpha\sigma}(k; \varepsilon)\right], \qquad (20)$$

where the minus and plus signs correspond to $\varepsilon > 0$ and $\varepsilon < 0$, respectively, and $\rho_{0\alpha\sigma}(\varepsilon)$ is the density of states of a pure Luttinger liquid.

For such a liquid, the Green function $\mathcal{G}_{\alpha\sigma}(x; t)$ was obtained in [28]. The spectral functions and, accordingly, the density of states $\rho_{0\alpha\sigma}(\varepsilon)$ were calculated almost twenty years later in [29, 30]. In particular, the total spectral function (density of states) for a pure Luttinger liquid in the long-wave limit is defined as

$$\rho_{0\alpha\sigma}(\varepsilon) = \frac{A}{\Gamma(1 + \eta_c + \eta_s)\varepsilon_0}\left(\frac{|\varepsilon|}{\varepsilon_0}\right)^{\eta_c + \eta_s}, \quad \eta_\nu = \frac{(K_\nu - 1)^2}{4K_\nu}. \qquad (21)$$

Here, $A \sim 1$; $\varepsilon_0 \sim \varepsilon_F$ is the cutoff energy in the Luttinger model; η_c, η_s are anomalous dimensions in the charge and pseudospin channels; and parameters K_c and K_s are

defined in (9). The energy is measured from the Fermi level.

The Green function of impurity degrees of freedom for low energies is defined by relation (17). Thus, to calculate the density of states (20) in the presence of a pseudospin impurity, we must know the quantity

$$\sum_k \left|V_{kd}^{\alpha}\right|^2 \mathcal{G}_{\alpha\sigma}^2(k;\varepsilon) \approx \left|V_{k_F d}^{\alpha}\right|^2 \sum_k \mathcal{G}_{\alpha\sigma}^2(k;\varepsilon),$$

where

$$\sum_k \mathcal{G}_{\alpha\sigma}^2(k;\varepsilon) = -\frac{\partial \Sigma_{0\alpha\sigma}(\varepsilon)}{\partial \varepsilon}.$$

Functions $\Sigma_{0\alpha\sigma}(\varepsilon)$ are defined by expressions (4) with the density of states $\rho_{\alpha\sigma}(\varepsilon) = \rho_{0\alpha\sigma}(\varepsilon)$ from (21); namely,

$$\Sigma_{0\alpha\sigma}^{(+)}(z) = \int_{-\varepsilon_0}^{+\varepsilon_0} d\varepsilon' \frac{\rho_{0\alpha\sigma}(\varepsilon')}{z-\varepsilon'} = \frac{1}{\Gamma(1+\eta_c)\varepsilon_0^{1+\eta_c}} \times \left[\int_0^{\varepsilon_0} d\varepsilon' \frac{(\varepsilon')^{\eta_c}}{z_+ - \varepsilon'} + \int_0^{\varepsilon_0} d\varepsilon' \frac{(\varepsilon')^{\eta_c}}{z_- + \varepsilon'}\right] = A_0(J_{0-} + J_{0+}). \quad (22)$$

Here, $z_\pm \equiv \varepsilon \pm i\gamma$, $\varepsilon > 0$. For $\varepsilon < 0$, we have $\Sigma_{0\alpha\sigma}^{(-)}(z) = -\Sigma_{0\alpha\sigma}^{(+)}(z)$. Integrals $J_{0\mp}(z)$, which are Gilbert transforms of the density of states for a Luttinger liquid, are defined as

$$J_{0\mp}(z) = \frac{\varepsilon_0^{\eta_c+1}}{(\eta_c+1)z} F\left(1, \eta_c+1; \eta_c+2; \pm\frac{\varepsilon_0}{z_\pm}\right),$$

where $F(\alpha, \beta; \gamma; x)$ is a hypergeometric function. Since we consider here only the case when $\varepsilon_0/|z| \gg 1$, we can use the transformation formulas for a hypergeometric function with $|x| \gg 1$ [31],

$$F(\alpha,\beta;\gamma;x) = \frac{\Gamma(\gamma)\Gamma(\beta-\alpha)}{\Gamma(\beta)\Gamma(\gamma-\alpha)}(-1)^{\alpha}x^{-\alpha} + \frac{\Gamma(\gamma)\Gamma(-\beta+\alpha)}{\Gamma(\alpha)\Gamma(\gamma-\beta)}(-1)^{\beta}x^{-\beta},$$

and obtain the following expressions for $I_3^{(\mp)}(z)$:

$$J_{0+} = -\frac{\pi z_+^{\eta_c}}{\sin(\pi\eta_c)} + \frac{\varepsilon_0^{\eta_c}}{\eta_c+1},$$

$$J_{0-} = -(-1)^{(\eta_c+1)}\frac{\pi z_-^{\eta_c}}{\sin(\pi\eta_c)} - \frac{\varepsilon_0^{\eta_c}}{\eta_c+1}.$$

We have used the familiar relations $\Gamma(1) = 1$, $\Gamma(\eta_c + 1) = \eta_c\Gamma(\eta_c)$, and $\Gamma(1-\eta_c)\Gamma(\eta_c) = \pi/\sin(\pi\eta_c)$. Thus, in the long-wave limit, for $\gamma \longrightarrow 0$, we obtain

$$\Sigma_{0\alpha\sigma}^{(+)}(\varepsilon) = \left(\frac{2i\pi\sin[(\pi/2)\eta_c]\exp[i(\pi/2)\eta_c]}{\sin(\pi\eta_c)\Gamma(1+\eta_c)\varepsilon_0^{1+\eta_c}}\right)\varepsilon^{\eta_c}. \quad (23)$$

Since $F(1, 1; 2; \mp x) = \mp x^{-1}\ln(1 \pm x)$, expression (23) for $\eta_c = 0$ corresponds to the principal term in the expansion in parameter $x = (\varepsilon_0/\varepsilon) \gg 1$. Using the expressions

$$\sum_k \mathcal{G}_{\alpha\sigma}^2(k;\varepsilon) = -\frac{\partial\Sigma_{0\alpha\sigma}^{(-)}(\varepsilon)}{\partial\varepsilon}, \quad \varepsilon > 0;$$

$$\sum_k \mathcal{G}_{\alpha\sigma}^2(k;\varepsilon) = -\frac{\partial\Sigma_{0\alpha\sigma}^{(-)}(\varepsilon)}{\partial\varepsilon} = -\frac{\partial\Sigma_{0\alpha\sigma}^{(+)}(\varepsilon)}{\partial|\varepsilon|}, \quad \varepsilon < 0,$$

we obtain with the help of formulas (22) and (23)

$$\sum_k \mathcal{G}_{\alpha\sigma}^2(k;\varepsilon) = \sum_k \mathcal{G}_{\alpha\sigma}^2(k;|\varepsilon|) = \left(\frac{-2i\exp[i(\pi/2)\eta_c]}{\Gamma(1+\eta_c)\varepsilon_0^{1+\eta_c}}\right)\left(\frac{\pi\eta_c\sin[(\pi/2)\eta_c]}{\sin(\pi\eta_c)}\right)|\varepsilon|^{\eta_c-1}. \quad (24)$$

In the Kondo regime, substituting this expression into (20), we arrive at the following expression for the impurity contribution to the density of states:

$$\rho_{d\alpha}(|\varepsilon|) = \left[\frac{2V^2\eta_c\sin[(\pi/2)\eta_c]}{\Gamma(1+\eta_c)\varepsilon_0^{1+\eta_c}\sin(\pi\eta_c)}\right] \times \left[\frac{\cos[(\pi/2)\eta_c]}{\Gamma_K^2}|\varepsilon|^{\eta_c} + \frac{\sin[(\pi/2)\eta_c]}{\Gamma_K}|\varepsilon|^{\eta_c-1} + \cos\left(\frac{\pi}{2}\eta_c\right)|\varepsilon|^{\eta_c-2}\right]. \quad (25)$$

Here, we have introduced the notation $V^2 \equiv |V_{dk_F}^{\alpha}|^2$. For low energies, the most singular term in the density of states is $\rho_{d\alpha}(|\varepsilon|) \propto |\varepsilon|^{\eta_c-2}$. Pay attention to the fact that this contribution is due to the term of the form $\mathrm{Re}\mathcal{G}_d(\varepsilon)\mathrm{Im}\sum_k \mathcal{G}_{\alpha\sigma}^2(k;\varepsilon)$; in $\mathrm{Re}\mathcal{G}_d(\varepsilon)$, we take the part corresponding to nonhybridized degrees of freedom (second term in relation (17)). While deriving formula (25), we took into account the fact that the term with the δ function in $\mathrm{Im}\mathcal{G}_d(\varepsilon)$ makes zero contribution to the density of states for $\eta_c \neq 0$.

Thus, we see that, in contrast to noninteracting Fermi gas, excitations in a Luttinger liquid consider-

ably modify the peaks at the Fermi level, which are generated by the interaction with a two-level impurity. The singularity in the impurity density of states is enhanced for $\eta_c < 1$; on the contrary, it is suppressed for $\eta_c > 2$.

2. The contribution to $\Sigma_\alpha(z)$ from the impurity term $\rho_{d\alpha}(\varepsilon)$ to the density of states is determined by the expression

$$\Sigma_\alpha(z) = \int_{-\infty}^{+\infty} d\varepsilon \frac{\rho_{d\alpha}(\varepsilon)}{z-\varepsilon} = \int_0^\infty d\varepsilon \frac{\rho_{d\alpha}(|\varepsilon|)}{z_+ - \varepsilon} + \int_0^\infty d\varepsilon \frac{\rho_{d\alpha}(|\varepsilon|)}{z_- + \varepsilon} \equiv \Sigma_\alpha^{(+)}(z_+) + \Sigma_\alpha^{(-)}(z_-). \quad (26)$$

Taking into account expression (25) for the impurity density of states, we see that $\Sigma_\alpha(z)$ contains the integrals of three types, which are defined in the complex plane:

$$I_1^{(\mp)} = \int_0^{+\infty} d\varepsilon \frac{\varepsilon^{\eta_c - 2}}{z \mp \varepsilon}, \quad I_2^{(\mp)} = \int_0^{+\infty} d\varepsilon \frac{\varepsilon^{\eta_c - 1}}{z \mp \varepsilon},$$

$$I_3^{(\mp)} = \int_0^{\varepsilon_0} d\varepsilon \frac{\varepsilon^{\eta_c}}{z \mp \varepsilon}.$$

To evaluate integrals $I_1^{(\mp)}$ and $I_2^{(\mp)}$ (which are singular for small z), in the complex plane we choose an auxiliary function of the form $f(z') = (z')^{\mu-1}/(z' - z)$ and the contour C consisting of a large circle C_R, $R \longrightarrow \infty$, a small circle c_r, $r \longrightarrow 0$, around zero, and a cut along the real axis (I and II are the upper and lower banks of the cut). Noting that $f(e^{2i\pi}\varepsilon) = e^{2i\pi\mu} f(\varepsilon)$ at the lower bank, considering that $\int_{c_r}, \int_{C_R} \longrightarrow 0$, and taking into account the fact that there is a pole for $z' = z$ within contour C, we obtain

$$J_- = \int_0^{+\infty} d\varepsilon \frac{\varepsilon^{\mu-1}}{z-\varepsilon}$$

$$= \int_C f(z')dz' = \int_{C_R} + \int_{c_r} + \int_I + \int_{II} = 2i\pi z^{\mu-1},$$

$$\int_I + \int_{II} = (1 - e^{2i\pi\mu}) J_-, \quad J_- = \frac{2i\pi z^{\mu-1}}{1 - e^{2i\pi\mu}}.$$

Substituting the expression obtained with different values of μ for the integrals in (26), we get

$$\Sigma_\alpha^{(\pm)} = A_1 z_\pm^{\eta_c - 2} P_\pm \quad (27)$$

for $\rho_{d\alpha} \sim |\varepsilon|^{\eta_c - 2}$, where $P_\pm = -e^{-i\pi\eta_c}$, (–1) for "±," and

$$A_1 = \frac{2\pi\eta_c \cos(\pi\eta_c/2)\sin(\pi\eta_c/2)V^2}{\varepsilon_0^{\eta_c+1}\sin(\pi\eta_c)\Gamma(\eta_c+1)},$$

$$\Sigma_\alpha^{(\pm)} = A_2 z_\pm^{\eta_c - 1} P_\pm \quad (28)$$

for $\rho_{d\alpha} \sim |\varepsilon|^{\eta_c - 1}$, where $P_\pm = -e^{-i\pi\eta_c}$, (+1) for "±," and

$$A_2 = \frac{2\pi\eta_c \sin^2(\pi\eta_c/2)V^2}{\Gamma_K \sin^2(\pi\eta_c)\Gamma(\eta_c+1)\varepsilon_0^{\eta_c+1}}.$$

Integrals $I_3^{(\mp)}(z)$ are defined by the same formulas that were derived while determining $\Sigma_{0\alpha}^{(\pm)}$.

3. In the X-ray-edge regime, expression (18) defines the retarded Green function $\mathcal{G}_d^R(\tilde{\varepsilon})$, where $\tilde{\varepsilon} = \varepsilon - \tilde{\Delta}$. Let us consider the case when the quantity

$$\left[\sum_k \mathcal{G}_{\alpha\sigma}^R(k;\varepsilon)\right]^2 \propto \varepsilon^{\eta_c - 1}$$

as a function of energy varies much more slowly than $\mathcal{G}_d^R(\tilde{\varepsilon})$. The definitions of these two quantities imply that this is possible for $\eta_c \gg \alpha_S$.

In this case, we can set

$$\left[\sum_k \mathcal{G}_{\alpha\sigma}^R(k;\varepsilon)\right]^2 \approx \left[\sum_k \mathcal{G}_{\alpha\sigma}^R(k;\tilde{\Delta})\right]^2.$$

Accordingly, it can easily be verified that the density of states for $\tilde{\varepsilon} > 0$ has the form

$$\rho_d(\tilde{\varepsilon}) = Q_d \left(\frac{2V^2}{\pi\varepsilon_0^3}\right)\left(\frac{\varepsilon_0}{\tilde{\varepsilon}}\right)^{1-\alpha_S}, \quad Q_d = \left(\frac{\pi^2}{\Gamma(\alpha_d)\Gamma(\eta_c)}\right)\left(\frac{\varepsilon_0}{\tilde{\Delta}}\right)^{1-\eta_c}. \quad (29)$$

The retarded self-energy functions are defined as

$$\Sigma_\alpha^R(z) = \int_{-\infty}^{+\infty} d\varepsilon \frac{\rho_d(\varepsilon)}{z_+ - \varepsilon} = \int_{\tilde{\Delta}}^{\infty} d\varepsilon \frac{\rho_d(\varepsilon)}{z_+ - \varepsilon} + \int_{-\infty}^{\tilde{\Delta}} d\varepsilon \frac{\rho_d(\varepsilon)}{z_+ - \varepsilon} = \int_0^\infty d\tilde{\varepsilon}\left[\frac{\rho_d(\tilde{\varepsilon})}{\tilde{z}_+ - \tilde{\varepsilon}} + \frac{\rho_d(-\tilde{\varepsilon})}{\tilde{z}_+ + \tilde{\varepsilon}}\right], \quad (30)$$

where $\tilde{z}_+ = z_+ - \tilde{\Delta}$ and $\mathrm{Re}\tilde{z}_+ = \tilde{\varepsilon} > 0$.

Using formulas (29) and the expressions for the integrals derived above, we obtain

$$\Sigma_\alpha^R(z) = -\tilde{Q}_d\left(\frac{4V^2}{\varepsilon_0^3}\right)\left(\frac{\varepsilon_0}{z}\right)^{1-\alpha_S}, \qquad \tilde{Q}_d = Q_d\cot(\pi\alpha_S). \tag{31}$$

Since $\alpha_S < 1$, the above contribution of the impurity density of states to $\Sigma_\alpha^R(z)$ is also singular for small $|z|$.

5. FERMI-LIQUID RESONANCES IN THE VICINITY OF THE FERMI LEVEL

Let us now consider the Fermi-liquid singularities in the total Green function (4), which are associated with the scattering of many-particle excitations, described by Hamiltonian H_{sc} in expression (2). In the presence of resonance and potential scattering, the position of the poles is determined from the solution of Eq. (5). Substituting Σ_{sc} into Eq. (5) and assuming that all matrix elements appearing in $\Sigma_\alpha^{ab}(z)$ are determined by their values for $k = k_F$, we obtain the equation

$$1 - \mathcal{T}^2\Sigma_L(z_r)\Sigma_R(z_r) - V^2\mathcal{T}\mathcal{G}_d(z_r)\Sigma_L(z_r)\Sigma_R(z_r) \times [2 + \mathcal{T}(\Sigma_L(z_r) + \Sigma_R(z_r))] = 0, \tag{32}$$

where V and $\mathcal{T}$ are the matrix elements of the resonance and potential scattering and $\Sigma_\alpha(z)$ are the self-energy functions calculated in (27), (28), and (31). The three terms on the left-hand side of Eq. (32), containing self-energy functions $\Sigma_\alpha(z)$, describe different scattering processes of many-particle excitations. The term proportional to $\mathcal{T}^2$ corresponds to scattering of right fermions from left fermions taking into account the fact that the density of states contains the impurity contribution $\rho_{d\alpha}$. The terms proportional to $\mathcal{T}V^2$ and $\mathcal{T}^2V^2$ describe two possible processes of resonant scattering, i.e., scattering of charge and pseudospin densities from many-particle impurity degrees of freedom, which are described by the Green function $\mathcal{G}_d(z_r)$ defined in (17).

1. In the Kondo regime, the impurity density of states is defined by expression (25); accordingly, the poles in the vicinity of the Fermi level are generated by the contributions to $\Sigma_\alpha(z)$, which are most singular in a certain range of parameters for small $|z| \ll \Gamma_K$. We will first consider the solutions to Eq. (32), which are associated with scattering of nonhybridized excitations with band collective excitations of impurity degrees of freedom. In accordance with relation (17), nonhybridized degrees of freedom are described by the Green function $\mathcal{G}_d(z_r) = 1/z_r$. We begin with the case when the main contribution to $\Sigma_\alpha(z)$ comes from the term proportional to $|\varepsilon|^{\eta_c - 2}$ in the density of states. Retaining in Eq. (32) the principal terms with small values of $|z|$, we obtain the following equation for the poles:

$$1 - V^2\mathcal{T}^2\mathcal{G}_d(z_r)\Sigma_L(z_L)\Sigma_R(z_r) \times (\Sigma_L(z_r) + \Sigma_R(z_r)) = 0. \tag{33}$$

For $z_{r\pm}$, we respectively substitute $\Sigma_\alpha^\pm(z_{r\pm})$ into Eq. (33). It should also be noted that $\Sigma_L(z_r) = \Sigma_R(z_r)$ since $\rho_L(|\varepsilon|) = \rho_R(|\varepsilon|)$.

Let us consider solutions with z_{r-}. We write z_{r-} in the form $z_{r-} = |z_{r-}|\exp(i\varphi)$; physical solutions correspond to values of $0 < \varphi < \pi/2$ for $\mathrm{Re}(z_{r-}) < 0$. Solving the imaginary and real parts of Eq. (33), we obtain the following solution with $|z_{r-}| \ll \Gamma_K$:

$$\frac{|z_{r-}|}{\varepsilon_0} \approx A_1\left(\frac{\mathcal{T}}{\varepsilon_0}\right)^{2/(7-3\eta_c)}\left(\frac{V}{\varepsilon_0}\right)^{8/(7-3\eta_c)}, \qquad \varphi = \frac{\pi}{7-3\eta_c}, \tag{34}$$

where A_1 is a factor on the order of unity. This solution exists when the following conditions are satisfied:

$$1 < \eta_c < \frac{5}{3}, \quad \left(\frac{\mathcal{T}}{\varepsilon_0}\right)^{1/4}\left(\frac{V}{\varepsilon_0}\right) \ll \left(\frac{\Gamma_K}{\varepsilon_0}\right)^{(7-3\eta_c)/8}. \tag{35}$$

The first system of inequalities emerges from the requirement of vanishing of the imaginary part of Eq. (33).

When conditions (35) are violated, Eq. (33) may have solutions due to two other contributions to the density of states. For $\rho_{d\alpha} \sim |\varepsilon|^{\eta_c - 1}$ and $\mathcal{G}_d(z_r) = 1/z_r$, there exists a solution $z_{r-} = \varepsilon_{r-} < 0$ corresponding to a localized level below the Fermi energy. The position of this level is determined by the energy

$$\frac{|\varepsilon_r|}{\varepsilon_0} = A_2\left(\frac{V}{\varepsilon_0}\right)^{8/(4-3\eta_c)}\left(\frac{\mathcal{T}}{\varepsilon_0}\right)^{2/(4-3\eta_c)}\left(\frac{\varepsilon_0}{\Gamma_K}\right)^{3/(4-3\eta_c)}, \qquad |\varepsilon_r| \ll \Gamma_K. \tag{36}$$

Finally, for $\rho_{d\alpha} \sim |\varepsilon|^{\eta_c}$ and $\mathcal{G}_d(z_r) = 1/z_r$, we have a resonance with z_{r+},

$$\frac{|z_{r+}|}{\varepsilon_0} \approx A_3\left(\frac{\mathcal{T}}{\varepsilon_0}\right)^{1/(1-2\eta_c)}\left(\frac{V}{\varepsilon_0}\right)^{6/(1-2\eta_c)}\left(\frac{\varepsilon_0}{\Gamma_K}\right)^{4/(1-2\eta_c)}, \qquad \varphi = \frac{2\pi\eta_c}{1-2\eta_c}, \tag{37}$$

which exists when the following conditions are satisfied:

$$\eta_c < \frac{1}{6}, \quad \left(\frac{\mathcal{T}}{\varepsilon_0}\right)^{1/4}\left(\frac{V}{\varepsilon_0}\right) \ll \left(\frac{\Gamma_K}{\varepsilon_0}\right)^{(5-2\eta_c)/8}. \tag{38}$$

It can be seen that all the resonance obtained above are formed as a result of scattering of nonhybridized

impurity degrees of freedom by many-particle excitations, which form the density of states at low energies.

The scattering of many-particle excitations with $\mathcal{G}_d(z_r) = 1/(z_{r\pm} + i\Gamma_K \operatorname{sgn}\varepsilon_{r\pm})$, which is described by the term in Σ_{sc} proportional to $\mathcal{T}^2V^2$, does not lead to the formation of resonances with $|z_r| \ll \Gamma_K$. In other words, Eq. (33) has no physical solutions in this case.

In the X-ray-edge regime, we substitute into Eq. (33) the expressions for $\mathcal{G}_d$ and Σ_α from relations (18) and (31), respectively. For $\mathrm{Re}(\tilde{z}_r) > 0$, we obtain a resonance whose position and width are defined by the formulas

$$\frac{|\tilde{z}_r|}{\varepsilon_0} \approx A_3\left(\frac{\mathcal{T}}{\varepsilon_0}\right)^{1/2(1-\alpha_S)}\left(\frac{V}{\varepsilon_0}\right)^{2/(1-\alpha_S)}\left(\frac{\tilde{\Delta}}{\varepsilon_0}\right)^{(\eta_c-1)/4(1-\alpha_S)}, \qquad \varphi = \frac{\pi}{4}, \tag{39}$$

where $A_3 \sim 1$. Since $\tilde{\Delta} \ll \varepsilon_0$ and $\eta_c \gg \alpha_S$, the resonance for $\eta_c > 1$ can be quite narrow with a width much smaller than $\tilde{J}_z^2/\varepsilon_F$.

2. In the range of parameters where Eq. (33) has no solutions, the poles can be due to other scattering processes apart from those making contribution to Σ_{sc}, proportional to $\mathcal{T}^2V^2$. Assuming that this contribution is small in the regions where Eq. (33) has no solutions, we consider the equation

$$1 - 2V^2\mathcal{T}\mathcal{G}_d(z_r)\Sigma_L(z_r)\Sigma_R(z_r) = 0. \tag{40}$$

In the Kondo regime, when the main contribution to Σ_α comes from the term proportional to $|\varepsilon|^{\eta_c-2}$ in the density of states, the Fermi-liquid resonance is formed due to scattering of many-particle excitations by the resonant level with

$$\mathcal{G}_d(z_r) = \frac{1}{z_{r\pm} + i\Gamma_K \operatorname{sgn}\varepsilon_{r\pm}} \approx \mp\frac{i}{\Gamma_K}$$

for $\mathrm{Re}(z_r) > 0$ and $\mathrm{Re}(z_r) < 0$, respectively. The position and width of the Fermi-liquid resonance are determined by the solution to Eq. (40),

$$\frac{|z_{r-}|}{\varepsilon_0} \approx A\left(\frac{\mathcal{T}}{\varepsilon_0}\right)^{1/2(2-\eta_c)}\left(\frac{V}{\varepsilon_0}\right)^{6/2(2-\eta_c)}\left(\frac{\varepsilon_0}{\Gamma_K}\right)^{1/2(2-\eta_c)}, \qquad \varphi = \frac{\pi}{4(2-\eta_c)}, \tag{41}$$

where A is a factor on the order of unity. This solution exists when the following conditions are satisfied:

$$1 < \eta_c < \frac{3}{2}, \quad \left(\frac{\mathcal{T}}{\varepsilon_0}\right)^{1/6}\left(\frac{V}{\varepsilon_0}\right) \ll \left(\frac{\Gamma_K}{\varepsilon_0}\right)^{(5-2\eta_c)/6}. \tag{42}$$

The same scattering process generates the Fermi-liquid resonance with z_{r-} in the case when the main contribution to Σ_α comes from the term proportional to $|\varepsilon|^{\eta_c-1}$ in the density of states. In the present case, the new resonance exists for

$$\eta_c < 1, \quad \left(\frac{\mathcal{T}}{\varepsilon_0}\right)^{1/6}\left(\frac{V}{\varepsilon_0}\right) \ll \left(\frac{\Gamma_K}{\varepsilon_0}\right)^{(3-2\eta_c)/6}. \tag{43}$$

Naturally, when the regions of existence of resonances intersect, the most singular contribution for low energies is decisive in Eq. (32).

Finally, new singularities may appear due to potential scattering of right fermions from left ones taking into account the impurity density of states (25). The positions of the poles in this case is determined by the solutions to the equation $1 - \mathcal{T}^2\Sigma_L(z_r)\Sigma_R(z_r) = 0$. It can easily be verified that a solution exists in the case when the main contribution to Σ_α comes from the term proportional to $|\varepsilon|^{\eta_c-2}$ and has the form of a localized level below the Fermi energy.

It can also be proved that potential scattering in the X-ray-edge regime generates a localized level above the Fermi energy.

If the density of states at the Fermi level is determined by expression $\rho_{0\alpha}$ for a pure Luttinger liquid, the potential scattering of right fermions from left ones does not lead to the formation of additional Fermi-liquid resonances or levels at low energies. This case was treated in [10].

Concluding the section, let us prove that the inequalities for parameter η_c derived in (35), (38), (42), and (43), which determine the ranges for new Fermi-liquid resonances, are in accordance with the inequality $K_c < 1/2$. It should be recalled that this is the condition of applicability of the model with $J_B = 0$. In particular, the conditions for η_c in (35) and (42) correspond to the inequalities $0.13 < K_c < 0.2$ and $0.11 < K_c < 0.2$, respectively. Finally, the condition $\eta_c < 1$ in (43) corresponds to values of $K_c < 0.2$. In this case, the mechanism considered in this study ensures instability in the region $1 > \eta_c > 1/8$, which is also in agreement with the conditions derived above.

6. CONCLUDING REMARKS: SINGULARITIES IN THE BEHAVIOR OF PHYSICAL QUANTITIES

Several types of non-Fermi-liquid excitations with different quantum numbers characterize a Luttinger liquid with two-level impurities at low energies. First, these are charge density fluctuations of the Luttinger liquid and, second, many-particle excitations due to two-channel exchange interactions, which are associated with band and impurity fermion states. The results described above indicate that the scattering of many-particle excitations of various types from one another leads to the emergence of an additional Fermi-liquid

singularity in the vicinity of the Fermi level in a 1D system. The conditions are determined under which the Fermi-liquid states with a new energy scale much smaller than the Kondo temperature is the ground state of the system.

1. The type of many-particle excitations at low energies and, accordingly, the type of the phase state are determined by the following parameters: width Γ_K or (in the strong-coupling limit) the Kondo temperature T_K; the initial energy E_d of a deep level, which appears in exchange constants; the value of abnormal dimensionality η_c in the charge channel of a Luttinger liquid (or, which is the same, the value of parameter K_c); and, finally, the impurity level splitting Δ. Let us define the states of the system depending on the values of these parameters.

New Fermi-liquid resonances are absent. In accordance with conditions (35), (37), (42), and (43), this means that, at any rate, η_c must be greater than 5/3 and/or the Kondo temperature is found to be very low in the Kondo regime. In the latter case, conditions of the type $|z_r| \ll T_K$ are violated and we obtain a phase state of the system, in which excitations of the Luttinger liquid take place in the charge channel and the excitations generated by two-channel exchange interaction occur in the pseudospin channel. In turn, if the initial values of the parameters are such that the condition $T_K \gg \Delta$ holds and the system is in the Kondo regime, we pass to the X-ray-edge regime by reducing the value of T_K by increasing the depth of the impurity level and, hence, by decreasing J. A similar transition can also be carried out by increasing the value of Δ at a fixed value of T_K.

The low-temperature behavior of the heat capacity C of a linear chain is determined by the product γT, where γ is inversely proportional to the velocity of sound.

This expresses the Debye law in the 1D case. In a pure Luttinger liquid, the heat capacity has two additive contributions of the same form with $\gamma_c \sim v_c^{-1}$ and $\gamma_s \sim v_s^{-1}$ owing to the separation of the pseudospin and charge degrees of freedom.

In the presence of two-level impurities, excitations in the pseudospin channel are determined by the two-channel exchange interaction. It is well known [24, 32] that γ in this case behaves as $\ln(T_K/T)$ in the second order of perturbation theory in $\tilde{J}_z/J_\perp$; this dependence plays a major role at low temperatures. At the crossover temperature $T_{c1} \sim \Delta$, the logarithmic dependence is transformed into the power dependence $\gamma \propto T^{-1+\alpha_s}$. The increase in the heat capacity upon cooling continues to $T \propto \max[\gamma_c^{-1}, \Delta^2/T_K]$. Accordingly, $\gamma_{max} \propto \max[\gamma_c, T_K/\Delta^2]$. The homogeneous static susceptibility behaves analogously on the EK line ($\tilde{J}_z = 0$); however, since the susceptibility has no constant term, the low-temperature increase continues until $T \propto \Delta^2/T_K$.

It should be noted that anomalous correlations exist at an impurity site. The divergence of the correlator $\langle S^+S^-\rangle(\omega = 0, T) \sim \ln(1/T)$ corresponds to free rotation of the impurity pseudospin. It should also be noted that the anomalous components $\langle dd\rangle$ and $\langle d^+d^+\rangle$ of the Green function $\mathcal{G}_d$ for the resonant level differ from zero in the Kondo regime since the number of fermions in the two-channel exchange interaction (15) is not conserved.

Fermi-liquid resonances exist in the Kondo regime for relatively small values of η_c and high values of T_K in accordance with the formulas derived above. If the value of η_c satisfies any condition of the existence of resonances, but the corresponding condition of the type $|z_r| \ll T_K$ is violated, we can obtain an additional Fermi-liquid resonance in the vicinity of the Fermi level by reducing the depth of the impurity level and, hence, by increasing the Kondo temperature.

Expression (40) shows that Fermi-liquid resonances exist in the X-ray-edge regime for all admissible values of α_S (i.e., impurity level depths), but only for strong interactions in the charge channel of a Luttinger liquid.

Both in the Kondo regime and in X-ray-edge regimes, transitions from an NFL state to an FL state can take place. The characteristic crossover temperatures $T_{c2} \propto \gamma_r$, where γ_r are the widths of Fermi-liquid resonances. In this case, the temperature transition to the Fermi liquid state, which took place in the absence of FL resonances at $T \sim \max[\gamma_c^{-1}, \Delta^2/T_K]$, can now occur at $T \sim \gamma_r$ if $\gamma_r > \max[\gamma_c^{-1}, \Delta^2/T_K]$.

2. Fermi gases in a trap are formed by fermion atoms of mass m with two intrinsic states [4–7]. The number of atoms in each state is the same. If the atoms are cooled to a temperature below the Fermi temperature T_F, they form a degenerate Fermi gas. The system can be treated as effectively one-dimensional if the characteristic energy of longitudinal motion is much smaller than the characteristic separation $\omega_\perp$ between transverse quantization levels. In other words, the condition $\varepsilon_F \ll \omega_\perp$ ($\hbar = 1$) must be satisfied. At low temperatures, only s-type collisions are possible between Fermi atoms with different spins. For this reason, the interactions are characterized by a single parameter, viz., scattering length $a \ll l_\perp$, where $l_\perp = (1/m\omega_\perp)^{1/2}$. The effective 1D interaction can be represented in the form of short-range potential with a characteristic value of $g = 2\pi a/m l_\perp^2$ [4, 5]. Taking these limitations into account, we assume that, in the longitudinal directions, atoms are in a box of length L with periodic boundary condi-

tions. Then the Hamiltonian of the atomic Fermi gas free of impurities has the form

$$H = \pi v_F \sum_{\alpha,\sigma} \int dx \rho_{\alpha\sigma}^2(x)$$

$$+ \sum_{\alpha,\sigma,\sigma'} \int dx \Big[g_4^{\sigma\sigma'} \rho_{\alpha\sigma}(x) \rho_{\alpha\sigma'}(x)$$

$$+ \sum_{\alpha' \neq \alpha} (g_2^{\sigma\sigma'} \rho_{\alpha\sigma}(x) \rho_{\alpha'\sigma'}(x)$$

$$+ g_1^{\sigma\sigma'} \psi_{\alpha\sigma}^+ \psi_{\alpha'-\sigma}^+ \psi_{\alpha\sigma} \psi_{\alpha'-\sigma}) \Big],$$

i.e., the Hamiltonian in the Luttinger model. The possibility of describing an atomic impurity-free Fermi gas with the help of the Luttinger model and the methods for experimental observation of the spin-charge separation were considered in [33]. In the presence of impurities, the results obtained above are applicable to an atomic gas if, in addition to the relations given above, the inequality $U < \omega_\perp$ holds. In an atomic Fermi gas, charge fluctuations correspond to fluctuations of the average density of the gas, while pseudospin fluctuations describe the fluctuations of the relative density at two levels corresponding to intrinsic states of atoms. The anomalous behavior of the one-site correlator $\langle S^+ S^- \rangle(\omega = 0, T) \sim \ln(1/T)$ for atomic Fermi gases, which is associated with the two-channel exchange interaction, corresponds to an anomalous increase (upon cooling) of the correlations between the occupancies for two intrinsic states of an impurity site. This may indicate a tendency to the formation of the superfluid state in the relative density of two components of the impurity subsystem in a quasi-one-dimensional system.

ACKNOWLEDGMENTS

This study was financed by the Russian Foundation for Basic Research.

REFERENCES

1. K. Schonhammer, cond-mat/9710330.
2. H. J. Schulz, G. Cuniberti, and P. Pieri, cond-mat/9807366.
3. U. Meirav, M. A. Kastner, M. Heiblum, and S. J. Wind, Phys. Rev. B **40**, 5871 (1989).
4. B. DeMarco and D. S. Jin, Phys. Rev. A **58**, R4267 (1998).
5. B. DeMarco and D. S. Jin, Science **285**, 1703 (1999).
6. M. O. Mewes, G. Ferrary, F. Schrenk, *et al.*, Phys. Rev. A **61**, 011403(R) (1999).
7. J. L. Roberts, N. R. Claussen, S. L. Cornish, *et al.*, Phys. Rev. Lett. **86**, 4211 (2001).
8. F. D. M. Haldane, J. Phys. C **14**, 2585 (1981).
9. A. M. Tsvelik, *Quantum Field Theory in Condensed Matter Physics* (Cambridge Univ. Press, Cambridge, 1995).
10. C. L. Kane and M. P. A. Fisher, Phys. Rev. B **46**, 15233 (1992).
11. T. Ogava, A. Furusaki, and N. Nagaosa, Phys. Rev. Lett. **68**, 3638 (1992).
12. D. K. K. Lee and Y. Chen, Phys. Rev. Lett. **69**, 1399 (1992).
13. A. O. Gogolin, Phys. Rev. Lett. **71**, 2995 (1993).
14. N. V. Prokof'ev, Phys. Rev. B **49**, 2148 (1994).
15. A. C. Hewson, *The Kondo Problem for Heavy Fermions* (Cambridge Univ. Press, Cambridge, 1993).
16. P. Nozieres and A. Blandin, J. Phys. (Paris) **41**, 193 (1980).
17. N. Andrei and S. Destri, Phys. Rev. Lett. **52**, 364 (1984).
18. Dung-Hai Lee and J. Toner, Phys. Rev. Lett. **69**, 3378 (1992).
19. A. Furusaki and N. Nagaosa, Phys. Rev. Lett. **72**, 892 (1994).
20. M. Fabrizio and A. O. Gogolin, Phys. Rev. B **51**, 17827 (1995).
21. M. Fabrizio, A. O. Gogolin, and Ph. Nozieres, Phys. Rev. Lett. **74**, 4503 (1995).
22. L. A. Manakova, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 1009 (1998) [JETP Lett. **67**, 1069 (1998)].
23. L. A. Manakova, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 721 (1999) [JETP Lett. **69**, 772 (1999)].
24. V. J. Emery and S. Kivelson, Phys. Rev. B **46**, 10812 (1992).
25. J. L. Black, K. Vladar, and A. Zawadowski, Phys. Rev. B **26**, 1559 (1982).
26. K. Vladar and A. Zawadowski, Phys. Rev. B **28**, 1564 (1983).
27. P. Nozieres and C. T. de Dominicis, Phys. Rev. **178**, 1097 (1969); K. D. Schotte and U. Schotte, Phys. Rev. **182**, 479 (1969).
28. A. Luther and I. Peschel, Phys. Rev. B **9**, 2911 (1974).
29. V. Meden and K. Schonhammer, Phys. Rev. B **46**, 15753 (1992).
30. K. Schonhammer and V. Meden, Phys. Rev. B **47**, 16205 (1993).
31. I. S. Gradshteĭn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 4th ed. (Fizmatlit, Moscow, 1962; Academic, New York, 1980).
32. A. M. Sengupta and A. Georges, Phys. Rev. B **49**, 1020 (1994).
33. A. Ricati, P. O. Fedichev, W. Zweger, and P. Zoller, cond-mat/0212195.

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 169–176.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 194–202.

SOLIDS
Electronic Properties

Magnetic Flux Annihilation Waves in Inhomogeneous High-Temperature Superconductors

I. A. Rudnev[a], A. E. Khodot[a], A. V. Eremin[a], and B. P. Mikhaĭlov[b]

[a]*Moscow Engineering Physics Institute (State University), Moscow, 115409 Russia*
[b]*Baikov Institute of Metallurgy and Materials Science, Russian Academy of Sciences, Moscow, 119991 Russia*
e-mail: rudnev@supercon.mephi.ru

Received November 12, 2003

Abstract—The process of magnetic field penetration into polycrystalline high-T_c superconductors of the $YBa_2Cu_3O_{7-x}$ and $Bi_2Sr_2Ca_2Cu_3O_{10-x}$ systems has been studied using traditional magnetooptical methods and scanning Hall probe microscopy. It is established that remagnetization of a sample is accompanied by the formation and propagation of a stationary magnetic flux annihilation (MFA) wave. Spatial inhomogeneity of the superconductors studied is manifested by a curvature of the MFA wave front. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Investigation of the process of magnetic flux penetration and propagation in type II superconductors, including high-T_c superconductors, is very important because these processes determine the magnetic and transport characteristics of superconductors, in particular, the local and integral magnetization $M(H)$. In turn, once the behavior of $M(H)$ is known, it is possible to determine both the basic characteristics of superconductors (e.g., the lower and upper critical fields) and the parameters important in practical applications (critical current J_c, hysteresis energy losses, and residual trapped magnetic flux).

The process of the magnetic flux penetration and trapping in the so-called hard superconductors of type II (i.e., high-T_c superconductors exhibiting strong vortex pinning) are frequently described in terms of the one-dimensional Bean model [1]. According to this model, the process of superconductor remagnetization is accompanied by motion of an interface between the regions where the magnetic induction has opposite signs. Within the framework of the Bean model, this boundary appears as a plane front parallel to the superconductor surface, which moves from the sample surface into bulk with increasing external field. However, in some cases (e.g., of weak pinning or strongly inhomogeneous superconductors), the Bean model is too rough and cannot reflect the real local distribution of magnetic induction and the integral magnetization of a superconductor.

Analytical description of the behavior of magnetization in a high-T_c superconductor with an arbitrary preset arrangement of pinning centers is very difficult and requires many parameters to be taken into account. This makes expedient the use of numerical methods for the description of magnetization. Numerical calculations of the magnetization and the local distribution of magnetic induction in a type II superconductor with an arbitrary preset local distribution of defects have been performed in [2–5] using a modified Monte Carlo method. As a result, we determined the behavior of magnetization of a model superconductor (with the parameters of a real high-T_c superconductor of the $Bi_2Sr_2CaCu_2O_x$ system) and calculated the patterns of magnetic flux penetration and distribution in superconductors with arbitrary (in particular, inhomogeneous) distribution of defects. The results of computer simulation showed that a change in the sign of applied magnetic field leads to the appearance of a macroscopic region of zero magnetic induction in the superconductor. The process of motion of this zone from the surface to the center of the superconductor plate was called the magnetic flux annihilation (MFA) wave because the leading front of the input magnetic flux features annihilation of the Abrikosov vortices with opposite directions. The magnetic flux annihilation wave is characterized by its "velocity" $V_{\text{flux}} = dx_{\text{flux}}/dH$ equal to the average distance x_{flux} traveled by the front of the input magnetic flux per unit change in the magnetic field H.

An analysis of the results of computer simulation of the magnetic flux penetration in stratified high-T_c superconductors with defect structures of various types [3–5] allowed us to predict the following peculiarities in the motion of MFA waves.

(1) The "velocity" V_{flux} of the MFA wave front depends on the temperature and defect structure of the

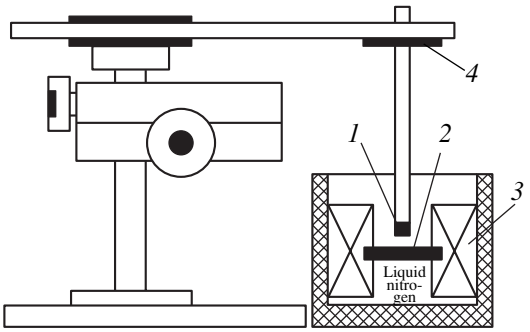


Fig. 1. Schematic diagram of a scanning Hall probe microscope: (*1*) Hall effect transducer; (*2*) sample; (*3*) resistive coil; (*4*) mechanical manipulator.

superconductor: V_{flux} increases with increasing temperature and decreasing defectness of the material and vice versa.

(2) The MFA wave front can exhibit a significant curvature in inhomogeneous superconductors. In the case of strong inhomogeneity, the MFA wave front can separate into several independently propagating fronts.

(3) The MFA wave front "velocity" V_{flux} exhibits a nonmonotonic dependence on the applied magnetic field.

(4) Cyclic variations of the sign of the applied magnetic field leads to periodic motion of the MFA wave front from the surface to the center of the superconductor.

Despite the large number of papers devoted to imaging of the magnetic induction distribution by magnetooptical methods (see review [6] and references therein) and scanning Hall probe microscopy [7–10], no direct observations of a curved MFA wave front in real polycrystalline high-T_c superconductors were reported so far.

This paper presents experimental data on the formation and propagation of MFA waves in inhomogeneous high-T_c superconductors.

2. EXPERIMENTAL

The MFA wave was observed by means of traditional magnetooptics and scanning Hall probe microscopy (SHPM). The experimental SHPM setup (Fig. 1) was based on a semiconductor Hall effect transducer (Hall probe) mounted on a three-coordinate mechanical manipulator. We employed the Hall probe with a sensor area of 450×150 μm^2 and a magnetic responsivity of several dozens microvolts per millitesla. The Hall probe monitored the normal component of the local magnetic induction B_z immediately at the surface of a high-T_c superconductor sample. By moving the Hall probe in the vertical direction, the distance between the sensor and the sample surface was set at 0.1–0.2 mm. Then the Hall probe was scanned along two coordinates in the plane to measure the local magnetic induction variation along the entire sample surface. External magnetic field was generated by a resistive coil. The manipulator allowed the Hall probe to be scanned over an area of 8×8 mm^2 line by line with a resolution of 24×24 points. Further increase in the resolution did not change the image obtained. The output voltage of the Hall effect transducer was amplified and fed via an analog-to-digital converter to a computer in the form of bit sequences. Then the data were rearranged into a two-dimensional array and displayed in the form of an image (map) showing the local magnetic induction $B_z(x, y)$ distribution on the sample surface. The measurements were performed at liquid nitrogen temperature (77 K).

The spatial distribution of the magnetic flux at liquid helium temperature was imaged by a magnetooptical technique [6] based on the Faraday effect in a ferrite garnet indicator film. The corresponding experimental setup is described in detail elsewhere [11]. For the magnetooptical measurements, a sample with the indicator film is fastened with a vacuum seal to the cooled sample holder in an optical flow-type cryostat. The sample was cooled down to liquid helium temperature in a zero magnetic field, after which an external magnetic field was applied along the normal to the sample surface.

We have studied polycrystalline high-T_c superconductors of the $YBa_2Cu_3O_{7-x}$ (Y123) and $Bi_2Sr_2Ca_2Cu_3O_{10-x}$ (Bi2223) systems in the form of tablets with a diameter of 8 mm and a height of 2 mm. Temperature T_c corresponding to the onset of the superconducting transition was determined from the magnetic susceptibility measurements and was equal to 92 K for Y123 and 108 K for Bi2223. The superconducting transition widths ΔT_c for the samples chosen for the experiments were in the range of $\Delta T_c \approx 5$–15 K. We intentionally selected the samples with large widths of the superconducting transition, which ensured a considerable inhomogeneity required for our experiments.

3. RESULTS AND DISCUSSION

The MFA waves were observed in Y123 high-T_c superconductor samples possessing strong spatial inhomogeneity. The sample was cooled in a zero magnetic field to liquid nitrogen temperature, after which a positive magnetic field of $H = 50$ mT was applied. Then the field was switched off and the magnetic flux trapped in the sample was studied by SHPM. Figure 2 shows the typical SHPM image of the residual local magnetic induction distribution. The image reveals three clearly resolved macroscopic superconducting grains possessing residual magnetic induction and a

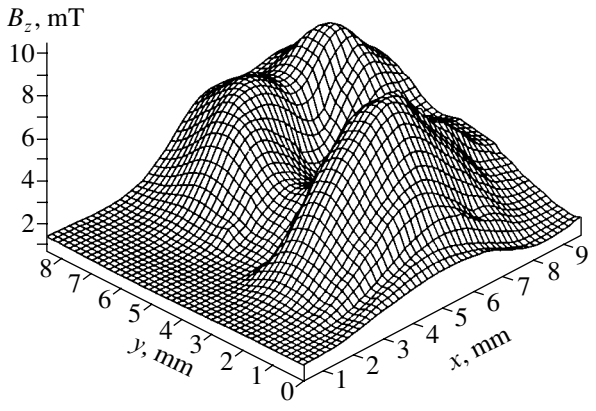


Fig. 2. Residual local magnetic induction $B_z(x, y)$ distribution over the surface of a Y123 sample measured by SHPM at $T = 77$ K.

region between these grains possessing lower critical characteristics.

Application and increase of the external field in the opposite direction leads to the appearance of a region with the opposite orientation of magnetic induction (denoted B_- to distinguish from the direct field B_+) and gives rise to the MFA wave, representing the motion of a zero magnetic induction zone at the leading front of the input magnetic flux. In a strongly inhomogeneous material, magnetic field from the opposite direction rapidly penetrates into the sample under a relatively weak external field to occupy the intergranular region. At a certain value of the negative field, the front separates into two parts surrounding regions of the positive trapped magnetic field (Fig. 3).

The formation of a curved MFA wave was also observed in polycrystalline Bi2223 samples studied by magnetooptical techniques. It should be noted that scanning of the trapped magnetic flux by a Hall probe did not reveal significant inhomogeneity of the samples. At the same time, the results of magnetooptical measurements at a high spatial resolution indicated the presence of inhomogeneities ranging from several units to dozens of microns, which accounted for the curvature of the MFA wave front.

A sample of Bi2223 was mounted in the magnetooptical setup and cooled in a zero magnetic down to $T = 4.0$ K, after which a positive magnetic field was applied. Even a small field was sufficient to obtain a

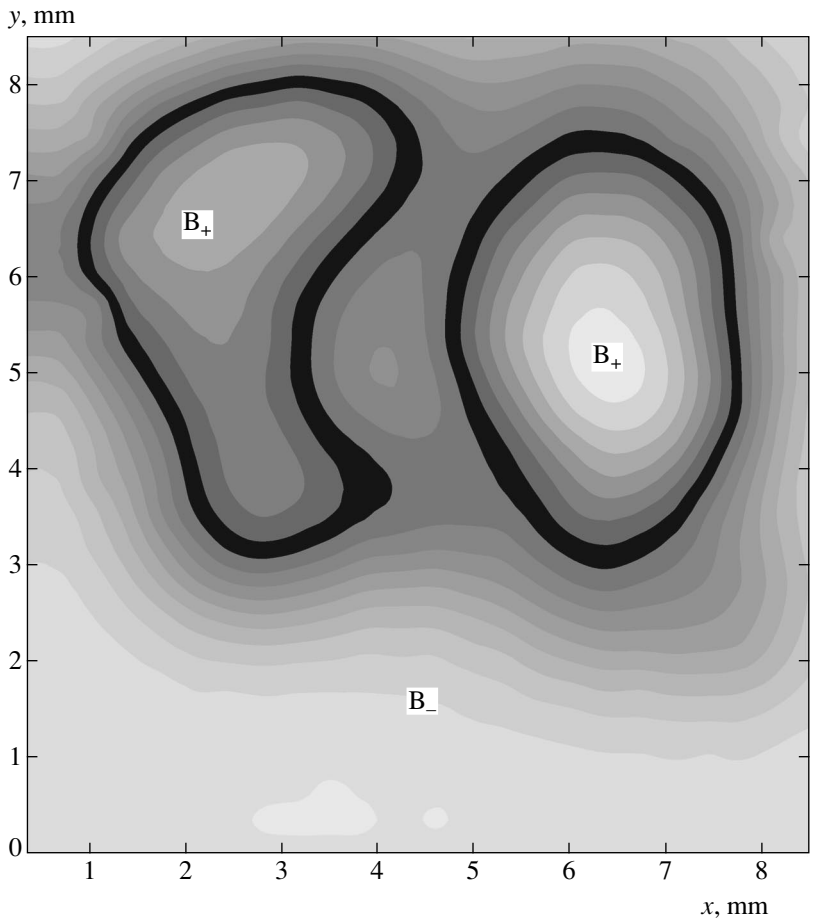


Fig. 3. SHPM image of MFA waves (black zones) surrounding regions of strong pinning in an Y123 sample measured at $T = 77$ K.

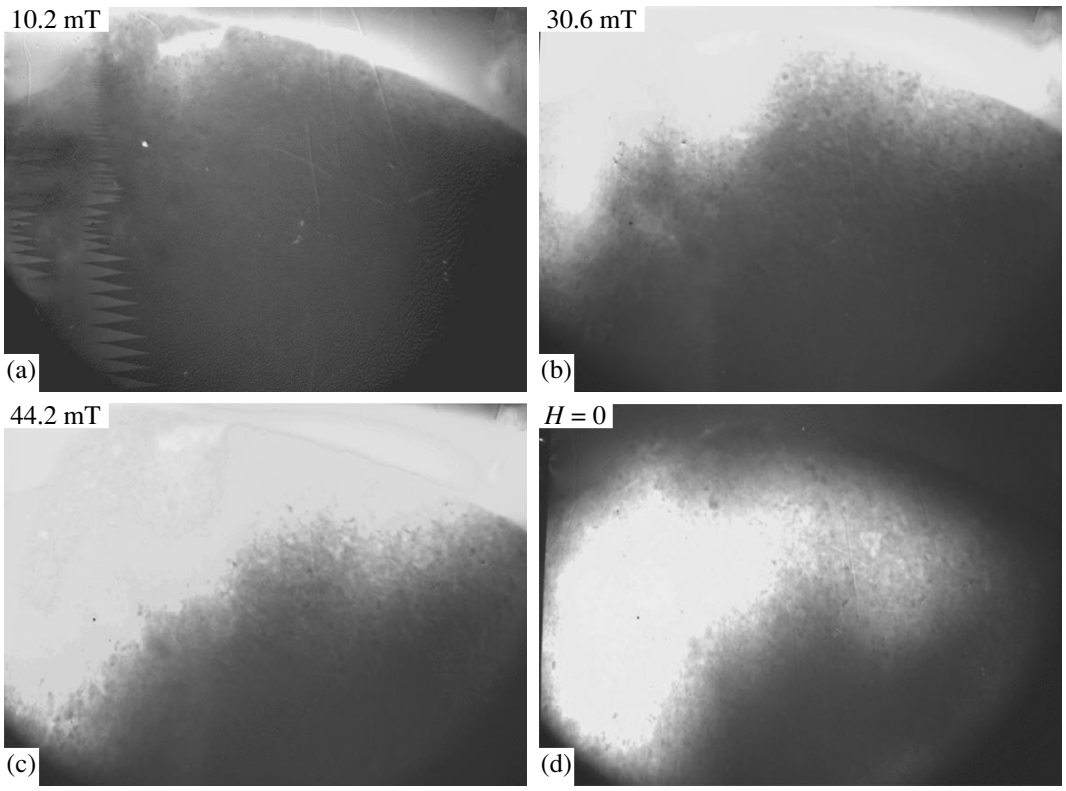


Fig. 4. Magnetooptical images of a polycrystalline Bi2223 sample at various (a–c) values of applied magnetic field and (d) trapped magnetic flux at the maximum field $H = 50$ mT ($T = 4.0$ K).

magnetooptical response in the indicator film, which was manifested by a bright contour at the tablet edge (Fig. 4a). This contour is due to the Meissner effect, whereby the enhanced brightness is caused by demagnetizing fields and the rapid drop in brightness in the direction inward the sample is caused by magnetic field screening. An increase in the applied magnetic field leads to deeper field penetration into the depth of the sample (Figs. 4b and 4c). The magnetooptical image clearly reveals inhomogeneity of the sample containing regions possessing lower critical characteristics. After removal of the maximum positive field ($H = 50$ mT), a positive trapped flux is retained in the sample (Fig. 4d). Note that the image brightness decreases on approaching the sample edge, which reflects a decrease in the magnetic induction down to zero. At the same time, the magnetooptical response arises again in the vicinity of the sample due to a magnetic flux of the opposite sign existing along the sample edge. It should be pointed out that the intensity of the magnetooptical response (image brightness) in these experiments depends only on the absolute value of the magnetic induction and is not influenced by its direction. For this reason, the positive and negative magnetic induction of the same amplitude is manifested by the regions of equal brightness in the magnetooptical image (in SHPM measurements, the direction of the magnetic field can be uniquely determined by the sign of the Hall response voltage).

An increase in the negative magnetic field (Fig. 5) leads to gradually increasing response near the sample edge and gives rise to an initially small ($H = 10.2$ mT) and then strong penetration of a negative flux into the sample. The formation of a curved MFA wave is slightly manifested at $H = 13.6$ mT. Then the MFA wave moves away from the sample edge ($H = 23.8$ mT) and propagates into the region of trapped positive magnetic field. The annihilation wave front is indicated in Fig. 5 by white arrows. The entire sample is gradually occupied by the negative field and the MFA wave disappears.

Figure 6 shows a series of linear profiles of the magnetic induction modulus obtained by digital processing of the magnetooptical images. In the interval of fields $H = 13.6$–34.6 mT, the profiles clearly reveal a local minimum corresponding to the MFA wave. Figure 7 presents the position of the leading MFA wave front x_{flux} and the front "velocity" V_{flux} as functions of the applied magnetic field. The curve of $x_{\text{flux}}(H)$ exhibits two inflections, which correspond to a sharp drop in the value of V_{flux} in the region of $H = 20$–27 mT. This decrease in the MFA wave "velocity," predicted in [3–5], is related to remagnetization of the region of a super-

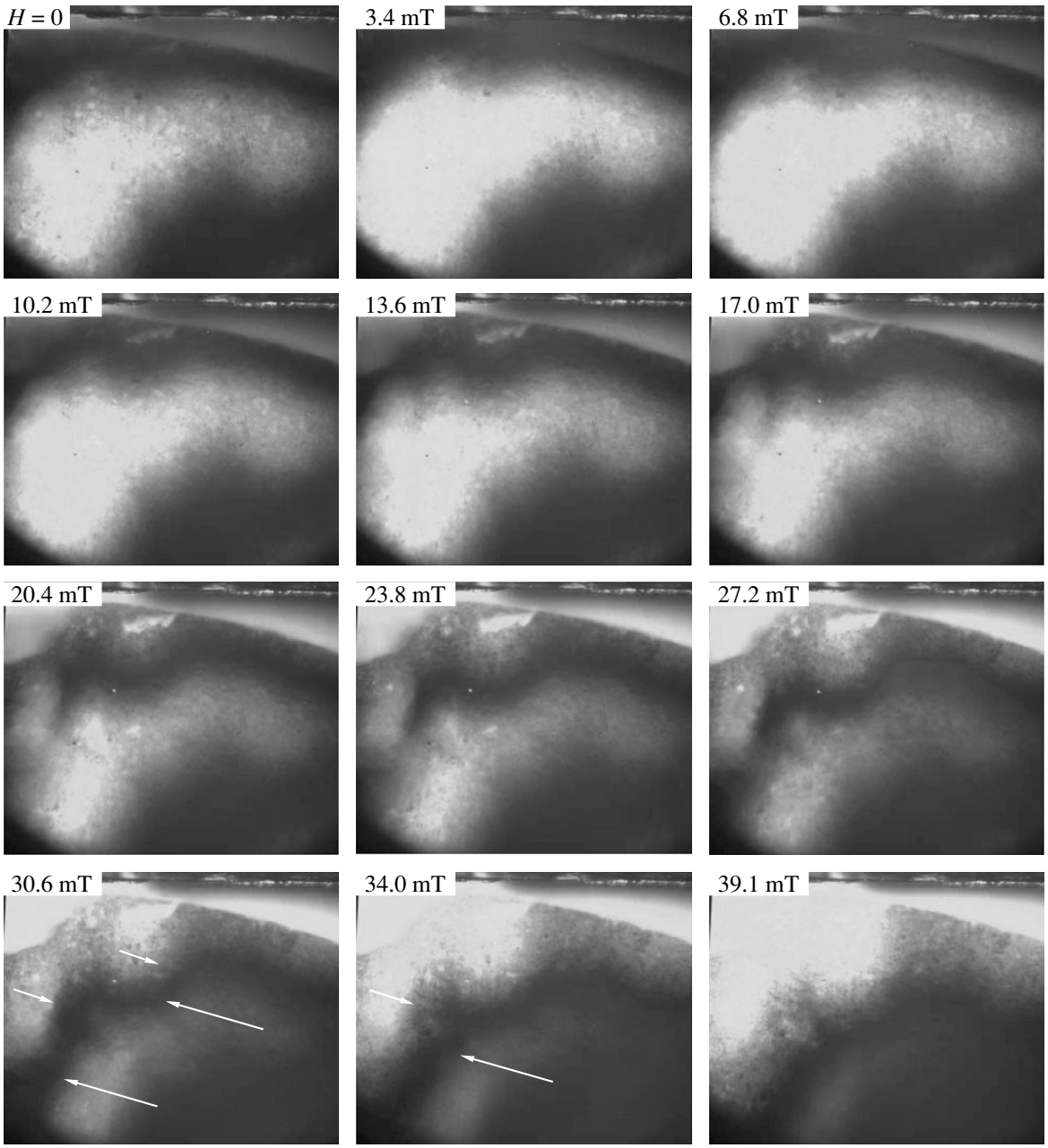


Fig. 5. Magnetooptical images illustrating the dynamics of the formation and propagation of the MFA wave front in a Bi2223 sample (white arrows indicate the leading and trailing fronts of the vortex annihilation wave).

conductor where the magnetic flux of the opposite sign is trapped.

In order to elucidate the nature of the MFA wave, let us consider the results of computer simulation of the process of magnetic field penetration into high-T_c superconductors. The process of remagnetization was studied by Monte Carlo method [4] in high-T_c superconductor plates with various defect configurations set by varying the concentration and potential of the pinning centers. The main result of these numerical calculations was equilibrium configuration of the probability density of Abrikosov vortices in the plate for the given temperature, applied field, and defect configuration. The overall pattern of the magnetic flux distribution in the sample was obtained by adding the magnetic fields of all vortices.

An increase in the applied magnetic field initially leads to manifestation of the Meissner effect. When the field exceeds the first critical value H_{c1}, vortices enter the plate and the magnetic flux starts penetrating deep into the superconductor. In the case of medium or strong pinning, the front of the magnetic flux gradually

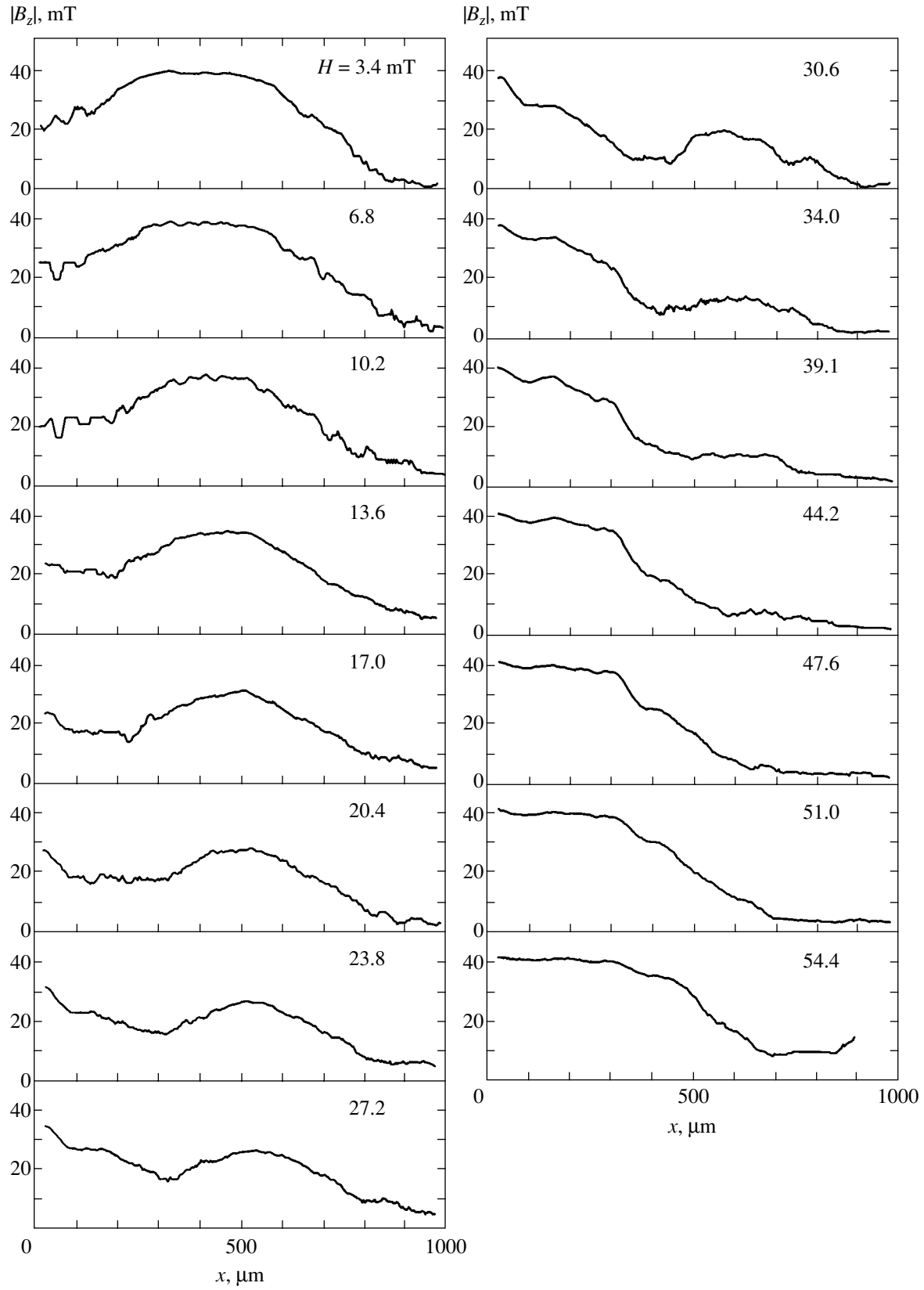


Fig. 6. Linear profiles of the local magnetic induction for various values of the applied magnetic field. In the range of H = 13.6–34.6 mT, the profiles exhibit a clear minimum corresponding to the MFA wave.

moves from the surface to the center of the plate. The magnetic flux front may exhibit curvature that reflects a local inhomogeneity in the distribution of pinning centers. In the case of weak pinning, the magnetic flux spreads over the superconductor volume almost immediately. On reaching the so-called full penetration field (depending on the sample defectness), the magnetic flux occupies the entire plate. Further increase in the applied field leads to a growth in the magnetic induction in the plate, which is manifested by an increase in the vortex density with a trend of triangular lattice formation.

When the external field decreases to zero, the superconductor retains the trapped magnetic flux formed due to the pinning of Abrikosov vortices on defects. As the applied field increases with changed sign, vortices of the opposite sense (antivortices) begin to enter the plate. Annihilation of the vortices and antivortices leads to the formation of a clearly pronounced zone of zero magnetic induction at the front of the entering magnetic flux. This zone, moving inward the superconductor with increasing amplitude of the external field, represents the MFA wave. The lower the pinning, the higher the MFA wave "velocity." In the case of weak pinning, annihilation is almost instantaneous. The MFA wave ceases when the magnetic field exceeds the full penetration value. Subsequently, the MFA waves appear with every change in the sign of the applied magnetic field. The MFA wave velocity also depends on the temperature: as the temperature increases, the wave velocity grows.

According to the results of numerical calculations, the width of the zone of zero magnetic induction at the MFA wave front is on the order of 1 μm. However, experimental data obtained in this study give an estimate on the order of 100 μm. This large value of the zero-induction zone width is probably explained by the polycrystalline structure of the samples studied. This structure is characterized by a large number of weak links acting as channels of accelerated field penetration into the polycrystalline superconductor, which is manifested by strong broadening of the zero-induction zone. This assumption is confirmed by a clearly pronounced cellular structure of the magnetooptical image at the leading front of the input magnetic flux (see Figs. 4 and 5).

It should be emphasized that the MFA wave is a stationary phenomenon and moves only with increasing external magnetic field. Every change in the field sign leads to the formation of the MFA wave moving from the surface to the center of a sample. At a relatively high frequency of the applied field and a complex, spatially inhomogeneous potential relief of the pinning centers, the magnetooptical images of MFA waves may appear as multiple moving dark spots, observed, for example, in [2].

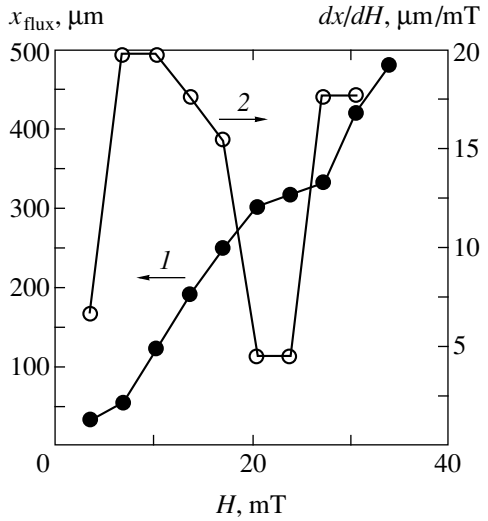


Fig. 7. Plots of the (*1*) position x_{flux} of the leading MFA wave front and (*2*) "velocity" V_{flux} of this front versus the applied magnetic field H.

4. CONCLUSIONS

Using magnetooptical imaging and SHPM techniques, we have demonstrated the MFA waves in inhomogeneous high-T_c superconductors. The obtained experimental data qualitatively agree with the results of numerical analysis [3–5]. Further investigations are required for a comparative analysis of the behavior of MFA waves depending on the temperature and defect state of a superconductor. We suggest modifying the defect state of Bi2223 samples in the stage of synthesis, for example, by introducing various nanodimensional inorganic inclusions playing the role of additional pinning centers. By changing the type and concentration of such inclusions, it is possible to provide for a controlled modification of the defect configuration. This possibility is confirmed by the results of measurements of the trapped flux magnitude in $(Bi,Pb)_2Sr_2Ca_2Cu_3O_{10+x}$ with nanodimensional inclusions of hafnium nitride [12].

ACKNOWLEDGMENTS

The authors are grateful to T.H. Johansen for kindly providing the opportunity for conducting magnetooptical experiments at the Laboratory of Superconductivity of Oslo University (Norway) and to V.A. Kashurnikov for fruitful discussions of the results.

This study was sponsored in part by the Ministry of Education of the Russian Federation within the framework of the Program of Support for Research Work of

Young Scientists at the Leading Scientific and Pedagogical Institutions (project RD02-1.2-152) and by the Federal Targeted Program "Integration" (project B-0048).

REFERENCES

1. C. P. Bean, Phys. Rev. Lett. **8**, 250 (1962).
2. A. V. Eremin, O. S. Esikov, V. A. Kashurnikov, *et al.*, Supercond. Sci. Technol. **14**, 690 (2001).
3. V. A. Kashurnikov, I. A. Rudnev, and M. V. Zubin, Supercond. Sci. Technol. **14**, 695 (2001).
4. V. A. Kashurnikov, I. A. Rudnev, and M. V. Zyubin, Zh. Éksp. Teor. Fiz. **121**, 442 (2002) [JETP **94**, 377 (2002)].
5. M. V. Zyubin, I. A. Rudnev, and V. A. Kashurnikov, Zh. Éksp. Teor. Fiz. **123**, 1212 (2003) [JETP **96**, 1065 (2003)].
6. Ch. Jooss, J. Albrecht, H. Kuhn, *et al.*, Rep. Prog. Phys. **65**, 651 (2002).
7. K. Kawan, J. S. Abell, A. Ohtake, and A. Oota, Supercond. Sci. Technol. **13**, 1373 (2000).
8. K. Kawano and A. Oota, Physica C (Amsterdam) **275**, 1 (1997).
9. A. V. Volkozub, A. D. Caplin, Y. Huang, *et al.*, Physica C (Amsterdam) **310**, 159 (1998).
10. A. N. Grigorenko, G. D. Howells, S. J. Bending, *et al.*, Phys. Rev. B **63**, 052504 (2001).
11. T. H. Johansen, M. Baziljevich, H. Bratsberg, *et al.*, Phys. Rev. B **54**, 16264 (1996).
12. B. P. Mikhaĭlov, G. S. Burkhanov, P. E. Kazin, *et al.*, Neorg. Mater. **39**, 462 (2003) [Inorg. Mater. **39**, 379 (2003)].

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 177–188.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 203–214.

SOLIDS
Electronic Properties

The Effect of the Character of Electron Reflection on the Electromagnetic Properties of an Inhomogeneous Spherical Particle

E. V. Zavitaev and A. A. Yushkanov

Moscow State University of Forestry, Mytishchi-5, Moscow oblast, 141005 Russia
e-mail: yushkanov@mtu-net.ru

Received November 24, 2003

Abstract—The magnetic absorption cross section of a small spherical particle with a dielectric core and a metallic shell is calculated. The general case is considered when the ratio of the radius of the dielectric core to the total radius of the particle may take arbitrary values. The condition of specular–diffuse reflection of conduction electrons from the surfaces of the metal layer of the particle is chosen as the boundary conditions of the problem. The limit cases are considered, and the results are discussed. © *2004 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

Electromagnetic properties of small metal particles have a number of distinctive features [1]. These features are associated with the fact that the electron mean free path in such particles is on the order of their linear dimensions. Here, nonlocal phenomena begin to play an essential role. The classical theory of interaction of electromagnetic waves with metal particles [2] (the Mie theory), which is based on local equations of macroscopic electrodynamics, is inapplicable in this case.

In [3, 4], magnetic dipole absorption of infrared radiation by spherical particles is considered. To describe the electromagnetic response of a particle, the authors apply the standard kinetic theory of degenerate Fermi gas of conduction electrons in metals [5]. In [3], the analysis is restricted to the case of pure diffuse reflection of conduction electrons from the internal surface of a particle, whereas, in [4], the authors carried out a detailed analysis of the magnetic dipole absorption of a spherical particle under the condition of mixed, specular–diffuse, reflection of electrons from the surface of the particle [5]. In all these studies, the authors considered only homogeneous particles; i.e., the question about the internal structure of absorbing particles was not raised.

Note also the studies in which the authors tried to take into account quantum-mechanical effects in the problem; these studies are especially important at low temperatures [6, 7].

Recently, publications have appeared about experimental investigations of particles with complex structure [8, 9]. These particles consist of a dielectric (or metallic) core surrounded by a metallic shell; such a structure naturally affects the optical properties of these particles.

In this paper, which is a logical continuation of [4], we construct a theory of interaction of electromagnetic radiation with an inhomogeneous spherical particle (a metal particle with a dielectric core) with regard to the mixed (specular–diffuse) reflection of electrons inside the metal layer.

2. STATEMENT OF THE PROBLEM

Consider a metal sphere with a dielectric core in the field of a plane electromagnetic wave. The radius of the core is R_1, and the radius of the shell is R_2. The wave frequency is bounded from above by the near-infrared band ($\omega < 2 \times 10^{15}$ s^{-1}). Magnetic absorption of such a particle is associated with eddy currents induced in its metallic shell. In the dipole approximation, neglecting the skin effect (assuming that $R < \delta$, where δ is the skin depth), we obtain the following expression for the eddy electric filed that induces eddy currents:

$$\mathbf{E} = \frac{1}{2c}\left[\mathbf{r} \times \frac{\partial \mathbf{H}}{\partial t}\right] = \frac{\omega}{2ic}[\mathbf{r} \times \mathbf{H}_0]\exp(-i\omega t), \quad (1)$$

where $\mathbf{H} = \mathbf{H}_0\exp(-i\omega t)$ is the magnetic field strength, $\mathbf{r}$ is the radius vector (the origin of coordinates is chosen at the particle center), $\mathbf{H}_0$ is the magnetic field amplitude, ω is the angular frequency of the wave, and c is the speed of light.

The mean dissipated power $\overline{Q}$ in the particle is determined by the formula [10]

$$\overline{Q} = \int \overline{(\mathrm{Re}\mathbf{E})(\mathrm{Re}\mathbf{j})}\, d^3r = \frac{1}{2}\mathrm{Re}\int \mathbf{j} \cdot \mathbf{E}^* d^3r. \quad (2)$$

Here, the bar denotes time averaging, the star denotes complex conjugation, and $\mathbf{j}$ is the eddy current.

When the particle radius R_2 is comparable to or less than the electron mean free path Λ in metal, the relation between **E** and **j** proves to be essentially nonlocal. To describe this relation, we apply a kinetic equation (in the relaxation-time approximation) to the degenerate Fermi gas of conduction electrons in the metallic shell of the particle.

For sufficiently weak external fields, this equation can be linearized with respect to the external field **E** and small deviations $f_1(\mathbf{r}, \mathbf{v})$ from the equilibrium Fermi distribution function f_0:

$$-i\omega f_1 + \mathbf{v}\frac{\partial f_1}{\partial \mathbf{r}} + e(\mathbf{v}\cdot\mathbf{E})\frac{\partial f_0}{\partial \varepsilon} = -\frac{f_1}{\tau}. \qquad (3)$$

Here, e and $\mathbf{v}$ are the charge and the mass of a conduction electron, respectively, and τ is the electron relaxation time.

Next, we consider a quadratic dependence of the energy ε of an electron on its velocity, $\varepsilon = mv^2/2$ (m is the electron effective mass) and apply a stepwise approximation for the equilibrium energy distribution function of electrons $f_0(\varepsilon)$ [11]:

$$f_0(\varepsilon) = \theta(\varepsilon_f - \varepsilon) = \begin{cases} 1, & 0 \le \varepsilon \le \varepsilon_f \\ 0, & \varepsilon_f < \varepsilon, \end{cases}$$

where $\varepsilon_f = m v_f^2/2$ is the Fermi energy (v_f is the Fermi velocity).

The distribution function of electrons is given by

$$f(\mathbf{r}, \mathbf{v}) = f_0(\varepsilon) + f_1(\mathbf{r}, \mathbf{v}), \quad \varepsilon = \frac{m\mathbf{v}^2}{2}.$$

The deviation $f_1(\mathbf{r}, \mathbf{v})$, due to the eddy electric field, of the electron distribution function $f(\mathbf{r}, \mathbf{v})$ from its equilibrium value $f_0(\varepsilon)$ gives rise to the eddy current

$$\mathbf{j} = en\langle\mathbf{v}\rangle = en\Big[\int f_0 d^3v\Big]^{-1}\int f_1 \mathbf{v} d^3v \qquad (4)$$

inside the particle.

The electron concentration n in the metal layer of the particle is determined by the standard formula

$$n = 2\frac{m^3}{h^3}\int f_0 d^3v = 2\frac{m^3}{h^3}\frac{4\pi v_f^3}{3}, \qquad (5)$$

where h is the Planck constant.

Substituting field **E** in the form (1) into Eq. (3), we determine $f_1(\mathbf{r}, \mathbf{v})$ as a solution to this equation. Then, we determine current by (4) and the absorption cross section of the external electromagnetic field of the particle:

$$\sigma = 8\pi\overline{Q}/cH_0^2. \qquad (6)$$

The problem under consideration has a unique solution when we impose boundary conditions on the unknown function $f_1(\mathbf{r}, \mathbf{v})$ on the spherical surfaces of the metallic shell and the dielectric core of the particle. As such boundary conditions, we take the condition of specular–diffuse reflection of electrons from these surfaces [4]. Since electrons can be reflected from the internal (R_1) and external (R_2) boundaries of the metal layer, we have to write out two boundary conditions:

$$f_{11}(\mathbf{r}, \mathbf{v}) = q_1 f_{11}(\mathbf{r}, \mathbf{v}') \text{ for } \begin{cases} |\mathbf{r}| = R_1 \\ \mathbf{rv} > 0, \end{cases} \qquad (7)$$

$$f_{12}(\mathbf{r}, \mathbf{v}) = q_2 f_{12}(\mathbf{r}, \mathbf{v}') \text{ for } \begin{cases} |\mathbf{r}| = R_2 \\ \mathbf{rv} < 0. \end{cases} \qquad (8)$$

The case $\mathbf{rv} > 0$ ($\mathbf{rv} < 0$) corresponds to the motion of electrons from the core (toward the core). Here,

$$\mathbf{v}' = \mathbf{v} - \frac{2\mathbf{r}(\mathbf{r}\cdot\mathbf{v})}{R^2}$$

is the velocity vector, which reduces to the vector $\mathbf{v}$ under the specular reflection from the inner or outer surface of the metal layer at the point $\mathbf{r}$ ($|\mathbf{r}| = R_1$ or $|\mathbf{r}| = R_2$), and q_1 and q_2 are specular reflection coefficients (the probabilities of specular reflection):

$$0 \le q_1 \le 1, \quad 0 \le q_2 \le 2.$$

For $q_1 = 0$ ($q_2 = 0$), we obtain a condition of diffuse reflection of conduction electrons from the inner or outer surface of the metal layer of the particle, whereas, for $q_1 = 1$ ($q_2 = 1$), we obtain a condition of pure specular reflection. When $q \ne 0$ and $q \ne 1$, we obtain various types of mixed (specular–diffuse) reflection of electrons.

3. DISTRIBUTION FUNCTION

The kinetic equation (3) is solved by the method of characteristics [12]. The variation of f_1 along a trajectory (characteristic)

$$d\mathbf{r} = \mathbf{v}dt$$

is determined by the equation

$$df_1 = -\Big(\nu f_1 + e(\mathbf{v}\cdot\mathbf{E})\frac{\partial f_0}{\partial \varepsilon}\Big)dt, \qquad (9)$$

where

$$\nu = \frac{1}{\tau} - i\omega$$

is a complex scattering rate.

The boundary conditions (7) and (8) allow one to trace the variation of the function $f_1(\mathbf{r}, \mathbf{v})$ along the specularly reflected trajectory. At the reflection point $t = t_n$ (from any surface), the function $f_1(t)$ has a jump

$$f_1(t_n + 0) = q f_1(t_n - 0). \quad (10)$$

The signs + and – indicate the limits of the function $f_1(\mathbf{r}, \mathbf{v})$ at the reflection point t_n on the left and right along the flight time, respectively. The quantity q stands for q_1 or q_2, depending on the place of scattering.

Under the specular reflection, the angular momentum $[\mathbf{r} \times \mathbf{v}] = [\mathbf{r} \times \mathbf{v}']$ is conserved; therefore,

$$[\mathbf{r} \times \mathbf{v}] = \text{const}$$

on the trajectory under consideration.

The difference $t_n - t_{n-1}$ is independent of the number n of the reflection point:

$$t_n = nT + \text{const}, \quad n \in Z,$$

where T is the flight time of an electron from point $\mathbf{r}_{n-1}$ to point $\mathbf{r}_n$ with velocity $\mathbf{v}$:

$$T = -\frac{2(\mathbf{v}_n \cdot \mathbf{r}_n)}{v^2}.$$

The product $\mathbf{v} \cdot \mathbf{E}$ is also constant on the trajectory,

$$\mathbf{v} \cdot \mathbf{E} = \frac{\omega}{2ic}[\mathbf{r} \times \mathbf{H}] \cdot \mathbf{v} = \frac{i\omega}{2c}[\mathbf{r} \times \mathbf{v}] \cdot \mathbf{H} = \text{const}.$$

A solution to Eq. (9) is given by the function

$$f_1 = C\exp(-\nu t) + A, \quad (11)$$

where

$$A = -\frac{e(\mathbf{v} \cdot \mathbf{E})}{\nu}\frac{\partial f_0}{\partial \varepsilon}.$$

The parameter t in (11) has the meaning of the motion time of an electron along the trajectory from the boundary where the reflection occurs to the point $\mathbf{r}$ at velocity $\mathbf{v}$.

Let us solve this equation on the interval (t_{n-1}, t_n) in the case when the electron moves along a trajectory that does not intersect the core of the particle under a specular reflection.

At the initial moment ($t = 0$),

$$f_1(t_{n-1} + 0) = C + A.$$

Hence, we seek a constant C such that

$$C = f_1(t_{n-1} + 0) - A.$$

Now, we obtain a relation between the initial values of the function f_1 on two adjacent segments of the trajectory. Since $t_n - 0 = t_{n-1} + T$, we have

$$f_1(t_n - 0) = (f_1(t_{n-1} + 0) - A)e^{-\nu T} + A$$
$$= A(1 - e^{-\nu T}) + f_1(t_{n-1} + 0)e^{-\nu T}.$$

Applying Eq. (10), we obtain

$$f_1(t_n + 0) = q_2\{A(1 - e^{-\nu T}) + f_1(t_{n-1} + 0)e^{-\nu T}\}. \quad (12)$$

Then, applying this recurrent relation to express $f_1(t_{n-1} + 0)$ in terms of $f_1(t_{n-2} + 0)$, etc., we express $f_1(t_n + 0)$ in terms of an infinite geometric progression with the ratio $q_2e^{-\nu T}$. Summing this progression, we obtain

$$f_1(t_n + 0) = \frac{q_2A(1 - e^{-\nu T})}{1 - q_2e^{-\nu T}}. \quad (13)$$

To determine a specific form of the solution to Eq. (9), we apply condition (13). When $t = 0$,

$$\frac{q_2A(1 - e^{-\nu T})}{1 - q_2e^{-\nu T}} = C + A.$$

Hence,

$$C = A\left\{\frac{q_2(1 - e^{-\nu T})}{1 - q_2e^{-\nu T}} - 1\right\} = A\left\{\frac{q_2 - 1}{1 - q_2e^{-\nu T}}\right\}.$$

Therefore,

$$f_{10}(t_2) = A\left\{\frac{q_2 - 1}{1 - q_2e^{-\nu T}}\right\}\exp(-\nu t_2) + A$$
$$= A\left\{\frac{(q_2 - 1)\exp(-\nu t_2)}{1 - q_2\exp(-\nu T_2)} + 1\right\}. \quad (14)$$

The parameters t_2 and T_2 can be related to the coordinates of the point $(\mathbf{r}, \mathbf{v})$ in the phase space (for $n = 0$, $\mathbf{v}_0 = \mathbf{v}$) by the conditions

$$\mathbf{r} = \mathbf{r}_0 + \mathbf{v}t_2, \quad \mathbf{v} \cdot \mathbf{r}_0 < 0,$$

$$r_0^2 = R_2^2, \quad T_2 = -\frac{2(\mathbf{v} \cdot \mathbf{r}_0)}{v^2}.$$

Eliminating $\mathbf{r}_0$, we obtain

$$t_2 = \{\mathbf{r}\cdot\mathbf{v} + [(\mathbf{r}\cdot\mathbf{v})^2 + (R_2^2 - r^2)v^2]^{1/2}\}/v^2, \quad (15)$$

$$T_2 = 2[(\mathbf{r}\cdot\mathbf{v})^2 + (R_2^2 - r^2)v^2]^{1/2}/v^2. \quad (16)$$

Formulas (14)–(16) completely define the function $f_1(\mathbf{r}, \mathbf{v})$ in the case when electrons move along the trajectories that do not intersect the core of the particle.

Now, we pass on to the case of double specular reflection of an electron (from the core and from the external boundary of the metallic shell). We solve the kinetic equation (9) on the interval (t_{n-1}, t_n) under the assumption that, at a certain moment of time, an electron is reflected from the boundary of the metal layer (previously, it was reflected from the core). Then (see (12)),

$$f_n^+ = q_1\{A(1 - e^{-\nu T}) + f_{n-1}^+ e^{-\nu T}\}.$$

Next, expressing f_{n-1}^+ in terms of f_{n-2}^+, etc., we obtain

$$f_n^+ = q_1\{A(1 - e^{-\nu T}) + q_2 e^{-\nu T}(A(1 - e^{-\nu T}) + f_{n-2}^+ e^{-\nu T})\}$$

$$= q_1 A(1 - e^{-\nu T}) + q_1 q_2 A(1 - e^{-\nu T})e^{-\nu T} + q_1 q_2 f_{n-2}^+ e^{-2\nu T}$$

$$= q_1 A(1 - e^{-\nu T}) + q_1 q_2 A(1 - e^{-\nu T})e^{-\nu T} + q_1^2 q_2 e^{-2\nu T}\{A(1 - e^{-\nu T}) + f_{n-3}^+ e^{-\nu T}\}$$

$$= q_1 A(1 - e^{-\nu T}) + q_1 q_2 A(1 - e^{-\nu T})e^{-\nu T} + A q_1^2 q_2 e^{-2\nu T}(1 - e^{-\nu T}) + q_1^2 q_2 f_{n-3}^+ e^{-3\nu T}$$

$$= q_1 A(1 - e^{-\nu T})\{1 + q_2 e^{-\nu T} + q_1 q_2 e^{-2\nu T} + q_1 q_2^2 e^{-3\nu T} + q_1^2 q_2^2 e^{-4\nu T} + \ldots\}.$$

Let

$$S_2 = 1 + q_2 e^{-\nu T} + q_1 q_2 e^{-2\nu T} + q_1 q_2^2 e^{-3\nu T} + q_1^2 q_2^2 e^{-4\nu T} + \ldots.$$

Then,

$$S_2 = 1 + q_2 e^{-\nu T}(1 + q_1 e^{-\nu T}) + q_1 q_2 e^{-2\nu T} + q_1^2 q_2 e^{-3\nu T} + \ldots = 1 + q_2 e^{-\nu T}\{1 + q_1 e^{-\nu T}(1 + q_2 e^{-\nu T} + q_1 q_2 e^{-2\nu T} + \ldots)\} = 1 + q_2 e^{-\nu T}(1 + q_1 e^{-\nu T} S_2).$$

Hence,

$$S_2 = \frac{1 + q_2 e^{-\nu T}}{1 - q_1 q_2 e^{-2\nu T}}$$

and

$$f_{n2}^+ = q_1 A(1 - e^{-\nu T})S_2 = \frac{q_1 A(1 - e^{-\nu T})(1 + q_2 e^{-\nu T})}{1 - q_1 q_2 e^{-2\nu T}}. \quad (17)$$

Let us write the solution of the kinetic equation (9) in explicit form. Using the initial condition (17), we have

$$\frac{q_1 A(1 - e^{-\nu T})(1 + q_2 e^{-\nu T})}{1 - q_1 q_2 e^{-2\nu T}} = C_2 + A.$$

Hence,

$$C_2 = A\frac{q_1(1 - e^{-\nu T} + q_2 e^{-\nu T}) - 1}{1 - q_1 q_2 e^{-2\nu T}}. \quad (18)$$

Substituting (18) into (11), we obtain the deviation $f_{12}(t)$ of the distribution function of electrons from its equilibrium value:

$$f_{12}(t_2) = A\left\{\frac{q_1(1 - e^{-\nu T_1} + q_2 e^{-\nu T_1}) - 1}{1 - q_1 q_2 e^{-2\nu T_1}} e^{-\nu t_2} + 1\right\}. \quad (19)$$

The deviation $f_{11}(t)$ of the distribution function of electrons reflected from the core of the particle is determined analogously:

$$f_{11}(t_1) = A\left\{\frac{q_2(1 - e^{-\nu T_1} + q_1 e^{-\nu T_1}) - 1}{1 - q_1 q_2 e^{-2\nu T_1}} e^{-\nu t_1} + 1\right\}. \quad (20)$$

The parameter t_1 in (20) is determined as

$$t_1 = \{\mathbf{r}\cdot\mathbf{v} - [(\mathbf{r}\cdot\mathbf{v})^2 + (R_1^2 - r^2)v^2]^{1/2}\}/v^2. \quad (21)$$

Indeed, from the obvious vector equality $\mathbf{r} = \mathbf{r}_0 + \mathbf{v}t_1$, where $\mathbf{r}_0$ is the radius vector of the electron at the moment of reflection from the core of the particle ($r_0^2 = R_1^2$), one can easily obtain (21) by squaring both sides of this equality and solving the equation obtained for t_1.

The parameter T_1 (the period of motion of an electron under double reflection, i.e., a period of time after which the electron is again reflected from the core or from the external boundary of the metal layer) can be determined from another vector equality, $\mathbf{r}^* = \mathbf{r}_0 + \mathbf{v}T_1$,

where $\mathbf{r}_0 = \mathbf{r} - \mathbf{v}t_1$, $|\mathbf{r}_0| = R_1$, and $|\mathbf{r}^*| = R_2$ (we assume that an electron moves from the core to the boundary of the particle). Squaring both sides of this equality, we obtain the quadratic equation

$$v^2T_1^2 + 2(\mathbf{r} - \mathbf{v}t_1)\mathbf{v}T_1 + (R_1^2 - R_2^2) = 0. \qquad (22)$$

A solution to this equation (which will be presented below) allows us to determine T_1.

Formulas (15) and (19)–(22) completely define the function $f_1(\mathbf{r}, \mathbf{v})$ in the case when electrons move along a trajectory and experience a double reflection, from the core and the external boundary of the particle.

4. ABSORPTION CROSS SECTION

The distribution function obtained in the previous section allows us to calculate the current (4), the mean dissipated power (2), and the absorption cross section (6) for the energy of the external electromagnetic field.

To calculate the integrals (4) and (2), it is convenient to pass to spherical coordinates both in the space of coordinates (r, θ, φ; the polar axis is z, and the vector $\mathbf{H}_0$ is parallel to the axis z) and the space of velocities (v, α, β; the polar axis is v_r). Field (1) in spherical coordinates has only the φ component:

$$\mathbf{E} = E_\varphi \mathbf{e}_\varphi, \quad E_\varphi = \frac{i\omega}{2c} r H_0 \sin\theta e^{-i\omega t}. \qquad (23)$$

Accordingly, current (4) also has only the φ component (the current lines are closed circles with the centers on the axis z in the planes perpendicular to the axis z).

When integrating expression (4), one should bear in mind that the place where electrons are reflected in the particle is determined by the polar angle α in the space of velocities.

(1) If the inequalities $\alpha_0 \le \alpha \le \pi - \alpha_0$ hold, where the angle α_0 is determined by

$$\alpha_0 = \arccos\left(\frac{\sqrt{r^2 - R_1^2}}{r}\right), \qquad (24)$$

then the electron trajectory does not intersect the core and the electron is reflected from the external boundary of the metal layer of the particle. In this case, by the function $f_1(\mathbf{r}, \mathbf{v})$, we mean $f_{10}(\mathbf{r}, \mathbf{v})$ ($t = t_2$, $T = T_2$) (see (14)).

(2) If $\pi - \alpha_0 < \alpha \le \pi$, then electrons travel toward the core of the particle and, by the function $f_1(\mathbf{r}, \mathbf{v})$, we mean $f_{12}(\mathbf{r}, \mathbf{v})$ ($t = t_2$, $T = T_1$) (see (19)).

(3) Finally, if $0 < \alpha \le \alpha_0$, then electrons travel from the core of the particle and, by the function $f_1(\mathbf{r}, \mathbf{v})$, we mean $f_{11}(\mathbf{r}, \mathbf{v})$ ($t = t_1$, $T = T_1$) (see (20)).

Then, the eddy current (4) inside the particle is calculated by the formula

$$j_\varphi = 2e\left(\frac{m}{h}\right)^3 \frac{1}{\nu}\int v_\varphi e(\mathbf{v}\cdot\mathbf{E})\frac{\partial f_0}{\partial \varepsilon}\left\{\frac{(q_2-1)e^{-\nu t_2}}{1-q_2e^{-\nu T_2}} + 1\right\} d^3v$$

$$+ 2e\left(\frac{m}{h}\right)^3 \frac{1}{\nu}\int v_\varphi e(\mathbf{v}\cdot\mathbf{E})\frac{\partial f_0}{\partial \varepsilon}$$

$$\times\left\{\frac{q_1(1-e^{-\nu T_1}+q_2e^{-\nu T_1})-1}{1-q_1q_2e^{-2\nu T_1}}e^{-\nu t_2} + 1\right\} d^3v$$

$$+ 2e\left(\frac{m}{h}\right)^3 \frac{1}{\nu}\int v_\varphi e(\mathbf{v}\cdot\mathbf{E})\frac{\partial f_0}{\partial \varepsilon}$$

$$\times\left\{\frac{q_2(1-e^{-\nu T_1}+q_1e^{-\nu T_1})-1}{1-q_1q_2e^{-2\nu T_1}}e^{-\nu t_1} + 1\right\} d^3v$$

$$= E_\varphi 2e^2\left(\frac{m}{h}\right)^3 \frac{1}{\nu}\int v_\varphi^2 \delta(\varepsilon - \varepsilon_f)\left\{\frac{(q_2-1)e^{-\nu t_2}}{1-q_2e^{-\nu T_2}} + 1\right\} d^3v$$

$$+ E_\varphi 2e^2\left(\frac{m}{h}\right)^3 \frac{1}{\nu}\int v_\varphi^2 \delta(\varepsilon - \varepsilon_f)$$

$$+\left\{\frac{q_1(1-e^{-\nu T_1}+q_2e^{-\nu T_1})-1}{1-q_1q_2e^{-2\nu T_1}}e^{-\nu t_2} + 1\right\} d^3v$$

$$+ E_\varphi 2e^2\left(\frac{m}{h}\right)^3 \frac{1}{\nu}\int v_\varphi^2 \delta(\varepsilon - \varepsilon_f)$$

$$\times\left\{\frac{q_2(1-e^{-\nu T_1}+q_1e^{-\nu T_1})-1}{1-q_1q_2e^{-2\nu T_1}}e^{-\nu t_1} + 1\right\} d^3v.$$

Taking into account that $v_\varphi = v\sin\alpha\cos\beta$ and substituting the integration limits, we arrive at the expression

$$j_\varphi = E_\varphi 2e^2\left(\frac{m}{h}\right)^3 \frac{1}{\nu}\int_0^{v_f}\int_{\alpha_0}^{\pi-\alpha_0}\int_0^{2\pi} v^4\delta(\varepsilon - \varepsilon_f)$$

$$\times\left\{\frac{(q_2-1)e^{-\nu t_2}}{1-q_2e^{-\nu T_2}} + 1\right\}\sin^3\alpha\cos^2\beta\, dv\, d\alpha\, d\beta$$

$$+ E_\varphi 2e^2\left(\frac{m}{h}\right)^3 \frac{1}{\nu}\int_0^{v_f}\int_{\pi-\alpha_0}^{\pi}\int_0^{2\pi} v^4\delta(\varepsilon - \varepsilon_f)$$

$$\times\left\{\frac{q_1(1-e^{-\nu T_1}+q_2e^{-\nu T_1})-1}{1-q_1q_2e^{-2\nu T_1}}e^{-\nu t_2}+1\right\}$$

$$\times\sin^3\alpha\cos^2\beta\,d\nu d\alpha d\beta$$

$$+E_\varphi 2e^2\left(\frac{m}{h}\right)^3\frac{1}{\nu}\int_0^{\nu_f}\int_0^{\alpha_0}\int_0^{2\pi}\nu^4\delta(\varepsilon-\varepsilon_f)$$

$$\times\left\{\frac{q_2(1-e^{-\nu T_1}+q_1e^{-\nu T_1})-1}{1-q_1q_2e^{-2\nu T_1}}e^{-\nu t_1}+1\right\}$$

$$\times\sin^3\alpha\cos^2\beta\,d\nu d\alpha d\beta.$$

Integrating with respect to ν and β, we obtain

$$j_\varphi=\frac{3nE_\varphi e^2}{4m\nu}\int_{\alpha_0}^{\pi-\alpha_0}\left\{\frac{(q_2-1)e^{-\nu t_2}}{1-q_2e^{-\nu T_2}}+1\right\}\sin^3\alpha\,d\alpha+\frac{3nE_\varphi e^2}{4m\nu}$$

$$\times\int_{\pi-\alpha_0}^{\pi}\left\{\frac{q_1(1-e^{-\nu T_1}+q_2e^{-\nu T_1})-1}{1-q_1q_2e^{-2\nu T_1}}e^{-\nu t_2}+1\right\}\sin^3\alpha\,d\alpha$$

$$+\frac{3nE_\varphi e^2}{4m\nu}\qquad(25)$$

$$\times\int_0^{\alpha_0}\left\{\frac{q_2(1-e^{-\nu T_1}+q_1e^{-\nu T_1})-1}{1-q_1q_2e^{-2\nu T_1}}e^{-\nu t_1}+1\right\}\sin^3\alpha\,d\alpha.$$

Here, we took into account that the concentration of conduction electrons in metals is determined by formula (5).

The absorption cross section of an inhomogeneous particle for electromagnetic radiation is given by

$$\sigma=\frac{1}{2}\frac{8\pi}{cH_0^2}\mathrm{Re}\{\int j_\varphi E_\varphi^* d^3r\};$$

taking into account (23) and (25), after straightforward calculations, we obtain

$$\sigma=\mathrm{Re}\left\{\frac{3ne^2\pi\omega^2}{4m\nu c^3}\int_{\alpha_0}^{\pi-\alpha_0}\left[\frac{(q_2-1)e^{-\nu t_2}}{1-q_2e^{-\nu T_2}}+1\right]\right.$$

$$\times r^2\sin^2\theta\sin^3\alpha\,d\alpha d^3r+\frac{3ne^2\pi\omega^2}{4m\nu c^3}$$

$$\times\int_{\pi-\alpha_0}^{\pi}\int\left\{\frac{q_1(1-e^{-\nu T_1}+q_2e^{-\nu T_1})-1}{1-q_1q_2e^{-2\nu T_1}}e^{-\nu t_2}+1\right\}$$

$$\times r^2\sin^2\theta\sin^3\alpha\,d\alpha d^3r+\frac{3ne^2\pi\omega^2}{4m\nu c^3}$$

$$\times\int_0^{\alpha_0}\int\left\{\frac{q_2(1-e^{-\nu T_1}+q_1e^{-\nu T_1})-1}{1-q_1q_2e^{-2\nu T_1}}e^{-\nu t_1}+1\right\}$$

$$\left.\times r^2\sin^2\theta\sin^3\alpha\,d\alpha d^3r\right\}.$$

Substituting the integration limits in the outer integral, we obtain

$$\sigma=\frac{3ne^2\pi\omega^2}{4mc^3}\mathrm{Re}\left\{\int_{R_1}^{R_2}\int_0^{2\pi}\int_0^{\pi-\alpha_0}\int_{\alpha_0}^{\pi}\left[\frac{(q_2-1)e^{-\nu t_2}}{\nu(1-q_2e^{-\nu T_2})}+\frac{1}{\nu}\right]\right.$$

$$\times r^4\sin^3\theta\sin^3\alpha\,d\alpha dr d\theta d\varphi$$

$$+\int_{R_1}^{R_2}\int_0^{2\pi}\int_0^{\pi}\int_{\pi-\alpha_0}^{\pi}\left[\frac{q_1(1-e^{-\nu T_1}+q_2e^{-\nu T_1})-1}{\nu(1-q_1q_2e^{-2\nu T_1})}e^{-\nu t_2}+\frac{1}{\nu}\right]$$

$$\times r^4\sin^3\theta\sin^3\alpha\,d\alpha dr d\theta d\varphi$$

$$\left.+\int_{R_1}^{R_2}\int_0^{2\pi}\int_0^{\alpha_0}\int_0^{\pi}\left[\frac{q_2(1-e^{-\nu T_1}+q_1e^{-\nu T_1})-1}{\nu(1-q_1q_2e^{-2\nu T_1})}e^{-\nu t_1}+\frac{1}{\nu}\right]\right.$$

$$\left.\times r^4\sin^3\theta\sin^3\alpha\,d\alpha dr d\theta d\varphi\right\}.$$

The integrals with respect to θ and φ are elementary; therefore, we finally obtain

$$\sigma=\frac{2ne^2\pi^2\omega^2}{mc^3}$$

$$\times\mathrm{Re}\left\{\int_{R_1}^{R_2}r^4dr\int_{\alpha_0}^{\pi-\alpha_0}\left[\frac{(q_2-1)e^{-\nu t_2}}{\nu(1-q_2e^{-\nu T_2})}+\frac{1}{\nu}\right]\sin^3\alpha\,d\alpha+\int_{R_1}^{R_2}r^4dr\right.\qquad(26)$$

$$\left.\times\int_{\pi-\alpha_0}^{\pi}\left[\frac{q_1(1-e^{-\nu T_1}+q_2e^{-\nu T_1})-1}{\nu(1-q_1q_2e^{-2\nu T_1})}e^{-\nu t_2}+\frac{1}{\nu}\right]\sin^3\alpha\,d\alpha\right.$$

$$+\int_{R_1}^{R_2} r^4 dr \int_0^{\alpha_0} \left[\frac{q_2(1-e^{-\nu T_1}+q_1 e^{-\nu T_1})-1}{\nu(1-q_1q_2e^{-2\nu T_1})}e^{-\nu t_1}+\frac{1}{\nu}\right]\sin^3\alpha d\alpha\Bigg\}.$$

For further calculations and the analysis of the results, it is convenient to represent the complex expression (26) for calculating the absorption cross section as

$$\sigma = \sigma_1 + \sigma_2 + \sigma_3, \tag{27}$$

where

$$\sigma_1 = \mathrm{Re}\Bigg\{\frac{2ne^2\pi^2\omega^2}{mc^3}\int_{R_1}^{R_2} r^4 dr \times \int_{\alpha_0}^{\pi-\alpha_0}\left[\frac{(q_2-1)e^{-\nu t_2}}{\nu(1-q_2e^{-\nu T_2})}+\frac{1}{\nu}\right]\sin^3\alpha d\alpha\Bigg\}, \tag{28}$$

$$\sigma_2 = \mathrm{Re}\Bigg\{\frac{2ne^2\pi^2\omega^2}{mc^3}\int_{R_1}^{R_2} r^4 dr \times \int_{\pi-\alpha_0}^{\pi}\left[\frac{q_1(1-e^{-\nu T_1}+q_2e^{-\nu T_1})-1}{\nu(1-q_1q_2e^{-2\nu T_1})}e^{-\nu t_2}+\frac{1}{\nu}\right]\sin^3\alpha d\alpha\Bigg\}, \tag{29}$$

$$\sigma_3 = \mathrm{Re}\Bigg\{\frac{2ne^2\pi^2\omega^2}{mc^3}\int_{R_1}^{R_2} r^4 dr \times \int_0^{\alpha_0}\left[\frac{q_2(1-e^{-\nu T_1}+q_1e^{-\nu T_1})-1}{\nu(1-q_1q_2e^{-2\nu T_1})}e^{-\nu t_1}+\frac{1}{\nu}\right]\sin^3\alpha d\alpha\Bigg\}. \tag{30}$$

Introduce new variables

$$\xi = \frac{r}{R_2},$$

$$z = \nu\frac{R_2}{v_\mathrm{f}} = \left(\frac{1}{\tau}-i\omega\right)\frac{R_2}{v_\mathrm{f}} = x - iy, \tag{31}$$

$$K = \frac{R_1}{R_2}$$

and reduce expressions (15), (16), (21), and (24) to

$$t_2 = \frac{R_2}{v_\mathrm{f}}\eta, \quad \eta = \xi\cos\alpha + \sqrt{1-\xi^2\sin^2\alpha},$$

$$T_2 = \frac{R_2}{v_\mathrm{f}}\eta_0, \quad \eta_0 = 2\sqrt{1-\xi^2\sin^2\alpha},$$

$$t_1 = \frac{R_2}{v_\mathrm{f}}\psi, \quad \psi = \xi\cos\alpha - \sqrt{K^2-\xi^2\sin^2\alpha},$$

$$\alpha_0 = \arccos\left(\sqrt{1-\frac{K^2}{\xi^2}}\right).$$

Here, we took into account that $\mathbf{r}\cdot\mathbf{v} = rv_\mathrm{f}\cos\alpha$ (all the electrons on the Fermi surface inside the metal layer of a particle move at velocity v_f).

Solving Eq. (22), we determine the parameter T_1:

$$T_1 = \frac{R_2}{v_\mathrm{f}}\psi_0, \quad \psi_0 = \sqrt{1-\xi^2\sin^2\alpha} - \sqrt{K^2-\xi^2\sin^2\alpha}.$$

Then, formulas (28)–(30) are rewritten as

$$\sigma_1 = \mathrm{Re}\Bigg\{\frac{2ne^2\pi^2 v_\mathrm{f}R_2^4}{mc^3}\frac{y^2}{z}\int_K^1 \xi^4 d\xi \times \int_{\alpha_0}^{\pi-\alpha_0}\left[\frac{(q_2-1)e^{-z\eta}}{1-q_2e^{-z\eta_0}}+1\right]\sin^3\alpha d\alpha\Bigg\},$$

$$\sigma_2 = \mathrm{Re}\Bigg\{\frac{2ne^2\pi^2 v_\mathrm{f}R_2^4}{mc^3}\frac{y^2}{z}\int_K^1 \xi^4 d\xi \times \int_{\pi-\alpha_0}^{\pi}\left[\frac{q_1(1-e^{-z\psi_0}+q_2e^{-z\psi_0})-1}{1-q_1q_2e^{-2z\psi_0}}e^{-z\eta}+1\right]\sin^3\alpha d\alpha\Bigg\},$$

$$\sigma_3 = \mathrm{Re}\Bigg\{\frac{2ne^2\pi^2 v_\mathrm{f}R_2^4}{mc^3}\frac{y^2}{z}\int_K^1 \xi^4 d\xi \times \int_0^{\alpha_0}\left[\frac{q_2(1-e^{-z\psi_0}+q_1e^{-z\psi_0})-1}{1-q_1q_2e^{-2z\psi_0}}e^{-z\psi}+1\right]\sin^3\alpha d\alpha\Bigg\}.$$

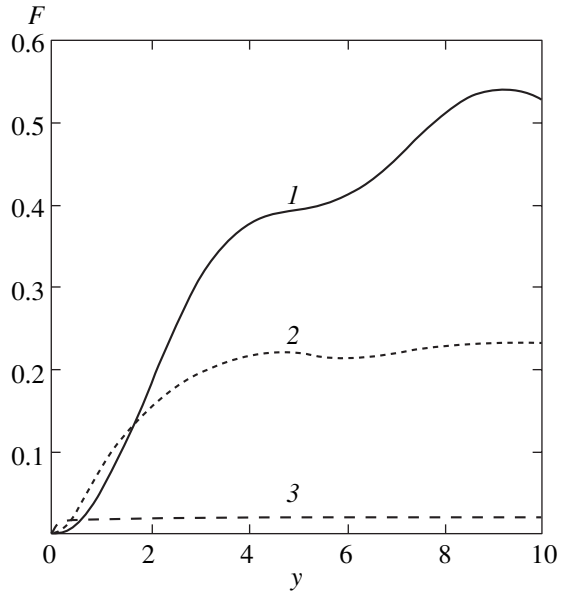


Fig. 1. The dimensionless absorption cross section F as a function of the dimensionless frequency y for fixed values of the dimensionless inverse mean free path $x = 0.1$ and the ratio of the core radius to the total radius of the particle $K = 0.8$ for equal values reflection coefficients q_1 and q_2 for each curve; (*1*) $q_1 = q_2 = 0$, (*2*) $q_1 = q_2 = 0.5$, and (*3*) $q_1 = q_2 = 1$.

The absorption cross section (27) can be represented as

$$\sigma = \sigma_0(F_1 + F_2 + F_3),$$

where

$$\sigma_0 = \frac{2\pi^2 n e^2 v_f R_2^4}{mc^3}, \tag{32}$$

$$F_1 = \mathrm{Re}\left\{\frac{y^2}{z}\int_K^1 \xi^4 d\xi \times \int_{\alpha_0}^{\pi-\alpha_0}\left[\frac{(q_2-1)e^{-z\eta}}{1-q_2 e^{-z\eta_0}}+1\right]\sin^3\alpha d\alpha\right\}, \tag{33}$$

$$F_2 = \mathrm{Re}\left\{\frac{y^2}{z}\int_K^1 \xi^4 d\xi \times \int_{\pi-\alpha_0}^{\pi}\left[\frac{q_1(1-e^{-z\psi_0}+q_2 e^{-z\psi_0})-1}{1-q_1 q_2 e^{-2z\psi_0}}e^{-z\eta}+1\right]\sin^3\alpha d\alpha\right\}, \tag{34}$$

$$F_3 = \mathrm{Re}\left\{\frac{y^2}{z}\int_K^1 \xi^4 d\xi \times \int_0^{\alpha_0}\left[\frac{q_2(1-e^{-z\psi_0}+q_1 e^{-z\psi_0})-1}{1-q_1 q_2 e^{-2z\psi_0}}e^{-z\psi}+1\right]\sin^3\alpha d\alpha\right\}. \tag{35}$$

Formulas (33)–(35) allow one to calculate the dimensionless absorption cross section of an inhomogeneous spherical particle

$$F(x, y, K, q_1, q_2) = F_1(x, y, K, q_1, q_2) + F_2(x, y, K, q_1, q_2) + F_3(x, y, K, q_1, q_2) \tag{36}$$

and the absorption cross section for electromagnetic radiation

$$\sigma = \sigma_0 F(x, y, K, q_1, q_2). \tag{37}$$

When $K \longrightarrow 0$ ($\alpha_0 \longrightarrow 0$) (i.e., a particle has no core), formula (36) yields

$$F(x, y) = \mathrm{Re}\left\{\frac{y^2}{z}\int_0^1 \xi^4 d\xi \int_0^{\pi}\left[\frac{(q_2-1)e^{-z\eta}}{1-q_2 e^{-z\eta_0}}+1\right]\sin^3\alpha d\alpha\right\}.$$

This expression coincides with the result obtained in [4] for a homogeneous spherical metal particle.

Numerical calculations of $F(x, y, K, q_1, q_2)$ are illustrated in Figs. 1–6.

5. DISCUSSION OF THE RESULTS

In the limit of pure specular reflection of electrons from the boundaries of the metal layer of a particle ($q_1 = 1$ and $q_2 = 1$), we obtain the following expression for the dimensionless absorption cross section $F(x, y, K)$ from formulas (33)–(35):

$$F(z, K) = \mathrm{Re}\left\{\frac{y^2}{z(x, y)}\frac{4}{15}(1-K^5)\right\}. \tag{38}$$

As a result, the absorption cross section (37) takes the form

$$\sigma(z, K) = \sigma_0 \mathrm{Re}\left\{\frac{y^2}{z(x, y)}\frac{4}{15}(1-K^5)\right\} = \sigma_0\frac{4}{15}(1-K^5)\frac{xy^2}{x^2+y^2}. \tag{39}$$

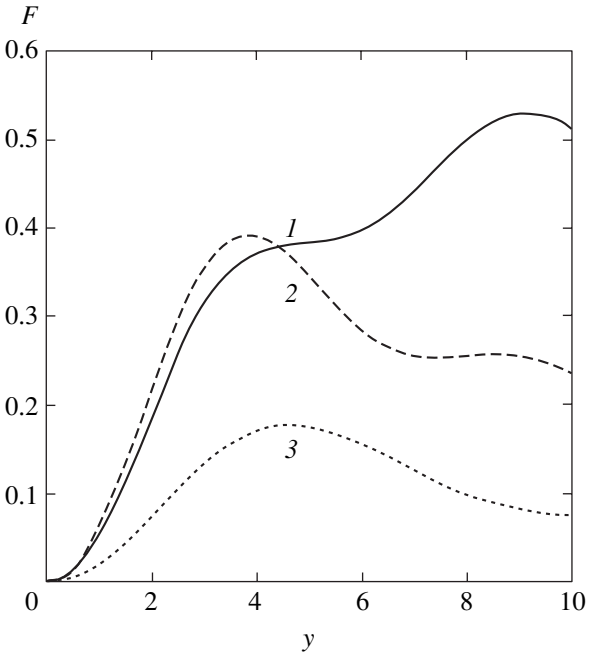


Fig. 2. The dimensionless absorption cross section F as a function of the dimensionless frequency y for fixed values of the dimensionless inverse mean free path $x = 0$ and the ratio of the core radius to the total radius of the particle $K = 0.8$ for different reflection coefficients q_1 and q_2 for each curve; (*1*) $q_1 = 0$ and $q_2 = 0$, (*2*) $q_1 = 1$ and $q_2 = 0$, and (*3*) $q_1 = 0$ and $q_2 = 1$.

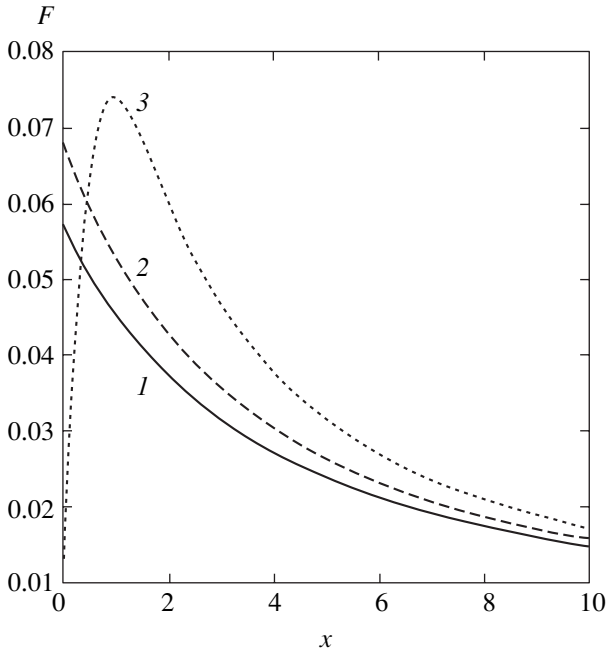


Fig. 3. The dimensionless absorption cross section F as a function of the dimensionless inverse mean free path x for fixed values of the ratio of the core radius to the total radius of the particle $K = 0.8$ and the dimensionless frequency $y = 1$ for different reflection coefficients q_1 and q_2 for each curve; (*1*) $q_1 = 0$ and $q_2 = 0$, (*2*) $q_1 = 1$ and $q_2 = 0$, and (*3*) $q_1 = 0$ and $q_2 = 1$.

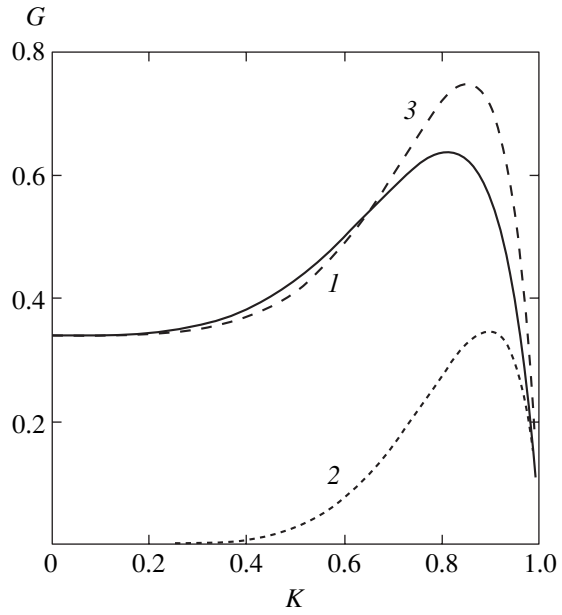


Fig. 4. The specific absorption cross section G as a function of the ratio of the core radius to the total radius of the particle K for fixed values of the dimensionless inverse mean free path $x = 0$ and the dimensionless frequency $y = 3$ for different reflection coefficients q_1 and q_2 for each curve; (*1*) $q_1 = 0$ and $q_2 = 0$, (*2*) $q_1 = 0$ and $q_2 = 1$, and (*3*) $q_1 = 1$ and $q_2 = 0$.

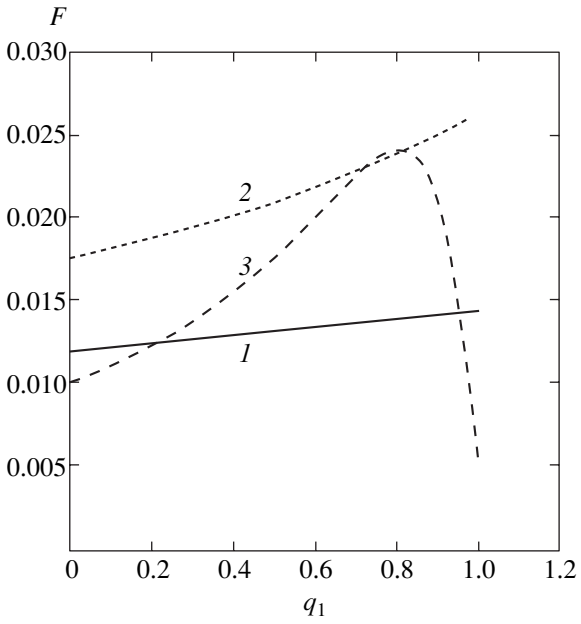


Fig. 5. The dimensionless absorption cross section F as a function of the reflection coefficient q_1 for fixed values of the dimensionless frequency $y = 1$, dimensionless inverse mean free path $x = 0$, and the ratio of the core radius to the total radius of the particle $K = 0.95$ for different values of the reflection coefficient q_2 for each curve; (*1*) $q_2 = 0$, (*2*) $q_2 = 0.5$, and (*3*) $q_2 = 1$.

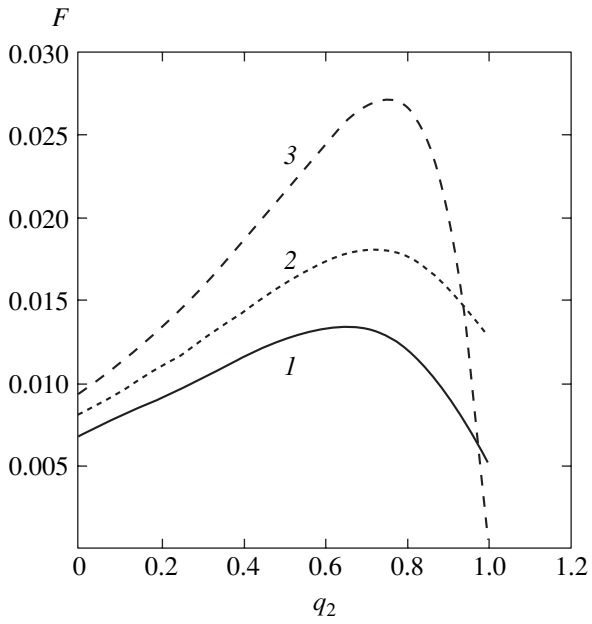


Fig. 6. The dimensionless absorption cross section F as a function of the reflection coefficient q_2 for fixed values of the dimensionless frequency $y = 1$, dimensionless inverse mean free path $x = 0$, and the ratio of the core radius to the total radius of the particle $K = 0.95$ for different values of the reflection coefficient q_1 for each curve; (*1*) $q_1 = 0$, (*2*) $q_1 = 0.5$, and (*3*) $q_1 = 1$.

In the case of a metal particle without core ($K \longrightarrow 0$), this expression corresponds to the classical result for a spherical particle (the Drude formula) [3]:

$$\sigma(z) = \sigma_0 \frac{4}{15} \frac{xy^2}{x^2 + y^2}.$$

With regard to (31) and (32), the absorption cross section (39) exactly coincides with the classical result for a spherical metal layer. This is associated with the fact that, for $q_1 = 1$ and $q_2 = 1$, the boundaries of the metal layer of the particle do not affect the distribution function of electrons $f(\mathbf{r}, \mathbf{v})$ The eddy current inside the specularly reflecting metal layer (see (25)) satisfies the local Ohm's law for any relation between the thickness of the layer l and the electron mean free path Λ. Thus, specular reflection does not exhibit any nonlocal (surface) phenomena.

Irrespective of the character of reflection of electrons from the boundaries of the metal layer (for any q_1 and q_2), the macroscopic asymptotics (38) also holds true as the particle size increases (for $x \gg 1$) (in this case, one can neglect the terms containing exponential functions in formulas (33)–(35) because they rapidly decay).

In the case of pure diffuse reflection of electrons from the boundaries of the metal layer of a particle, when $q_1 = q_2 = 0$, the dimensionless absorption cross section of the inhomogeneous spherical particle (36) is calculated by the formula

$$F(x, y, K) = y^2 \mathrm{Re}\left\{\int_K^1 \xi^4 d\xi \int_{\alpha_0}^{\pi} \frac{1 - e^{-z\eta}}{z} \sin^3\alpha d\alpha + \int_K^1 \xi^4 d\xi \int_0^{\alpha_0} \frac{1 - e^{-z\psi}}{z} \sin^3\alpha d\alpha \right\}.$$

If the probabilities of specular reflection of electrons are equal, i.e., $q_1 = q_2 = q$, then the dimensionless absorption cross section of the inhomogeneous spherical particle (36) is calculated as

$$F(x, y, K, q) = y^2 \times \mathrm{Re}\left\{\int_K^1 \xi^4 d\xi \int_{\alpha_0}^{\pi - \alpha_0} \left[\frac{(q-1)e^{-z\eta}}{z(1 - qe^{-z\eta_0})} + \frac{1}{z}\right] \sin^3\alpha d\alpha \right. \quad (40)$$
$$+ \int_K^1 \xi^4 d\xi \int_{\pi - \alpha_0}^{\pi} \left[\frac{q(1 - e^{-z\psi_0} + qe^{-z\psi_0}) - 1}{z(1 - q^2 e^{-2z\psi_0})} e^{-z\eta} + \frac{1}{z}\right] \sin^3\alpha d\alpha$$
$$\left. + \int_K^1 \xi^4 d\xi \int_0^{\alpha_0} \left[\frac{q(1 - e^{-z\psi_0} + qe^{-z\psi_0}) - 1}{z(1 - q^2 e^{-2z\psi_0})} e^{-z\psi} + \frac{1}{z}\right] \sin^3\alpha d\alpha \right\}.$$

The cases of $q_1 = 0$, $q_2 \neq 0$ and $q_1 \neq 0$, $q_2 = 0$ are also of interest; in these cases, the dimensionless absorption cross section of the inhomogeneous spherical particle (36) is described by formulas (41) and (42):

$$F(x, y, K, q_2) = y^2 \times \mathrm{Re}\left\{\int_K^1 \xi^4 d\xi \int_{\alpha_0}^{\pi - \alpha_0} \left[\frac{(q_2 - 1)e^{-z\eta}}{z(1 - q_2 e^{-z\eta_0})} + \frac{1}{z}\right] \sin^3\alpha d\alpha \right. \quad (41)$$
$$+ \int_K^1 \xi^4 d\xi \int_{\pi - \alpha_0}^{\pi} \frac{1 - e^{-z\eta}}{z} \sin^3\alpha d\alpha$$
$$\left. + \int_K^1 \xi^4 d\xi \int_0^{\alpha_0} \left[\frac{q_2(1 - e^{-z\psi_0}) - 1}{z} e^{-z\psi} + \frac{1}{z}\right] \sin^3\alpha d\alpha \right\},$$

$$F(x, y, K, q_1) = y^2 \mathrm{Re}\left\{\int_K^1 \xi^4 d\xi \int_{\alpha_0}^{\pi-\alpha_0} \frac{1-e^{-z\eta}}{z} \sin^3\alpha d\alpha\right.$$

$$+ \int_K^1 \xi^4 d\xi \int_{\pi-\alpha_0}^{\pi} \left[\frac{q_1(1-e^{-z\psi_0})-1}{z} e^{-z\eta} + \frac{1}{z}\right] \sin^3\alpha d\alpha \quad (42)$$

$$\left. + \int_K^1 \xi^4 d\xi \int_0^{\alpha_0} \frac{1-e^{-z\psi}}{z} \sin^3\alpha d\alpha \right\}.$$

Figures 1 and 2 show the dimensionless absorption cross section F as a function of the dimensionless frequency y of the external field. Figure 1 corresponds to the case of equal reflection coefficients of electrons from the surfaces of the particle (see (40)). For each curve, the ratio K of the core radius to the total radius of the particle is assumed to be fixed. One can see that, in a narrow interval of low dimensionless frequencies y (when $y \ll 1$), the dimensionless absorption cross section F may be greater for particles in which pure specular reflection of conduction electrons occurs. In another region of dimensionless frequencies ($y > 2$), the dimensionless absorption cross section is greater for particles in which the reflection of conduction electrons from each surface is purely diffuse. Figure 2 corresponds to the case when the particle size is small compared with the electron mean free path, when $R \ll \Lambda$ ($x = 0$). Each curve is drawn for various values of the reflection coefficients q_1 and q_2. The oscillations of the frequency response are attributed to the effect of the ratio of the electron flight time between collisions with the surfaces to the variation period of the external electromagnetic field on the energy dissipation inside the metallic shell of the particle. This effect is very significant for the diffuse reflection of electrons from the boundaries of the metal layer of the particle (when either $q_1 = 0$ or $q_2 = 0$). It decreases as the reflection coefficient from the surfaces increases. An increase in the particle radius results in the smoothing of the oscillations of the frequency response due to the increased effect of bulk collisions of electrons. As the reflection coefficients increase, the absorption cross section decreases due to the decreased role of surface phenomena in energy dissipation.

Figure 3 shows the dimensionless absorption cross section F as a function of the dimensionless inverse mean free path x. The function is plotted for a fixed dimensionless frequency y and various values of the reflection coefficients q_1 and q_2. Curve *3* emanates from the origin and has a maximum. This curve is close to the classical result (39) because the main contribution to the absorption cross section is made by electrons that are specularly reflected from the boundary of the inhomogeneous particle. For intermediate values of the reflection coefficients ($q \neq 0$ and $q \neq 1$), the dimensionless absorption cross section is different from zero even for a very pure metal, when $x = 0$. As the particle radius increases, all the curves merge into a single curve that reduces to the classical result. As the frequency increases, the absorption cross section also increases because the strength of the eddy electric field is directly proportional to the frequency of the external field.

To analyze the dependence of the dimensionless absorption cross section F on the ratio K of the core radius to the particle radius, we use Fig. 4, which represents the dimensionless absorption cross section of a metal particle with a dielectric core per unit volume of metal in the particle, $G(K)$,

$$G(K) = \frac{F(K)}{1-K^3}.$$

Let us restrict the analysis to the case of a pure metal particle ($x = 0$) and a fixed value of the dimensionless frequency y of the external field. The specific absorption cross section of such particles (electrons in pure metals are characterized by the large mean free path) can be greater when electrons are specularly reflected from the external surface of the metal layer of a particle in a wide range of K. When K is close to unity, the absorption cross section is small for any frequencies and any values of the reflection coefficients because the metallic shell of a particle is very thin and electrons cannot be significantly accelerated by the external electromagnetic wave as they move between the surfaces of the shell (in this case, the current density in the shell tends to zero).

Figures 5 and 6 illustrate the effect of the reflection coefficients q_1 and q_2 on the dimensionless absorption cross section for particles with a thin metallic shell (when K is close to unity). Figure 5 shows that, in the absence of bulk scattering of electrons in the metal (the thickness of the metal layer of a particle is extremely small), the dimensionless absorption cross section F as a function of the reflection coefficient q_1 exhibits especially complicated behavior in the case of specular reflection of electrons from the external boundary of the metal layer ($q_2 = 1$). In this case, the absorption cross section tends to zero for any y when the scattering of electrons by two reflecting surfaces of the metal layer becomes pure specular ($q_1 = 1$ and $q_2 = 1$). Figure 6 represents the dependence of the dimensionless absorption cross section F on the reflection coefficient q_2. Virtually for any values of q_2 (except for a narrow interval near unity), the absorption cross section increases as the degree of specularity of electron reflection from the internal boundary of the metal layer of the particle increases.

REFERENCES

1. Yu. I. Petrov, *Physics of Small Particles* (Nauka, Moscow, 1984), Chap. 7.
2. M. Born and E. Wolf, *Principles of Optics*, 6th ed. (Nauka, Moscow, 1973; Pergamon, Oxford, 1980), Chap. 13.
3. A. G. Lesskis, V. E. Pasternak, and A. A. Yushkanov, Zh. Éksp. Teor. Fiz. **83**, 310 (1982) [Sov. Phys. JETP **56**, 170 (1982)].
4. A. G. Lesskis, A. A. Yushkanov, and Yu. I. Yalamov, Poverkhnost, No. 11, 115 (1987).
5. J. M. Ziman, *Electrons and Phonons* (Clarendon, Oxford, 1960; Mir, Moscow, 1962), Chap. 11.
6. R. Kubo, J. Phys. Soc. Jpn. **17**, 975 (1962).
7. É. A. Manykin, P. P. Poluéktov, and Yu. G. Rubezhnyĭ, Zh. Éksp. Teor. Fiz. **70**, 2117 (1976) [Sov. Phys. JETP **43**, 1105 (1976)].
8. R. D. Averitt, S. L. Westcott, and N. J. J. Halas, J. Opt. Soc. Am. B **16**, 1824 (1999).
9. A. Henglein, J. Phys. Chem. B **104**, 2201 (2000).
10. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 3rd ed. (Pergamon, Oxford, 1984; Nauka, Moscow, 1992).
11. W. A. Harrison, *Solid State Theory* (McGraw-Hill, New York, 1970; Mir, Moscow, 1972).
12. R. Courant, *Partielle Differentialgleichungen* (Göttingen, 1932; Mir, Moscow, 1962).

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 189–196.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 215–223.

SOLIDS
Electronic Properties

The Potential Barrier Height for a Jahn–Teller Center

V. N. Vasyukov[a], A. D. Prokhorov[a], V. P. D'yakonov[a], and H. Szymczak[b]

[a]*Galkin Physicotechnical Institute, National Academy of Sciences of Ukraine, Donetsk, 83114 Ukraine*
e-mail: vasyukov@levch.fti.ac.donetsk.ua
[b]*Institute of Physics, Polish Academy of Sciences, 02-668 Warsaw, Poland*

Received December 2, 2003

Abstract—The EPR spectrum of the Cu^{2+} ion in a $ZnSiF_6 \cdot 6H_2O$ crystal is studied in the temperature range T = 5–300 K. It is shown that the EPR spectrum can be represented in the form of a superposition of three contributions with essentially different properties. The first contribution is characterized by the maximal intensity at low temperatures and is described by a spin Hamiltonian with a large anisotropy of parameters. The second contribution has the maximal intensity at high temperatures and is described by a spin Hamiltonian with a low anisotropy of parameters. The third contribution cannot be described by a spin Hamiltonian and has the form of a partly orientationally averaged EPR spectrum. The reason for the emergence of these contributions is substantiated along with the form of the temperature dependence of their intensities on the basis of variation of the populations of vibronic states upon a change in temperature. The height ($E_0 = 4 \pm 1$ cm^{-1}) of the potential barrier separating three equivalent Jahn–Teller potential wells of the Cu^{2+} ion is determined from analysis of the temperature dependence of the integrated intensity of the EPR spectrum. The obtained value of the barrier height substantially differs from the estimate (100 cm^{-1}) obtained earlier [2, 3] for the Cu^{2+} ion in $ZnSiF_6 \cdot 6H_2O$ on the basis of the tunneling model. It is shown that the forms of the temperature dependences of the linewidth of the low- and high-temperature EPR spectra are essentially different. This difference indicates that the contributions of the low- and high-temperature EPR spectra are associated with quantum-mechanical transitions between these states. It is proposed that the low- and high temperature contributions to the EPR spectrum are associated with the filling of under-the-barrier and above-the-barrier vibronic states, respectively. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The Jahn–Teller (JT) effect of a magnetic ion possessing the fundamental orbital doublet has been studied in a large number of experimental and theoretical works. References to the classical works devoted to this question can be found in [1].

The potential surface of a JT magnetic center has three potential wells separated by a barrier. The shape of potential wells and the potential barrier height are the main parameters determining the features of the effect. In spite of considerable number of publications devoted to the JT effect, these parameters of magnetic centers have not been actually investigated.

In the publications [2–4] devoted to analysis of the Cu^{2+} ion magnetic center in $ZnSiF_6 \cdot 6H_2O$ and $ZnZrF_6 \cdot 6H_2O$ crystals, an attempt was made at filling this gap. To estimate the height of the potential barrier, the authors of these publications used the result obtained in the framework of the tunneling model [5], according to which the potential barrier height is associated with the difference $g_x - g_y$ in the g factors. The estimates of the barrier height obtained in [2–4] are dubious as regards the application of the tunneling model. It is well known that this model is applicable only for infinitely high barriers (the barrier height must be much higher than the kinetic energy of oscillations). In addition, the experimental fact that the difference $g_x - g_y$ of the g factors for most of the centers studied is small and even zero for a number of centers. Due to this peculiarity, the barrier height can be determined using this method only to a low degree of accuracy. This method cannot be applied for determining the barrier height for the centers for which $g_x = g_y$.

The method worked out in [6] appears to be more reliable. In this method, the barrier height is determined from analysis of the temperature dependence of the peak intensity of resonance lines for the low-temperature (LT) and high-temperature (HT) spectra in the crystal $Ca(OH)_2$. If, following [7, 8], we assume that the HT spectrum appears as a result of populating excited vibronic states, the integrated intensity of the HT EPR spectrum is proportional to $\exp(-E_0/kT)$. Energy E_0 is equal to the potential barrier height. To determine the barrier height, Wilson *et al.* [6] studied the peak intensity of resonance lines. The temperature dependences of the peak and integrated intensities can noticeably differ as a result of temperature variation of the resonance linewidth.

This study aims at determining the potential barrier height for a Jahn–Teller ion from an analysis of the temperature dependence of the integrated intensity of

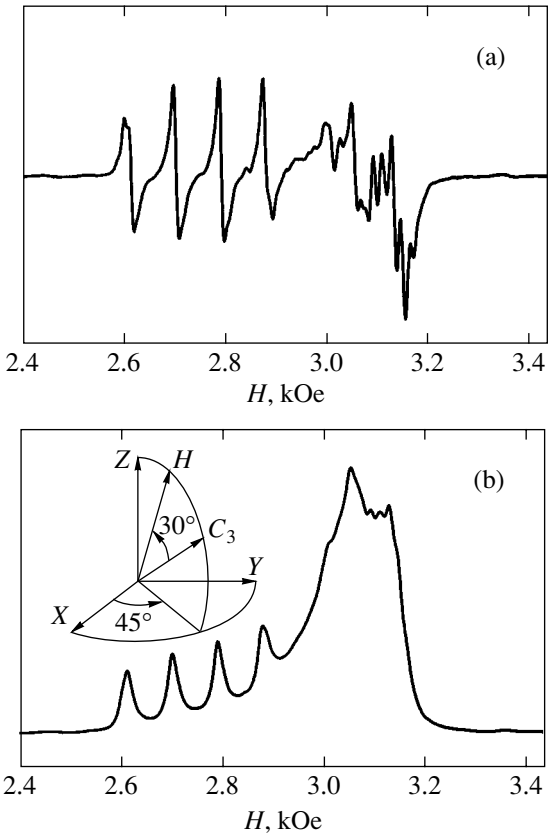


Fig. 1. The EPR spectrum of the Cu^{2+} ion in a $ZnSiF_6 \cdot 6H_2O$ crystal at $T = 5.4$ K: (a) derivative of the absorption line; (b) absorption line. The angle between the C_3 axis and the magnetic field direction is 30°.

the LT EPR spectrum. We have chosen a $ZnSiF_6 \cdot 6H_2O$ crystal for our experiments, for which the magnetic center of the Cu^{2+} ion has been investigated thoroughly in a number of works [2, 9, 10].

To obtain quantitative characteristics for the temperature dependence of the LT spectrum, we decomposed the experimental absorption line of the EPR spectrum into components. For this purpose, we simulated the shape of the absorption line of the LT spectrum on the basis of application of the method of Green functions. The technique for calculating the model shape of the absorption line is described in greater detail in Section 3.

2. EXPERIMENTAL RESULTS

Crystals of $ZnSiF_6 \cdot 6H_2O$ with 0.1% of Cu^{2+} ions were grown from aqueous solution using the evaporation technique. The crystal belongs to the trigonal system, space group C_{3i} [11]. The Cu^{2+} ion isovalently substitutes the Zn^{2+} ion. The nearest neighbors of the Cu^{2+} ion are six water molecules forming a nearly regular octahedron with a small distortion along the C_3 axis.

The EPR spectra are analyzed on an EPR spectrometer with the microwave field frequency $\nu = 9.24$ GHz.

The EPR spectrum is a superposition of contributions of different origins. One of the contributions is predominant at a low (helium) temperature and hence can be referred to as the LT contribution. The second contribution dominates at a high (room) temperature and can be referred to as the HT contribution. Each of these spectra can be described by a spin Hamiltonian. In addition to these two contributions, the EPR spectrum of the Cu^{2+} ion contains an additional contribution which is observed at a low temperature, but cannot be described by a spin Hamiltonian.

2.1. Low-Temperature EPR Spectrum

The main singularities of the spectrum at low temperature are determined by the contribution that is usually referred to as the LT EPR spectrum. In turn, the LT spectrum is a superposition of three EPR spectra. The principal symmetry axes of these spectra are mutually orthogonal. These axes are arranged symmetrically relative to the trigonal axis of the crystal and are turned through the same angle equal to 54.7°.

If we choose the coordinate system so that it coincides with the three mutually perpendicular axes of the EPR spectra, we will for brevity refer to the magnetic center whose symmetry axis is directed along the *X* axis as the *X* center. Analogously, the *Y* and *Z* centers have anisotropy axes directed along the *Y* and *Z* axes, respectively.

Figure 1a shows the EPR spectrum in the case when the direction of the magnetic field is determined by angles $\theta = 24.7°$ and $\varphi = 45°$ of the spherical system of coordinates.

To decompose the spectrum into components, it is convenient to represent the EPR spectrum not in the form of a derivative of the absorption line (as is usually done when the phase detection method is employed), but in the form of the absorption line as shown in Fig. 1b.

The first (in the field) four resonance lines of width $\Delta H = 28$ Oe are the lines of hyperfine splitting of the *Z* center. The partly split broad line in strong fields ($H \approx 3.1$ kOe) is a superposition of spectral lines from *X* and *Y* centers.

The LT EPR from the magnetic center with the principal symmetry axis along the *Z* axis can be described by spin Hamiltonian (1) with an electron spin $S = 1/2$ and a nuclear spin $I = 3/2$:

$$H_{\mathrm{LT}} = g_{\parallel}\mu_{\mathrm{B}}H_zS_z + g_{\perp}\mu_{\mathrm{B}}(H_xS_x + H_yS_y) + AS_zI_z + B(S_xI_x + S_yI_y) + P\{I_z^2 + I(I+1)/3\}, \quad (1)$$

where H_x, H_y, and H_z are the magnetic field components; S_x, S_y, S_z, I_x, I_y, and I_z are the components of the electron and nuclear spin operators; $g_{\parallel}$ and $g_{\perp}$ are the

elements of the g tensor; A and B are the hyperfine splitting parameters; and P is the parameter of nuclear quadrupole splitting.

A peculiar feature of the EPR spectrum of the Cu^{2+} ion is that parameters $g_{\|} = 2.460 \pm 0.004$, $g_{\perp} = 2.105 \pm 0.004$, and $|A| = (100 \pm 9) \times 10^{-4}$ cm^{-1} can be determined relatively easily and to a high degree of accuracy on the basis of the LT EPR spectrum for $H \| Z$. The form of the EPR spectrum for this magnetic field orientation is shown in Fig. 2.

As a rule, parameters B and P are small and are manifested in the EPR spectrum only when the magnetic field deviates from the principal anisotropy axis. We could not determine these parameters in view of the relatively large width of the resonance line. While simulating the profile of the absorption line of the LT spectrum, we used the values of parameters $|B| = 14 \times 10^{-4}$ cm^{-1} and $P = +8 \times 10^{-4}$ cm^{-1}, which were obtained earlier in [10] on a deuterium-enriched sample.

2.2. Temperature Variation of the Spectrum

The temperature variation of the EPR spectrum is due to the operation of the following two main mechanisms. The first mechanism is familiar and is determined by the change in the difference between the occupancies of resonant states and leads to a decrease in the integrated intensity of resonance spectral lines in accordance with the relation

$$I(T) = I_0 \tanh(h\nu/2kT), \qquad (2)$$

where $h\nu$ is the microwave quantum, k is the Boltzmann constant, and T is the temperature.

The second mechanism is typical of magnetic centers possessing a potential with many minima. As a result of operation of this mechanism, the spectral intensity is redistributed between the contribution from the LT and HT EPR spectra. At a temperature close to the liquid helium temperature, the contribution of the LT spectrum prevails. An increase in temperature leads to a decrease in the intensity of the LT spectrum and to an increase in the intensity of the HT spectrum. At room temperature, only the HT spectrum is observed.

Figures 3 and 4 show the redistribution of the intensities. To separate the operation of the second mechanism in pure form, Figs. 3 and 4 depict the temperature dependences of the reduced integrated intensity, which was obtained by dividing the experimental value of the intensity by $\tanh(h\nu/2kT)$.

It can be seen from Fig. 3 that an increase in temperature reduces the peak intensity of the lines from the Z center of the LT spectrum. This change is determined by two factors: the broadening of the resonance line and the redistribution of the intensities between the LT ad HT spectra.

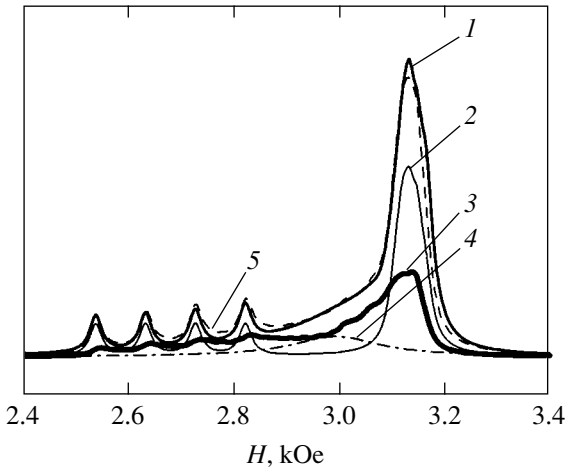


Fig. 2. The EPR spectrum of the Cu^{2+} ion in a $ZnSiF_6 \cdot 6H_2O$ crystal at $T = 9$ K in a magnetic field directed along the symmetry axis of the Z-center: experimental absorption line (*1*), the profile of the absorption line of the LT EPR spectrum (*2*), the shape of the absorption line of the LT EPR spectrum averaged as a result of rotation about the C_3 axis (*3*), the shape of the absorption line of the HT EPR spectrum (*4*), and the sum of contributions *2*, *3*, and *4* (*5*).

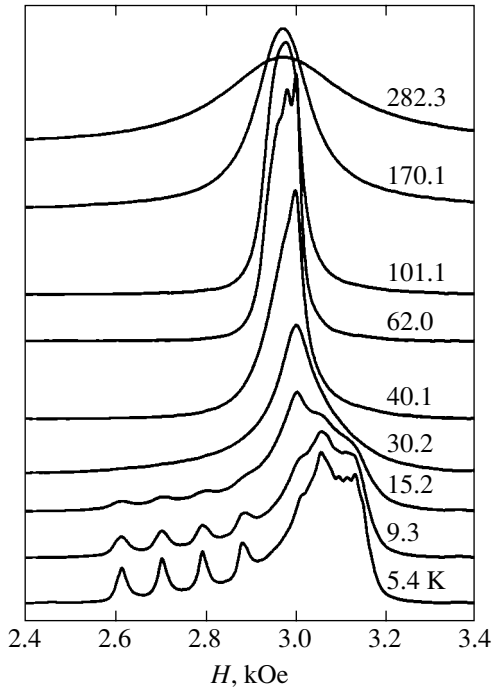


Fig. 3. Variation of the shape of the absorption line of the EPR spectrum for the Cu^{2+} ion in a $ZnSiF_6 \cdot 6H_2O$ crystal with temperature. The magnetic field orientation is the same as in Fig. 1.

The temperature dependence of the resonant linewidth is shown in Fig. 5. At temperatures $T < 20$ K, we can experimentally determine the linewidth of the LT spectrum of the Z center. Points in Fig. 5 were obtained

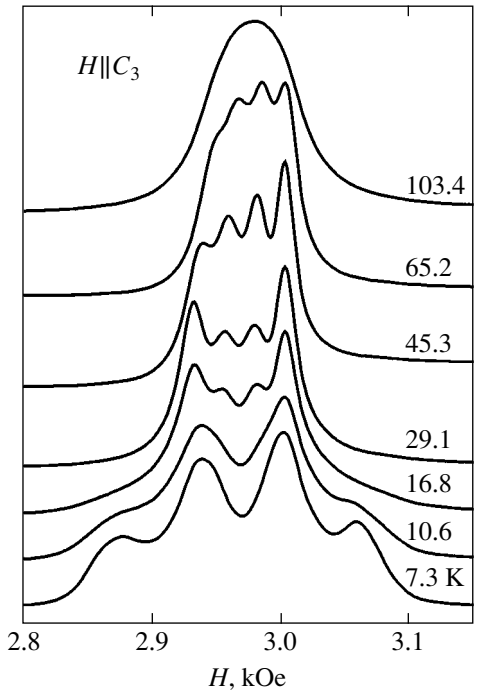


Fig. 4. Variation of the shape of the absorption line of the EPR spectrum for the Cu^{2+} ion in a $ZnSiF_6 \cdot 6H_2O$ crystal with temperature. The magnetic field is directed along the ternary axis of the crystal.

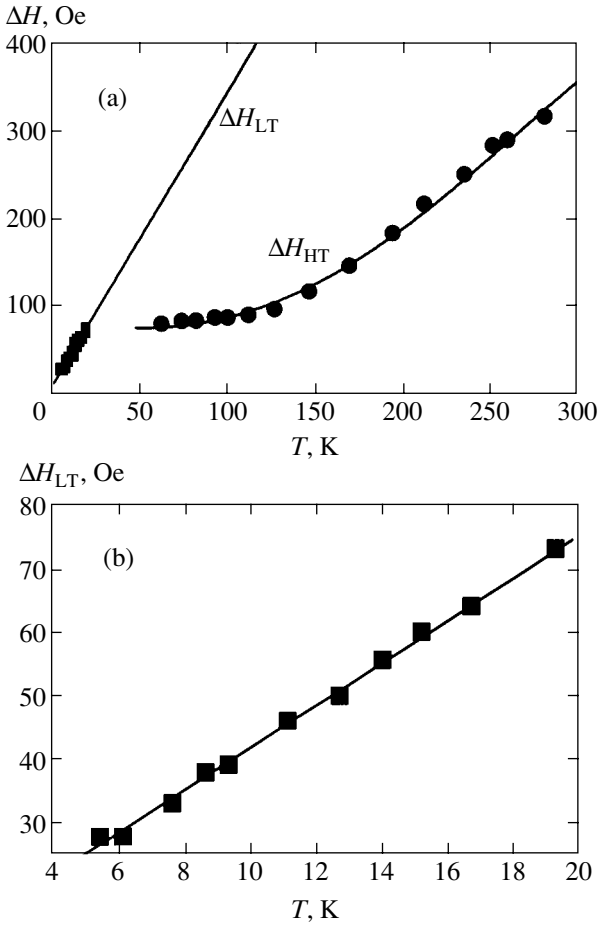


Fig. 5. (a) Temperature dependences of the resonance linewidths of the HT and LT EPR spectra for the Cu^{2+} ion in a $ZnSiF_6 \cdot 6H_2O$ crystal. (b) Temperature dependence of the resonance linewidth of the LT EPR spectrum, $\Delta H = 9.55 + 3.289T$.

from a comparison of the model absorption line of the LT spectrum with the experimental absorption line, which are successfully described by a linear dependence (see Fig. 5b). At temperatures $T > 20$ K, the linewidth of the *Z* center cannot be determined since the line intensity becomes commensurate with or smaller than the noise. At temperatures $T > 100$ K, the linewidth of the HT spectrum can be experimentally determined. Pay attention to the fact that the temperature dependences of ΔH_{LT} and ΔH_{HT} differ significantly.

A detailed analysis of the spectrum with digital processing of the data presented in Fig. 3 shows that the LT spectrum of the *Z* center is observed up to a temperature of $T = 40$ K. The contribution of the LT spectrum of the *X* and *Y* centers in Fig. 3 is observed at a higher temperature and is manifested in the asymmetry of the resonance lines of the *X* and *Y* centers.

It can be seen from Fig. 3 that an increase in temperature increases the integrated intensity of the line of the HT spectrum.

When the magnetic field is directed along the ternary axis ($H \| C_3$), hyperfine splitting of both the LT and the HT spectra is observed. Hyperfine splitting of the HT spectrum in Fig. 4 is observed for $16.8\ \text{K} \le T \le 65$ K.

2.3. High-Temperature EPR spectrum

It can be seen from Fig. 4 that the HT spectrum for $16\ \text{K} \le T \le 65$ K has a resolved hyperfine structure. The principal symmetry axis of the HT spectrum is the ternary axis of the crystal. The HT EPR spectrum can be described by the spin Hamiltonian

$$H_{HT} = g_{\|}\mu_B H_z S_z + g_{\perp}\mu_B (H_x S_x + H_y S_y) + A S_z I_z + B(S_x I_x + S_y I_y). \quad (3)$$

If we use the experimental shape of the absorption line for $H \| C_3$ corresponding to temperature $T = 29.1$ K, the parameters of spin Hamiltonian (3) are $g_{\|} = 2.225 \pm 0.004$, $g_{\perp} = 2.197 \pm 0.004$, and $A = (24 \pm 4) \times 10^{-4}$ cm^{-1}. These values coincide with the results obtained in [10] to within the error of measurements.

3. SIMULATION OF THE SHAPE OF ABSORPTION LINE

The EPR spectrum of the Cu^{2+} ion is a superposition of different contributions whose resonance lines can overlap. For this reason, it is important to compare the

experimental absorption line with the model absorption line obtained on the basis of spin Hamiltonians (1) and (3).

Here, we use the method of Green functions for calculating the model absorption line. The absorbed power W in the EPR is proportional to frequency ω of the microwave field ($\omega = 2\pi\nu$), the square of the varying magnetic field H_1, and the imaginary part of susceptibility χ'':

$$W = \omega H_1^2 \chi''/2. \quad (4)$$

The susceptibility can be represented using the retarded commutator Green function [12]: $\chi(\omega) = -\langle\langle M_\lambda | M_\lambda \rangle\rangle$.

In this study, experiments are made for mutually perpendicular constant and varying magnetic fields. The main results are given for the orientation of constant magnetic field H in a plane parallel to the Z axis and symmetric relative to the X and Y axes. For the X, Y, and Z magnetic centers, the projections of the operators of magnetic moment M onto the λ axis with a direction coinciding with the direction of varying magnetic field H_1 can be presented in the form

$$M_\lambda = \mu_B(g_\parallel S_x - g_\perp S_y)/\sqrt{2} \text{ for the } X \text{ center},$$
$$M_\lambda = \mu_B(g_\perp S_x - g_\parallel S_y)/\sqrt{2} \text{ for the } Y \text{ center}, \quad (5)$$
$$M_\lambda = \mu_B g_\perp (S_x - S_y)/\sqrt{2} \text{ for the } Z \text{ center}.$$

To calculate the Green functions, it is convenient to pass from the spin operators S_x and S_y to the projection operators $P_{mn} \equiv |m\rangle\langle n|$. The wave function of state $|m\rangle$ is the product of spin functions $|m_S\rangle$ and the function of nuclear spin $|m_I\rangle$. Parameters $|m_S\rangle$ and $|m_I\rangle$ assume the values $m_S = \pm 1/2$ and $m_I = \pm 3/2, \pm 1/2$. States $|m\rangle$ can be assigned number m from 1 to 8. For example, $m_S = 1/2$, and $m_I = 3/2$ correspond to $m = 1$, while $m_S = -1/2$ and $m_I = -3/2$ correspond to $m = 8$.

Operator M_λ can be represented in the form of the linear combination $M_\lambda = \sum_{m,n} G_{mn} P_{mn}$ of projective operators. Parameters G_{mn} assume different values for centers X, Y, and Z. For example, $M_\lambda = \mu_B\{(g_\parallel + ig_\perp)(P_{15} + P_{26} + P_{37} + P_{48}) + (g_\parallel - ig_\perp)(P_{51} + P_{62} + P_{73} + P_{84})\}/(2\sqrt{2})$ for the X center.

Further, we numerically diagonalize spin Hamiltonian (1) and pass to new wave functions $|\psi_\alpha\rangle$, which can be written in the form of a linear combination of the initial states $|m\rangle$,

$$|\psi_\alpha\rangle = \sum_m a_{\alpha m} |m\rangle. \quad (6)$$

Further, we must pass to new projective operators $P_{\alpha\beta} = |\psi_\alpha\rangle\langle\psi_\beta|$, which are connected with operators P_{mn} via the relation

$$P_{\alpha\beta} = \sum_{n,m} a_{\alpha n} a^*_{\beta m} P_{mn}. \quad (7)$$

Spin Hamiltonian (1) and operators M_λ can be written in the form of linear combinations of operators $P_{\alpha\beta}$:

$$H_{LT} = \sum_\alpha \varepsilon_\alpha P_{\alpha\beta}, \quad M_\lambda = \sum_{\alpha,\beta} B_{\alpha\beta} P_{\alpha\beta}. \quad (8)$$

Indices α and β label states and assume values from 1 to 8.

The green function $\langle\langle M_\lambda | M_\lambda \rangle\rangle$ expressed in terms of operators $P_{\alpha\beta}$ is given by

$$\langle\langle M_\lambda | M_\lambda \rangle\rangle = \sum_{\alpha,\beta,\gamma,\delta} B_{\alpha\beta} B_{\gamma\delta} \langle\langle P_{\alpha\beta} | P_{\gamma\delta} \rangle\rangle. \quad (9)$$

In the case of a diagonal spin Hamiltonian (8), the Green function can be calculated by directly solving the equation for the Green function [12]. As a result, the anticommutator Green function assumes the form

$$\langle\langle P_{\alpha\beta} | P_{\gamma\delta} \rangle\rangle_E = \delta_{\gamma\beta}\delta_{\alpha\delta}\langle P_{\alpha\alpha} + P_{\beta\beta}\rangle/(E + \varepsilon_\alpha - \varepsilon_\beta), \quad (10)$$

where ε_α is the energy of state $|\psi_\alpha\rangle$, $\langle P_{\alpha\alpha}\rangle = \exp(-\varepsilon_\alpha/kT)/Z$ is the mean value of operator $P_{\alpha\alpha}$, and Z is the partition function,

$$Z = \sum_\alpha \exp(-\varepsilon_\alpha/kT).$$

Using the properties of the Green functions [12], we can obtain the imaginary part of the retarded commutator Green function,

$$\mathrm{Im}\langle\langle M_\lambda | M_\lambda \rangle\rangle_\omega = -\frac{\pi}{2\hbar}\tanh\left(\frac{\hbar\omega}{2kT}\right) \times \sum_{\alpha,\beta,\gamma,\delta} B_{\alpha\beta} B_{\gamma\delta} \langle P_{\alpha\alpha} + P_{\beta\beta}\rangle \delta\left(\omega + \frac{\varepsilon_\alpha - \varepsilon_\beta}{\hbar}\right). \quad (11)$$

In this expression, the δ function $\delta(\omega + (\varepsilon_\alpha - \varepsilon_\beta)/\hbar)$ represents the profile of the resonant line corresponding to the transition between the states ε_α and ε_β. To calculate the actual EPR spectrum, it is insufficient to take into account spin Hamiltonian (1) alone. We must also take into account the spin–phonon interaction. The inclusion of the spin–phonon interaction leads to resonant line broadening. In this case, the δ function in relation (11) is replaced by the Lorentz function.

Here, we calculate the model shape of the absorption line using expression (11) in which the δ function is replaced by the Lorentz function with the same reso-

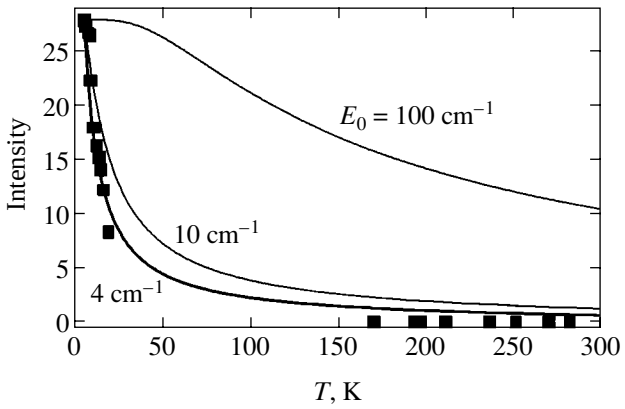


Fig. 6. Temperature dependence of the reduced integrated intensity ($I = I(1 - \exp(-E_0/kT))$) of the LT EPR spectrum for the Cu^{2+} ion in a $ZnSiF_6 \cdot 6H_2O$ crystal.

nance conditions. The linewidth is determined from a comparison with the experimental absorption line.

The advantage of this method for calculating the absorption line primarily lies in how easy it is to separate the mechanisms of the EPR spectrum's dependence on temperature. To single out the contribution associated with the redistribution of the intensity between the LT and HT spectra, it is sufficient to divide expression (11) by $\tanh(\hbar\omega/2kT)$. The second advantage is that relation (11) takes into account both the "allowed" quantum-mechanical transitions and the "forbidden" ones, which can make a noticeable contribution to the shape of the absorption line, especially for the magnetic field orientation deviating from the principal anisotropy axis of a magnetic center.

4. DISCUSSION

It should be mentioned above all that the EPR spectrum of the Cu^{2+} ion in a $ZnSiF_6 \cdot 6H_2O$ crystal is a superposition of several contributions with different properties. Our results make it possible to separate three contributions. The first contribution exhibits a predominant intensity at low temperatures close to liquid helium temperature. This contribution is referred to as the LT spectrum and is described by spin Hamiltonian (1). The second contribution dominates at a high (room) temperature. This contribution is referred to as the HT EPR spectrum and is described by spin Hamiltonian (3). In addition to these contributions, a third contribution exists, which cannot be described by a spin Hamiltonian in the conventional sense. The existence of this contribution is demonstrated in Fig. 2.

Curve *1* in Fig. 2 is the experimental absorption line. Curve *2* is the LT EPR spectrum, which was calculated on the basis of spin Hamiltonian (1) by using the technique described in Section 3. A comparison of these two curves leads to the conclusion that the experimental EPR spectrum at $T = 9$ K contains an additional contribution equal to the difference of intensities of the experimental absorption curve and the low-temperature spectrum. It should be emphasized that the additional contribution in Fig. 2 can be described neither by Hamiltonian (1) nor by Hamiltonian (3); for this reason, this contribution should be treated as an individual additional contribution to the EPR spectrum.

The shape of the absorption curve describing this contribution resembles the LT EPR spectrum averaged as a result of rotation about the ternary axis of the crystal. The shape of absorption line *3* shown in Fig. 2 is obtained by averaging the LT EPR spectrum as a result of rotation about the ternary axis. Curve *4* is the HT EPR spectrum calculated on the basis of spin Hamiltonian (3). Curve *5* is the sum of the contributions described by absorption lines *2*, *3*, and *4*. The difference between the experimental absorption line *1* and the model curve describing the three contributions does not exceed 1.5%.

A possible reason for the emergence of the additional contribution is the population of near-barrier vibronic states with an energy commensurate with the barrier height. For vibronic state energies close to the barrier height, the properties of vibronic states change from the properties typical of under-the-barrier states to that characteristic of over-the-barrier states [7, 8]. In this case, it is clear why the shape of the absorption line for the additional contribution resembles the spectrum of partly disoriented magnetic centers.

With increasing temperature, the reduced integrated intensity of the additional contribution decreases. The reason for such a decrease is analogous to the reason for the decrease of the LT spectrum and is described in detail in [7, 8]. It is associated with a change in the population of vibronic states. At low temperature ($kT < E_0$), over-the-barrier states are virtually empty and, hence, the HT spectrum is not observed in experiments. The predominant contribution comes from the LT spectrum, corresponding to the ground vibronic state. The additional contribution corresponds to near-barrier states and can be observed at low temperatures only for a barrier height E_0 smaller than kT. With increasing temperature, magnetic centers pass to excited over-the-barrier states. The intensity of the HT spectrum increases. The number of magnetic centers in the ground state and in near-barrier states decreases. As a result, the intensities of the LT spectrum and of the additional contributions decrease.

The barrier height was determined from the temperature dependence of the reduced integrated intensity of the LT spectrum. This dependence is shown in Fig. 6.

Experimental points on this curve were obtained by comparing the experimental absorption line and the model absorption line of the LT spectrum. The shape of the absorption line of the LT spectrum was calculated on the basis of spin Hamiltonian (1) using the method described in Section 3. The intensity of the LT spectrum and the linewidth were determined from a comparison

with experiment. As the criterion for the comparison, we chose the first line of hyperfine splitting of the *Z* center. This line is separated from the HT spectral line by the largest distance and, hence, is distorted by the contribution from the HT spectrum to a lesser extent. The intensity of the LT spectrum and the linewidth were chosen so that the first line of hyperfine splitting for the *Z* center of the experimental line and the model line coincides in shape and intensity. The temperature dependence of the LT spectral linewidth is shown in Fig. 5b. This dependence is linear.

The obtained temperature dependence of the LT spectral intensity is essentially nonlinear and can be approximated by the relation

$$I(T) = I_0\left(1 - \exp\left(-\frac{E_0}{kT}\right)\right). \quad (12)$$

If we assume that all under-the-barrier vibronic states make a contribution to the LT spectrum, while all over-the-barrier vibronic states contribute to the HT spectrum, parameter E_0 has the meaning of the barrier height. In the case of equidistant splitting of vibronic states, relation (12) for the LT spectral intensity can be derived theoretically.

For comparison, Fig. 6 shows three curves corresponding to three different values of barrier height E_0. A comparison of these curves with experimental results indicates that the barrier height for the Cu^{2+} ion in a $ZnSiF_6 \cdot 6H_2O$ crystal is $E_0 = 4 \pm 1$ cm^{-1}.

The obtained value of the barrier height is smaller by a factor of 25 than the barrier height obtained in [2, 3] for the Cu^{2+} ion in a $ZnSiF_6 \cdot 6H_2O$ crystal on the basis of the tunneling model.

The temperature dependences of the linewidth of the LT and HT spectra presented in Fig. 5 are worth noting. It should be observed above all that these dependences differ significantly. A qualitative estimate of these dependences leads to the conclusion that the temperature dependence of the linewidth of the HT spectrum is more traditional, while the temperature dependence of the linewidth of the LT spectrum exhibits some peculiarities.

The dependence for the HT spectrum contains a low-temperature segment on which the linewidth weakly depends on temperature. As the temperature increases, the linewidth smoothly increases. Such a form of the dependence is typical of the relaxation broadening mechanism associated with Raman scattering of phonons. It should be noted that the experimental value of the linewidth of the HT spectrum contains a contribution associated with hyperfine splitting of the HT spectrum. The value of this contribution is approximately equal to 50 Oe.

In the temperature interval 5 K < *T* < 20 K, the temperature dependence of the linewidth in the LT spectrum is correctly described by a linear dependence to within the error of measurements. Peculiar features of the temperature dependence include absence of a tendency towards a decrease in the steepness of the curve with decreasing temperature. In this case, coefficient $d\Delta H/dT$ of the temperature dependence is quite large for the given temperature interval.

These features suggest the existence of an additional mechanism of resonance line broadening. A possible reason for this additional broadening is the increase in the population of near-barrier vibronic states. The higher the temperature, the larger the occupancy of the near-barrier states and the broader the resonance line of the LT spectrum. For a low barrier height ($E_0 = 4$ cm^{-1}), this mechanism can make the largest contribution.

The difference in the temperature dependences of the LT and HT spectral lines points to different mechanisms of broadening of these lines. This peculiarity is important for verifying the validity of the existing models of the Jahn–Teller effect. For example, the transition from the LT to the HT EPR spectrum as a result of averaging of the EPR spectrum or even the parameters of the spin Hamiltonian is often explained in the literature as a consequence of jumps of a magnetic center from one potential well to another. In this model, quantum-mechanical transitions occur between the same energy levels both for the LT and for the HT spectrum. For this reason, the LT spectral linewidth must be smoothly transformed into the HT spectral linewidth. It is shown in Fig. 5 that a smooth transition is not observed in experiments. The different forms of the temperature dependences depicted in Fig. 5 indicate that quantum-mechanical transitions in the LT and HT spectra occur between different energy states. In the case of the LT spectrum, these are vibronic under-the-barrier states, while in the case of the HT spectrum, these are over-the-barrier vibronic states.

5. CONCLUSIONS

Analysis of the EPR spectrum for the Cu^{2+} ion in a $ZnSiF_6 \cdot 6H_2O$ crystal in a wide temperature range has made it possible to separate three contributions to the total absorption: the LT EPR spectrum consisting of three mutually perpendicular spectra with identical parameters of spin Hamiltonian, which corresponds to the contribution of under-the-barrier vibronic states; the HT EPR spectrum, which is a nearly isotropic spectrum corresponding to the contribution from over-the-barrier vibronic states; and an additional contribution that can be attributed to near-barrier states.

It was found that the linewidth of the LT and HT spectra exhibit essentially different temperature dependences, indicating that quantum-mechanical transitions in the LT and HT spectra occur between different energy states.

The height of the potential barrier (4 cm^{-1}) separating three equivalent Jahn–Teller potential wells of the Cu^{2+} ion is determined. The obtained value of the bar-

rier height for the Cu^{2+} ion in a $ZnSiF_6 \cdot 6H_2O$ crystal differs considerably from the estimate (100 cm^{-1}) reported earlier [2, 3] for zinc fluorosilicate crystals.

REFERENCES

1. A. Abragam and B. Bleaney, *Electron Paramagnetic Resonance of Transition Ions* (Clarendon, Oxford, 1970; Mir, Moscow, 1973).
2. V. N. Vasyukov, S. N. Lukin, and G. A. Tsintsadze, Fiz. Nizk. Temp. **10**, 742 (1984) [Sov. J. Low Temp. Phys. **10**, 390 (1984)].
3. S. N. Lukin, Fiz. Tverd. Tela (Leningrad) **33** (1), 47 (1991) [Sov. Phys. Solid State **33**, 25 (1991)].
4. S. N. Lukin, Fiz. Tverd. Tela (Leningrad) **31** (8), 244 (1989) [Sov. Phys. Solid State **31**, 1415 (1989)].
5. K. P. Lee and D. Walsh, Can. J. Phys. **49**, 1620 (1971).
6. R. G. Wilson, F. Holuj, and N. E. Hedgecock, Phys. Rev. B **1**, 3609 (1970).
7. V. N. Vasyukov, Phys. Status Solidi B **137**, 623 (1986).
8. I. N. Vasyukov and B. Ya. Sukharevskiĭ, Fiz. Nizk. Temp. **20**, 821 (1994) [Low Temp. Phys. **20**, 644 (1994)].
9. B. Bleaney, K. D. Bowers, and R. S. Trenam, Proc. R. Soc. London, Ser. A **228**, 157 (1955).
10. Le Si Dang, R. Buisson, and F. I. B. Williams, J. Phys. (Paris) **35**, 46 (1974).
11. R. W. G. Wyckoff, *Crystal Structures*, 2nd ed. (Interscience, New York, 1965), Vol. 3.
12. D. N. Zubarev, *Nonequilibrium Statistical Thermodynamics* (Nauka, Moscow, 1971; Consultants Bureau, New York, 1974).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 19–27.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 26–35.

ATOMS, SPECTRA, RADIATION

Femtosecond Spectroscopy of Coherent Anti-Stokes Raman Scattering with Frequency-Tunable Chirped Pulses Generated in a Microstructure Fiber

S. O. Konorov[a], D. A. Akimov[a], A. A. Ivanov[b], M. V. Alfimov[b], and A. M. Zheltikov[a,c]

[a]*Physics Department, Moscow State University, Vorob'evy gory, Moscow, 119992 Russia*
[b]*Center of Photochemistry, Russian Academy of Sciences, ul. Novatorov 7a, Moscow, 117421 Russia*
[c]*International Laser Center, Moscow State University, Vorob'evy gory, Moscow, 119992 Russia*
e-mail: zheltikov@top.phys.msu.su

Received November 11, 2003

Abstract—A new scheme of chirped-pulse femtosecond coherent anti-Stokes Raman scattering spectroscopy is proposed and experimentally implemented. A theory of this modification of coherent nonlinear spectroscopy is developed. We use this approach to show that a linear time–frequency mapping defined by linearly chirped pulses allows the spectra of nonlinear response of a medium to be measured by varying the delay time between the pump pulses. Microstructure fibers with a special dispersion profile are at the heart of the experimental implementation of this technique. Such fibers are ideally suited for the generation of frequency-tunable pulses with a smooth envelope and a controlled chirp. We present the results of experimental characterization of the envelope, spectrum, and chirp of anti-Stokes pulses generated in microstructure fibers by femtosecond Cr:forsterite-laser pulses. These frequency-tunable anti-Stokes pulses produced and shaped in microstructure fibers are then employed for coherent anti-Stokes Raman scattering spectroscopy of toluene solution. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Coherent anti-Stokes Raman scattering (CARS) [1, 2] is one of the most powerful and widespread techniques in nonlinear spectroscopy. CARS spectroscopy provides a high spatial, temporal, and spectral resolution [3] in the investigation of excited gas media, plasmas, flames, and combustion and offers new solutions for the coherent microscopy of biological objects [4] and ionized spatially nonuniform gas media [5]. Femtosecond CARS spectroscopy [6] gives access to ultrafast processes and the dynamics of vibrational wave packets in molecular systems in gas, liquid, and solid phases. Polarization [1, 3] and phase [7] control of four-wave mixing (FWM) substantially enhances the potential of CARS spectroscopy and CARS microscopy, improving the sensitivity, as well as the temporal, spatial, and spectral resolution of these methods.

Conceptual and technical progress of CARS spectroscopy has largely followed the development of tunable radiation sources. In many of the early CARS experiments, including the pioneering works on time-resolved CARS [8], the Stokes signal was generated through stimulated Raman scattering in a cell with a Raman-active gas. The advent of frequency-tunable sources has opened a new phase in nonlinear spectroscopy [1]. Optical parametric oscillators and dye lasers have substantially expanded the area of applications of CARS spectroscopy, allowing this technique to be employed, in particular, as a convenient and practical engineering tool for the diagnostics of combustion and flames in automobile and rocket engines [9, 10].

In this work, we propose and experimentally implement a new scheme of femtosecond CARS spectroscopy based on the use of chirped pulses. We will develop a theory of this modification of coherent nonlinear spectroscopy (Section 2) to show that a linear time–frequency mapping defined by linearly chirped pulses allows the spectra of nonlinear response of a medium to be measured by varying the delay time between the pump pulses. Microstructure fibers [11–14] with a special dispersion profile are at the heart of the experimental implementation of this technique (Section 3). Such fibers will be shown to be ideally suited for the generation of frequency-tunable pulses with a smooth envelope and a controlled chirp. Due to the strong confinement of electromagnetic radiation in the

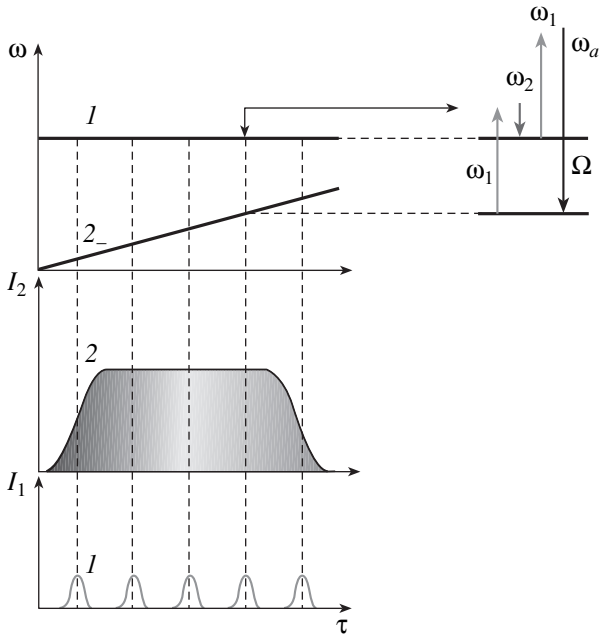


Fig. 1. A diagram of femtosecond CARS spectroscopy with chirped pulses.

core of a microstructure fiber and the possibility of tailoring the dispersion of guided modes, such fibers can radically enhance nonlinear-optical frequency conversion of femtosecond pulses [15–19], opening ways for the creation of conceptually new sources of frequency-tunable radiation for nonlinear spectroscopy. In Section 4, we show that femtosecond pulses frequency-upconverted and shaped in microstructure fibers can be employed for the CARS spectroscopy of toluene molecules in solution.

2. COHERENT ANTI-STOKES RAMAN SCATTERING SPECTROSCOPY WITH CHIRPED PULSES

Chirped ultrashort pulses are widely used in time-resolved spectroscopy and quantum control of ultrafast processes [20–27]. In this section, we examine the potential of chirped pulses for measuring the spectrum of nonlinear response of a Raman-active medium. Physically, the possibility of performing spectral measurements by means of time-resolved spectroscopy with the use of chirped pulses is based on linear frequency–time mapping, defined by linearly chirped pulses, which gives a simple linear relation between the frequency and time variables (Fig. 1). The spectrum of the nonlinear-optical susceptibility of a medium can be then found by measuring the nonlinear signal as a function of the delay time between the pump and probe pulses [28–31].

Consider a generic situation where the nonlinear signal is produced through the CARS process

$$\omega_4 = \omega_1 - \omega_2 + \omega_3$$

involving light pulses propagating along the z-axis,

$$E_i = A_i(t - z/v_i, z)\exp[i(k_i z - \omega_i t)] + \text{c.c.}, \qquad (1)$$

where A_i, ω_i, k_i, and v_i are the slowly varying envelope, the frequency, the wave number, and the group velocity of the ith pulse ($i = 1, 2, 3, 4$), respectively.

The equation for the slowly varying envelope of the CARS signal can then be written in the first order of dispersion theory as [32]

$$\left(\frac{\partial}{\partial z} - \frac{1}{v_4}\frac{\partial}{\partial t}\right)A_4(t - z/v_4, z) = -\frac{2\pi\omega_4^2}{ik_4c^2}P^{NL}\exp[-i(k_4 z - \omega_4 t)]. \qquad (2)$$

This equation involves the third-order nonlinear polarization of the medium [18]

$$P^{NL}(t, z) = \int_0^\infty\int_0^\infty\int_0^\infty \chi(t_1, t_2, t_3, z)E_1(t - t_1, z) \times E_2(t - t_2, z)E_3(t - t_3, z)dt_3dt_2dt_1, \qquad (3)$$

where $\chi(t_1, t_2, t_3, z)$ is the time-domain cubic nonlinear-optical susceptibility of the medium responsible for the CARS process.

In the case of a Raman-active transition, probed with the frequency difference $\omega_1 - \omega_2$, the Raman-resonant cubic susceptibility is written as

$$\chi(t_1, t_2, t_3, z) = \chi(t_1, z)\delta(t_1 - t_2)\delta(t_3). \qquad (4)$$

In the frequency domain, the cubic susceptibility then depends only on the frequency difference $\omega_1 - \omega_2$. Substituting Eq. (4) into Eq. (3) and integrating in t_1 and t_3, we arrive at the following expression for the third-order polarization of the medium:

$$P^{NL}(t, z) = \int_0^\infty \chi(t_1, z)E_1(t - t_1, z)E_2(t - t_1, z)dt_1E_3(t, z). \qquad (5)$$

Formula (5) and Eq. (2) govern a generic-type Raman-resonant FWM of plane waves in a situation when the dispersion spreading of short pulses is negligible.

We consider now the generation of a CARS signal by a triad of laser pulses assuming that one of these

pulses (for definiteness, the pulse with the frequency ω_2) is linearly chirped, while the other two pulses are transform-limited. Let the chirped pulse be delayed in time with respect to transform-limited pulses and scan the frequency difference $\omega_1 - \omega_2$ of the pump pulses through a resonance with a Raman-active transition in the nonlinear medium under study (Fig. 1). We represent the envelopes of the light pulses employed to pump and probe the medium as

$$A_1(\theta, z) = B_1(\theta), \quad (6)$$

$$A_2(\theta, z) = B_2(\theta)\exp[i\alpha(\theta - \tau)^2], \quad (7)$$

$$A_3(\theta, z) = B_3(\theta), \quad (8)$$

where $\theta = t - z/v_j$ is the retarded time, B_j is the envelope of the jth pump pulse involving no chirp ($j = 1, 2, 3$), α is the chirp parameter, and τ is the delay time between the first and second pulses.

Using Eqs. (6)–(8) in the retarded frame of references defined by θ and z, we derive the following expression for the envelope of the CARS signal:

$$\frac{\partial A_4(\theta, \tau, z)}{\partial z} = \frac{2i\pi\omega_4^2}{k_4 c^2}$$

$$\times \int_0^\infty \chi(t_1, z) B_1(\theta - t_1) B_2^*(\theta - \tau - t_1) \quad (9)$$

$$\times \exp[i(\Delta\omega t_1 - \alpha(\theta - \tau - t_1)^2 - \Delta k z)] dt_1 B_3(\theta),$$

where $\Delta k = k_4 - k_1 + k_2 - k_3$ is the phase mismatch and $\Delta\omega = \omega_1 - \omega_2$. The central frequency of the CARS signal is dictated by energy conservation:

$$\omega_4 = \omega_1 - \omega_2 + \omega_3.$$

In a spatially uniform medium with $\chi(t_1, z) = \chi(t_1)$, the following simplifications can be made for pulse envelopes whose variations are small on a time scale of the characteristic relaxation time T of nonlinear polarization, $t_1 \le T$:

$$B_{1,3}(\theta - t_1) \approx B_{1,3}(\theta), \quad B_2(\theta - \tau - t_1) \approx B_2(\theta - \tau).$$

Integration of Eq. (9) then yields

$$A_4(\theta, \tau, z) = \frac{2i\pi\omega_4^2}{k_4 c^2}$$

$$\times K(\Delta\omega + 2\alpha(\theta - \tau), \alpha) B_1(\theta) B_2^*(\theta - \tau) B_3(\theta) \quad (10)$$

$$\times \exp[-i\alpha(\theta - \tau)^2] \frac{\exp[-i\Delta k z] - 1}{-i\Delta k},$$

where

$$K(\omega, \alpha) = \int_0^\infty \chi(t_1)\exp[i\omega t_1]\exp[-i\alpha t_1^2] dt_1. \quad (11)$$

As long as the condition $\alpha T^2 \ll 1$ is satisfied, we have

$$K(\omega, \alpha) \approx K(\omega, 0) \approx \chi'(\omega). \quad (12)$$

Thus, the detection of the envelope of the CARS pulse, i.e., the measurement of the amplitude of the CARS signal as a function of the retarded time θ or the delay time τ, visualizes the spectrum of the relevant Raman-resonant third-order nonlinear susceptibility of the medium. In particular, the energy and the spectrum of the CARS signal as functions of the delay time τ in the regime of phase matching ($\Delta k = 0$) are given by

$$W_4(\tau) \propto \left| \int_{-\infty}^\infty K(\Delta\omega + 2\alpha(\theta - \tau)) \right.$$

$$\left. \times B_1(\theta) B_2^*(\theta - \tau) B_3(\theta) [-i\alpha(\theta - \tau)^2] l d\theta \right|^2, \quad (13)$$

$$S_4(\Delta\omega, \tau) = \left| \int A_4(\theta, \tau, z)\exp(i\Delta\omega\theta) d\theta \right|^2, \quad (14)$$

where l is the nonlinear interaction length.

In the standard scheme of two-color CARS spectroscopy, a single laser pulse is used instead of a pair of pulses with envelopes $B_1(\theta)$ and $B_3(\theta)$:

$$B_1(\theta) = B_3(\theta), \quad k_1 = k_3, \quad \omega_1 = \omega_3,$$

$$\Delta k = k_4 - 2k_1 + k_2.$$

Let the duration of the transform-limited pulse τ_1 be much less than the duration of the chirped pulse τ_2,

$$\tau_1 \ll \tau_2.$$

As can be seen from Eqs. (10), (13), and (14), the spectral resolution of CARS spectroscopy with a single chirped pulse is then determined by the duration of the short, transform-limited pulse with the envelope $B_1(\theta)$ and the parameter α of the chirped pulse. In the limiting case (Fig. 1) of a transform-limited pulse in the form of a delta function,

$$B_1(\theta) = B_1\delta(\theta),$$

and a chirped pulse whose amplitude remains constant on the considered time scale, $B_2(\theta) = B_2 = \text{const}$,

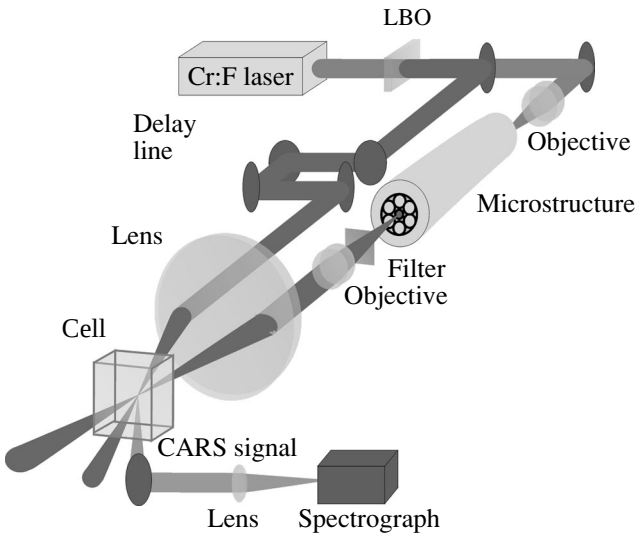


Fig. 2. Experimental setup for femtosecond CARS spectroscopy with the use of ultrashort pulses frequency-upconverted and chirped in a microstructure fiber.

Eqs. (13) and (14) yield the following expressions for the energy and the spectrum of the CARS signal:

$$W_4(\tau) \propto \left|K(\Delta\omega - 2\alpha\tau)B_1^2 B_2^* l\right|^2, \qquad (15)$$

$$S_4(\Delta\omega, \tau) \propto |K(\Delta\omega - 2\alpha\tau)|^2 \left|B_1^2 B_2\right|^2 l^2. \qquad (16)$$

Measuring the spectrum or the energy of the CARS signal as a function of the delay time τ, we can thus find the spectrum of the relevant nonlinear-optical susceptibility. A finite duration of the short, transform-limited pulse restricts the spectral resolution of this technique to $\delta\omega \approx \alpha\tau_1$, leading to spectrally nonuniform phase distortions of the signal. In Section 4 of this paper, we present an experimental implementation of the above-described cross-correlation CARS spectroscopy using transform-limited femtosecond pulses of a Cr: forsterite laser and chirped pulses produced and shaped through the frequency up-conversion of femtosecond Cr: forsterite-laser pulses in a microstructure fiber.

3. EXPERIMENTAL

The laser system employed in our experiments (Fig. 2) consisted of a Cr^{4+}: forsterite master oscillator, a stretcher, an optical isolator, a regenerative amplifier, and a compressor. The master oscillator, pumped with a fiber ytterbium laser, generated 30–50-fs light pulses with a central wavelength of approximately 1270 nm and a mean power of about 180 mW at a repetition rate of 120 MHz. These 30–50-fs pulses were then stretched up to 700 ps in a grating stretcher to be amplified in a Nd: YLF-laser-pumped regenerative amplifier. Amplified pulses with an energy up to 100 μJ were recompressed to a 75–150-fs duration in a grating compressor. Approximately 50% of the pulse energy was lost at this stage. An LBO crystal was employed for the frequency doubling of femtosecond pulses of the Cr: forsterite laser.

Microstructure fibers employed in our experiments were fabricated of soft glass and fused silica using technology described in detail elsewhere [33–38]. Cross-sectional images of these microstructure fibers are shown in the insets to Figs. 3a and 4. Glass microstructure fibers were designed in such a way as to provide optimal dispersion profiles of guided modes [17, 39–41] for the maximum efficiency of anti-Stokes frequency conversion of femtosecond Cr: forsterite-laser pulses to the range of wavelengths from 650 to 730 nm.

A propagation of femtosecond laser pulses through microstructure fibers was accompanied by nonlinear-optical interactions, giving rise to new frequency components in the spectrum of radiation coming out of the fiber. Parametric four-wave mixing

$$2\omega_p = \omega_s + \omega_a$$

(ω_p is the frequency of pump radiation and ω_s and ω_a are the frequencies of the Stokes and anti-Stokes signals, respectively) in the central core of a fused silica microstructure fiber (see the inset to Fig. 3a) resulted in the efficient generation of an anti-Stokes signal with a central wavelength around 530 nm and a spectral bandwidth of about 35 nm (Fig. 3a). We also observed efficient generation of the third harmonic of pump radiation. The amplitude of the third harmonic in these experiments was comparable to the amplitude of the anti-Stokes signal (the inset to Fig. 3b). Nonlinear-optical spectral transformation of femtosecond Cr: forsterite-laser pulses in submicron fused silica waveguiding channels of different diameters resulted in the generation of new frequency components within a broad spectral range, including spectral components blue-shifted with respect to the third harmonic of the pump field (Fig. 3b).

The wavelength range where the efficiency of anti-Stokes nonlinear-optical frequency conversion reaches its maximum is determined by the dispersion properties of guided modes involved in the nonlinear-optical process [18, 42]. The size of the waveguide channel is thus the key parameter for process of nonlinear-optical spectral transformation, controlling the frequencies of new spectral components generated as a result of nonlinear-optical frequency conversion of ultrashort laser pulses. Submicron waveguide channels with different diameters microstructure-integrated in the developed fibers allow efficient frequency conversion of femtosecond laser pulses through the generation of wavelength-tunable anti-Stokes radiation. For the CARS spectroscopy of Raman-active modes of toluene molecules (see Section 4), we employed anti-Stokes radiation generated within the range of wavelengths from 650 to 730 nm in the central core of a glass microstructure fiber (see the inset to Fig. 4). The efficiency of anti-Stokes frequency

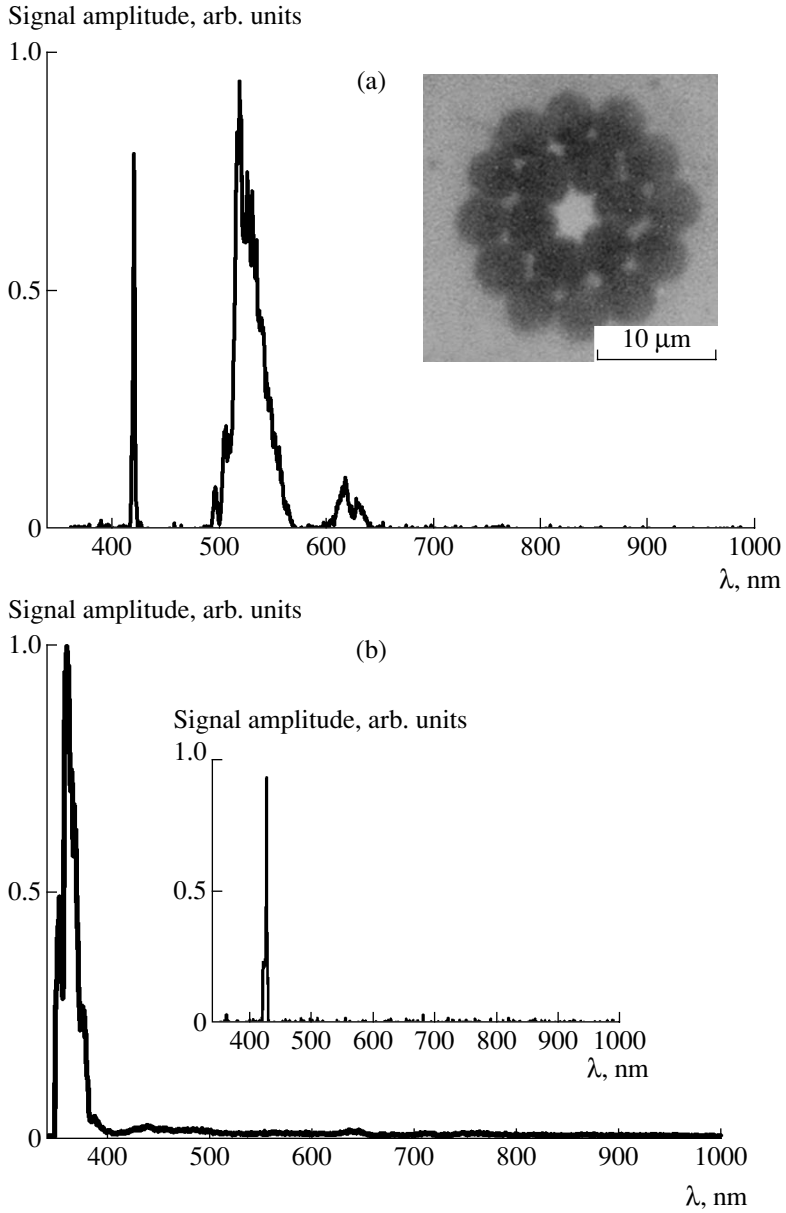


Fig. 3. (a) Generation of an anti-Stokes signal and the third harmonic in the central core of a fused silica microstructure fiber with a diameter of 3 μm. The inset shows the cross section image of the microstructure fiber. (b) Generation of a high-frequency spectral component with a central wavelength around 370 nm in a submicron waveguide channel of a microstructure fiber. The fiber length is 7 cm. The energy of laser pulses coupled into the fiber is about 50 nJ. The inset shows the spectrum of the third harmonic generated in submicron waveguide channels of the microstructure fiber.

conversion to this spectral range was estimated as 15% for pump pulses with an initial pulse duration of about 100 fs and an energy of 250 nJ. The central wavelength of the anti-Stokes signal was tuned by varying the input intensity of pump radiation (Fig. 4).

4. RESULTS AND DISCUSSION

Cross-correlation frequency-resolved optical gating (XFROG) [43] was used to characterize anti-Stokes pulses generated in the microstructure fiber. An XFROG signal was generated by mixing the anti-Stokes signal from the fiber, E_a, with the second-harmonic output of the Cr: forsterite laser, E_{SH}, in a BBO crystal. A two-dimensional XFROG sonogram,

$$S(\omega, \tau) \propto \left| \int_{-\infty}^{\infty} E_a(t) E_{SH}(t-\tau) \exp(-i\omega t) dt \right|^2,$$

was then plotted in a standard way by measuring the XFROG signal as a function of the delay time τ

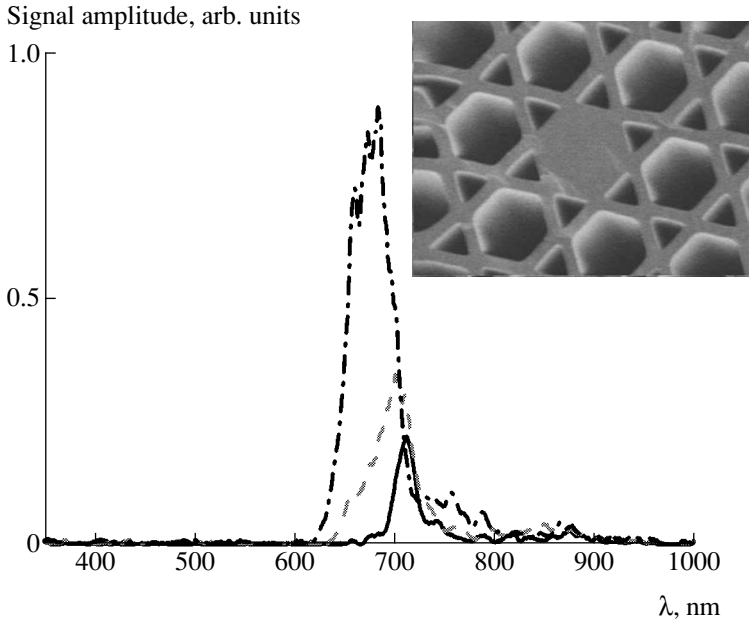


Fig. 4. The spectrum of the anti-Stokes signal generated in a microstructure fiber by 1.24-μm 90-fs Cr: forsterite-laser pulses with an input energy of (solid line) 200 nJ, (dashed line) 250 nJ, and (dash–dotted line) 275 nJ. The inset shows an image of the cross section of the microstructure fiber.

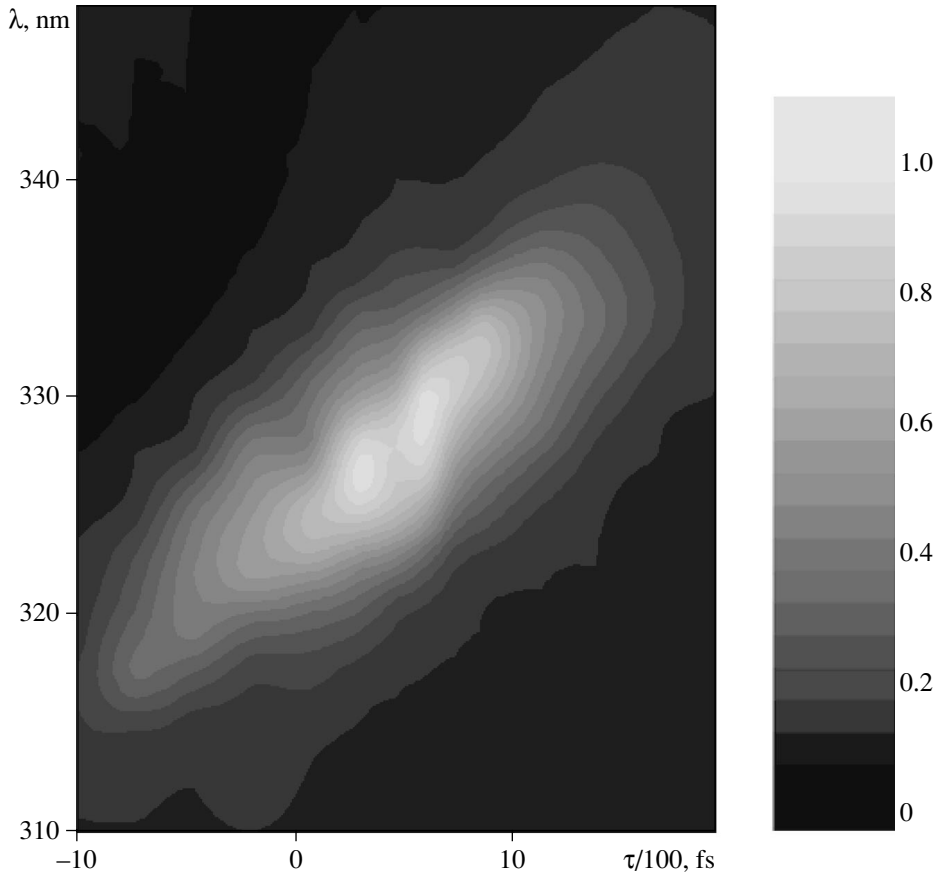


Fig. 5. The intensity of the sum-frequency signal generated in a BBO crystal by the second-harmonic pulse from the Cr: forsterite laser and the anti-Stokes pulse from the microstructure fiber as a function of the wavelength and the delay time τ between the second-harmonic and anti-Stokes pulses.

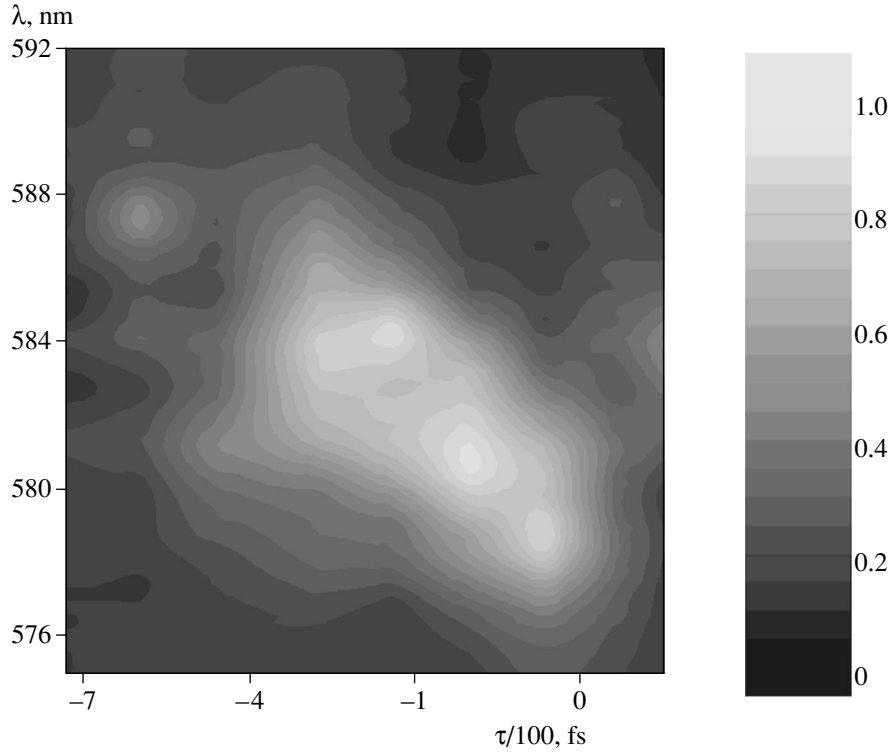


Fig. 6. The intensity of the CARS signal generated in a noncoplanar boxcars geometry from toluene solution as a function of the wavelength and the delay time τ between the second-harmonic and anti-Stokes pulses used as a biharmonic pump.

between the second-harmonic and anti-Stokes pulses and spectrally dispersing the XFROG signal. The XFROG sonogram visualizes the temporal envelope, the spectrum, and the chirp of the anti-Stokes signal generated in the microstructure fiber. Figure 5 presents an XFROG sonogram of the anti-Stokes signal from the microstructure fiber measured with the use of 620-nm 90-fs second-harmonic pulses. As can be seen from this map, the anti-Stokes pulse has a duration on the order of 1 ps, a smooth temporal envelope, and a virtually linear positive chirp.

Different frequency components of the anti-Stokes pulse generated and shaped in the microstructure fiber are characterized by different group delays. The instantaneous frequency ω in this pulse, as can be seen from the results presented in Fig. 5, is related to the delay time τ by a one-to-one mapping, defined by the chirp. Such a one-to-one τ – ω mapping allows spectral measurements to be performed by varying the delay time between the pump pulses [28–31]. In our experiments, we used the linearly chirped anti-Stokes pulse from the microstructure fiber as one of the biharmonic pump fields for the CARS spectroscopy of toluene solution.

Second-harmonic pulses of the Cr: forsterite laser (the frequency ω_1) and the anti-Stokes pulses from microstructure fibers (the frequency ω_2) were employed in our CARS scheme to coherently excite and probe Raman-active modes of toluene molecules in solution, generating the CARS signal at the frequency

$$\omega_{CARS} = 2\omega_1 - \omega_2$$

through the scattering off the coherently excited Raman-active vibrations. The structure of the fiber was optimized for efficient frequency conversion of Cr: forsterite-laser pulses to the wavelength region allowing Raman-active modes of toluene molecules to be probed within the range of wavenumbers from 1000 up to 1200 cm^{-1}. The light beams with frequencies ω_1 and ω_2 were focused into a cell with toluene solution at a small angle with respect to each other (Fig. 2). The CARS signal generated in the area of beam interaction in this noncoplanar boxcars geometry had a form of a sharply directed light beam with a low, phase-matching-controlled angular divergence spatially separated from the pump beams. This signal was then filtered and detected with a spectrometer.

XFROG CARS sonograms of Raman-active modes of toluene are presented in Fig. 6. The frequency difference $\omega_1 - \omega_2$ was scanned through the frequencies of Raman-active modes of toluene molecules by tuning the delay time τ between the second-harmonic output of the Cr: forsterite laser and the anti-Stokes pulse from the microstructure fiber. In the positively chirped pulse from the microstructure fiber, the instantaneous frequency increases from the leading to the trailing edge

(see Fig. 5). Small delay times τ, therefore, correspond to the excitation of low-frequency Raman-active modes ($\tau \approx -200$ fs in Fig. 6). In particular, the 1004-cm^{-1} Raman mode of toluene is well-resolved in the XFROG CARS sonogram. This mode is excited with the second harmonic of Cr: forsterite-laser radiation and the spectral slice around the wavelength of $\tau_2 \approx 661$ nm picked with an appropriate delay time from the positively chirped pulse frequency-upconverted in the microstructure fiber, giving rise to a CARS signal with a wavelength $\lambda_{CARS} \approx 584$ nm. Raman modes with higher frequencies are probed with larger delay times ($\tau \approx$ 100–200 fs in Fig. 6), with larger frequency differences $\omega_1 - \omega_2$ leading to the generation of a higher frequency CARS signal.

5. CONCLUSIONS

Experiments presented in this paper demonstrate that microstructure fibers with a specially designed dispersion offer ways to create efficient sources of ultrashort pulses for coherent nonlinear spectroscopy. Such fibers allow the generation and shaping of short pulses of anti-Stokes radiation with a smooth temporal envelope and a controlled chirp, providing a high efficiency of frequency upconversion even for nano- and subnanojoule femtosecond pulses. This allows requirements on the energy of pump radiation to be substantially reduced relative to optical parametric amplifiers and dye lasers, conventionally used as tunable light sources for coherent nonlinear spectroscopy and quantum-control experiments. Dye lasers are often employed for the generation of tunable-wavelength nano- and picosecond pulses for CARS spectroscopy. Such sources typically require millijoule pump pulses. Similar requirements on the pump have to be met in the case of nanosecond optical parametric oscillators. Modern femtosecond optical parametric amplifiers, which have received wide acceptance for femtosecond CARS spectroscopy, are usually pumped with laser pulses having typical energies on the order of tens or hundreds of microjoules. Microstructure fibers can efficiently generate ultrashort pulses with an unprecedentedly low level of pump energies. We have shown in this work that microstructure fibers can be optimized for the generation of anti-Stokes pulses ideally suited for time-resolved CARS spectroscopy.

ACKNOWLEDGMENTS

We are grateful to V.I. Beloglazov, N.B. Skibina, Yu.N. Kondrat'ev, V.S. Shevandin, K.V. Dukel'skiĭ, and A.V. Khokhlov for fabricating microstructure fibers for these experiments.

This study was supported in part by President of Russian Federation Grant MD-42.2003.02, the Russian Foundation for Basic Research (projects nos. 03-02-16929 and 02-02-17098), Award no. RP2-2558 of the U.S. Civilian Research and Development Foundation for the Independent States of the Former Soviet Union (CRDF), and the European Research Office of the US Army (contract no. 62558-03-M-0033).

REFERENCES

1. S. A. Akhmanov and N. I. Koroteev, *Methods of Nonlinear Optics in Light Scattering Spectroscopy* (Nauka, Moscow, 1981).
2. G. L. Eesley, *Coherent Raman Spectroscopy* (Pergamon, Oxford, 1981).
3. A. M. Zheltikov and N. I. Koroteev, Usp. Fiz. Nauk **169**, 385 (1999) [Phys. Usp. **42**, 321 (1999)].
4. A. Zumbusch, G. R. Holtom, and X. Sunney Xie, Phys. Rev. Lett. **82**, 4142 (1999).
5. D. A. Akimov, S. O. Konorov, D. A. Sidorov-Biryukov, *et al.*, Proc. SPIE **4749**, 101 (2002).
6. *Femtosecond Coherent Raman Spectroscopy*, Ed. by W. Kiefer, Special Issue of J. Raman Spectrosc. **31** (1/2) (2000).
7. N. Dudovich, D. Oron, and Y. Silberberg, Nature **418**, 512 (2002).
8. D. von der Linde, A. Lauberau, and W. Kaiser, Phys. Rev. Lett. **26**, 954 (1971).
9. *Nonlinear Raman Spectroscopy*, Ed. by P. Radi and A. M. Zheltikov, Special Issue of J. Raman Spectrosc. **33** (11/12) (2002).
10. *Nonlinear Raman Spectroscopy*, Ed. by P. Radi and A. M. Zheltikov, Special Issue of J. Raman Spectrosc. **34** (11/12) (2003).
11. J. C. Knight, T. A. Birks, P. St. J. Russell, and D. M. Atkin, Opt. Lett. **21**, 1547 (1996).
12. J. C. Knight, J. Broeng, T. A. Birks, and P. St. J. Russell, Science **282**, 1476 (1998).
13. P. St. J. Russell, Science **299**, 358 (2003).
14. A. M. Zheltikov, Usp. Fiz. Nauk **170**, 1203 (2000) [Phys. Usp. **43**, 1125 (2000)].
15. *Nonlinear Optics of Photonic Crystals*, Ed. by C. M. Bowden and A. M. Zheltikov, Feature Issue of J. Opt. Soc. Am. B **19** (9) (2002).
16. *Supercontinuum Generation*, Ed. by A. M. Zheltikov, Special Issue of Appl. Phys. B **77** (2/3) (2003).
17. D. A. Akimov, E. E. Serebryannikov, A. M. Zheltikov, *et al.*, Opt. Lett. **28**, 1948 (2003).
18. S. O. Konorov and A. M. Zheltikov, Opt. Express **11**, 2440 (2003).
19. A. M. Zheltikov, Usp. Fiz. Nauk **174**, 73 (2004) [Phys. Usp. **47**, 69 (2004)].
20. E. T. J. Nibbering, D. A. Wiersma, and K. Duppen, Phys. Rev. Lett. **68**, 514 (1992).
21. B. Kohler, V. V. Yakovlev, J. Che, *et al.*, Phys. Rev. Lett. **74**, 3360 (1995).
22. V. V. Lozovoy, S. A. Antipin, F. E. Gostev, *et al.*, Chem. Phys. Lett. **284**, 221 (1998).
23. T. Lang and M. Motzkus, J. Opt. Soc. Am. B **19**, 340 (2002).
24. C. J. Bardeen, Q. Wang, and C. V. Shank, Phys. Rev. Lett. **75**, 3410 (1995).
25. A. Assion, T. Baumert, M. Bergt, *et al.*, Science **282**, 919 (1998).

26. J. L. Herek, W. Wohlleben, R. J. Cogdell, *et al.*, Nature **417**, 533 (2002).
27. D. Zeidler, S. Frey, W. Wohlleben, *et al.*, J. Chem. Phys. **116**, 5231 (2002).
28. A. M. Zheltikov and A. N. Naumov, Quantum Electron. **30**, 606 (2000).
29. A. M. Zheltikov and A. N. Naumov, Laser Phys. **10**, 887 (2000).
30. A. N. Naumov and A. M. Zheltikov, J. Raman Spectrosc. **32**, 960 (2001).
31. A. N. Naumov and A. M. Zheltikov, Appl. Phys. B **77**, 369 (2003).
32. Y. R. Shen, *The Principles of Nonlinear Optics* (Wiley, New York, 1984; Nauka, Moscow, 1989).
33. M. V. Alfimov, A. M. Zheltikov, A. A. Ivanov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 714 (2000) [JETP Lett. **71**, 489 (2000)].
34. A. N. Naumov, A. B. Fedotov, A. M. Zheltikov, *et al.*, J. Opt. Soc. Am. B **19**, 2183 (2002).
35. A. B. Fedotov, Ping Zhou, A. P. Tarasevitch, *et al.*, J. Raman Spectrosc. **33**, 888 (2002).
36. D. A. Akimov, M. Schmitt, R. Maksimenka, *et al.*, Appl. Phys. B **77**, 299 (2003).
37. A. M. Zheltikov, Ping Zhou, V. V. Temnov, *et al.*, Quantum Electron. **32**, 542 (2002).
38. S. O. Konorov, O. A. Kolevatova, A. B. Fedotov, *et al.*, Zh. Éksp. Teor. Fiz. **123**, 975 (2003) [JETP **96**, 857 (2003)].
39. S. Coen, A. Hing Lun Chau, R. Leonhardt, *et al.*, J. Opt. Soc. Am. B **19**, 753 (2002).
40. J. M. Dudley, L. Provino, N. Grossard, *et al.*, J. Opt. Soc. Am. B **19**, 765 (2002).
41. A. B. Fedotov, I. Bugar, D. A. Sidorov-Biryukov, *et al.*, Appl. Phys. B **77**, 313 (2003).
42. S. O. Konorov, A. B. Fedotov, W. Boutu, *et al.*, Laser Phys. **14**, 100 (2004).
43. J. M. Dudley, X. Gu, L. Xu, *et al.*, Opt. Express **10**, 1215 (2002).

Translated by A. Zheltikov

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 197–201.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 224–228.

SOLIDS
Electronic Properties

Effect of a Strong Magnetic Field on the Quadrupole and Magnetic Orders and Crossover in the Jahn–Teller Magnet $DyVO_4$

A. A. Demidov[a], Z. A. Kazeĭ[b], N. P. Kolmakova[a,*], J.-M. Broto[c], and H. Rakoto[c]

[a] *Bryansk State Technical University, Bryansk, 241035 Russia*
**e-mail: npk@bitmcnit.bryansk.su*
[b] *Moscow State University, Vorob'evy gory, Moscow, 119992 Russia*
[c] *Laboratoire National des Champs Magnétiques Pulsés, 31432, Toulouse, France*
Received December 3, 2003

Abstract—The effect of a strong magnetic field $H \parallel [001]$ on magnetic properties of the Jahn–Teller compound $DyVO_4$ is studied. New phase transitions discovered and investigated experimentally and theoretically include the breaking of quadrupole order (enhancement of the crystal symmetry) and breaking of antiferromagnetic order as well as the effect of convergence of energy levels of the Dy^{3+} ion (crossover). The differential magnetic susceptibility of a $DyVO_4$ crystal is measured in fields up to 36 T in the temperature interval 1.4–15 K. The observed magnetic susceptibility anomalies and phase transitions are described using a unified theoretical approach. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Jahn–Teller and magnetic phase transitions and their description on the basis of various theoretical models have always been an object of interest in solid-state physics. Breaking of the initial magnetic order and enhancement of crystal symmetry by a magnetic field at temperatures below the critical value, which are also phase transitions, are of considerable interest both as physical phenomena and as an object for verifying the applicability of theoretical models. Such phase transitions are studied here for dysprosium vanadate ($DyVO_4$) belonging to systems with two types of ordering. This compound occupies a special place among Jahn–Teller compounds since the molecular field approximation, which is successfully used for compounds of this type, has limited applicability for it. This is attributed to almost complete compensation of long-range contributions to the interionic interaction responsible for the cooperative Jahn–Teller effect (quadrupole ordering). The studies devoted to $DyVO_4$ are reviewed in [1]. Spectroscopic and thermodynamic characteristics of this compound were also studied in subsequent publications, but information on its behavior in strong magnetic fields is not available.

Dysprosium vanadate experiences a structural phase transition at $T_C = 14$ K; as a result, the crystal symmetry changes from tetragonal D_{4h}^{19} (zirconium-type structure) to orthorhombic D_{2h}^{28}. The transition is a cooperative Jahn–Teller effect (ordering of the quadrupole moments of Dy^{3+} ions) associated with the specific arrangement of energy levels of two lower Kramers doublets. At $T_N \approx 3$ K, $DyVO_4$ becomes antiferromagnetic with the orientation of the magnetic moments of Dy^{3+} ions along the **a** axis. For a field directed along this axis, at approximately 0.21 T, compound $DyVO_4$ exhibits a metamagnetic transition with a peculiar phase diagram (see, for example, [2]). This study is devoted to experimental and theoretical analysis of the effect of a strong magnetic field on the magnetic and quadrupole ordering as well as on the rearrangement of the energy spectrum of the Dy^{3+} ion, which is accompanied by considerable anomalies in the magnetic properties at low temperatures. It should be noted that suppression of quadrupole order by a magnetic field in a zircon structure was experimentally investigated for elastic properties of $TmVO_4$ ($T_C = 2.15$ K, $\mu_0 H_C \approx 0.6$ T) [3].

2. RESULTS AND DISCUSSION

The differential magnetic susceptibility $dM/dH(H)$ of a $DyVO_4$ crystal was measured in the temperature range 1.4–15 K for the field orientation $H \parallel [001]$ by the inductive method using the equipment of the National Laboratory of Pulsed Magnetic Fields (Toulouse, France). The field was generated during the discharge of a capacitor bank through a copper coil. The maximal field of 36 T was attained in 100 ms, the time of field removal was about 1 s. The field disorientation relative to the [001] axis under the experimental conditions was

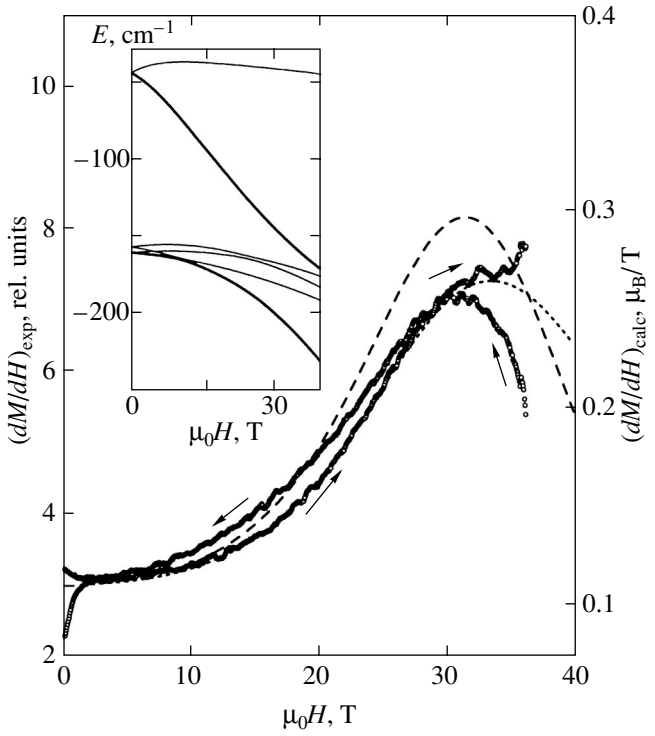


Fig. 1. Experimental curves of differential magnetic susceptibility $dM/dH(H)$ of $DyVO_4$ at $T = 15$ K recorded during the increase and decrease of the field $H \parallel [001]$. Theoretical curves are given for the isothermal (dashed curve) and adiabatic (dotted curve) regimes. The inset shows a fragment of the Zeeman effect.

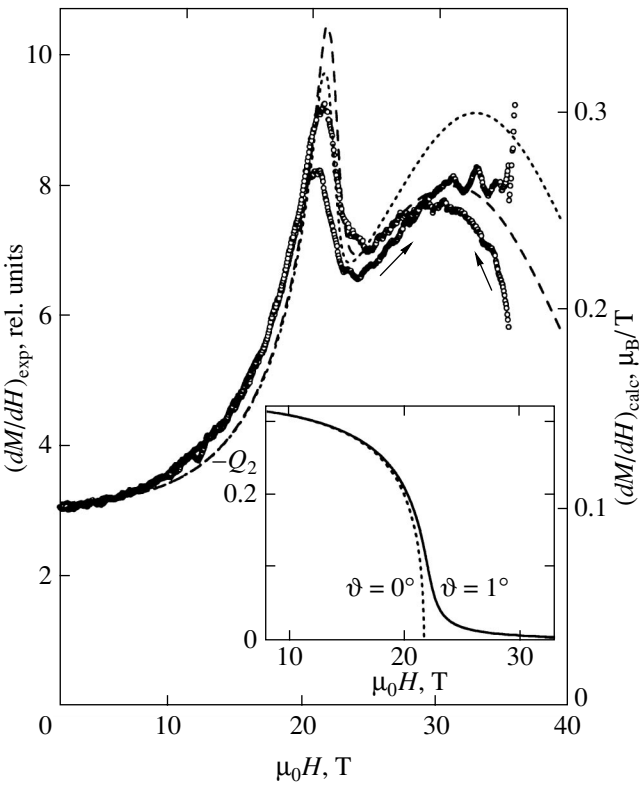


Fig. 2. Experimental curves of differential magnetic susceptibility $dM/dH(H)$ of $DyVO_4$ at $T = 7$ K recorded during the increase and decrease of the field. Theoretical curves are given for the isothermal (dashed curve) and adiabatic (dotted curve) regimes, $\vartheta = 1°$. The inset shows the field dependence of the quadrupole order parameter $Q_2 = \alpha_J \langle O_2^2 \rangle$ for two field orientations.

the same at all temperatures and did not exceed 1°–2°. Our calculations revealed that a disorientation angle of 1° corresponds to the best agreement for the effects sensitive to field disorientation.

The experimental curves of differential magnetic susceptibility $dM/dH(H)$ of a $DyVO_4$ crystal for $H \parallel [001]$ are shown in Figs. 1–3 for temperatures from the three characteristic intervals. For $T = 15\ \text{K} > T_C$, the dM/dH curve displays a broad peak in the region of strong fields, which is seen most clearly for field removal (Fig. 1). At $T_N < T = 7\ \text{K} < T_C$, two anomalies can be distinguished on the susceptibility curve (Fig. 2). AT $T = 1.4\ \text{K} < T_N$, three anomalies are observed (Fig. 3): sharp peaks followed by a less clearly manifested anomaly in a strong field.

To calculate the effect of a strong magnetic field in the low-temperature magnetic properties of $DyVO_4$, we used the Hamiltonian $\mathcal{H}$ including the crystal-field Hamiltonian $\mathcal{H}_{CF}$ written in terms of equivalent operators O_n^m, the Zeeman term $\mathcal{H}_Z$, and the Hamiltonian of the exchange and dipole interactions $\mathcal{H}_M$ responsible for antiferromagnetic ordering of the magnetic moments of Dy^{3+} ions:

$$\mathcal{H} = \mathcal{H}_{CF} + \mathcal{H}_Z + \mathcal{H}_M, \tag{1}$$

$$\mathcal{H}_{CF} = \alpha_J(B_2^0 O_2^0 + B_2^2 O_2^2) + \beta_J(B_4^0 O_4^0 + B_4^4 O_4^4) + \gamma_J(B_6^0 O_6^0 + B_6^4 O_6^4),$$

$$\mathcal{H}_Z = g_J \mu_B \mathbf{H} \cdot \mathbf{J}.$$

In the molecular field approximation for an ion from the ith sublattice ($i = 1, 2$), $\mathcal{H}_{Mi}$ has the form

$$\mathcal{H}_{Mi} = \lambda_x M_{jx} J_{ix} + \lambda_z M_{jz} J_{iz}, \quad \mathbf{M}_j = g_J \mu_B \langle \mathbf{J}_j \rangle, \quad j = 1, 2, \quad j \neq i. \tag{2}$$

In these expressions, B_n^m are the crystal field parameters; α_J, β_J, and γ_J are the Stevens coefficients, g_J is the Lande factor, and $\mathbf{J}$ is the angular momentum operator for the rare-earth ion.

For the tetragonal symmetry, the crystal-field parameter B_2^2 equals zero. In the orthorhombic phase, the interactions responsible for the cooperative Jahn–Teller effect in the quadrupole approximation give $B_2^2 \neq 0$ since structure distortion below T_C possesses the B_{1g} symmetry and the quadrupole moment $Q_2 = \alpha_J\langle O_2^2\rangle$ is the corresponding order parameter (see [4]). In this paper, we use the model proposed in [5] in the framework of the pseudospin formalism and which takes into account the increase in the effective interaction with increasing ordering (compressible model). A generalization to the case when the quadrupole moment calculated on the principal multiplet of the rare-earth ion is the order parameter gives

$$B_2^2 = B_2^2(0)\frac{Q_2(T,H)}{Q_2(0,0)}\left\{1+\xi\left[\frac{Q_2(T,H)}{Q_2(0,0)}\right]^2\right\}, \qquad (3)$$

where $\xi = 0$ corresponds to the ordinary molecular field approximation. The value of B_2^2 for $H = 0$ and low temperatures was found from the splitting of lower Kramers doublets, which is known to be equal to 27 cm^{-1} under these conditions [6, 7].

The crystal field parameters for $DyVO_4$ are actually unknown since the spectroscopic data obtained for the Dy^{3+} ion in $DyVO_4$ in the tetragonal and orthorhombic phases are insufficient. We determined the crystal field on the basis of our experimental results on the temperature dependences of the initial magnetic susceptibility along the principal crystallographic axes, the values of *g*-tensor components, and available information on the structure of the fundamental multiplet [1]. The obtained values of crystal field parameters are as follows: $B_2^0 = -92$ cm^{-1}, $B_4^0 = 47.3$ cm^{-1}, $B_6^0 = -40.6$ cm^{-1}, $B_4^4 = 900$ cm^{-1}, and $B_6^4 = -75.8$ cm^{-1}. These parameters well describe the initial magnetic susceptibility of $DyVO_4$ and are close to those obtained for $HoVO_4$ [8], whose crystal field is regarded as reliably established. The gap between two Kramers doublets in the tetragonal phase amounts to approximately 3.5 cm^{-1}, which is in good agreement with the Raman and IR studies of the energy spectrum of the Dy^{3+} ion in $DyVO_4$ above 14 K [6, 7] and contradicts the results obtained in [9], according to which this splitting is approximately equal to 9 cm^{-1}.

The pulse duration in our experiment is such that the regimes of magnetization can be regarded as close to adiabatic conditions for increasing field and to isothermal conditions for decreasing field except for the transient process observed after the sign reversal of *dH*/*dt*. Magnetization *M* and differential magnetic susceptibility *dM*/*dH* were calculated from the spectrum and wave functions of the Dy^{3+} ion, which were calculated by numerical diagonalization of Hamiltonian $\mathcal{H}$ for each value of *H* and *T* by solving the corresponding self-consistent problems. For adiabatic magnetization, allowance was made for the magnetocaloric effect, which was calculated by the formula

$$\Delta T = -T\left(\frac{\partial M}{\partial T}\right)_H \frac{\Delta H}{C_H} \equiv -T\left(\frac{\partial S_{\text{mag}}}{\partial H}\right)_T \frac{\Delta H}{C_H}, \qquad (4)$$

for field variation from *H* to $H + \Delta H$; the field step ΔH was chosen small enough to ensure the convergence of the numerical procedure. The heat capacity C_H in formula (4) includes, apart from the magnetic heat capacity, the lattice heat capacity $C_{\text{lat}} = (12/5)\pi^4 k_B \nu (T/T_D)^3$ (the Debye temperature for zircons is $T_D = 275$ K [10], $\nu = 6$).

In the tetragonal phase of $DyVO_4$, field $H \parallel [001]$ leads to convergence of energy levels of the Dy^{3+} ion (see the inset to Fig. 1), accompanied by an increase in

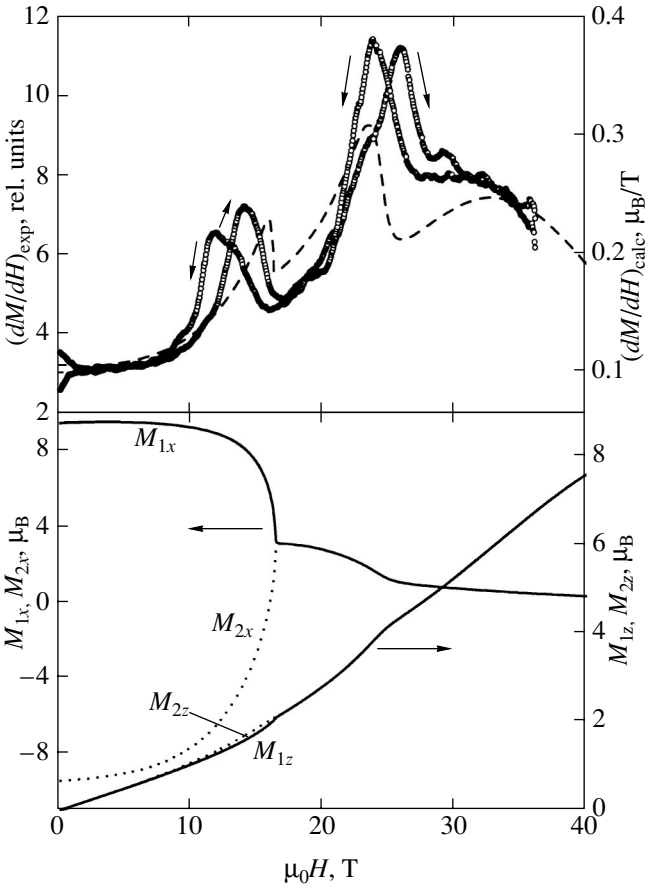


Fig. 3. Experimental curves of differential magnetic susceptibility $dM/dH(H)$ of $DyVO_4$ at $T = 1.4$ K recorded during the increase and decrease of the field $H \parallel [001]$. Theoretical curves for the isothermal and adiabatic regimes coincide (dashed curve). The field dependences of the projections of magnetic moments $\mathbf{M}_1$ and $\mathbf{M}_2$ of Dy^{3+} ions belonging to different sublattices at $T = 1.4$ K and $\vartheta = 1°$.

the susceptibility in the vicinity of the field $\mu_0 H_{c1} \approx 31.5$ T, which depends on temperature and disorientation only slightly. Matching of experimental and calculated isothermal and adiabatic curves $dM/dH(H)$ for $T = 15$ K (see Fig. 1) serves as a verification of the value obtained for the crystal field. (The scales in the figures are chosen so that the experimental and theoretical susceptibility curves in weak fields coincide.)

For $T_N < T < T_C$, magnetic field $H \parallel [001]$ first breaks the quadrupole order (see the inset to Fig. 2) and transforms the crystal from the orthorhombic to the tetragonal phase (this corresponds to a sharp peak on the $dM/dH(H)$ curve in a field of $\mu_0 H_{c2} \approx 21$ T) and then leads to the convergence of the energy levels of the Dy^{3+} ion and susceptibility anomaly similar to that described above for $T = 15$ K. This situation is depicted in Fig. 2 for $T = 7$ K, indicating a good agreement between the experimental data and calculations performed without any fitting parameters (the value of $\xi = 0.3$ is taken from [5]). The calculation of the magnetocaloric effect for this initial temperature gives an insignificant change in temperature in fields smaller than H_{c2}; for this reason, the isothermal and adiabatic curves $dM/dH(H)$ differ only slightly for $H < H_{c2}$ in the region of quadrupole order breaking also. The appreciable heating for $H > H_{c2}$ (in a field of 36 T, $T \approx 16$ K) obtained in calculations makes it possible to explain the mutual arrangement of the susceptibility curves in the crossover region. The isotherm $dM/dH(H)$ calculated for $T = 7$ K in the crossover region lies above the experimental curve obtained for field removal, which corresponds to a higher temperature of the sample in the initial stage of field removal.

When the field is oriented strictly along the axis [001], quadrupole order breaking occurs as a second-order phase transition; disorientation blurs the phase transition (see the inset to Fig. 2). The mechanism of quadrupole order suppression by an external field can be visualized as follows. The magnetic field splits Kramers doublets and rearranges the entire energy spectrum. Jahn–Teller deformations only change the spacing between the doublets. If the splitting in the magnetic field is such that the gain in energy as a result of the Jahn–Teller distortion of the structure becomes insufficient for overcoming the effect corresponding to the decrease in entropy, the low-symmetry phase vanishes.

At $T < T_N$, a magnetic field perpendicular to the antiferromagnetism vector ($H \parallel [001]$) breaks the antiferromagnetic order along the [100] axis and $DyVO_4$ passes to the paramagnetic state with a magnetic moment orientation along the field. This occurs in the field H_{c3} corresponding to the first peak of the susceptibility curve. In strong fields, the effects described above are observed. A considerable hysteresis that takes place for field application and removal at $T = 1.4$ K (Fig. 3) suggests that relaxation processes affect the shape and position of the peaks on the $dM/dH(H)$ curve. An analysis of the magnetocaloric effect shows that in the field range under investigation, the temperature varies insignificantly (by tenths of a kelvin); consequently, the hysteresis cannot be due to the magnetocaloric effect.

Figure 3 demonstrates a good qualitative (but not quantitative) agreement between the experimental and calculated $dM/dH(H)$ curves in $DyVO_4$ at $T = 1.4$ K (the adiabat and the isotherm coincides in view of a weak magnetocaloric effect). The same figure also shows the field dependences of the projections of magnetic moments $\mathbf{M}_1$ and $\mathbf{M}_2$ of Dy^{3+} ions that form an antiferromagnetic structure along axis $\mathbf{a} \parallel x$ for $H = 0$ for a field deviation from axis $\mathbf{c} \parallel z$ by 1° in the xz plane. The calculations were made for the total Hamiltonian by solving the self-consistent problem in five order parameters: Q_2, M_{1z}, M_{1x}, M_{2z}, and M_{2x}. The exchange interaction constants in Eq. (2) ($\lambda_x = \lambda_z = 0.03$ cm$^{-1}/\mu_B$) are estimated from the known value of T_N. The calculations proved that the inclusion of the exchange interaction anisotropy ($\lambda_x \neq \lambda_z$) does not lead to any significant changes. The field dependences of projections of $\mathbf{M}_1$ and $\mathbf{M}_2$ for a small disorientation angle $\vartheta = 1°$ demonstrate a huge anisotropy formed by the crystal field and typical of many rare-earth compounds. In particular, this dependence is manifested in considerable anisotropy of the g-tensor components for $DyVO_4$ at low temperatures ($g_a \approx 19$ and $g_c \approx 0.5$ [1]). In field H_{c3}, the sublattices collapse, but the orientation of the resultant magnetic moment is still far from the field direction; e.g., $\vartheta_M(H = H_{c3}) \approx 56°$ for $\vartheta = 1°$. In the model used here, the collapse of the sublattices in the field and the transition of the system to the paramagnetic state occur as a second-order phase transition. The field disorientation relative to the axis does not protract the transition in contrast to the retardation taking place in quadrupole order breaking. In particular, this constitutes the difference in the role of the crystal field for quadrupole and magnetic ordering.

3. CONCLUSIONS

Thus, the experimental and theoretical investigations of low-temperature magnetic properties of the Jahn–Teller compound $DyVO_4$ in a strong field proved that the field directed along the [001] axis at $T < T_N$ first breaks the antiferromagnetic order of the magnetic moments of Dy^{3+} ions along the [100] axis and transforms the crystal to the paramagnetic phase. Then quadruple order breaking takes place and the crystal symmetry is elevated to tetragonal, for which the quadrupole moment $Q_2 = \alpha_J \langle O_2^2 \rangle$ vanishes. A further increase in the field induces one more considerable rearrangement of the energy spectrum of Dy^{3+} ions, viz., the convergence of energy levels in the lower part of the multiplet of the rare-earth ion. The second and third event from those listed above take place at $T_N < T < T_C$, while only the third event occurs at $T > T_C$. All these transi-

tions/effects are accompanied by considerable anomalies in magnetic properties (peaks on the field dependence of the differential magnetic susceptibility), which are successfully described using a unified approach.

ACKNOWLEDGMENTS

This study was partly financed by the Russian Foundation for Basic Research (project no. 03-02-16809).

REFERENCES

1. A. Kasten, Z. Phys. B **38**, 65 (1980).
2. M. Sieger, A. Kasten, and W. Paul, Solid State Commun. **53**, 909 (1985).
3. R. L. Melcher, E. Pytte, and B. A. Scott, Phys. Rev. Lett. **31**, 307 (1973).
4. Z. A. Kazeĭ, N. P. Kolmakova, I. B. Krynetskiĭ, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **42**, 278 (2000) [Phys. Solid State **42**, 285 (2000)].
5. J. H. Page and D. R. Taylor, Solid State Commun. **40**, 907 (1981).
6. R. T. Harley, W. Hayes, and S. R. P. Smith, Solid State Commun. **9**, 515 (1971).
7. A. H. Cooke, C. J. Ellis, K. A. Gehring, *et al.*, Solid State Commun. **8**, 689 (1970).
8. B. Bleaney, J. F. Gregg, P. Hansen, *et al.*, Proc. R. Soc. London **416**, 63 (1988).
9. G. A. Gehring, A. P. Malozemoff, W. Staude, and R. N. Tyte, J. Phys. Chem. Solids **33**, 1499 (1972).
10. A. Kasten, H. G. Kahle, P. Klofer, *et al.*, Phys. Status Solidi B **144**, 423 (1987).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 202–210.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 229–238.

SOLIDS
Electronic Properties

Spin Dynamics and Internal Motion in Magnetically Dilute Manganites Probed by EPR

V. A. Atsarkin[a], V. V. Demidov[a], D. G. Gotovtsev[a], N. E. Noginova[a,b], D. Bitok[b], and R. Bah[b]

[a]*Institute of Radio Engineering and Electronics, Russian Academy of Sciences, Moscow, 125009 Russia*
e-mail: atsarkin@mail.cplire.ru
[b]*Norfolk State University, 23504 Norfolk VA, USA*

Received January 12, 2004

Abstract—EPR spectra of a series of single crystal samples of $LaGa_{1-x}Mn_xO_3$ manganites were studied in a broad range of temperatures (20–300 K) and compositions ($0.02 \le x \le 1$). Evolution of the spectra under the action of exchange interactions of Mn^{3+} ions leads to the formation of a single Lorentzian line for $x \ge 0.2$. The results confirm the transition from antiferromagnetic to ferromagnetic spin ordering during diamagnetic dilution of the manganite with gallium, beginning with $x = 0.8$. For $x = 0.1$, the EPR spectra exhibit an anomalous broadening and splitting on cooling of the samples, which is interpreted as manifestations of a thermoactivated internal motion with a characteristic energy of about 50 meV. The results are compared with published data on the nuclear spin–lattice relaxation of gallium. The model of thermoactivated reorientations of the electron orbitals of Mn^{3+} ions subject to the Jahn–Teller effect is discussed. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In recent years, rare earth manganites—substances with the general formula $La_{1-y}Me_yMnO_3$, where Me = Ca, Sr, … and La is a rare earth element—have received much attention (see, e.g., reviews [1–4]). These materials possess a perovskite structure, are characterized by very rich phase diagrams, and exhibit unique magnetic and transport properties, the most interesting of which is the so-called colossal magnetoresistance (CMR). The nature of the CMR phenomenon is commonly related to the "double exchange" mechanism involving electron hopping between Mn^{3+} and Mn^{4+} ions, which results in the effective exchange of the charged states. Such hopping is more probable for the parallel orientation of spins of the interacting ions, which qualitatively explains the CMR effect whose maximum is observed near the point of ferromagnetic ordering (T_C) coinciding with the insulator–metal transition.

However, the double exchange mechanism cannot account for all peculiarities in the behavior of manganites. Indeed, Mn^{3+} ions comprising the main sublattice have an electron configuration of d^4 corresponding to an integer effective spin of $S = 2$. The orbital state of the outer electron (e_g) is subject to the Jahn–Teller effect, whereby the octahedron of the nearest-neighbor oxygen ions is distorted in one axis according to the $d_{3x^2-r^2}$ or $d_{3y^2-r^2}$ orbital selected in the crystallographic *ab* plane. In contrast, the Mn^{4+} ion with a d^3 configuration and a spin of $S = 3/2$ is characterized by a frozen orbital moment and is not subject to the Jahn–Teller effect. Therefore, the electron hopping between Mn^{3+} and Mn^{4+} ions not only implies a change in the charge and spin, but also involves rearrangement of the nearest environment with the formation of an object conventionally referred to as the Jahn–Teller polaron. The physical picture is far from being completely clear and gives rise to animated discussions between specialists.

In the investigation of doped manganites featuring the CMR phenomenon, it is expedient to concentrate on and proceed from the properties of $LaMnO_3$—the parent compound containing only triply charged manganese ions. This compound does not exhibit a transition to the metallic state, while the magnetic ordering observed at temperatures below $T_N = 140$ K (Néel temperature) renders it an antiferromagnet comprising alternating ferromagnetic *ab* planes with oppositely oriented magnetizations. It was established (see, e.g., [5, 6]) that, in a broad range of temperatures up to $T_{JT} = 750$ K, the Jahn–Teller distortions of the crystal field for the adjacent Mn^{3+} ions in the *ab* plane of $LaMnO_3$ are correlated and the outer orbitals of these ions are mutually perpendicular (i.e., directed alternatively along *x* and *y* axes). At $T < T_{JT}$, this orbital order is rigid and corresponds to the cooperative static Jahn–Teller effect. As the temperature increases to $T > T_{JT}$, the phase transition to a more symmetric state takes place in which the orbitals exhibit rapid and continuous reorientation: this corresponds to the cooperative dynamic Jahn–Teller effect. The transition from static orbital ordering to the

dynamic regime is manifested, in particular, by a considerable decrease in the linewidth in the electron paramagnetic resonance (EPR) spectra [7–9].

It is interesting to note that adding any impurities—not only divalent alkaline earth ions, but diamagnetic trivalent ions of the Ga^{3+} type (not influencing the charged state of manganese ions)—makes the static orbital ordering in $LaMnO_3$ less stable, which is manifested, in particular, by a sharp decrease in the characteristic temperature T_{JT} [10–15]. It was suggested that this is related to violation of the strict periodicity of the crystal lattice as a result of replacement of the host components by foreign atoms, leading eventually to the order–disorder transition. In this case, two alternative mechanisms of the exchange interaction between manganese ions are proposed: (i) vibronic superexchange under the conditions of correlated reorientations of the mutually perpendicular e_g orbitals [10, 11]; (ii) double exchange due to the real electron hopping in the Mn^{2+}–Mn^{4+} pair formed from the Mn^{3+}–Mn^{3+} pair as a result of the disproportionation reaction [15]. Both mechanisms predict a ferromagnetic character of the exchange interaction. Further experimental investigations of isomorphous solid solutions of the $LaMnO_3$–$LaGaO_3$ system must provide answers to the questions that arise. Such investigations in the entire range of Mn–Ga substitution will show evolution of the spin exchange and orbital correlations from their appearance on the level of elementary processes up to the establishment of a collective character, when the picture corresponds to that observed in $LaMnO_3$ and CMR materials.

The magnetic properties, structure, and electric conductivity of the $LaMnO_3$–$LaGaO_3$ system have been extensively studied [10–17], but only using integral methods. Information concerning the spin dynamics on a local (atomic) level can be gained by means of NMR and EPR techniques. The results of recent NMR and nuclear spin relaxation measurements for ^{69}Ga, ^{71}Ga, and ^{139}La nuclei in single crystals of $LaGa_{1-x}Mn_xO_3$ ($0 \le x \le 0.2$) [18] showed that these materials feature thermoactivated internal motion with an activation energy of about 50 meV, which can be related to reorientations of the Jahn–Teller distortions in the clusters or pinned centers of Mn^{3+}.

We have studied $LaGa_{1-x}Mn_xO_3$ crystals by EPR. As will be shown below, the results agree with the NMR data and provide new supplementary information about the formation and evolution of the exchange-coupled spin system in manganites.

2. EXPERIMENTAL METHODS AND RESULTS

The experiments were performed on a series of single crystal samples of the $LaGa_{1-x}Mn_xO_3$ system with $0 < x \le 1$. The samples with $x \le 0.2$ were grown using the Czochralski method (see [16, 18]); the crystals with higher manganese concentrations were obtained by zone melting (the compositions are indicated for the initial charge). The results of X-ray diffraction measurements demonstrated single-phase character of all samples and confirmed the orthorhombic structure as belonging to the *Pbnm* point symmetry group of type *O*' ($c/\sqrt{2} < a, b$) for $x \ge 0.5$ and *O** ($a, b \sim c/\sqrt{2}$) for $x < 0.5$.Variation of the crystal lattice parameters *a*, *b*, and *c* depending on *x* corresponded to the published data [11, 13], which was evidence of the correct determination of manganese content in the crystals studied.

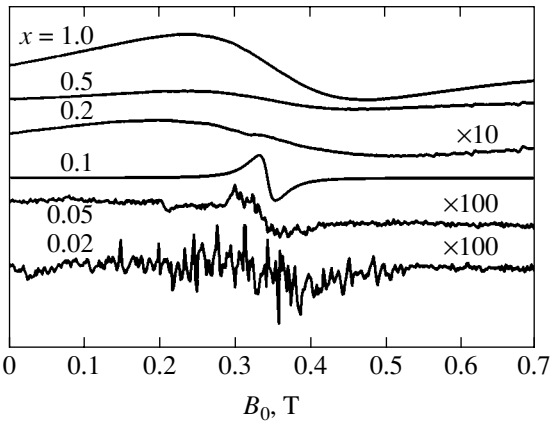


Fig. 1. Room-temperature EPR spectra (derivative of the absorption signal in arbitrary units) of $LaGa_{1-x}Mn_xO_3$ crystals with variable manganese content *x* (indicated at the curves).

The EPR spectra were measured in the X band with a Bruker ER-200 spectrometer using magnetic modulation and a lock-in detection at 100 kHz. The spectra were recorded in the form of a derivative of the resonance absorption with respect to the external magnetic field. The temperature was controlled in a broad range (20–300 K) to within 1 K by a cryogenic system (Oxford Instruments). Analytical expressions were fit to the experimental signals with the aid of a standard program (MicroCal Origin).

Figure 1 shows the EPR spectra measured at room temperature for a series of $LaGa_{1-x}Mn_xO_3$ samples with increasing *x*. As can be seen, the spectrum of a sample with small manganese content ($x = 0.02$) displays a vast number of narrow, partly overlapping lines with strongly anisotropic positions (note that this is not noise: the signal to noise ratio in this record is sufficiently large). Quantitative interpretation of this spectrum is hardly possible. The complex shape of the EPR spectrum observed in this case is related to the fact that manganese in the compositions with small *x* may enter into the perovskite lattice in various charged states. In addition to Mn^{3+} ($S = 2$) ions isomorphously substituted for Ga^{3+} ions, this manganite probably contains Mn^{4+} ($S = 3/2$), Mn^{2+} ($S = 5/2$), and Mn^{5+} ($S = 1$) ions, which is confirmed by the optical spectra [19]. The EPR sig-

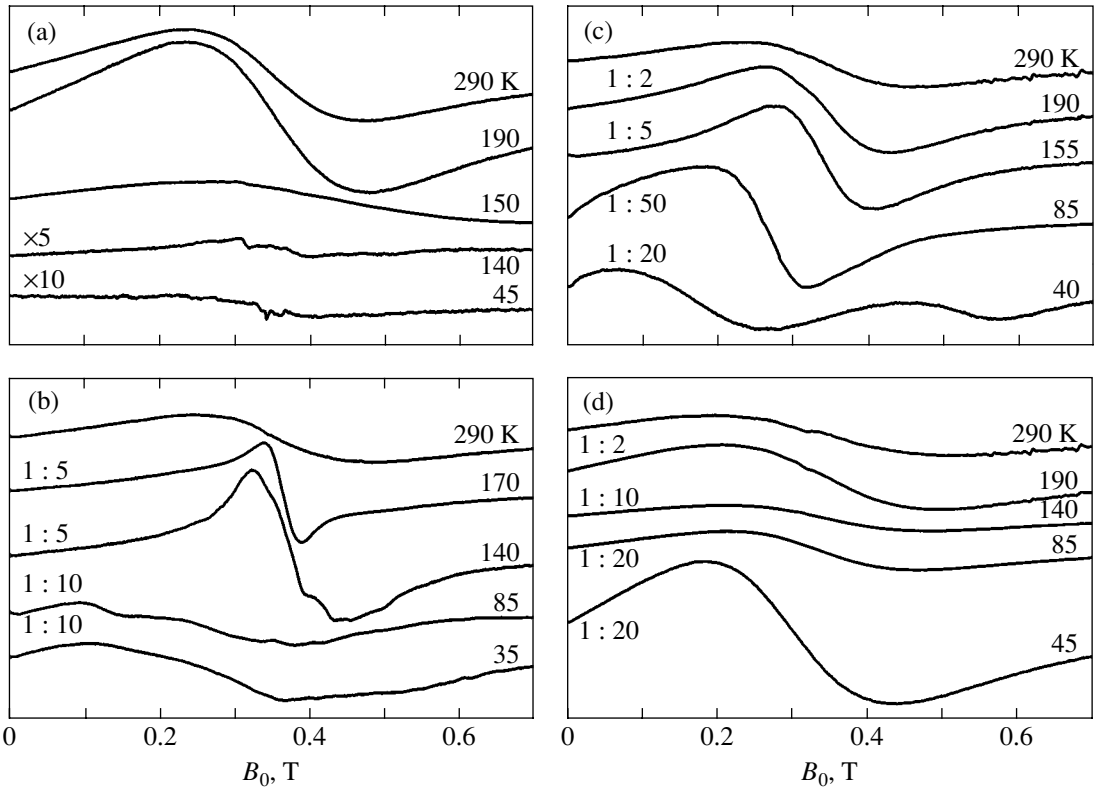


Fig. 2. EPR spectra of $LaGa_{1-x}Mn_xO_3$ crystals with $x = 1.0$ (a), 0.8 (b), 0.5 (c), and 0.2 (d) measured at various temperatures (indicated at the curves).

nals from each ion exhibit the corresponding fine structure. In addition, the unit cell of the perovskite lattice contains four positions differing in directions of the principal axes of the spin Hamiltonian. Moreover, the number of spectral lines increases by a factor of six due to the hyperfine structure arising as a result of the interaction with ^{55}Mn nuclei (nuclear spin $I = 5/2$). All these factors account for the appearance of hundreds of lines in the manganite spectrum with $x = 0.02$, presented in Fig. 1.

As the manganese content increases, the shape of the spectrum gradually simplifies. Narrow components exhibit broadening and partly merge with each other, while a characteristic signal of approximately Lorentzian shape appears at the center of gravity, grows in intensity, and becomes the predominant line in the spectrum. For $x \geq 0.2$, the area of this line determined by double integration of the experimental spectrum approximately corresponds to the nominal concentration of paramagnetic ions in the sample.

Figures 2 and 3 show the temperature variation of the EPR spectra of samples with the manganese content $x \geq 0.1$. At a sufficiently high temperature, these spectra are well approximated by one or two overlapping Lorentzian lines. A sharp decrease in the intensity of EPR absorption in $LaMnO_3$ ($x = 1$) at $T < 140$ K is related to the antiferromagnetic phase transition. The shift and distortion of lines observed on cooling below 140 K for the samples with $x = 0.8$ and below 85 K for $x = 0.5$ is evidence of a ferromagnetic ordering at these temperatures, in agreement with the data of magnetic investigations [11, 13]. No direct evidence of the magnetic ordering for $x = 0.2$ was observed in the entire temperature range studied, but an increase in the linewidth on cooling this sample below 50 K (see also Fig. 4 below) is indicative of the appearance of a short-range order.

Very interesting features were found in the EPR spectra of the samples with $x = 0.1$. As can be seen from Figs. 1 and 3a, the spectra measured at $T \geq 250$ K display a single line with $g = 2$ (signal *A*). The width of this line is much smaller than that observed for a higher manganese content. Upon cooling below 250 K, another line (signal *B*) appears and rapidly grows, whose width also rapidly increases with decreasing temperature. A comparison of the EPR spectra of samples with $x = 0.1$ cut from various parts of the initial single crystal ingot or from various crystals grown under the same nominal conditions showed that the shapes of these spectra and their temperature dependences may somewhat differ (cf. Figs. 3a and 3b). Annealing of these samples in air for 1–2 h at 800–900°C leads to significant additional broadening and decrease in intensity of the "narrow" signal *A*. Nevertheless, the main characteristic features—broadening and/or subsequent splitting on cooling—are reproduced in all samples with $x = 0.1$.

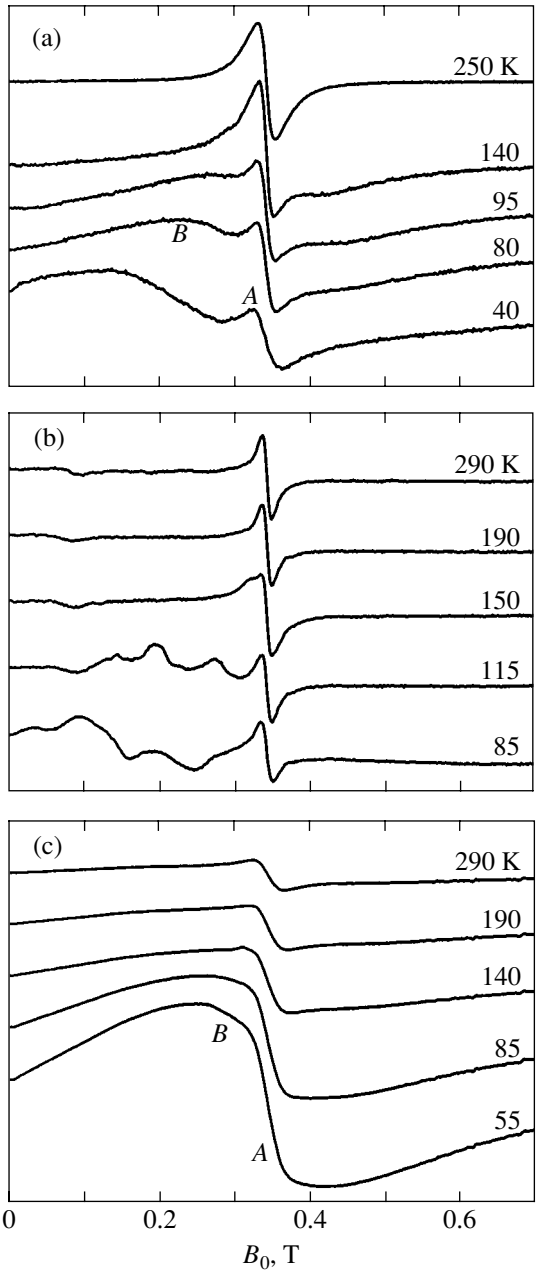


Fig. 3. EPR spectra of (a, b) two different $LaGa_{1-x}Mn_xO_3$ samples with $x = 0.1$ and (c) a strontium-substituted sample with the composition $La_{0.9}Sr_{0.1}Ga_{0.8}Mn_{0.2}O_3$ measured at various temperatures (indicated at the curves).

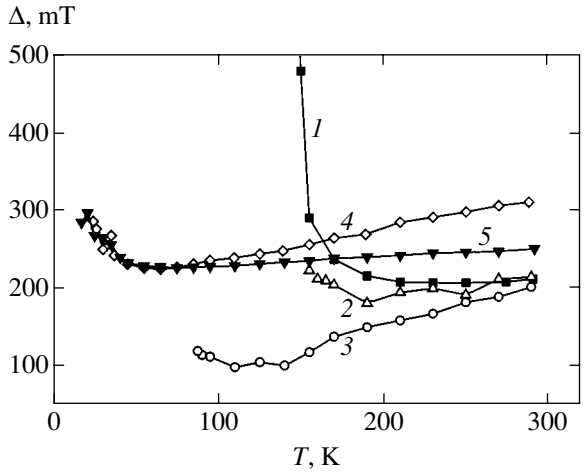


Fig. 4. Temperature dependences of the EPR linewidth in (*1–4*) $LaGa_{1-x}Mn_xO_3$ crystals with $x = 1.0$ (*1*), 0.8 (*2*), 0.5 (*3*), and 0.2 (*4*); curve (*5*) refers to a crystal with $x = 0.2$ containing 10% of strontium (curves are drawn via experimental points).

Interpretation of these data (see Section 3 below) leads to a hypothesis about the possible role of Mn^{4+} ions. Such ions can form during the synthesis of single crystals as a result of a small nonstoichiometry with respect to oxygen (one extra O^{2-} ion leads to the recharge of two Mn^{3+} to Mn^{4+}). In order to verify this assumption, we have synthesized and studied a strontium-substituted sample with the composition $La_{0.9}Sr_{0.1}Ga_{0.8}Mn_{0.2}O_3$, in which the content of Mn^{4+} ions is fixed and (according to the nominal formula) amounts to 0.1. The temperature variation of the EPR spectrum of this sample is illustrated in Fig. 3c.

Figures 4 and 5 show the temperature dependence of the halfwidth Δ of the Lorentzian EPR lines in the samples with $x \geq 0.2$ and signal *B* in one sample with $x = 0.1$. The weak monotonic broadening of the EPR lines at high temperatures, which is characteristic of the samples with $x \geq 0.2$ (Fig. 4), qualitatively agrees with the data reported for concentrated manganites [7–9, 20–22], although this dependence somewhat weakens with increasing diamagnetic dilution. A sharp broadening of the EPR line observed for $LaMnO_3$ in the vicinity of the Néel temperature agrees with the published data [7–9, 21] and is typical of the critical behavior [7, 8]. Some increase in the EPR linewidth near the temperature of ferromagnetic transition in the samples with $x = 0.8$ and 0.5 can be also considered as analogous to the effect observed in concentrated manganites [21, 22]. However, it was shown [23–26] that the EPR line broadening near T_C in the latter case exhibits an inhomogeneous character and is not related to the critical acceleration of relaxation. As for the sample with $x = 0.1$ (Fig. 5), the significant monotonic broadening of the EPR signal observed on cooling in a broad temperature range is absolutely untypical of manganites studied so far.

We have also studied the dependence of the EPR linewidth on the orientation of external magnetic field $\mathbf{B}_0$ relative to the crystal axes. The anisotropy of Δ reaches 40% for the concentrated manganite $LaMnO_3$ (see also [27]), decreases by half for $x = 0.8$, and practically disappears for $x = 0.2$.

The intensity of EPR absorption (proportional to the magnetic susceptibility χ_{EPR}) is determined by double numerical integration of the spectra or by processing of the corresponding analytically fitted Lorentzian lines. These data were calibrated using standard samples of

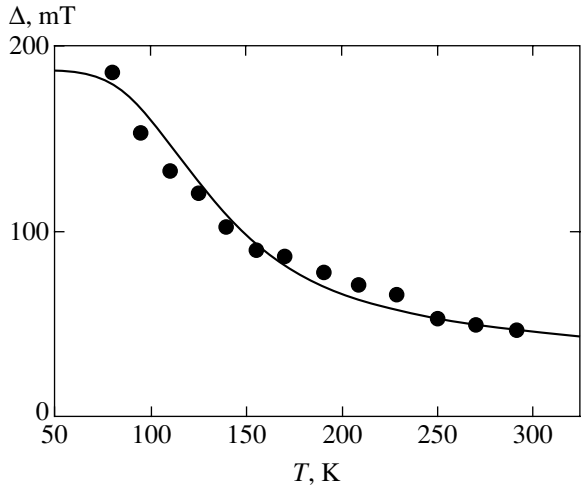


Fig. 5. Temperature dependence of the width of signal *B* in the EPR spectrum of $LaGa_{1-x}Mn_xO_3$ with $x = 0.1$. Points show the experimental data; solid curve was calculated by formulas (3) and (4) with the values of parameters indicated in the text.

MgO:Mn. Figure 6a presents the temperature dependence of $1/\chi_{EPR}$; Fig. 6b shows (in expanded temperature scale) the behavior of $\chi_{EPR}(T)$ for $x \geq 0.2$ in the region of magnetic phase transitions. Straight lines in Fig. 6a approximate the inverse susceptibility in the high-temperature limit in accordance with the Curie–Weiss law:

$$\chi_{EPR} = \frac{C}{T - \Theta},$$

where C is the paramagnetic Curie constant and Θ is the Curie–Weiss temperature corresponding to the point of intersection of a straight line in Fig. 6a with the abscissa axis. Using the measured values of susceptibilities and the Curie–Weiss law, it is possible to evaluate the number of paramagnetic Mn^{3+} ions contributing to the experimental EPR spectrum. Our calculation performed for manganite samples with $x \geq 0.2$ showed that this numbers agree to within ±10% with the nominal content of manganese, while the value calculated for the sample with $x = 0.1$ was about one-tenth of the nominal.

3. DISCUSSION AND CONCLUSIONS

The exchange-induced narrowing of the EPR linewidth has been thoroughly studied in concentrated paramagnets and, in particular, in manganites including $LaMnO_3$ [7, 8, 21, 27]. It is established that the width of the exchange-narrowed EPR line is well described by the formula [21, 28]

$$\Delta = \frac{M_2^0}{J} f(T), \qquad (1)$$

where J is the exchange integral (taking into account all neighbors of the given paramagnetic center), M_2^0 is the second spectral moment related to the single-particle spin Hamiltonian and anisotropic spin–spin interactions, and $f(T)$ is a factor taking into account a relatively weak temperature dependence determined by deviation of the magnetic susceptibility $\chi(T)$ from the Curie law [21]. It should be recalled that formula (1) is valid for $J \gg \sqrt{M_2^0}$. It was demonstrated [7, 8, 21, 27] that, in the case of concentrated manganites, the main contribution to M_2^0 is due to the fine structure of the single-par-

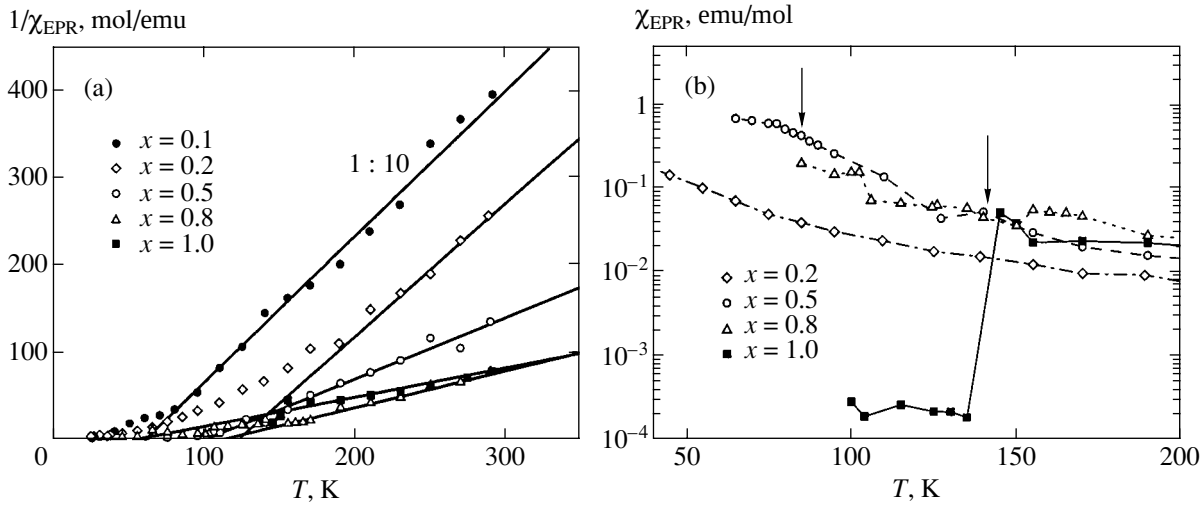


Fig. 6. The temperature dependence of the magnetic susceptibility χ_{EPR} determined from the EPR absorption in $LaGa_{1-x}Mn_xO_3$ crystals with various manganese concentrations: (a) $1/\chi_{EPR}$ values fit to the Curie–Weiss law (straight lines; data for $x = 0.1$ scaled 1 : 10); (b) χ_{EPR} in the region of magnetic phase transitions (indicated by arrows); lines are drawn via experimental points.

ticle EPR spectrum of Mn^{3+} ion described by the spin Hamiltonian

$$\hat{H} = \gamma B_0 \hat{S}_z + D\hat{S}_z^2 + E(\hat{S}_x^2 - \hat{S}_y^2), \quad (2)$$

and a somewhat lower contribution is due to the anisotropic part of the Dzialoshinsky–Moriya exchange interaction. The axial (D) and orthorhombic (E) parameters evaluated in [7, 8, 27] had comparable values on the order of 20 GHz, so that the full width of the fine structure (with allowance for $S = 2$) would be about 100 GHz (3–4 T in the magnetic field units). The EPR linewidth for $x = 1$ (see Figs. 1 and 2) is significantly smaller and obeys formula (1).

Quantitative description of the exchange-induced narrowing of the EPR linewidth in magnetically dilute crystals meets considerable difficulties. In this case, instead of the exchange integral J common for all manganese ions, we have a set of J_{ij} values scattered within a broad range as a result of the random distribution of the paramagnetic centers. The centers involved in paramagnetic clusters (i.e., possessing one or several nearest neighbors) participate to a more or less pronounced extent in the exchange interaction, while the other centers must exhibit a single-particle spectrum.

The question as to how many paramagnetic centers are necessary to form an exchange cluster possessing a spectrum approaching the many-body limit representing a single Lorentzian line is still open. It is well known that the spectra of exchange pairs, triads, and so on, up to "molecular magnets" containing Fe_8 and Mn_{12} groups, are not described by a single Lorentzian line and possess complicated anisotropic shapes (see, e.g., [29]). On the other hand, it was shown [30] that already the spectra of five- and six-member clusters in their central part (in the vicinity of $g = 2$) may approximate (with a certain error) the Lorentzian line. We may suggest that the real exchange-induced narrowing takes place only in the infinite (percolation) cluster. For a simple cubic lattice, such clusters are formed for $x = 0.31$ [31]. As can be seen from Fig. 1, a complete transition to an exchange-narrowed line characteristic of concentrated systems in our samples is observed for $x = 0.2$, which is somewhat below the percolation value. This fact may be indicative of the important role played by exchange interactions both in the first and in the second coordination sphere. It should be recalled that the values of exchange integrals for the nearest neighbors in $LaMnO_3$ amount to $J_{ab}/k = 6.6$ K (k is the Boltzmann constant) in the ab plane (ferromagnetic superexchange) and $J_c/k = -4.4$ K in the direction of c axis plane (antiferromagnetic superexchange) [9]. We failed to find reliable data on the exchange integrals for the next neighbors in manganites. Using published data for the exchange pairs in the structurally close compound $LaAlO_3$:Cr^{3+} [32], we may expect that exchange interaction in the next coordination sphere decreases by approximately one order of magnitude. It is also not excluded that Mn^{3+} ions can be inhomogeneously distributed in the volume and the percolation conditions can be realized in a part of the sample.

Using data on the temperature dependence of the magnetic susceptibility χ_{EPR} presented in Fig. 6, it is possible to trace variations in the type of dominant magnetic ordering with a decrease in the manganese content. According to these data, the Curie–Weiss temperature Θ is positive in all samples, which corresponds to the ferromagnetic exchange interaction. At the same time, the relatively small value of this temperature ($\Theta \approx 50$ K) for undoped $LaMnO_3$ ($x = 1$), which agrees with the results of static magnetic measurements [9], is unambiguously indicative of a competition between ferromagnetic interaction in the ab planes and antiferromagnetic interaction between the adjacent planes. This competition is also clearly manifested in the behavior of χ_{EPR} on approaching the Néel temperature on the side of the paramagnetic phase (see Fig. 6b). As can be seen, the initially sharp increase in the magnetic susceptibility (evidencing ferromagnetic correlations in the ab planes) changes to a sharp decrease and a nearly complete vanishing of the resonance absorption as a result of transition to the antiferromagnetic phase. No such behavior is observed in the samples with gallium partly substituted for manganese: even for $x = 0.8$, the magnetic susceptibility χ_{EPR} monotonically increases ($T_C \approx 140$ K), which is characteristic of ferromagnetic ordering observed previously in analogous manganites studied by static methods [11, 13]. It should be noted that the sample with $x = 0.8$ still retains the static orbital ordering in the ab plane and the corresponding orthorhombic structure of the O' type in this temperature interval [11, 13, 15].

The increase in χ_{EPR} characteristic of the ferromagnetic ordering is also observed for a sample with $x = 0.5$ (Fig. 6b). Judging by the EPR line shift (Fig. 2), the value of T_C for this sample is about 85 K. For a still lower manganese content ($x = 0.2$), no clear evidence of magnetic transitions in the temperature interval studied was observed. However, a significant deviation from the Curie–Weiss law at $T < 200$ K (Fig. 6a) is indicative of a considerable superparamagnetic clusterization.

Now let us consider the EPR spectrum of a sample with $x = 0.1$. According to Fig. 1, the room-temperature linewidth in this sample is significantly smaller than in the manganites with higher manganese content. At first glance, this is a paradoxical result, since we might expect that a decrease in x from 0.2 to 0.1 will not influence the M_2^0 value determined by the single-particle Hamiltonian (2). On the other hand, the exchange interaction may only decrease, which, according to formula (1), must result in a broadening, rather than narrowing, of the EPR spectrum.

In interpreting these data, it is important to take into account the temperature dependence of the EPR spectrum for $x = 0.1$ (see Figs. 3a, 3b, and 5). Since the iso-

tropic exchange interaction is temperature-independent, the broadening and/or splitting of the EPR line observed with a decrease in the temperature (or narrowing of the spectrum and disappearance of the structure on heating) can be explained only if the presence of thermoactivated internal motion in the system under consideration is assumed. Irrespective of the nature of this motion (the possibilities will be discussed below), this motion must produce averaging of the fine structure in the EPR spectrum of Mn^{3+} ions at the expense of effective decrease in one or both parameters, D and/or E, of the spin Hamiltonian. If the characteristic velocities of the internal motion (for a given manganese content) exceed a certain effective value ($\langle J \rangle$), the averaging effect will be manifested prior to the effect of exchange. In this case, M_2^0 in formula (1) should be replaced by a certain partly averaged value M_2^* ($M_2^* < M_2^0$), which decreases with increasing temperature.

A key to determining the physical nature of this effect is provided by recently published data on the narrowing of the EPR spectrum of $LaMnO_3$ observed on heating the sample to $T > T_{JT}$ [7–9]. It was found that the role of internal motion averaging the anisotropic part of the spin Hamiltonian can be played by fast reorientations of the Jahn–Teller configurations on the passage from rigid orbital ordering to the dynamic cooperative Jahn–Teller effect.

Application of this idea to the magnetically dilute case ($x = 0.1$) requires certain modification. It is obvious that, in the case of low manganese content (in particular, below the percolation threshold for exchange interactions), we may speak only of limited exchange clusters where the concept of many-body phase transition has no physical meaning. However, mutually perpendicular arrangement of the Jahn–Teller distortions (and orbitals) in the exchange cluster can be energetically favorable in this case as well [10, 11]. At very low temperatures, we may expect a static effect resembling a kind of local orbital ordering [33]. As the thermal energy kT approaches the height of the barrier separating the competing orbital orientations, thermoactivated reorientations become possible and their rate exceeds the exchange rate. This leads to an effective decrease in the fine structure parameters and eventually, in the width of the exchange-narrowed EPR linewidth.

Figure 5 shows an attempt at fitting the temperature dependence of the width Δ of signal B in a sample with $x = 0.1$ by using the Arrhenius law with a certain activation energy E_a. We assumed that the linewidth is determined by fast fluctuations of the internal field with an rms amplitude of ω_L and a correlation time

$$\tau_e = \tau_e^0 \exp\left(\frac{E_\alpha}{kT}\right), \qquad (3)$$

as well as by the additive temperature-independent term Δ_0, reflecting inhomogeneous broadening. Using the standard theory of narrowing of the magnetic resonance spectra as a result of internal motion [34], we obtain

$$\gamma_e \Delta = \left(\frac{1}{\omega_L^2 \tau_e} + \frac{1}{\gamma_e \Delta_\infty}\right)^{-1} + \gamma_e \Delta_0, \qquad (4)$$

where γ_e is the gyromagnetic ratio and Δ_∞ is the linewidth in the static limit (this term eliminates divergence for $\tau_e \longrightarrow \infty$). The result of fitting formula (4) to the experimental data for the parameters $E_a/k = 550$ K, $\Delta_0 = 27$ mT, $\Delta_\infty = 160$ mT, and $\tau_e^0 \omega_L^2 = 5 \times 10^5$ s^{-1} is shown by the solid curve in Fig. 5. Although the large number of fitting parameters and a limited interval of variation of the linewidth in Fig. 5 do not allow unambiguous quantitative conclusions to be made, a qualitative agreement with the proposed model seems to be quite reasonable. It should be emphasized that the value of E_a adopted in the above fitting procedure was independently obtained for the same samples from an analysis of the temperature dependence of the rate of nuclear spin–lattice relaxation in ^{69}Ga and ^{71}Ga isotopes [18].

The obtained estimate of the activation energy ($E_a \sim 50$ meV) agrees in the order of magnitude with the minimum height of a potential barrier characteristic of reorientations of the tetragonal Jahn–Teller configurations in the hydrated complexes of copper [35, 36]. It should be noted that this barrier, appearing due to the quadratic vibronic interactions with allowance for anharmonicity and corresponding to by-pass of the potential energy peak [36], is significantly lower than the nominal energy of the Jahn–Teller splitting of the e_g state (the authors are grateful to V.S. Vikhnin for attracting their attention to this circumstance).

An alternative hypothesis concerning the nature of internal motion in the manganite crystals under consideration is offered by the thermoactivated hopping of small Jahn–Teller polarons related to the $Mn^{3+} \longrightarrow Mn^{4+}$ recharge. Although the holes (Mn^{4+} ions) were not intentionally introduced during the synthesis of our $LaGa_{1-x}Mn_xO_3$ crystals, we cannot exclude the presence of a certain amount of these ions related to small nonstoichiometry with respect to oxygen (see, e.g., [21]) or to the disproportionation reaction [15]. It should be noted that thermoactivated conductivity discovered previously in these crystals was successfully interpreted within the framework of the polaron model [16]. However, the corresponding activation energies were about 0.5 eV, which is one order higher than the above estimate of E_a. Note also that, according to theoretical estimates [37], the hopping of small polarons must lead to accelerated spin–lattice relaxation and, hence, to broadening (rather than narrowing) of the EPR line with increasing temperature.

For additional verification of the possible role of small polarons, it is expedient to compare the EPR spectra of gallium-doped manganite with $x = 0.1$ to the data for a sample of $La_{0.9}Sr_{0.1}Ga_{0.8}Mn_{0.2}O_3$ containing not less than 10% of Mn^{4+} ions (holes). As can be seen from Figs. 3a and 3c, the EPR spectra of both samples contain two components, narrow (signal *A*) and relatively wide (signal *B*). However, the EPR spectrum of the strontium-substituted sample (in contrast to that of $LaGa_{0.9}Mn_{0.1}O_3$) exhibits no broadening (and even somewhat narrows) on cooling. The natural explanation of the "doublet" spectrum in the sample with strontium is to assume the presence of spatially separated regions characterized by different intensity of exchange interaction. The regions enriched with holes (Mn^{4+} ions) feature strong double exchange decreasing the EPR linewidth to 20–30 mT (typical of optimum-doped CMR manganites). It should be noted that similar local "droplets" were also observed in concentrated manganites [33]. The regions depleted of holes exhibit only the superexchange between Mn^{3+} ions, and the corresponding EPR linewidth is close to that observed in $LaGa_{1-x}Mn_xO_3$ with $x \geq 0.2$. Thus, the control experiment with strontium additives showed that the internal motion revealed in $LaGa_{1-x}Mn_xO_3$ can hardly be explained by the presence of Mn^{4+} ions and the related small polarons.

The ferromagnetic character of the exchange interaction confirmed by our experimental data is in agreement with a model of the correlated quasi-static Jahn–Teller regime in Mn^{3+} clusters [10, 11]. A complete theory of the magnetic resonance and relaxation for such systems has not yet been developed. We can only emphasize that the large variety of possible cluster formations in the magnetically dilute system makes the existence of a single, clearly defined E_a value hardly probable, so that the above estimates rather refer to an averaged parameter. This is confirmed by the fact that only a small part of the manganese ions in the sample contribute to the EPR line observed in the spectrum of a sample with $x = 0.1$ (see Section 2 above). We may also suggest that potential barriers for the thermoactivated reorientations of e_g orbitals are created with the participation of the structural defects acting as the pinning centers [18]. Various samples with the same nominal manganese content $x = 0.1$ exhibit a characteristic scatter (cf. Figs. 3a and 3b). Apparently, the course of clusterization of manganese ions is highly sensitive to the conditions of crystal growth (in particular, to the cooling rate), which is confirmed by the effect of additional annealing (described in Section 2) on the EPR spectrum.

In conclusion, let us briefly consider the temperature and concentration dependence of the EPR linewidth in the samples with $x \geq 0.2$ (see Fig. 4) where the exchange interaction involves virtually all manganese ions. In view of the above considerations, the observed Δ values reflect the interplay of two opposite trends. On the one hand, the growth of manganese content x leads to an increase in the average value $\langle J \rangle$ of the exchange integral (i.e., of the denominator in formula (1)). On the other hand, this is accompanied by a growing degree of orbital ordering, leading to an increase in the values of parameters D and E determining the numerator in the same formula. An analysis of our experimental data (Fig. 4) shows that the temperature dependence of the EPR linewidth (in a region sufficiently far from both T_C and T_N) for $x \geq 0.5$ agrees well with the Huber law $\Delta \propto (\chi T)^{-1}$ [21] observed in concentrated CMR manganites. However, the dependence of Δ on T for $x = 0.2$ is much weaker, which can be at least partly related to the thermoactivated decrease in M_2^* on heating. In addition, the diamagnetic dilution leads to an additional contribution of the temperature-independent inhomogeneous broadening related to disorientation of the principal axes of the spin Hamiltonian of Mn^{3+} ions. This effect is clearly manifested by a sharp decrease in anisotropic broadening of the EPR line with decreasing x (see Section 2).

To summarize, our investigation of the EPR spectra of a series of magnetically dilute single crystal samples of $LaGa_{1-x}Mn_xO_3$ manganites, as well as the previous study of the NMR and nuclear spin relaxation [18], provides a basis for speaking on the existence of an internal thermoactivated motion with an activation energy on the order of 50 meV. This motion can be considered as an intermediate stage on the passage from rigid static orbital ordering of the Jahn–Teller e_g orbitals to a dynamic disorder. The ferromagnetic character of the exchange interaction inherent in this phenomenon agrees with the theoretical model of the vibronic superexchange under conditions of correlated reorientations of mutually perpendicular e_g orbitals [10, 11]. However, exhaustive quantitative description of the proposed model requires further investigations using various experimental methods and consistent theoretical analysis of spin effects under conditions of the thermoactivated Jahn–Teller effect.

ACKNOWLEDGMENTS

The authors are grateful to G.B. Loutts (Norfolk State University, USA) for kindly providing samples for the investigations and to V.S. Vikhnin and F.S. Dzheparov for fruitful discussions.

This study was supported by the Russian Foundation for Basic Research (project no. 02-02-16219), the Program "Spin-Dependent Effects in Solids and Spintronics" of the Russian Academy of Sciences, and the NSF CREST (project HRD-9805059).

REFERENCES

1. J. M. D. Coey, M. Viret, and S. von Molnar, Adv. Phys. **48**, 167 (1999).
2. E. L. Nagaev, Phys. Rep. **346**, 387 (2001).

3. E. Dagotto, J. Hotta, and A. Moreo, Phys. Rep. **344**, 1 (2001).
4. K. I. Kugel' and D. I. Khomskiĭ, Usp. Fiz. Nauk **136**, 621 (1982) [Sov. Phys. Usp. **25**, 231 (1982)].
5. Q. Huang, A. Santoro, J. W. Lynn, *et al.*, Phys. Rev. B **55**, 14987 (1997).
6. T. Chatterji, B. Ouladdiaf, P. Mandal, *et al.*, Phys. Rev. B **66**, 054403 (2002).
7. B. I. Kochelaev, E. Shilova, J. Deisenhofer, *et al.*, Mod. Phys. Lett. B **17**, 459 (2003).
8. J. Deisenhofer, B. I. Kochelaev, E. Shilova, *et al.*, Phys. Rev. B **68**, 214427 (2003).
9. M. T. Causa, G. Alejandro, R. Zysler, *et al.*, J. Magn. Magn. Mater. **136–137**, 506 (1999).
10. J. B. Goodenough, A. Wold, R. J. Arnott, and N. Menyuk, Phys. Rev. **124**, 373 (1961).
11. J.-S. Zhou, H. Q. Yin, and J. B. Goodenough, Phys. Rev. B **63**, 184423 (2001).
12. A. I. Coldea, S. J. Blundell, I. M. Marshall, *et al.*, Phys. Rev. B **65**, 054402 (2002).
13. J. Blasco, J. Garcia, J. Campo, *et al.*, Phys. Rev. B **66**, 174431 (2002).
14. S. Hébert, C. Martin, A. Maignan, *et al.*, Phys. Rev. B **65**, 104420 (2002).
15. J.-S. Zhou and J. B. Goodenough, Phys. Rev. B **68**, 144406 (2003).
16. N. Noginova, G. B. Loutts, E. S. Gillman, *et al.*, Phys. Rev. B **63**, 174414 (2001).
17. S. M. Yusuf, M. Sahana, K. Dörr, *et al.*, Phys. Rev. B **66**, 064414 (2002).
18. N. Noginova, E. Arthur, T. Weaver, *et al.*, Phys. Rev. B **69**, 024406 (2004).
19. M. A. Noginov, N. Noginova, G. B. Loutts, and R. R. Rakhimov, in *Magnetoresistive Oxides and Related Materials*, Ed. by M. S. Rzchowski, M. Kawasaki, A. J. Millis, S. von Monlar, and M. Rajewari (MRS, Warrendale, Pennsylvania, 2000), MRS Symp. Proc., Vol. 602, p. 107.
20. M. T. Causa, M. Tovar, A. Caneiro, *et al.*, Phys. Rev. B **58**, 3233 (1998).
21. D. L. Huber, G. Alejandro, A. Caneiro, *et al.*, Phys. Rev. B **60**, 12155 (1999).
22. V. A. Ivanshin, J. Deisenhofer, H.-A. Krug von Nidda, *et al.*, Phys. Rev. B **61**, 6213 (2000).
23. F. Rivadulla, M. A. Lopez-Quintela, L. E. Hueso, *et al.*, Phys. Rev. B **60**, 11922 (1999).
24. V. A. Atsarkin, V. V. Demidov, G. A. Vasneva, and K. Conder, Phys. Rev. B **63**, 092405 (2001).
25. V. A. Atsarkin, V. V. Demidov, G. A. Vasneva, and D. G. Gotovtsev, Appl. Magn. Reson. **21**, 147 (2001).
26. V. A. Atsarkin, V. V. Demidov, F. Simon, *et al.*, J. Magn. Magn. Mater. **258–259**, 256 (2003).
27. J. Deisenhofer, M. V. Eremin, D. V. Zakharov, *et al.*, Phys. Rev. B **65**, 104440 (2002).
28. P. W. Anderson and P. R. Weiss, Rev. Mod. Phys. **25**, 269 (1953).
29. S. Hill, S. Maccagnano, K. Park, *et al.*, Phys. Rev. B **65**, 224410 (2002).
30. G. A. Korteweg and L. L. van Reijen, J. Magn. Reson. **44**, 159 (1981).
31. N. Jan and D. Stauffer, Int. J. Mod. Phys. C **9**, 341 (1998).
32. K. W. Blazey and G. Burns, Helv. Phys. Acta **37**, 638 (1964).
33. B. B. van Aken, O. D. Jurchesku, A. Meetsma, *et al.*, Phys. Rev. Lett. **90**, 066403 (2003).
34. C. P. Slichter, *Principles of Magnetic Resonance*, 3rd ed. (Springer, Berlin, 1990; Mir, Moscow, 1981), Chap. 5.
35. U. Opik and M. H. L. Pryce, Proc. R. Soc. London, Ser. A **238**, 425 (1957).
36. A. Abragam and B. Bleaney, *Electron Paramagnetic Resonance of Transition Ions* (Clarendon, Oxford, 1970; Mir, Moscow, 1973), Vol. 2.
37. A. Shengelaya, G.-M. Zhao, H. Keller, *et al.*, Phys. Rev. B **61**, 5888 (2000).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 211–216.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 239–245.

SOLIDS
Electronic Properties

On the Role of Tunneling in Metal–Semiconductor Nanocontacts

N. V. Vostokov and V. I. Shashkin

Institute for the Physics of Microstructures, Russian Academy of Sciences, Nizhni Novgorod, 603950 Russia
e-mail: vostokov@ipm.sci-nnov.ru

Received November 27, 2003

Abstract—The shape of the contact potential that arises at the interface between semiconductor and a metal nanoparticle is calculated in the approximation of complete depletion. The particle represents a sphere of radius $a \ll S$, where S is the thickness of the depletion layer in the semiconductor in the case of an infinite plane contact with the metal. A WKB approximation is applied to develop a theory of thermal-field current transfer through such a contact. It is shown that, as the radius of the metal nanoparticle decreases, the component of thermal field emission current plays an increasing role in the current transfer, while the backward current density increases and may become comparable to the density of forward current. In this case, the current–voltage characteristics (CVCs) become more symmetric. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Currently, the study of the properties of metal–semiconductor nanocontacts is of great interest [1]. This interest has been stimulated by the development of probe methods for investigating semiconductor structures, by the development and fabrication of nanosize Schottky barriers for microwave and terahertz-band applications, and by attempts to construct an artificial nonlinear medium in the form of two- and three-dimensional arrays of nanocontacts [2–13]. Here, some questions arise as to whether the description of current transfer is changed when the size of the contacts is reduced and what kind of effects should one expect in this case. There are reports that the measured height of a Schottky barrier depends on the contact size [4, 5] due to the effect of the surface bending of semiconductor bands outside the metal. In [2], the authors pointed out that the role of tunneling increases as the contact size decreases. When describing optoelectronic phenomena in the structures involving low-temperature gallium arsenide, one applies the model of a spherical contact with a Schottky barrier. The Schottky barrier in such structures arises around 2–10-nm arsenic clusters [10]. The short lifetime of photocarriers, low conductivity, high mobility of electrons, and high breakdown field allow one to apply low-temperature gallium arsenide to generate and detect terahertz-band radiation [14, 15]. Another possibility for constructing a semiconductor medium with implanted metal nanoclusters (which is fabricated by the method of metalorganic vapor-phase epitaxy) is discussed in [9].

Important factors in metal–semiconductor nanocontacts are either the existence of boundary effects or the nonzero curvature of the interface. In this case, the potential and the electric field depend on more than one coordinate, and the field magnitude near the boundary with a metal may be much greater than in the plane case. The transmission probability of the potential barrier that arises in this contact may be large, and the role of tunneling in current transfer may be essential. This paper is devoted to the discussion of the role of tunneling.

We consider a contact that arises between a metal sphere of radius a and a surrounding homogeneous semiconductor (for definiteness, of n type). Figure 1 shows a section of a metal sphere in a semiconductor by a plane passing through its center. The distribution of

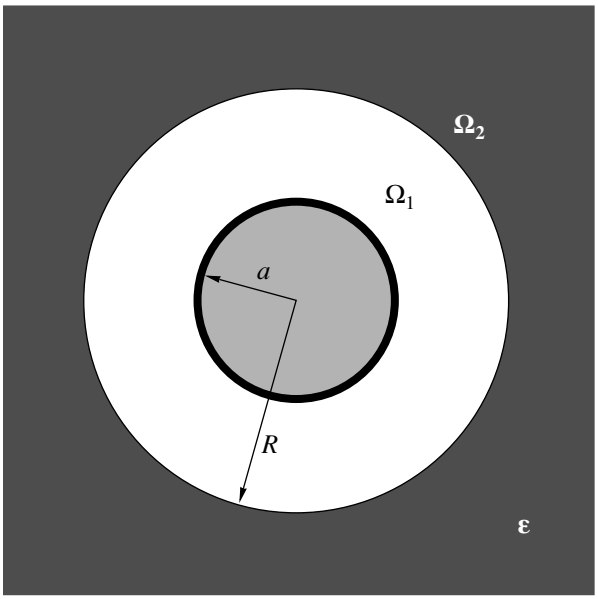


Fig. 1. Model of a nanocontact: a metal sphere of radius a in an infinite semiconductor.

the electrostatic potential φ of the contact is determined from the Poisson equation

$$\Delta\varphi = -\frac{4\pi eN}{\varepsilon}, \quad (1)$$

where N is the impurity concentration in a uniformly doped semiconductor, ε is its dielectric permittivity, and e is the electron charge. We apply the approximation of complete depletion [16], which, in view of the symmetry of the problem, leads to the following boundary conditions on two spherical surfaces, namely, on the metal surface Ω_1 and on the surface Ω_2 that bounds the complete depletion layer of the semiconductor:

$$\varphi(r)|_{\Omega_1} = 0, \quad (2)$$

$$\varphi(r)|_{\Omega_2} = u_c - u, \quad (3)$$

$$\nabla\varphi(r)|_{\Omega_2} = 0. \quad (4)$$

Here, u_c is the band bending in the semiconductor at zero voltage and u is the voltage across the contact. We assume that u_c is independent of the size and the shape of the contact. In a situation when conducting leads cannot be connected to the metal sphere surrounded by semiconductor and voltage cannot be applied to it, one can mean by u a photoelectric force, induced by illumination, that arises due to the redistribution of nonequilibrium charge carriers. This statement of the problem allows one to calculate the potential distribution around the nanocontact.

When the surface bending of the semiconductor bands at the boundary with air (vacuum) can be neglected, a solution to the problem of a nanocontact embedded into the bulk of a semiconductor with $\varepsilon \gg 1$ provides an approximate description of the electric properties of metal nanocontacts on the surface of the semiconductor [17]. A relevant example is contact atomic-force microscopy combined with simultaneous measurements of the current–voltage characteristics (CVCs) through a conducting probe and a sample.

2. POTENTIAL DISTRIBUTION AROUND A SPHERICAL NANOCONTACT

For a metal sphere of radius a embedded into a semiconductor, we will solve Eq. (1) in a spherical system of coordinates with the origin at the center of the sphere and with conditions (2)–(4). In this case, Ω_1 is a sphere of radius $r = a$ and, due to the symmetry of the problem, Ω_2 is also a sphere, but of radius $r = R$ (Fig. 1). A solution is readily obtained:

$$\varphi(r) = (u_c - u)\left[\frac{a^2}{3S^2}\left(1 - \frac{r^2}{a^2}\right) + \frac{2R^3}{3aS^2}\left(1 - \frac{a}{r}\right)\right]. \quad (5)$$

Here, $S = \sqrt{\varepsilon(u_c - u)/2\pi eN}$ is the thickness of the complete depletion layer in the semiconductor for a semi-infinite plane contact, and R is a solution to the equation $2R^3 - 3aR^2 - 3aS^2 + a^3 = 0$. The only real solution of this equation is given by

$$R = \frac{a}{2} + \frac{a^2}{2(2\sqrt{3}\sqrt{3a^2S^4 - a^4S^2} + 6aS^2 - a^3)^{1/3}} + \frac{1}{2}(2\sqrt{3}\sqrt{3a^2S^4 - a^4S^2} + 6aS^2 - a^2)^{1/3}. \quad (6)$$

Using (5) and taking into account the interaction potential of an electron with its image on the metal sphere, we obtain the following expression for the potential energy of the electron:

$$V(x) = e(u_c - u) \times \left[\frac{a^2}{3S^2}\frac{x}{a}\left(2 + \frac{x}{a}\right) - \frac{2R^3}{3aS^2}\frac{x}{x + a} - \frac{2c}{x\left(2 + \frac{x}{a}\right)}\right]. \quad (7)$$

Here, $x = r - a$ is the distance measured from the surface of the sphere and $c = e/4\varepsilon(u_c - u)$ is a parameter with a dimension of length that characterizes the potential of image forces. Under backward and small forward voltages, $c < 1$ nm in semiconductors. The third term in square brackets in (7) is essential only for $x \le c$. Since $c \ll a$, the term x/a in the denominator can be neglected. Furthermore, if we focus on the shape of the potential energy only near the metal sphere for x less than or on the order of a, then, for $a \ll S$, we can neglect the first term in square brackets in (7). This means that, in the vicinity of a sufficiently small metal sphere, one can neglect the electric field induced by the space charge in the semiconductor compared with the field induced by the surface charge of the sphere. Thus, for x less than or on the order of a, we obtain the following approximate expression for the potential energy of an electron:

$$V(x) = \mu + \Delta + eu_c - d(u)\left[g(u)\frac{x}{x + a} + \frac{c(u)}{x}\right]. \quad (8)$$

Here, the energy is measured from the bottom of the conduction band of the metal, μ is the Fermi energy in the metal, Δ is the distance from the Fermi level to the bottom of the conduction band of the semiconductor, $d(u) = e(u_c - u)$, and $g(u) = 2R^3/3aS^2$. Figure 2 gives a schematic view of the potential shape in a nanocontact (solid curve) in comparison with that for a quasi-plane case (dashed curve). It is essential that the width and the height of the barrier decrease for the nanocontact.

3. AN EXPRESSION FOR A CURRENT THROUGH A SPHERICAL NANOCONTACT

We will calculate the barrier transmission probability in a WKB approximation [18]:

$$D(E) \approx \exp[-A(E)], \tag{9}$$

where

$$A(E) = \frac{2\sqrt{2m}}{\hbar}\int_{x_1}^{x_2}\sqrt{V(x)-E}\,dx, \tag{10}$$

for energies E less than the maximum of the barrier height and x_1 and x_2 are the zeros of the radicand in (10). The integral in (10) is calculated exactly; as a result, we obtain

$$A(z) = p(z^2 + 2\beta z - 4\beta + \beta^2) f(z), \tag{11}$$

where

$$p = \frac{20\pi a\sqrt{2mdg}}{3h}, \quad z = \frac{\mu + \Delta + eu_c - E}{dg}$$

is a dimensionless energy measured down from the level $\mu + \Delta + eu_c$, the maximum of the barrier height calculated without taking the image forces into consideration; $\beta = c/ag$; and $f(z)$ is a slowly varying function of z expressed in terms of the Appel function. The formulas and the graph of $f(z)$ are presented in the Appendix. The top of the potential barrier (8) corresponds to $z = z_m = 2\sqrt{\beta} - \beta$.

Assuming that the energy distribution of electrons transmitted through the barrier region attains its maximum above the Fermi level in the semiconductor $\mu + eu$ [18] by at least several kT and that the transmission coefficient over the barrier equals unity ($A(z < z_m) = 0$), we can write out the following expression for the current density through the contact as a sum of thermionic j_{TE} (electrons with energies greater than the maximum of the potential barrier) and thermal-field j_T (electrons with energies below the maximum of the potential barrier) components:

$$j = j_{TE} + j_T = \frac{4\pi mekT}{h^3}\exp\left(-\frac{\Delta + eu_c}{kT}\right)$$

$$\times\left(\exp\left(\frac{eu}{kT}\right) - 1\right)dg\left[\int_{-\infty}^{z_m}\exp\left(z\frac{dg}{kT}\right)dz\right. \tag{12}$$

$$\left. + \int_{z_m}^{1/g}\exp\left(z\frac{dg}{kT} - A(z)\right)dz\right].$$

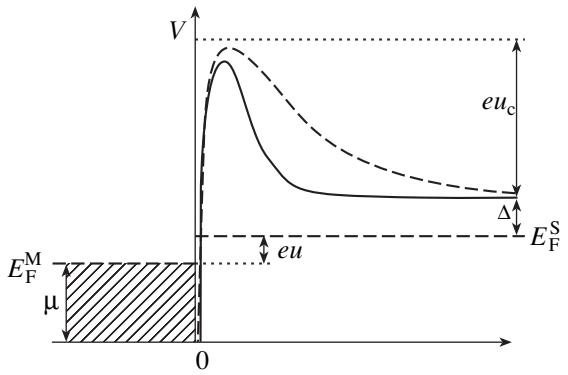


Fig. 2. The shape of potential for (solid line) a nanocontact and (dashed line) a plane case.

Here, m is the electron effective mass in the semiconductor and T is absolute temperature. The forward voltage across the contact should not be too high in order not to violate the inequality $a \ll S$, which is the applicability condition for (8). According to the Appendix, $f(z)$ is a slowly varying function. Therefore, when integrating (12), we can assume that this function is a constant equal to its value at the maximum of the integrand. In view of the fact that f = const, the integrand attains its maximum at $z = z_0$, where z_0 is a solution to the equation

$$z_0 = \frac{dg}{2pkTf(z_0)} - \beta.$$

Taking this fact into consideration, from (12) we obtain

$$j = \frac{4\pi mekT}{h^3}\exp\left(-\frac{\Delta + eu_c}{kT}\right)\left(\exp\left(\frac{eu}{kT}\right) - 1\right)dg$$

$$\times\left\{\frac{kT}{dg}\exp\left(z_m\frac{dg}{kT}\right) + \frac{\sqrt{\pi}}{2\sqrt{pf}}\right.$$

$$\times\left[\mathrm{Erf}\left[\sqrt{pf}\left(\frac{1}{g} - z_0\right)\right] + \mathrm{Erf}[\sqrt{pf}(z_0 - z_m)]\right] \tag{13}$$

$$\left.\times\exp\left(\frac{1}{4pf}\left(\frac{dg}{kT}\right)^2 + 4pf\beta - \frac{dg}{kT}\beta\right)\right\},$$

where $f = f(z_0)$.

We can find a domain of parameters where formula (13) is justified. For example, we construct the boundaries of the required domain for GaAs ($eu_c + \Delta$ = 0.7 eV, ε = 13.1). In Fig. 3, this domain is bounded by three lines: *1*, *2*, and *3*. On line *1*, the thickness of the depletion layer around the metal is equal to the radius of the metal sphere ($l = R - a = a$). Below this line (where $a \ll l$), the impurity concentration can be ignored.

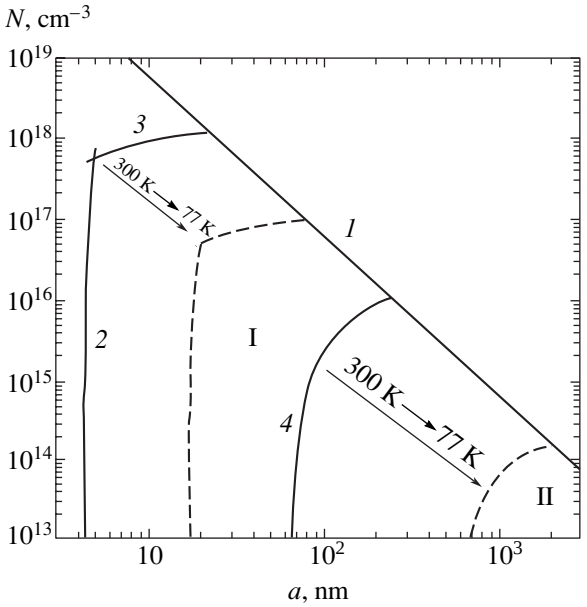


Fig. 3. The plane of parameters N, a with the domain (bounded by lines *1*, *2*, and *3*) in which the approximation used is applicable; line *4* is the equality line of the thermal-field and thermoelectric components of current. The arrows indicate the displacement of boundaries *2*, *3*, and *4* as temperature reduces to 77 K.

Conversely, above this line (where $a \gg l$), the curvature of the contact and edge effects can be ignored; this is a plane case approximation. The approximate expression (8) for the potential is valid below line *1*. On line *2*, the function under the sign of the second integral in (12) attains its maximum at $z_0 = 0.7$. To the right of this curve, $z_0 < 0.7$, and we can assume that $f(z) = \text{const}$ here. On line *3*, the maximum of the energy distribution of electrons transmitted through the barrier coincides with the Fermi energy; below this line, the maximum lies above the Fermi level, and here formula (12) is justified. The boundaries described above are constructed for zero bias voltage ($u = 0$); for simplicity, we assume that $\Delta = 0$ and is independent of temperature and the concentration of the doping impurity. On line *4*, $\lim_{u \to 0} j_T/j_{TE} = 1$, which corresponds to the equality of the two terms in curly brackets in (13). This line divides the domain of parameters where formula (13) holds into two parts: to the left from line *4* (parameter domain I), the current density is largely determined by the thermal-field component, whereas, to the right from line *4* (parameter domain II), by the thermionic component. Boundaries *2*, *3*, and *4*, which are indicated by solid lines, are constructed for temperature $T = 300$ K. As the temperature decreases to 77 K, these boundaries move to the positions shown by dashed lines. The transfer of thermionic current through nanocontacts is described in [19]. The main difference from the plane case lies in the stronger decrease of the barrier height due to the image forces and weak dependence of all effects on the doping level. Below, we consider in detail the nanocontacts where an essential role is played by thermal-field current transfer.

4. DISCUSSION OF THE TUNNELING EFFECTS IN NANOCONTACTS

Consider a nanocontact with the parameters belonging to domain I (Fig. 3) and far from its boundaries. In this case, we can neglect the thermionic component of current and simplify the expression for the current. Using Eq. (13); substituting p, β, d, and c into this equation; assuming that the sum of the probability integrals is equal to two, $f \approx 1$, and $g \approx 1$; and neglecting the small term $\sim\beta$, we obtain the following approximate expression for the current density through the nanocontact:

$$j \approx j_T \approx \frac{4\pi mekT}{h^3}\exp\left(-\frac{\Delta + eu_c}{kT}\right)$$

$$\times\left(\exp\left(\frac{eu}{kT}\right) - 1\right)\frac{\sqrt{3}he^{3/4}(u_c - u)^{3/4}}{2\sqrt{5}(2m)^{1/4}\sqrt{a}} \qquad (14)$$

$$\times \exp\left(\frac{3he^{3/2}(u_c - u)^{3/2}}{80\pi\sqrt{2m}a(kT)^2}\right).$$

This expression retains its strong temperature dependence, which is characteristic of thermal field emission processes. The possibility to neglect the terms $\sim\beta$ in the exponent in (13) for nanocontacts with radii less than or on the order of 50 nm is associated with the reduced role of the image forces in the barrier transmission probability.

Let us write out a current calculated by formula (14) for small forward voltages, $3kT/e < u < u_c$, in the form characteristic of Schottky barriers [16]:

$$j = AT^2\exp\left[-\frac{\Phi_{eff}}{kT}\right]\exp\left[\frac{eu}{nkT}\right]. \qquad (15)$$

Of course, the values of Φ_{eff} and n depend on temperature and the radius of the nanocontact and weakly depend on the forward voltage. For a fixed temperature, this formula allows one to evaluate the "apparent" effective height of the barrier Φ_{eff} and the nonideality factor n; for example, for GaAs, $eu_c + \Delta = 0.7$ eV, $\varepsilon = 13.1$, and an impurity concentration of $N = 10^{16}$ cm^{-3}. Figure 4 represents Φ_{eff} and n as a function of the radius of the nanocontact for two values of temperature: $T = 300$ K (solid curves) and $T = 77$ K (dashed curves). One can see that, as the nanocluster size and temperature decrease, the effective height of the barrier decreases, while the nonideality factor increases and may become greater than two. The squares and circles in this figure represent the values of Φ_{eff} and n calculated by exact

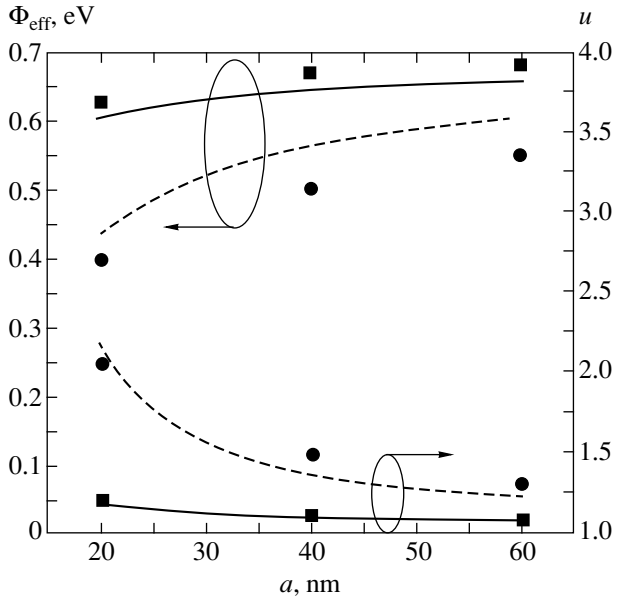


Fig. 4. The effective height of the barrier Φ_{eff} and the nonideality factor n versus the nanocontact radius.

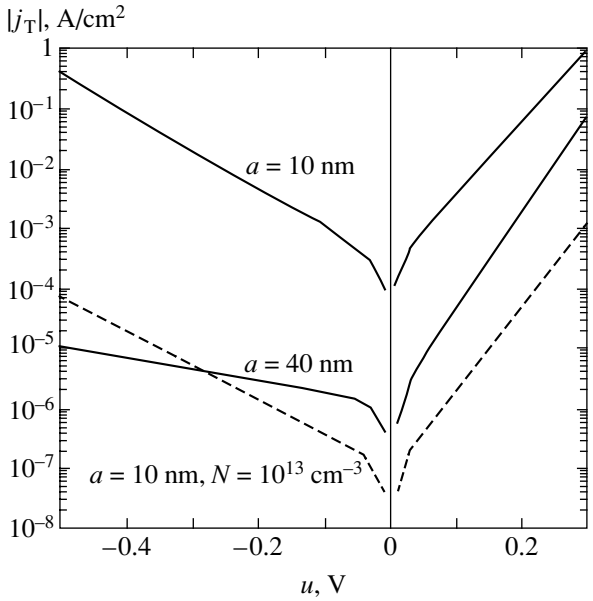


Fig. 5. The CVCs of nanocontacts with two different radii at temperature 300 K.

CVCs obtained numerically at T = 300 K (squares) and T = 77 K (circles), respectively. One can see that even the oversimplified formula (14) gives results close to exact.

The strong dependence of the barrier height on the backward voltage leads to the increase in the backward current as the nanocontact radius decreases. The estimates by formula (14) show that the forward and backward currents become equal at a voltage of $\pm u$, when the nanocontact radius decreases to 5 nm for T = 300 K and 21 nm for T = 77 K. Thus, the CVCs become more symmetric as the radius decreases. Figure 5 shows the CVCs of nanocontacts with two different radii at T = 300 K. One can see that the backward current density exhibits exponential behavior. The dashed lines represent the CVC of a nanocontact with a low concentration of doping impurity $N = 10^{13}$ cm^{-3}; the densities of the forward and backward currents remain comparable.

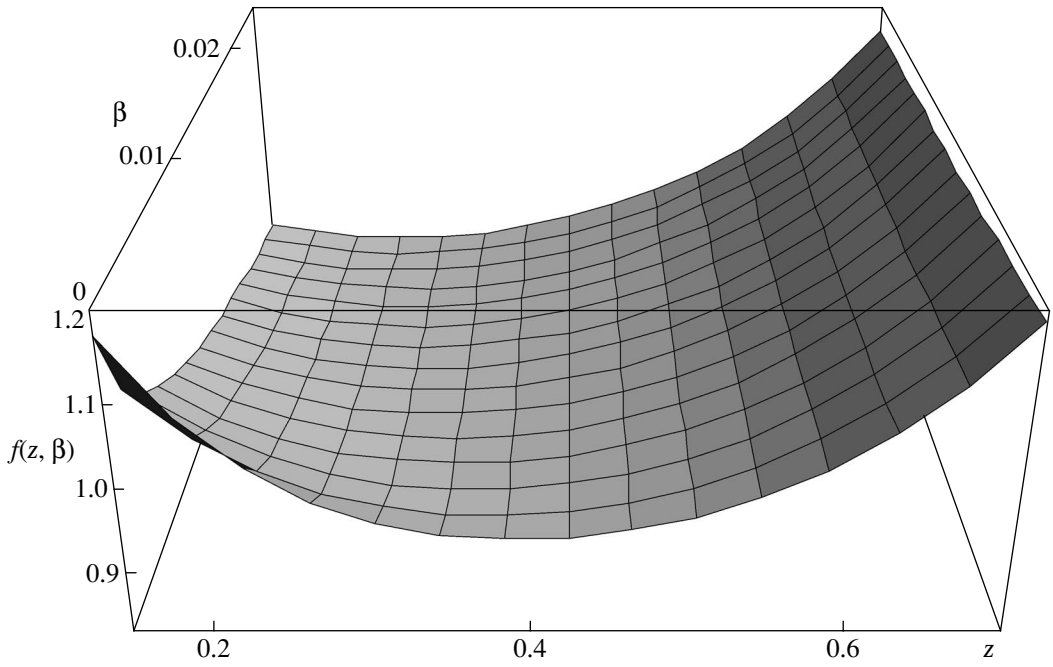


Fig. 6. The graph of the function $f(z, \beta)$.

5. CONCLUSIONS

A theory of thermal-field current transfer through a metal–semiconductor nanocontact has been constructed on the basis of a WKB approximation with regard to the reduction of the barrier height due to the image forces. Analytic expressions have been obtained for the CVC in terms of the Appel function. A domain of parameters has been constructed in which the current transfer through a barrier is determined by the thermal-field component. It has been shown that both the forward and the backward current densities in the case of thermal-field current transfer mainly exhibit exponential behavior. As the nanocontact radius decreases to $a \sim 5$ nm at $T = 300$ K, the backward current increases and may become comparable to the forward current. In this case, the CVCs become more symmetric. The effective height of the barrier substantially reduces as the nanocontact radius decreases, by tenth fractions of an electron-volt. All the phenomena listed above occur even for an arbitrarily small doping level of the semiconductor.

APPENDIX

$$f(z) = \frac{3\pi}{40(1-z)^{3/2}}\frac{a}{x_2}\left(1+\frac{a}{x_2}\right)^{-1/2}$$

$$\times F_1\left[\frac{3}{2}, \frac{1}{2}, \frac{1}{2}, 3; 1-\frac{x_1}{x_2}, \frac{1-x_1/x_2}{1+a/x_2}\right],$$

where

$$\frac{x_1}{x_2} = \frac{z-\beta-\sqrt{(z-\beta)^2-4\beta(1-z)}}{z-\beta+\sqrt{(z-\beta)^2-4\beta(1-z)}},$$

$$\frac{a}{x_2} = \frac{2(1-z)}{z-\beta+\sqrt{(z-\beta)^2-4\beta(1-z)}},$$

and F_1 is the Appel function [20]. It turns out that $f(z)$ is a slowly varying function of z in the relevant range of energies below the top of the barrier for any reasonable values of the parameter β. This result is corroborated by the fact that

$$\frac{\frac{d}{dz}[f(z)](z^2+2\beta z-4\beta+\beta^2)}{f(z)\frac{d}{dz}[z^2+2\beta z-4\beta+\beta^2]} \ll 1$$

for any values of $z_m \le z \le 0.7$. In addition, $f(z, \beta) \approx 1$ in this range of parameters, which is illustrated in Fig. 6.

REFERENCES

1. Takhee Lee, Jia Liu, Nien-Po Chen, *et al.*, J. Nanopart. Res. **2**, 345 (2000).
2. G. D. J. Smit, S. Rogge, and T. M. Klapwijk, Appl. Phys. Lett. **81**, 3852 (2002).
3. G. D. J. Smit, M. G. Flokstra, S. Rogge, and T. M. Klapwijk, Microelectron. Eng. **64**, 429 (2002).
4. Hideki Hasegawa, Taketomo Sato, and Chinami Kaneshiro, J. Vac. Sci. Technol. B **17**, 1856 (1999).
5. Hideki Hasegawa, Surf. Rev. Lett. **7**, 583 (2000).
6. Ichiro Tanaka, I. Kamiya, and H. Sakaki, J. Cryst. Growth **201**–**202**, 1194 (1999).
7. Kian-Giap Gan, Jin-Wei Shi, Yen-Hung Chen, *et al.*, Appl. Phys. Lett. **80**, 4054 (2002).
8. I. Aberg, K. Deppert, M. H. Magnusson, *et al.*, Appl. Phys. Lett. **80**, 2976 (2002).
9. V. Shashkin, V. Daniltsev, M. Drozdov, Yu. Drozdov, A. Murel, N. Vostokov, and S. Rushworth, in *Booklet of Extended Abstracts of 10th European Workshop on Metalorganic Vapour Phase Epitaxy* (Lecce, Italy, 2003), p. 79.
10. A. C. Warren, J. M. Woodall, J. L. Freeout, *et al.*, Appl. Phys. Lett. **57**, 1331 (1990).
11. C. Kadow, A. W. Jackson, and A. C. Gossard, Appl. Phys. Lett. **76**, 3510 (2000).
12. Keiichiro Kumada, Tomohiro Murata, Yutaka Ohno, *et al.*, Jpn. J. Appl. Phys. **42**, 2250 (2003).
13. A. Dorn, M. Peter, S. Kicin, *et al.*, Appl. Phys. Lett. **82**, 2631 (2003).
14. I. S. Gregory, C. Baker, W. R. Tribe, *et al.*, Appl. Phys. Lett. **83**, 4199 (2003).
15. C. Baker, I. S. Gregory, W. R. Tribe, *et al.*, Appl. Phys. Lett. **83**, 4113 (2003).
16. S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. (Wiley, New York, 1981; Mir, Moscow, 1984).
17. C. Donolato, J. Appl. Phys. **95**, 2184 (2004).
18. *Tunneling Phenomena in Solids*, Ed by E. Burstein and S. Lundqvist (Plenum, New York, 1969; Mir, Moscow, 1973).
19. N. V. Vostokov and V. I. Shashkin, in *Abstracts of VI Russian Conference on the Physics of Semiconductors* (St. Petersburg, 2003), p. 257.
20. *Higher Transcendental Functions (Bateman Manuscript Project)*, Ed. by A. Erdelyi (McGraw-Hill, New York, 1953; Nauka, Moscow, 1965), Vol. 1.

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 217–223.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 246–252.

SOLIDS
Electronic Properties

Thermoelectric Effects in Layered Conductors in a Strong Magnetic Field

O. V. Kirichenko[a], D. Krstovska[a,b], and V. G. Peschansky[a,c]

[a]*Verkin Institute for Low Temperature Physics and Engineering, National Academy of Sciences of Ukraine, Kharkov, 61103 Ukraine*

[b]*Faculty of Natural Sciences and Mathematics, Physical Institute, Cyril–Methodium University, Skopje, 9100 Republic of Macedonija*

[c]*Karazin Kharkov National University, Kharkov, 61077 Ukraine*

e-mail: vpeschansky@ilt.kharkov.ua

Received December 25, 2003

Abstract—Thermoelectric effects are investigated theoretically in layered conductors with a quasi-two-dimensional electron energy spectrum of arbitrary type in a strong magnetic field. It is shown that, at temperatures sufficiently low for quantization of the orbital motion of charge carriers in a magnetic field to be required, there exist giant quantum oscillations of the thermoelectric field. Thermoelectric emf is studied as a function of the orientation of the magnetic field with respect to the layers; experimental investigation of this function allows one to determine the velocity distribution of conduction electrons on the Fermi surface. © *2004 MAIK "Nauka/Interperiodica".*

The kinetic characteristics and the magnetic permeability of degenerate conductors in a strong magnetic field **H** are quite sensitive to the form of the energy spectrum of electrons. This fact enables one to design a reliable spectroscopic method for restoring the Fermi surface from experimental data [1]. Experimental investigation of galvanomagnetic and thermoelectric phenomena in a strong magnetic field provides a very convenient method for investigating the energy spectrum of charge carriers, in particular, for determining the topological structure of the Fermi surface of conduction electrons [2–4].

In this paper, we consider thermoelectric phenomena in layered conductors with a quasi-two-dimensional electron energy spectrum of arbitrary type that are placed in a strong magnetic field where an electron can perform many rotations with frequency Ω during its mean lifetime τ.

A quasi-two-dimensional character of the energy spectrum of charge carriers is inherent in virtually all layered conductors of organic origin, a large family of manganites, graphite, and other materials. In these layered structures, the ratio of electric conductivity along the layers, $\sigma_\parallel$, to the conductivity across the layers, $\sigma_\perp$, is on the order of 10^3–10^5 in the absence of a magnetic field. Such sharp anisotropy of electric conductivity of metal-like layered conductors is associated with the sharp anisotropy of the velocities of conduction electrons $\mathbf{v} = \partial\varepsilon(\mathbf{p})/\partial\mathbf{p}$ on the Fermi surface $\varepsilon(\mathbf{p}) = \varepsilon_F$. The maximal velocity of electrons with the Fermi energy ε_F along the normal **n** to the layers, $v_z = \mathbf{v} \cdot \mathbf{n}$, is much lower than the characteristic velocity of electrons along the layers, v_F, so that $v_z \leq \eta v_F \ll v_F$. The quasi-two-dimensionality parameter η of the energy spectrum of electrons is the smaller, the smaller the ratio $\sigma_\perp/\sigma_\parallel$, and can be defined by the formula

$$\eta^2 = \sigma_\perp/\sigma_\parallel \tag{1}$$

to a sufficiently high degree of accuracy. The weak dependence of the energy of charge carriers on the momentum projection $p_z = \mathbf{p} \cdot \mathbf{n}$ can be represented by a rapidly converging series

$$\varepsilon(\mathbf{p}) = \sum_{k=0}^{\infty} \varepsilon_k(p_x, p_y)\cos\left\{\frac{akp_z}{\hbar} + \alpha_k(p_x, p_y)\right\}, \tag{2}$$

$$\alpha_k(p_x, p_y) = -\alpha_k(-p_x, -p_y). \tag{3}$$

We assume that the functions $\varepsilon_k(p_x, p_y)$ are arbitrary and appreciably decrease as the number k increases.

In the quasi-classical approximation, when the interval $\Delta\varepsilon$ between quantized energy levels of conduction electrons is much less than ε_F, it suffices to apply the rule of quantizing areas,

$$S(\varepsilon, p_H) = \frac{2\pi\hbar eH}{c}\left(n + \frac{1}{2}\right), \tag{4}$$

in order to determine the energy spectrum of electrons; here, n are nonnegative integers, and $S(\varepsilon, p_H)$ is the area

of the section of the isoenergetic surface $\varepsilon(\mathbf{p}) = \varepsilon$ by the plane $p_H = \text{const}$.

The linear response of the electron system to an external action in the form of an electric field **E** or a temperature gradient ∇T,

$$j_i = \sigma_{ij}E_j - \alpha_{ij}\frac{\partial T}{\partial x_j}, \tag{5}$$

$$q_i = \beta_{ij}E_j - \kappa_{ij}\frac{\partial T}{\partial x_j} \tag{6}$$

should be determined from the solution of the kinetic equation for the statistical operator $\hat{f}$.

There is a large number of theoretical studies devoted to the derivation of quantum kinetic equations for the statistical operator $\hat{f}$ in the case of spatially inhomogeneous systems describing, in particular, thermomagnetic phenomena in a quantizing magnetic field (see, for example, [5–8] and survey [9], which contains an extensive bibliography). Here, the main difficulty is associated with the fact that, in addition to dynamic forces, which can be included in the Hamiltonian of conduction electrons, one should correctly take into consideration other forces of different natures in the kinetic equation, such as the gradients of the chemical potential μ and temperature, which cannot be included in the Hamiltonian of the electron system.

In spatially inhomogeneous systems of charge carriers, the conduction current density **j** differs from the total current $\mathrm{Sp}\{e\hat{\mathbf{v}}\hat{f}\}$ by the vector $c\mathrm{curl}\mathbf{M}$ [10],

$$\mathbf{j} = \mathrm{Sp}\{e\hat{\mathbf{v}}\hat{f}\} - c\mathrm{curl}\mathbf{M}, \tag{7}$$

and, in the expression for the heat flux due to charge particles,

$$\mathbf{q} = \mathrm{Sp}\{\hat{\mathbf{v}}(\varepsilon - \mu)\hat{f}\} + \mathbf{q}_M(\mathbf{r}), \tag{8}$$

one should eliminate the magnetic energy flux $\mathbf{q}_M(\mathbf{r})$, which contains, in addition to the Poynting vector, the derivatives of the magnetic moment $\mathbf{M}(\mathbf{r})$ with respect to **r**. The consideration of the magnetization of the electron system guarantees the fulfillment of the Einstein formula and the Onsager symmetry principle for kinetic coefficients, which follows from the maximum principle for entropy in the equilibrium state [6, 9].

Let us represent the statistical operator $\hat{f}$ as $\hat{f} = \hat{f}_0 + \hat{f}_1 + \hat{f}_2$, where $\hat{f}_0$ is the equilibrium statistical operator whose diagonal matrix elements f_0^{nn} are equal to the Fermi distribution function $f_0\{\varepsilon_n(p_H)\}$ and the operators $\hat{f}_1$ and $\hat{f}_2$ describe the perturbations of the electron system by the electric field and the temperature gradient, respectively. In a linear approximation with respect to a small perturbation of the system of conduction electrons, the kinetic equation is rewritten as [9, 11]

$$\frac{i}{\hbar}(\varepsilon_n - \varepsilon_{n'})f_1^{nn'} + \hat{W}_{nn'}\{\hat{f}_1\} = e\mathbf{E}\cdot\mathbf{v}_{nn'}\frac{f_0(\varepsilon_n) - f_0(\varepsilon_{n'})}{\varepsilon_n - \varepsilon_{n'}}, \tag{9}$$

$$\frac{i}{\hbar}(\varepsilon_n - \varepsilon_{n'})f_2^{nn'} + \hat{W}_{nn'}\{\hat{f}_2\} = \mathbf{v}_{nn'}\nabla T\frac{\mu - \varepsilon_{n'}}{T}\frac{\partial f_0(\varepsilon_{n'})}{\partial\varepsilon_{n'}}, \tag{10}$$

where $\mathbf{v}_{nn'}$ are the matrix elements of the velocity operator $\hat{\mathbf{v}}$ of conduction electrons and $\hat{W}\{\hat{f}_1\}$ and $\hat{W}\{\hat{f}_2\}$ are linear operators that describe the momentum and energy relaxation of charge carriers, respectively. The momentum (τ_p) and energy (τ_ε) relaxation times at temperature T below the Debye point T_D are essentially different in sufficiently perfect samples, where the charge carriers are scattered mainly by lattice vibrations. In this case, the system of equations (9), (10) should be complemented by the kinetic equation for nonequilibrium phonons. However, a strong magnetic field ($\Omega\tau \gg 1$) can usually be created at liquid helium temperature, where it is more important to take into account the scattering of conduction electrons by impurity atoms and other crystal defects in the collision integral. In this case, the relaxation times τ_p and τ_ε are on the same order of magnitude; therefore, if one does not distinguish between these times, the kinetic coefficients that relate the electric current density **j** and the heat flux **q** to the electric field and the temperature gradient satisfy the Kelvin–Onsager formulas

$$\alpha_{ij} = T\beta_{ij}$$

and the Wiedemann–Franz relations

$$\kappa_{ij} = \frac{\pi^2 T}{3e^2}\sigma_{ij}.$$

The thermoelectric field due to the temperature gradient,

$$E_i = \rho_{il}\alpha_{lj}\frac{\partial T}{\partial x_j}, \tag{11}$$

where ρ_{ij} is the electric resistivity tensor, inverse to the tensor of electric conductivity σ_{ij}, is quite sensitive to the form of the energy spectrum of electrons.

Consider a thermoelectric effect in layered conductors placed in a strong magnetic field $\mathbf{H} = (0, H\sin\vartheta, H\cos\vartheta)$ in which the interval between quantized levels is much greater than the width of the levels $\hbar/\tau$ but is

much less than $\eta\varepsilon_F$. In this case, for a given Fermi energy, there are sufficiently many levels with a discrete momentum projection p_H onto the direction of the magnetic field and the periodic dependence of thermoelectric coefficients on $1/H$ has a harmonic form. In the inverse limit case, when $\hbar\Omega \geq \eta\varepsilon_F$, the quantum oscillations of thermoelectric coefficients, just like the magnetoresistance of layered organic conductors [12], exhibits complicated behavior.

The Fermi surface $\varepsilon(\mathbf{p}) = \varepsilon_F$ of quasi-two-dimensional conductors represents an open surface with small corrugation along the axis p_z; this surface may be multisheeted and may consist of topologically different elements in the form of cylinders and planes that are weakly corrugated along the axis p_z. The Shubnikov–de Haas quantum oscillations of magnetoresistance, which have been observed nearly in all layered degenerate conductors, provide evidence for the fact that at least one sheet of the Fermi surface represents a weakly corrugated cylinder. We will assume that the Fermi surface of a layered conductor does not have any sheets other than a weakly corrugated cylinder. The Fermi surface of most organic conductors based on tetrathiafulvalene, restored by the measurements of the galvanomagnetic characteristics (see, for example, the surveys [13–15] and the references therein), is exactly of this type.

If the temperature gradient is orthogonal to the magnetic field vector, then the thermoelectric field is mainly directed along the temperature gradient. In particular, when $\partial T/\partial y = \partial T/\partial z = 0$, the asymptotic expression for the thermoelectric field E_x for $\gamma \ll 1$ and $\eta \ll 1$,

$$E_x = \rho_{xy}\alpha_{yx}\partial T/\partial x, \tag{12}$$

depends only on the "Hall" components of the tensors σ_{ik} and α_{ik}, which are different from zero in the collision-free limit ($\tau = \infty$).

The quantum oscillations of $\rho_{xy} = 1/\sigma_{xy}$ occur only in the higher order terms in the power series expansion in the parameter γ because the asymptote of σ_{xy} is the same in the classical and quantum considerations [11],

$$\sigma_{xy} = \frac{N(\mu)ec}{H\cos\vartheta} = \frac{Nec}{H\cos\vartheta}, \tag{13}$$

where $N(\mu)$ is the number of electron states with energy less than or equal to μ.

The conservation of the number N of charge carriers in a unit volume,

$$N(\mu) = \frac{eH}{c(2\pi\hbar)^2} \times \sum_{\pm}\sum_{n=0}^{\infty}\int dp_H \frac{1}{1+\exp\{[\varepsilon_n(p_H)-\mu_\pm]/T\}} = N \tag{14}$$

is guaranteed by the dependence of the chemical potential on the magnitude of the magnetic field. For convenience, we ascribe the spin splitting of energy levels to the chemical potential,

$$\mu_\pm = \mu \pm \frac{e\hbar H}{2mc},$$

so that $\varepsilon_n(p_H)$ should be determined from (4).

Let us apply the Poisson formula

$$\sum_{n=0}^{\infty}\phi_n = \int_{-1/2}^{\infty} dn\phi(n)\sum_{k=-\infty}^{\infty}\exp(2\pi ikn) \tag{15}$$

and, using the relation

$$dn = d\varepsilon\frac{c}{2\pi\hbar eH}\frac{\partial S}{\partial\varepsilon},$$

replace in (15) the integration with respect to n by integration with respect to energy. Then, after straightforward calculations, we obtain an oscillating (as $1/H$) correction to the chemical potential,

$$\frac{\mu_{osc}}{\mu} = \frac{2}{N(2\pi\hbar)^3}\sum_{k=1}^{\infty}(-1)^k\left(\frac{e\hbar H}{kc}\right)^{3/2}\sum_e\left|\frac{1}{2\pi}\frac{\partial^2 S_e}{\partial p_H^2}\right|^{-1/2} \times\frac{ku}{\sinh(ku)}\sin\left(\frac{kcS_e}{e\hbar H}+\frac{\pi}{4}s\right)\cos\frac{\pi km^*}{m}, \tag{16}$$

where

$$m^* = \frac{1}{2\pi}\frac{\partial S}{\partial\varepsilon}$$

is the cyclotron effective mass of conduction electrons, m is the free electron mass,

$$u = 2\pi^2 T/\hbar\Omega,$$

and

$$s = \operatorname{sgn}\frac{\partial^2 S_e}{\partial p_H^2}.$$

The main contribution to the quantum oscillations of μ for $\hbar\Omega \ll \eta\varepsilon_F$ is made by a small part of the conduction electrons in the neighborhood of the extremal section S_e of the Fermi surface; therefore, the summation in (16) should be performed over electron states with all possible sections S_e.

At sufficiently low temperatures $T \ll \hbar\Omega$, the amplitude of oscillations $\Delta\mu_{osc}$, which is proportional to $\hbar\Omega(\hbar\Omega/\eta\mu)^{1/2}$, is much greater than the part of the chemical potential that monotonically varies with magnetic field, $\Delta\mu_{mon} \approx (\hbar\Omega)^2/\mu$; however, this amplitude is

substantially less than the amplitude of the Shubnikov–de Haas magnetoresistance oscillations and other kinetic coefficients; therefore, in all the formulas below, we omit the dependence of μ on H.

Compared with σ_{xy}, the Hall thermoelectric coefficient α_{xy} is more sensitive to variations in the magnetic field. Even the rough estimate

$$\alpha_{xy} = -\alpha_{yx} = \frac{\pi^2 cT}{3H\cos\vartheta}\frac{\partial N(\mu)}{\partial\mu} \tag{17}$$

points to a strong oscillating dependence of α_{yx} on $1/H$ because the density of states $\nu(\mu) = \partial N(\mu)/\partial\mu$ of conduction electrons in a quantizing magnetic field has a singularity that is periodic in $1/H$.

The oscillating part of α_{yx},

$$\alpha_{yx} = \frac{e^2 H}{c(2\pi\hbar)^2}\sum_{\pm}\sum_{n=0}^{\infty}\int dp_{\mathrm{H}} \times \frac{v_y^{nn'}v_x^{n'n}[\varepsilon_n(p_{\mathrm{H}})-\mu_{\pm}]\exp[\varepsilon_n(p_{\mathrm{H}})-\mu_{\pm}]}{T^2\{1+\exp[\varepsilon_n(p_{\mathrm{H}})-\mu_{\pm}]/T\}^2}, \tag{18}$$

can be calculated as usual, by applying the Poisson formula and replacing integration with respect to n by integration with respect to energy. Then, to calculate $\alpha_{xy}^{\mathrm{osc}}$, it is convenient to analytically continue the integrand

$$\alpha_{yx}^{\mathrm{osc}} = -\frac{c}{H(2\pi\hbar)^3}2\mathrm{Re}\sum_{k=1}^{\infty}(-1)^k\sum_{\pm}\int dp_{\mathrm{H}} \times\int d\xi\xi\frac{S(\mu_{\pm}+T\xi, p_{\mathrm{H}})}{[1+\exp\xi][1+\exp(-\xi)]} \times\exp\left\{\frac{ikcS(\mu_{\pm}+T\xi, p_{\mathrm{H}})}{e\hbar H}\right\}, \tag{19}$$

where $\xi = (\varepsilon-\mu_{\pm})/T$, to the region of complex values of energy and apply the residue theorem at the poles of the function $\partial f_0(\varepsilon)/\partial\varepsilon$ for $\varepsilon_q = \mu + \pi iT(2q+1)$, where q is a nonnegative integer.

Straightforward calculations yield the following expression for $\alpha_{yx}^{\mathrm{osc}}$:

$$\alpha_{yx}^{\mathrm{osc}} = \frac{2c}{H(2\pi\hbar)^3}\sum_{k=1}^{\infty}(-1)^k\chi(ku)\left|\frac{2\pi e\hbar H}{kc}\right|^{1/2} \times\sum_e S_e\left|\frac{\partial^2 S_e}{\partial p_{\mathrm{H}}^2}\right|^{-1/2}\sin\left(\frac{kcS_e}{e\hbar H}+\frac{\pi}{4}s\right)\cos\frac{\pi km^*}{m}, \tag{20}$$

where

$$\chi(ku) = \frac{\partial}{\partial u}\left\{\frac{u}{\sinh(ku)}\right\} = \frac{\sinh(ku)-ku\cosh(ku)}{\sinh^2(ku)}.$$

At sufficiently low temperatures, when $2\pi^2 T \ll \hbar\Omega$, the amplitude $\alpha_{yx}^{\mathrm{osc}}$ is proportional to T, just as the part of thermoelectric field that is independent of the magnetic filed,

$$E_x = \frac{\pi m^*\cos\vartheta}{3Nea\hbar^2} \times T\left\{1+\frac{aA}{T(m^*)^{1/2}\hbar}\sum_{k=1}^{\infty}(-1)^k\chi(ku)\left|\frac{\hbar\Omega}{\eta k}\right|^{1/2} \times\sum_e S_e\sin\left(\frac{kcS_e}{e\hbar H}+\frac{\pi}{4}s\right)\cos\frac{\pi km^*}{m}\right\}\frac{\partial T}{\partial x}, \tag{21}$$

but is greater than $\alpha_{yx}^{\mathrm{mon}}$ by a factor of $(\mu/\hbar\Omega\eta)^{1/2}$. In a sufficiently strong magnetic field, when $\hbar\Omega$ is comparable with $\eta\mu$, nearly all conduction electrons are involved in quantum oscillations, and $\alpha_{yx}^{\mathrm{osc}}$ is greater than $\alpha_{yx}^{\mathrm{mon}}$ by a factor of $\mu/\hbar\Omega$. The coefficient A in (21) is on the order of unity and depends on the specific form of the energy spectrum of electrons.

As temperature increases, the oscillation amplitude of thermoelectric emf exponentially decays for $T \geq \hbar\Omega$ according to

$$\frac{2\pi^2 T}{\hbar\Omega}\exp\left(-\frac{2\pi^2 T}{\hbar\Omega}\right);$$

i.e., the temperature dependence of this amplitude is the same as that of the Shubnikov–de Haas magnetoresistance oscillations in degenerate conductors.

To determine the asymptotic values of α_{yy}, α_{yz}, and α_{zz} in a quantizing magnetic field, it suffices to know only the diagonal matrix components of the operator $\hat{f}_2$. At sufficiently low temperatures, where the dominant scattering mechanism is elastic scattering of charge carriers by impurity atoms, one can take into account the collision integral in the τ approximation to a sufficiently high degree of accuracy. After simple calculations, we obtain the following asymptotic expression for α_{yy} and α_{zz} in a strong magnetic field:

$$\alpha_{yy} = \alpha_{zz}\tan^2\vartheta = \frac{2e}{(2\pi\hbar)^3}\sum_{k=-\infty}^{\infty}\int d\varepsilon\frac{\mu-\varepsilon}{T}\frac{\partial f_0(\varepsilon)}{\partial\varepsilon}\tau_{\varepsilon} \times\int dp_{\mathrm{H}}\left\{\frac{\partial S}{\partial p_{\mathrm{H}}}\right\}^2\frac{\sin^2\vartheta}{\partial S/\partial\varepsilon}(-1)^k\exp\left\{\frac{ikcS(\varepsilon, p_{\mathrm{H}})}{eH\hbar}\right\}. \tag{22}$$

Here, we used the formula

$$\bar{v}_y = \bar{v}_z \tan\vartheta = \bar{v}_H \sin\vartheta = \frac{\partial S/\partial p_H}{\partial S/\partial\varepsilon}\sin\vartheta, \quad (23)$$

where $\bar{v}_H$ is the drift velocity of electrons along the magnetic field averaged over the states on the electron orbit $\varepsilon = \text{const}$, $p_H = \text{const}$, and the summation over n is replaced, as usual, by the integration with respect to the energy of charge carriers by the Poisson formula.

When studying quantum oscillation phenomena, it is very important to take into account the quantum oscillations of the relaxation time τ_{osc} of conduction electrons that arise during the summation over the electron states in the income term $\hat{W}$ of the collision integral. This fact was first pointed out by Adams and Holstein [16]. In the case of scattering of charge carriers by impurity atoms and crystal defects with short-range potential, Kosevich and Andreev [17] calculated the oscillatory dependence on $1/H$ of the collision integral, obtained by the Bogolyubov method, in the Born approximation. Later, this problem was studied by many authors; in particular, the quantum oscillations of the scattering amplitude of electrons by impurities in layered organic conductors were calculated in [12, 18–20]. The most clear derivation of τ_{osc} is presented in Abrikosov's monograph ([21, Subsection 11.1, p. 177]). At sufficiently low temperatures $T \ll \hbar\Omega \ll \eta\varepsilon_F$, the relaxation times τ_p and τ_ε in the Born approximation are given by

$$\frac{1}{\tau_p(\varepsilon)} = \frac{1}{\tau_\varepsilon(\varepsilon)} = \frac{1}{\tau_0}(1 + \Delta_{osc}), \quad (24)$$

$$\Delta_{osc} = \left(\frac{e\hbar H}{m^* c\varepsilon}\right)^{1/2} \sum_e \left|\frac{\partial^2 S_e}{\partial p_H^2}\right|^{-1/2} g,$$

$$g = \sum_k a_k (-1)^k k^{1/2} \frac{ku}{\sinh(ku)} \times \cos\left\{\frac{kcS_e}{e\hbar H} + \frac{\pi}{4}s\right\}\cos\frac{\pi k m^*}{m}.$$

Here, a_k are numerical coefficients that depend on the specific form of the dispersion law of charge carriers, and τ_0 is the mean free path time of charge carriers in the absence of magnetic field.

The main contribution to the oscillating $1/H$ part of the tensor component α_{zz},

$$\alpha_{zz}^{osc} = \frac{2e}{(2\pi\hbar)^3}\int d\varepsilon \frac{\varepsilon - \mu}{T}\frac{\partial f_0(\varepsilon)}{\partial\varepsilon} \times \left\{\int dp_H \left\{\frac{\partial S}{\partial p_H}\right\}^2 \frac{\cos^2\vartheta}{\partial S/\partial\varepsilon}\tau_{osc} + 2\mathrm{Re}\sum_{k=1}^{\infty}(-1)^{k-1}\tau_0 \quad (25)$$

$$\times \int dp_H \left\{\frac{\partial S}{\partial p_H}\right\}^2 \frac{\cos^2\vartheta}{\partial S/\partial\varepsilon}\exp\left(\frac{ikcS(\varepsilon, p_H)}{eH\hbar}\right)\right\}$$

is made by the first term, which is associated with the oscillatory dependence of the mean free path time on $1/H$,

$$\tau_{osc} = \frac{\tau_0\Delta_{osc}}{1 + \Delta_{osc}},$$

which is greater than the contribution of the terms in (25) with nonzero k to the thermoelectric field oscillations by a factor of $\varepsilon_F/\hbar\Omega$. After integrating with respect to ε in (25), one can easily verify that the oscillation amplitude of thermoelectric emf at low temperatures is greater than the thermoelectric field for $H = 0$ by a factor of $(\eta\varepsilon_F/\hbar\Omega)^{1/2}$.

In the above formulas for the oscillating part of the thermoelectric field, we omitted the Dingle factor $I_D = \exp(-\gamma)$, which takes into account the broadening of quantized levels due to the scattering of charge carriers [22, 23]; for $\gamma \ll 1$, this factor is close to unity. As the angle ϑ increases, the cross-section area of the Fermi surface and the cyclotron effective mass of charge carriers increase. For $\cos\vartheta \le \gamma_0 = 1/(\Omega_0\tau) \ll 1$, conduction electrons can hardly make a full rotation in a magnetic field, so that the Dingle factor becomes substantially less than unity. Here, Ω_0 is the rotation frequency of conduction electrons at $\vartheta = 0$. As ϑ approaches $\pi/2$, the amplitude of quantum oscillations of the kinetic coefficients exponentially decays. At $\vartheta = \pi/2$, there is a small part of closed sections of the Fermi surface in the form of a weakly corrugated cylinder about which charge carriers can rotate at frequency $\Omega \le \Omega_0\eta^{1/2}$. In a magnetic field directed along the layers, quantum oscillations of thermoelectric emf are only possible when $\gamma_0 \ll \eta^{1/2}$; these oscillations are formed by the charge carriers from the neighborhood of a self-intersecting orbit. According to Azbel' [24], the amplitude of such oscillations is very small.

Giant quantum oscillations of thermoelectric field versus $1/H$, whose amplitude is much greater than the part of the thermoelectric emf that monotonically varies with magnetic field, may occur only for $\cos\vartheta \gg \gamma_0$.

When the magnetic field substantially deviates from the temperature gradient, which is directed along the normal to the layers, for certain values of the angle ϑ, narrow minima appear in the diagrams of σ_{zz} and α_{zz}

plotted versus ϑ; these minima appear not only in σ_{zz}^{mon} and α_{zz}^{mon},

$$\sigma_{zz}^{\text{mon}} = \frac{ae^2\tau m^*\cos\vartheta}{2\pi\hbar^4}\sum_{n=1}^{\infty} n^2 I_n^2(\tan\vartheta), \tag{26}$$

$$I_n(\tan\vartheta) = \frac{1}{2\pi}\int_0^{2\pi} d\phi\varepsilon_n(\phi)\cos\left\{anp_y(\phi)\frac{\tan\vartheta}{\hbar}\right\}, \tag{27}$$

$$\alpha_{zz}^{\text{mon}} = \frac{\pi^2 T}{3e}\frac{\partial}{\partial\mu}\sigma_{zz}^{\text{mon}}(\mu), \tag{28}$$

but also in the amplitude of quantum oscillations of σ_{zz}^{osc} and α_{zz}^{osc}.

When $\gamma_0 \ll \cos\vartheta \ll 1$, these minima are periodic with period

$$\Delta(\tan\vartheta) = 2\pi\hbar/aD_{\text{p}}, \tag{29}$$

where D_{p} is the diameter of the Fermi surface along the axis p_y [25–27].

When $\tan\vartheta \gg 1$, the integrand in (27) rapidly oscillates as ϕ varies, and the main contribution to the integral is made by small neighborhoods of the stationary-phase points, where v_x vanishes [27, 28]. The functions $\varepsilon_n(\phi, p_{\text{H}})$ and $p_y(\phi, p_{\text{H}})$ weakly depend on the projection of momentum onto the direction of the magnetic field; therefore, this dependence need not be taken into account in the asymptotic expressions for σ_{zz}^{mon} and α_{zz}^{mon} with respect to the parameter η in (26)–(28). When $T \geq \hbar\Omega$, the effect of quantum oscillations of α_{zz} can also be neglected; therefore, simple calculations yield the following asymptotic expression for the thermoelectric field along the normal to the layers for $\gamma_0 \ll \gamma_0\tan\vartheta \ll 1$:

$$E_z = \frac{2\pi^2 aT}{3e\hbar v_\perp\cos\vartheta}\Phi(\beta)\frac{\partial T}{\partial z}, \tag{30}$$

$$\Phi(\beta) = \frac{\sin\beta - 4\delta_2\cos 2\beta - 9\delta_3\sin 3\beta + 16\delta_4\cos 4\beta}{(1-\cos\beta) + \delta_2(1-\sin 2\beta) + \delta_3(1+\cos 3\beta) + \delta_4(1+\sin 4\beta)}, \tag{31}$$

where the diameter of the Fermi surface D_{p} is related to β by the formula

$$\frac{aD_{\text{p}}}{\hbar}\tan\vartheta = \beta + 2\pi q - \frac{\pi}{2}. \tag{32}$$

Here, $v_\perp$ is the velocity of conduction electrons in the plane of layers at the stationary-phase point, where $v_x(0) = v_x(\pi) = 0$, and $\delta_n = n\varepsilon_n^2/\varepsilon_1^2$ also contains the functions $\varepsilon_n(\phi)$ only at the stationary-phase points $\varepsilon_n(0) = \varepsilon_n(\pi) = \varepsilon_n$. Since ε_n decreases rapidly as the number n increases and $\delta_{n+1} \ll \delta_n$, we retain only the first four terms in expression (26).

For small $\beta = \beta_0$, the magnetoresistance to the current flowing across the layers has a sharp maximum, which repeats periodically with period (29) as the tilt angle of the magnetic field is changed, and $\Phi(\beta_0) = -2$ exactly up to small terms proportional to δ_3 and δ_4.

At low temperatures, $T \ll \hbar\Omega$, when, conversely, α_{zz}^{osc} is much greater than α_{zz}^{mon}, the thermoelectric field is proportional to $\Phi(\beta)$ as before, and it becomes possible to determine the velocity distribution of the charge carriers with the Fermi energy more accurately.

Thus, by measuring the thermoelectric field E_z along the normal to the layers for $\gamma_0 \ll \gamma_0\tan\vartheta \ll 1$ and for the orientations of the magnetic field $\mathbf{H} = (H\sin\vartheta\cos\varphi, H\sin\vartheta\sin\varphi, H\cos\vartheta)$ when the magnetoresistance ρ_{zz} has a narrow maximum, one can determine the velocity distribution $v_\perp(\phi)$ of charge carriers on the Fermi surface to a sufficiently high degree of accuracy.

ACKNOWLEDGMENTS

One of the authors (V.G.P.) acknowledges the support of INTAS (grant no. 01-0791).

REFERENCES

1. I. M. Lifshits, M. Ya. Azbel', and M. I. Kaganov, *Electron Theory of Metals* (Nauka, Moscow, 1971; Consultants Bureau, New York, 1973).
2. I. M. Lifshits and V. G. Peschanskiĭ, Zh. Éksp. Teor. Fiz. **35**, 1251 (1958) [Sov. Phys. JETP **8**, 875 (1958)].
3. I. M. Lifshits and V. G. Peschanskiĭ, Zh. Éksp. Teor. Fiz. **38**, 188 (1960) [Sov. Phys. JETP **11**, 137 (1960)].
4. Yu. A. Bychkov, L. É. Gurevich, and G. N. Nedlin, Zh. Éksp. Teor. Fiz. **37**, 534 (1959) [Sov. Phys. JETP **10**, 377 (1959)].
5. P. S. Zyryanov and V. P. Silin, Zh. Éksp. Teor. Fiz. **46**, 537 (1964) [Sov. Phys. JETP **19**, 366 (1964)].
6. A. I. Akhiezer, V. G. Bar'yakhtar, and S. V. Peletminskiĭ, Zh. Éksp. Teor. Fiz. **48**, 204 (1965) [Sov. Phys. JETP **21**, 136 (1965)].

7. P. S. Zyryanov and V. I. Okulov, Fiz. Tverd. Tela (Leningrad) **7**, 2666 (1965) [Sov. Phys. Solid State **7**, 2157 (1965)].
8. A. I. Ansel'm, Yu. N. Obraztsov, and R. T. Tarkhanyan, Fiz. Tverd. Tela (Leningrad) **7**, 2837 (1965) [Sov. Phys. Solid State **7**, 2293 (1965)].
9. P. S. Zyryanov and G. I. Guseva, Usp. Fiz. Nauk **95**, 565 (1968) [Sov. Phys. Usp. **11**, 538 (1969)].
10. L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Fizmatgiz, Moscow, 1959; Pergamon, Oxford, 1960).
11. I. M. Lifshits, Zh. Éksp. Teor. Fiz. **32**, 1509 (1957) [Sov. Phys. JETP **5**, 1227 (1957)].
12. T. Champel and V. P. Mineev, Phys. Rev. B **66**, 195111 (2002).
13. J. Wosnitza, *Fermi Surfaces of Low-Dimensional Organic Metals and Superconductors* (Springer, Berlin, 1996), Springer Tracts Mod. Phys.
14. M. V. Kartsovnik and V. N. Laukhin, J. Phys. (France) **6**, 1753 (1996).
15. J. Singelton, Rep. Prog. Phys. **63**, 1111 (2000).
16. E. Adams and T. Holstein, J. Phys. Chem. Solids **10**, 254 (1959).
17. A. M. Kosevich and V. V. Andreev, Zh. Éksp. Teor. Fiz. **38**, 882 (1960) [Sov. Phys. JETP **11**, 631 (1960)].
18. V. M. Gvozdikov, Fiz. Nizk. Temp. **27**, 956 (2001) [Low Temp. Phys. **27**, 704 (2001)].
19. P. Grigoriev, M. V. Kartsovnik, W. Biberacher, *et al.*, Phys. Rev. B **65**, 060403 (2002).
20. M. V. Kartsovnik, P. D. Grigoriev, W. Biberacher, *et al.*, Phys. Rev. Lett. **89**, 126802 (2002).
21. A. A. Abrikosov, *Fundamentals of the Theory of Metals* (Nauka, Moscow, 1987; North-Holland, Amsterdam, 1988).
22. R. B. Dingle, Proc. R. Soc. London, Ser. A **211**, 517 (1952).
23. Yu. A. Bychkov, Zh. Éksp. Teor. Fiz. **39**, 1401 (1960) [Sov. Phys. JETP **12**, 971 (1960)].
24. M. Ya. Azbel', Zh. Éksp. Teor. Fiz. **39**, 878 (1960) [Sov. Phys. JETP **12**, 608 (1960)]; Zh. Éksp. Teor. Fiz. **39**, 1276 (1960) [Sov. Phys. JETP **12**, 891 (1960)].
25. K. Yamaji, J. Phys. Soc. Jpn. **58**, 1520 (1989).
26. R. Yagi, Y. Iye, T. Osada, and S. Kagoshima, J. Phys. Soc. Jpn. **59**, 3069 (1990).
27. V. G. Peschansky, J. A. Roldan Lopez, and Toyi Gnado Yao, J. Phys. I **1**, 1469 (1991).
28. V. G. Peschansky, Phys. Rep. **299**, 305 (1997).

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 28–36.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 36–46.

ATOMS, SPECTRA, RADIATION

Phase-Matched Third-Harmonic Generation in Anisotropically Nanostructured Silicon

S. V. Zabotnov[a], S. O. Konorov[a], L. A. Golovan'[a], A. B. Fedotov[a], A. M. Zheltikov[a], V. Yu. Timoshenko[a], P. K. Kashkarov[a], and H. Zhang[b]

[a]*Moscow State University, Vorob'evy gory, Moscow, 119992 Russia*
e-mail: s_zabotnov@mail.ru
[b]*Department of Physics, Peking University, 100871 Beijing, China*

Received December 17, 2003

Abstract—A frequency-tunable laser system based on an optical parametric oscillator is used to implement phase-matched third-harmonic generation in porous silicon films with a strong form birefringence. Phase matching, orientation dependences, and the behavior of the third harmonic as a function of the thickness of an absorbing film are confirmed by calculations and linear-optical measurements. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Nanostructured silicon is an interesting object in the physics of low-dimensional systems. This material provides intense luminescence in the visible [1], features photoinduced bleaching [2] and saturable absorption [3], and displays a high spectral sensitivity to the adsorption of active molecules of various gases [4], allowing the creation of compact components for optics, opto-electronics, and biomedical applications.

Nanostructured silicon is usually produced via electrochemical etching of single-crystal Si wafers. This process gives rise to a qualitatively new material—porous silicon (p-Si), with the sizes of silicon nanocrystal remainders and pore diameters ranging from a few up to tens of nanometers, depending on the specific resistance of the substrate and etching conditions. The porosity and the thickness of p-Si layers can be broadly tuned by varying the etching current density and the etching time [5].

Due to the smallness of nanocrystals and pores, optical properties of p-Si can be described in terms of an effective dielectric function [5]. With a considerable fraction of the material removed, the effective dielectric function may substantially differ from the dielectric function of crystal silicon.

Recent studies [6–9] have shown that p-Si may display a strong optical anisotropy (birefringence), which is not observed for crystal silicon, belonging to the cubic syngony. The origin of birefringence in p-Si is the anisotropy of pore growth rate along different crystallographic directions in the process of electrochemical etching [8], which induces a form anisotropy of silicon remainders in p-Si and, hence, form birefringence. In particular, p-Si layers grown from single-crystal silicon wafers with a (110)-oriented surface feature properties of a negative uniaxial crystal with an optical axis oriented along the [001] crystallographic direction and lying in the plane of the sample [9]. Fourier-transform infrared measurements of birefringence for p-Si films have shown that the difference of refractive indices for the ordinary and extraordinary waves in the transparency region substantially exceeds the refractive-index variation related to the normal material dispersion [9]. This effect suggests interesting solutions for nonlinear optics, offering the ways to phase-match three- and four-wave mixing. In contrast to its porous counterpart, single-crystal silicon displays normal dispersion with no birefringence in the visible and near-infrared ranges, leaving no way for phase-matched nonlinear-optical to interact.

Earlier studies have demonstrated that p-Si photonic crystals and microcavities [10, 11], fabricated by modulating etching current densities, can radically enhance nonlinear-optical processes. Strong birefringence of p-Si was employed to phase-match second-harmonic generation [12]. The third harmonic generated in mesoporous silicon with typical nanocrystal sizes on the order of 10 nm has been recently shown [13] to substantially exceed the third-harmonic response of both crystal silicon and porous silicon with smaller sizes of nanocrystals.

The purpose of this work is to investigate phase-matching solutions for third-harmonic generation (THG) in optically anisotropic p-Si films.

2. EXPERIMENTAL

Films of porous silicon were fabricated via the electrochemical etching of (110)-oriented single-crys-

tal silicon wafers with a specific resistance of 1.5 and 3 mΩ cm in a 1 : 1 ethanol solution of HF (48%). The etching current densities were 50 mA/cm^2 and 50 and 100 mA/cm^2 for 1.5- and 3-mΩ cm wafers, respectively. The choice of these regimes of etching was based on the results of earlier studies of optical properties of birefringent p-Si [9], which showed that these parameters are best suited for the fabrication of p-Si films with the required refractive indices. The etching time ranged from 2.5 up to 20 min, leading to variations in layer thicknesses in the range from 10 up to 80 μm. The film was cleaved from the substrate by abruptly increasing the current density up to 700 mA/cm^2 within a time interval of several seconds. Layer thicknesses were measured with the use of an optical microscope.

The films fabricated of single-crystal silicon wafers with a specific resistance of 1.5 mΩ cm by applying an etching current with a density of 50 mA/cm^2 will be referred to as samples *A*. The films fabricated of single-crystal silicon wafers with a specific resistance of 3 mΩ cm by applying etching currents with densities of 100 and 50 mA/cm^2 will be referred to as samples *B* and *C* (Table 1).

Transmission spectra of thin p-Si films were measured in the visible and near-infrared regions, within the range of wavelengths from 0.47 to 1.6 μm, by using an incandescence lamp, an MDR-12 monochromator, silicon (within the wavelength range of 0.47 to 1.0 μm) and germanium (for wavelengths from 1.0 up to 1.6 μm) photodiodes, and an automated detection system. A Perkin Elmer Spectrum RXI FT-IR spectrometer was employed to measure infrared spectra within the range of wavelengths from 1.6 up to 8.0 μm. Dispersion dependences of the refractive indices for the ordinary and extraordinary waves were calculated from the interference fringes observed in transmission spectra of thin films. With etching conditions kept constant, the birefringence remained unchanged from sample to sample with an accuracy of about 5%.

Experiments on THG were performed with the use of a Solar Laser Systems source, consisting of a master Nd:YAG oscillator, a nonlinear crystal for the frequency tripling of 1.06-μm radiation, and an optical parametric oscillator (OPO), allowing the wavelength of the idler signal to be smoothly tuned within the range of 1 to 2.1 μm. This idler signal was used as a pump field in our THG experiments. Longer pump wavelengths in our studies corresponded to a weaker absorption of the third harmonic in p-Si samples.

The laser pulses had a duration of about 3 ns and were generated at a repetition rate of 20 Hz. The maximum total energy of the idler and signal waves was equal to 30 mJ. The third harmonic was generated in p-Si films by the pump field normally incident on the film surface. The third harmonic, transmitted through a Glan prism and an MDR-6 monochromator, was detected with a Hamamatsu photomultiplier. Porous silicon films were placed in a holder. Rotating the holder around the axis of pump radiation propagation, we were able to change the angle between the polarization vector of the pump field and the optical axis of the sample, which lay in the plane of the sample and corresponding to the [001] crystallographic direction.

Table 1. Specific resistances of single-crystal Si wafers and etching current densities used for the fabrication of different types of porous silicon films for experiments on third-harmonic generation. The birefringence values $\Delta n = n_o - n_e$ are presented for $\lambda = 2$ μm

Film type	Specific resistance of a substrate, mΩ cm	Etching current density, mA/cm^2	Δn ($\lambda = 2$ μm)
A	1.5	50	0.16
B	3	100	0.14
C	3	50	0.09

3. ANALYSIS OF PHASE-MATCHING STRATEGIES FOR THIRD-HARMONIC GENERATION IN OPTICALLY ANISOTROPIC FILMS OF POROUS SILICON

Phase matching in harmonic generation in optically anisotropic materials is controlled by the form of the nonlinear susceptibility tensor and the magnitude of birefringence. The first factor determines the possibility of implementing a specific type of interaction between ordinary and extraordinary waves. The second factor controls the phase mismatch for a given type of interaction. This gives us a clue to analyze phase-matching solutions for THG in p-Si samples. Such an analysis is presented below in this section.

Crystal silicon, which belongs to the *m*3*m* point-group symmetry, is a centrosymmetric material. Its nonlinear-susceptibility tensor components $\chi^{(3)}_{ijkl}$ with an odd number of identical indices vanish. Since silicon nanocrystals constituting p-Si inherit the arrangement of atoms characteristic of crystal Si [1], p-Si can be considered as a centrosymmetric material. Four-photon processes (including THG) involving an odd number of photons with the same polarization are forbidden in such materials. Following the earlier qualitative analysis of symmetry elements in birefringent p-Si [14], we assign p-Si to the 4/*mmm* point-group symmetry. The nonlinear susceptibility tensor for THG in p-Si is then written as [15]

$$\chi^{(3)}_{ijkl}(3\omega;\,\omega,\,\omega,\,\omega) = \begin{pmatrix} \chi_{1111} & \chi_{1122} & \chi_{1133} & 0 & 0 & 0 & 0 & 0 & 0 \\ \chi_{1122} & \chi_{1111} & \chi_{1133} & 0 & 0 & 0 & 0 & 0 & 0 \\ \chi_{3311} & \chi_{3311} & \chi_{3333} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \chi_{1133} & 0 & 0 & \chi_{1133} & 0 & 0 \\ 0 & 0 & 0 & 0 & \chi_{3311} & 0 & 0 & \chi_{3311} & 0 \\ 0 & 0 & 0 & 0 & 0 & \chi_{1122} & 0 & 0 & \chi_{1122} \\ 0 & 0 & 0 & \chi_{3311} & 0 & 0 & \chi_{3311} & 0 & 0 \\ 0 & 0 & 0 & 0 & \chi_{1133} & 0 & 0 & \chi_{1133} & 0 \\ 0 & 0 & 0 & 0 & 0 & \chi_{1122} & 0 & 0 & \chi_{1122} \end{pmatrix}. \quad (1)$$

This tensor features only five independent nonvanishing components, which noticeably simplifies the analysis of the nonlinear-optical response.

The slowly varying envelope approximation yields the following equation for the projections of the third-harmonic electric field **E** [16]:

$$\frac{\partial E^i(3\omega)}{\partial y} = i\frac{6\pi\omega}{cn(3\omega)}\chi^{(3)}_{ijkl}E^j(\omega)E^k(\omega)E^l(\omega) \times \exp(i(k_1^j + k_1^k + k_1^l - k_3^i)y). \quad (2)$$

Here, subscripts 1 and 3 of the wave vectors k stand for the pump wave and the third harmonic, ω is the pump frequency, and y is the propagation direction of the third harmonic. The y axis in our experiments on THG by the pump field normally incident on p-Si films coincides with the [110] crystallographic direction. Cartesian x and z axes are chosen along the $[1\bar{1}0]$ and [001] crystallographic directions, respectively. According to the results of earlier studies [9], these directions are the propagation directions of the ordinary and extraordinary waves, respectively. We use Eq. (1) in this system of coordinates to derive the following solution to Eq. (2) for the projections of the third-harmonic field from the bulk of the medium under study:

$$E^x(3\omega) = \frac{3\pi\omega}{cn(3\omega)}\sin\psi \times \left[\frac{(\chi_{1111} + 3\chi_{1122})\sin^2\psi(\exp(i(3k_1^o - k_3^o)d) - 1)}{3k_1^o - k_3^o}\right. + \left.\frac{6\chi_{1133}\cos^2\psi(\exp(i(2k_1^e + k_1^o - k_3^o)d) - 1)}{2k_1^e + k_1^o - k_3^o}\right]|\mathbf{E}(\omega)|^3, \quad (3)$$

$$E^y(3\omega) = 0, \quad (4)$$

$$E^z(3\omega) = \frac{6\pi\omega}{cn(3\omega)}\cos\psi \times \left[\frac{\chi_{3333}\cos^2\psi(\exp(i(3k_1^e - k_3^e)d) - 1)}{3k_1^e - k_3^e}\right. + \left.\frac{3\chi_{3311}\sin^2\psi(\exp(i(2k_1^o + k_1^e - k_3^e)d) - 1)}{2k_1^o + k_1^e - k_3^e}\right]|\mathbf{E}(\omega)|^3, \quad (5)$$

where ψ is the angle between the polarization of the pump field and the [001] optical axis of the sample, and d is the thickness of the p-Si film. This solution shows that the E^y projection of the third harmonic along the direction of propagation vanishes, while ordinary and extraordinary components E^x and E^z of the third-harmonic wave are rather complicated functions of the angle ψ and the phase mismatches, appearing in the denominators of the terms governing a specific type of interaction between ordinary and extraordinary components of the pump wave in the THG process. To provide at least a qualitative analysis of our solution, we need to know the dispersion of the refractive indices for the ordinary and extraordinary waves.

Preliminary calculation of dispersion for ordinary and extraordinary waves in the samples under study was based on the measurement and analysis of infrared transmission spectra with the use of the procedure described in [9]. This analysis has shown that all the samples possess normal dispersion and negative birefringence. The central symmetry of p-Si prohibits THG involving an odd number of photons with the same polarization. An *ooe–e* interaction is, therefore, the only possibility for phase-matched THG. The birefringence in samples *A* and *B* at $\lambda = 2$ μm was equal to $\Delta n = n_o - n_e = 0.16$ and $\Delta n = 0.14$, respectively (Table 1). Such a birefringence is sufficient to phase-match THG for this type of interaction within the range of pump wavelengths from 1 to 2.1 μm. Optical anisotropy for samples *C*, on the other hand, was $\Delta n = 0.09$ at $\lambda = 2$ μm (Table 1), which is not sufficient to phase-match THG within the above-specified wavelength range. A

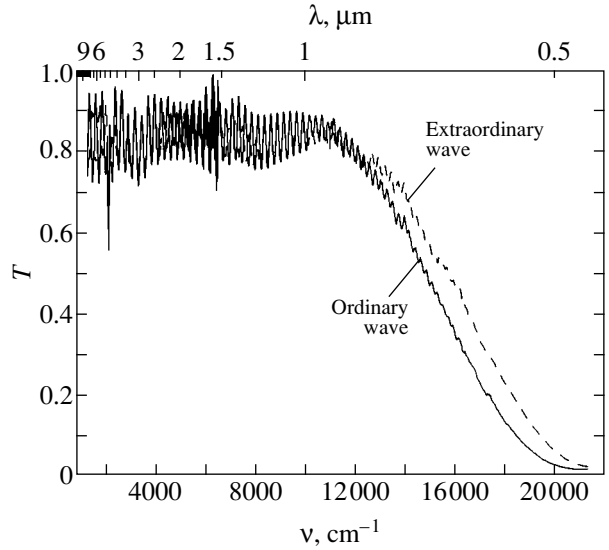


Fig. 1. Transmission spectra for the ordinary and extraordinary waves in a *B*-type porous silicon film with a thickness of 12 μm.

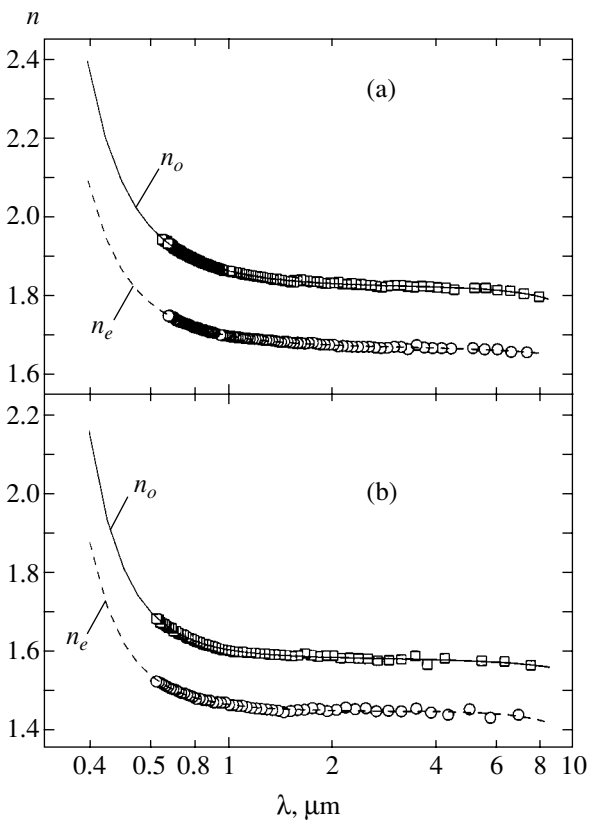


Fig. 2. Dispersion dependences of the refractive indices for the ordinary and extraordinary waves in (a) *A*- and (b) *B*-type porous silicon films.

more accurate calculation of the absolute values of phase mismatches for samples *A* and *B* was based on the analysis of transmission spectra for the ordinary and extraordinary waves ($[1\bar{1}0]$ and [001] crystallographic directions, respectively). These spectra were measured within the range of wavelengths from 0.47 to 8 μm. Figure 1 presents typical transmission spectra for the ordinary and extraordinary waves in a *B*-type p-Si film with a thickness of 12 μm.

Dispersion dependences of the refractive indices for ordinary and extraordinary waves were calculated from interference fringes in transmission spectra using the formula

$$2dn(\lambda) = N\lambda, \tag{6}$$

where d is the film thickness, λ is the wavelength, $n(\lambda)$ is the refractive index, and N is the natural integer numerating interference fringes. Dispersion dependences calculated with the use of this procedure for p-Si films with strong birefringence are shown in Figs. 2a and 2b. The data for refractive indices extracted from measurements were extrapolated with a generic expression [17]

$$n(\lambda) = a + bl + cl^2 + d\lambda^2 + e\lambda^4, \tag{7}$$

where a, b, c, d, and e are the variable extrapolation parameters, $l = 1/(\lambda^2 - 0.028)$, and λ is the wavelength in micrometers. Extrapolation parameters for ordinary and extraordinary waves in our samples are summarized in Table 2. Using this extrapolation, we can iden-

Table 2. Extrapolation parameters a, b, c, d, and e for the dispersion dependences $n(\lambda) = a + bl + cl^2 + d\lambda^2 + e\lambda^4$ of ordinary and extraordinary waves in porous silicon films with a strong birefringence. Parameters are given for $l = 1/(\lambda^2 - 0.028)$ and λ in micrometers

Film type	Polarization of the wave	Extrapolation parameter				
		a	*b*	*c*	*d*	*e*
A	*o*	1.82	0.0371	0.0051	−0.0001	-4.86×10^{-6}
	e	1.66	0.0268	0.0039	−0.00025	-1.66×10^{-7}
B	*o*	1.58	0.0154	0.0081	−0.00026	-8.90×10^{-7}
	e	1.44	0.0133	0.0058	0.00004	-6.26×10^{-6}

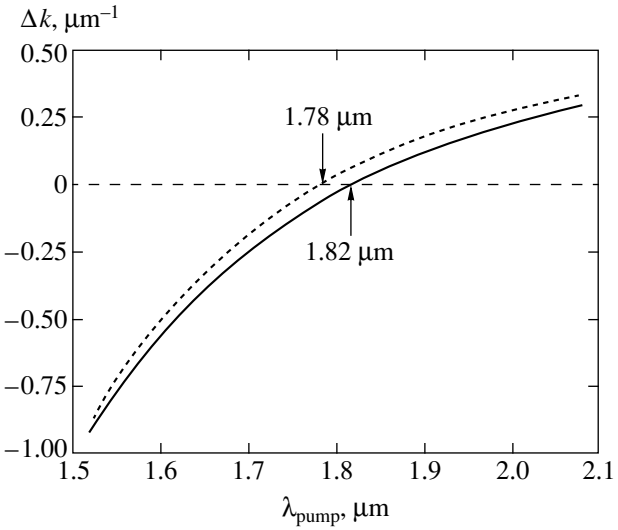


Fig. 3. The phase mismatch Δk as a function of the pump wavelength for third-harmonic generation in the *ooe–e* geometry in porous silicon films of types *A* (solid line, 1.5 mΩ cm, $j = 50$ mA/cm^2) and *B* (dashed line, 3 mΩ cm, $j = 100$ mA/cm^2).

tify the phase-matching area, where the mismatch Δk meets the condition

$$\Delta k = 2k_1^o + k_1^e - k_3^e = 0. \quad (8)$$

Here, k_1^o, k_1^e, and k_3^e are the wave numbers of the ordinary and extraordinary waves at the pump frequency and the extraordinary wave of the third harmonic. Figure 3 displays the mismatch as a function of the pump wavelength for *A*- and *B*-type p-Si films, allowing phase matching at $\lambda_{pump} = 1.82$ and 1.78 μm, respectively.

Thus, we demonstrated the possibility of phase-matching an *ooe–e* THG process in p-Si samples fabricated of single-crystal Si wafers with a specific resistance of 1.5 and 3 mΩ cm and (110)-oriented surface. The birefringence necessary for the phase matching is induced by choosing appropriate etching currents. The second term in parentheses in Eq. (5), which is proportional to $\cos\psi\sin^2\psi$ and which differs from other terms contributing to the nonlinear-optical response by the dependence on the angle ψ, provides a dominant contribution to phase-matched THG. The area of phase-matched THG can be, therefore, defined not only from the higher intensity of the third harmonic, but also from the analysis of orientation dependences of this signal.

4. EXPERIMENTAL RESULTS AND DISCUSSION

In view of the possibility of phase-matching *ooe–e*-type THG in p-Si films, predicted above, we started with the experimental study of this geometry of THG. The optical axis of the studied samples, corresponding to the [001] crystallographic direction was oriented at an angle $\psi = 55°$ with respect to the polarization of the pump field normally incident upon the film plane. The Glan prism was set to transmit the polarization component directed along the optical axis and corresponding to the *e*-wave (the inset to Fig. 4a). This polarization arrangement is optimal for the *ooe–e* geometry of THG since the maximum signal in phase-matched THG is determined, according to Eq. (5), by the extremum of the function $\cos\psi\sin^2\psi$, which is achieved with $\psi = 55°$. Figures 4a–4c display the intensity of the third harmonic from the bulk of p-Si films as a function of the pump wavelength. These dependences were corrected to include the experimentally measured spectrum of the response function of optical components placed behind the OPO and the intensity of pump radiation (the idler wave at the output of the OPO), taken as a cubic function of the wavelength. Depending on the intensity of the detected signal, the widths of spectral ranges of experimental measurements, presented in Figs. 4a–4c, were chosen for each plot in such a way as to meet the criterion of comparability of the signal level and the noise level. In other words, the boundary points in each plot correspond to the wavelength where the signal-to-noise ratio is approximately equal to unity. Dimensionless units in all the plots correspond to the normalization to the intensity at 1.2 μm in the *ooe–e* geometry.

Figure 4c shows the intensity of the third harmonic generated in the *ooe–e* geometry from sample *C* as a function of the pump wavelength. This plot reveals a relatively uniform growth in the intensity of the third harmonic as a function of the pump wavelength without any pronounced extrema. This result can be attributed to the fact that no phase matching can be achieved for this sample because of its weak birefringence. The growth in the intensity of the third harmonic as a function of the pump wavelength observed in the dependences presented in Figs. 4a and 4b is apparently due to the absorption of the third harmonic, which decreases with decreasing radiation frequency (see, e.g., Fig. 1). Small-amplitude extrema in Figs. 4a–4c may originate from the weak dispersion of the nonlinear susceptibility tensor $\chi_{ijkl}^{(3)}$ and experimental errors.

We consider now in greater detail the results of *ooe–e* THG experiments for *A*- and *B*-type p-Si films (Figs. 4a and 4b). The intensity of the third harmonic from these samples displays clearly pronounced maxima as a function of the pump wavelength. These maxima are observed at 1.95 μm for sample *A* and at 1.635 μm for sample *B*. The maximum level of third-harmonic intensity exceeds the third-harmonic intensity at 1.2 μm by a factor of approximately 500 in the former case and a factor of about 70 in the latter case. We will show below that these maxima in the dependence of the third-harmonic intensity as a function of the pump wavelength can be explained by phase match-

ing. The pump wavelengths providing maximum efficiencies of THG in our experiments are close to the optimal pump wavelengths predicted from the analysis of transmission spectra ($\lambda = 1.82$ μm for sample *A* and $\lambda = 1.78$ μm for sample *B*). Even with a margin allowing for the errors of the extrapolation procedure used to recover the refractive indices, the results of linear-optical measurements provide a rather convincing explanation to the observed extrema in the third-harmonic intensity measured as a function of the pump wavelength, permitting these extrema to be attributed to phase matching.

We measured the third-harmonic intensity as a function of the pump wavelength for other geometries of THG in all our samples. No signs of phase matching has been observed for those geometries. As an example, Fig. 4b displays the results of measurements for *ooo–o*- and *eee–e*-type THG in sample *B*. No phase matching for these types of THG in our samples is consistent with the above analysis of dispersion of optically anisotropic p-Si.

It should be noted, however, that as the pump wavelength increases, the efficiency of THG in both samples allowing phase matching is improved also due to the lower absorption at the frequency of the third harmonic. Phase matching alone would certainly give a less significant effect. As can be seen from Fig. 4b, the third harmonic generated in *ooo–o* and *eee–e* geometries around $\lambda = 1.635$ μm is only 5 to 7 times lower than the intensity of the third harmonic produced in the *ooe–e* geometry. On the other hand, THG enhancement due to phase matching is well pronounced and reliable and can be used to improve the efficiency of laser frequency tripling either jointly with or independently of absorption.

We proceed now with our studies of orientation dependences of the third harmonic, i.e., measurements of the third-harmonic intensity as a function of the angle between the [001] optical axis and the polarization of the pump field for different pump wavelengths. Orientation of the optical axis of the sample was varied by rotating the p-Si film holder. The plane of the sample in these rotations remained perpendicular to the wave vector of the pump field, while the Glan prism was adjusted to transmit the polarization component parallel to the direction of pump polarization (the inset to Fig. 5c). Typical orientation dependences measured for sample *A* with a thickness $d = 20$ μm in the phase-matching regime and with no phase matching are shown in Figs. 5a–5c. The zero of the abscissa axis in these plots corresponds to the orientation of the [001] axis along the direction of pump polarization. The intensity in all the plots was normalized to the maximum of the third-harmonic signal for a given dependence.

Using the solution (3)–(5) to Eq. (2) for the slowly varying envelope of the third harmonic, we represent the intensity of the third-harmonic signal detected in our experiments as

$$I_{TH}(\psi) = \frac{9\pi\omega^2}{8cn(3\omega)} \times \left| \left(\sin^4\psi \frac{(\chi_{1111} + 3\chi_{1122})(\exp(i(3k_1^o - k_3^o)d) - 1)}{3k_1^o - k_3^o} + 0.25\sin^2 2\psi \frac{6\chi_{1133}(\exp(i(2k_1^e + k_1^o - k_3^o)d) - 1)}{2k_1^e + k_1^o - k_3^o} \right) \times \exp(ik_3^o d) + \left(\cos^4\psi \frac{2\chi_{3333}(\exp(i(3k_1^e - k_3^e)d) - 1)}{3k_1^e - k_3^e} + 0.25\sin^2 2\psi \frac{6\chi_{3311}(\exp(i(2k_1^o + k_1^e - k_3^e)d) - 1)}{2k_1^o + k_1^e - k_3^e} \right) \times \exp(ik_3^e d) \right|^2 I_{\text{pump}}^3. \quad (9)$$

In the regime of phase matching, the square of the last term on the right-hand side of Eq. (9), which is proportional to $\sin^4 2\psi$, makes the dominant contribution to the intensity of the detected third harmonic. Indeed, the orientation dependence measured in our experiments in the regime of phase matching at $\lambda_{\text{pump}} = 1.95$ μm can be rather accurately approximated with such a function (Fig. 5a). Orientation dependences measured off the phase-matching condition (two of such dependences are presented in Figs. 5b and 5c) feature maxima coinciding with the extrema of the functions $\sin^4 2\psi$ and $\cos^4 2\psi$, appearing in the first and third terms on the right-hand side of Eq. (9). A smaller contribution of the second and fourth terms off the phase-matching condition is explained by the structure of these terms. All the exponential factors control the phase shift, changing the amplitude ratio of the extrema in orientation dependences with varying pump wavelength due to the dispersion of refractive indices of ordinary and extraordinary waves. Such a transformation of dependences was distinctly observed in our experiments (Figs. 5b and 5c). All the phase mismatches in the denominators off the phase-matching condition are monotonic functions of the wavelength, making no significant contribution to the relative variation of different terms.

Thus, the value of each term is mainly determined by the factors appearing in front of each trigonometric function in Eq. (9) and the factors $\chi_{1111} + 3\chi_{1122}$, $6\chi_{1133}$, $2\chi_{3333}$, and $6\chi_{3311}$. The latter factors are, rigorously speaking, not equal to each other (except for the $\infty\infty$ and $\infty\infty m$ point-group symmetries [15]), and their differences may become quite substantial. The factors in

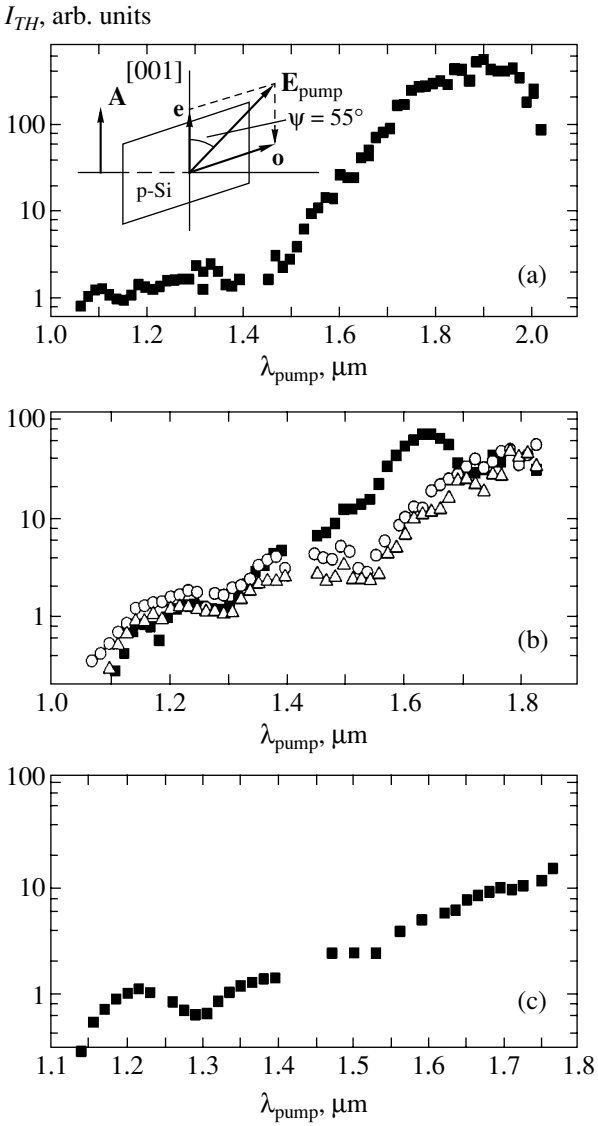


Fig. 4. The third-harmonic intensity as a function of the pump wavelength for porous silicon films of types *A* (a), *B* (b), and *C* (c) in (a, c) the *ooe–e* and (b) *ooe–e* (■), *ooo–o* (o), and *eee–e* (△) geometries. The third-harmonic intensity is normalized to its value in the *ooe–e* geometry measured at the pump wavelength of 1.2 μm. The inset to Fig. 4a illustrates orientation of the sample and polarization arrangement in experiments on third-harmonic generation in the *ooe–e* geometry; *A* shows the orientation of the Glan prism, used as a polarization analyzer.

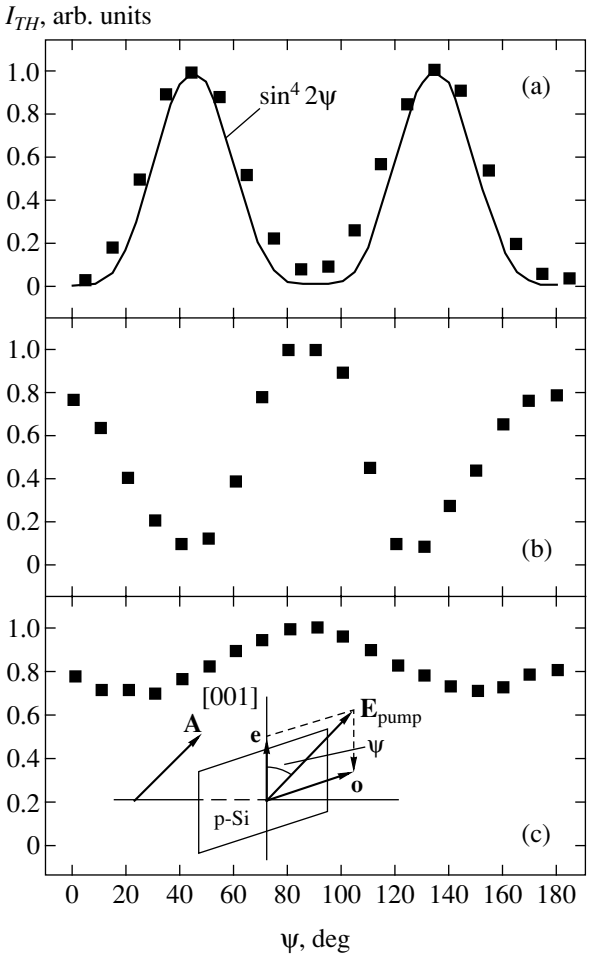


Fig. 5. Dependences of the third-harmonic intensity on the angle between the [001] optical axis and the polarization of the pump field for pump wavelengths of (a) 1.95, (b) 1.35, and (c) 1.1 μm. The third harmonic is generated in an *A*-type porous silicon film with a thickness $d = 20$ μm. The zero of the abscissa axis corresponds to the orientation of the [001] axis along the direction of pump polarization. The intensity in all the plots is normalized to the maximum of the third-harmonic signal for a given dependence. The experimental data for the intensity of phase-matched third harmonic (a) are fitted (solid curve) with a function proportional to $\sin^4 2\psi$. The inset to Fig. 5c illustrates orientation of the sample and polarization arrangement in experiments; *A* shows the orientation of the Glan prism, used as a polarization analyzer and set to transmit third-harmonic radiation polarized along the polarization of the pump field.

front of the trigonometric functions in the first and third terms are equal to 1. In the second and fourth terms, these factors are 0.25, which explains to a large extent the smaller contribution of these terms to the total signal.

We do not provide quantitative estimates on the ratios of nonlinear-susceptibility tensor components in this work. In view of the strong phase sensitivity of the solution (3)–(5) to Eq. (2), such a quantitative analysis would require highly accurate measurements of the film thickness and dispersion (within 0.1%) of the refractive indices for ordinary and extraordinary waves. It is important, however, that the behavior of the orientation dependences in the regime of phase matching and off the phase-matching condition is generally consistent with the classification of our p-Si samples as 4/*mmm* point-group symmetry materials and with the predictions based on the solution (3)–(5) to Eq. (2).

Orientation dependences for sample *B* are qualitatively similar to the orientation dependences for sample *A* presented in Figs. 5a–5c. In the case of the weakly birefringent sample *C*, we observed only orientation

dependences characteristic of the non-phase-matched regime (similar to the dependences shown in Figs. 5b and 5c). Thus, characteristic features observed in orientation dependences on and off the phase matching prove that THG in optically anisotropic p-Si films is enhanced due to phase-matching effects (Figs. 4a and 4b).

Yet another evidence of THG enhancement in the *ooe–e* geometry in strongly birefringent p-Si films due to phase matching was provided by the following experiment. We fabricated six *B*-type p-Si films with different thicknesses: 10, 12, 20, 44, 60, and 80 μm. Because of the identical etching conditions, all the films featured approximately the same birefringence, $\Delta n = 0.140 \pm 0.005$ at $\lambda = 2$ μm, measured from Fourier-transform infrared spectra [9]. We examined phase matching for *ooe–e* THG in these samples by searching for the maximum of the third harmonic with identical focusing of pump radiation on all the films, which was achieved by using a long-focal-length lens. The third-harmonic intensities measured in these experiments as functions of the sample thickness are presented by dots in Fig. 6. The maxima of the third-harmonic intensity were observed at slightly different pump wavelengths, varying from $\lambda_{pump} = 1.56$ to 1.77 μm in these experiments (which is apparently due to variations in optical anisotropy from sample to sample). To reflect this difference, we use different types of dots in Fig. 6 to represent the experimental data. For the third harmonic, these changes were observed within the spectral range of $\lambda_{TH} = 0.52$–0.59 μm, where absorption noticeably decreases with the growth in the wavelength (Fig. 1). We therefore analyze our experimental results for the third-harmonic intensity as a function of the film thickness by plotting different theoretical curves for different wavelengths corresponding to phase-matched THG in our experiments (Fig. 6). Absorption was included in our calculations via the formula [18]

$$I_{TH} \propto I_{\text{pump}}^3 \left(\frac{1 - \exp(-\alpha(\lambda) d/2)}{\alpha(\lambda)/2} \right)^2, \qquad (10)$$

where I_{TH} and I_{pump} are the intensities of the third harmonic and the pump field, $\alpha(\lambda)$ is the intensity absorption coefficient at the harmonic wavelength λ, and d is the thickness of the p-Si film. The absorption coefficient $\alpha(\lambda)$ was calculated for the *e*-wave from the transmission spectrum presented in Fig. 1. The experimental data for the intensity of the third harmonic generated in p-Si films of different thickness, as can be seen from Fig. 6, agree well with the results of calculations performed for phase-matched THG in an absorbing medium at different pump wavelengths. This agreement provides yet another proof of the phase-matching-induced THG enhancement in p-Si films with a strong form birefringence.

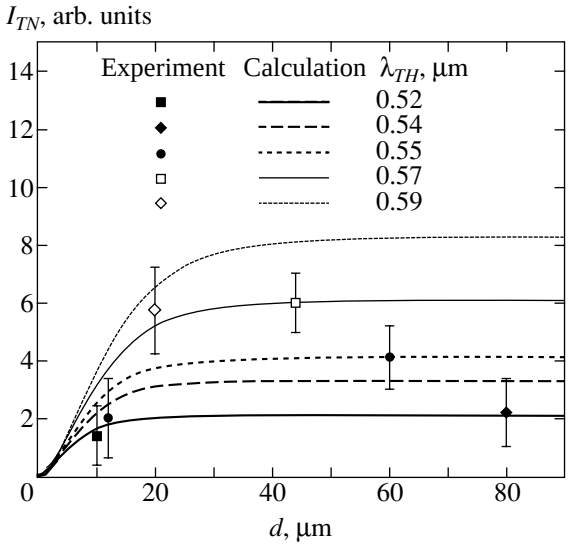


Fig. 6. Third-harmonic intensity as a function of the thickness of a porous silicon film for phase-matched *ooe–e* interaction in *B*-type samples. Theoretical curves, calculated with the inclusion of absorption at the frequency of the third harmonic, are shown for each wavelength of the third harmonic corresponding to phase-matched third-harmonic generation in one of the samples in our experiments.

5. CONCLUSIONS

In conclusion, we used a nanosecond laser system based on an optical parametric oscillator frequency tunable within the near-infrared range to study third-harmonic generation in porous silicon films with an artificial optical anisotropy. These experiments have demonstrated the possibility of phase-matched THG in the *ooe–e* geometry in the case of sufficiently strong birefringence and properly chosen pump wavelength. We found no other types of phase-matched THG in the studied samples, which is explained by normal dispersion and the central symmetry of p-Si. The maxima of third-harmonic intensity measured as a function of the pump wavelength are observed at $\lambda_{pump} = 1.95$ μm for *A*-type films and $\lambda_{pump} = 1.635$ μm for *B*-type films. These maxima have been explained by using the solution to the equation for the slowly varying envelope of the third harmonic and calculating the phase mismatch with refractive indices of the ordinary and extraordinary waves determined from the relevant transmission spectra. No phase matching found for THG in sample *C* is attributed to a low birefringence of this sample.

Transformation of orientation dependences of the third-harmonic on and off the phase matching can be also explained (at least qualitatively) in terms of the solution to the equation for the slowly varying envelope of the third harmonic. This broadly demonstrates the

applicability of the theory developed in this work for the description of THG in p-Si.

The intensity of the phase-matched third harmonic measured as a function of the film thickness agrees well with the results of calculations for the nonlinear-optical response in phase-matched THG with absorption included at the frequency of the third harmonic.

The results of this work suggest that porous silicon with a strong artificial optical form anisotropy is a promising material for the creation of compact optical frequency converters based on phase-matched nonlinear-optical interactions.

ACKNOWLEDGMENTS

This work was supported in part by the President of Russian Federation Grant MD-42.2003.02, the Russian Foundation for Basic Research (projects nos. 02-02-17259, 03-02-16929, and 02-02-17098), U.S. Civilian Research and Development Foundation for the Independent States of the Former Soviet Union (CRDF grants nos. RE2-2369 and RP2-2558), Volkswagen Foundation (grant no. I/76 869), the European Research Office of the US Army (Contract no. 62558-03-M-0033), and programs of the Ministry of Industry, Science, and Technologies of the Russian Federation.

REFERENCES

1. A. G. Cullis, L. T. Canham, and P. D. J. Calcott, J. Appl. Phys. **82**, 909 (1997).
2. V. I. Klimov, V. S. Dneprovskii, and V. A. Karavanskii, Appl. Phys. Lett. **64**, 2691 (1994).
3. V. I. Klimov, D. McBranch, and V. A. Karavanskii, Phys. Rev. B **52**, 16989 (1995).
4. P. K. Kashkarov, E. A. Konstantinova, and V. Yu. Timoshenko, Fiz. Tekh. Poluprovodn. (St. Petersburg) **30**, 1479 (1996) [Semiconductors **30**, 778 (1996)].
5. W. Theiß, Surf. Sci. Rep. **29**, 91 (1997).
6. M. E. Kompan, Ya. Salonen, and I. Yu. Shabanov, Zh. Éksp. Teor. Fiz. **117**, 368 (2000) [JETP **90**, 324 (2000)].
7. O. G. Sarbei, E. K. Frolova, R. D. Fedorovich, and D. B. Dan'ko, Fiz. Tverd. Tela (St. Petersburg) **42**, 1205 (2000) [Phys. Solid State **42**, 1240 (2000)].
8. D. Kovalev, G. Polisski, J. Diener, *et al.*, Appl. Phys. Lett. **78**, 916 (2001).
9. L. P. Kuznetsova, A. I. Efimova, L. A. Osminkina, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **44**, 780 (2002) [Phys. Solid State **44**, 811 (2002)].
10. L. A. Golovan', A. M. Zheltikov, P. K. Kashkarov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 274 (1999) [JETP Lett. **69**, 300 (1999)].
11. T. V. Dolgova, A. I. Maidykovskii, M. G. Martem'yanov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **75**, 17 (2002) [JETP Lett. **75**, 15 (2002)].
12. L. A. Golovan', V. Yu. Timoshenko, A. B. Fedotov, *et al.*, Appl. Phys. B **73**, 31 (2001).
13. L. A. Golovan', L. P. Kuznetsova, A. B. Fedotov, *et al.*, Appl. Phys. B **76**, 429 (2003).
14. A. M. Zheltikov, Laser Phys. **11**, 1024 (2001).
15. Yu. I. Sirotin and M. P. Shaskol'skaya, *Fundamentals of Crystal Physics* (Nauka, Moscow, 1975).
16. J. A. Armstrong, N. Bloembergen, J. Ducuing, *et al.*, Phys. Rev. **127**, 1918 (1962).
17. E. M. Voronkova, B. N. Grechushnikov, G. I. Distler, and I. P. Petrov, *Optical Materials for Infrared Technology* (Nauka, Moscow, 1965) [in Russian].
18. J. Reintjes, *Nonlinear Optical Parametric Processes in Liquids and Gases* (Academic, New York, 1984; Mir, Moscow, 1987).

Translated by A. Zheltikov

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 37–42.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 47–53.

ATOMS, SPECTRA, RADIATION

Interference of Scattered Waves and Mixing Rules for Group Velocities in Nanocomposite Materials

A. M. Zheltikov

International Laser Center, Physics Department, Moscow State University, Vorob'evy gory, Moscow, 119992 Russia

e-mail: zheltikov@top.phys.msu.su

Received December 22, 2003

Abstract—Mixing rules for group velocities in nanocomposite materials with different architecture, including lamellar-inhomogeneous nanotextures, Maxwell Garnett structures, and one-dimensional photonic crystals, are derived and analyzed. The group velocity can be controlled for such composite structures by changing nanocrystal sizes and varying the dielectric properties and the volume-filling fractions of the constituent materials. The interference of scattered waves in structures with a spatial scale of optical inhomogeneities comparable to the radiation wavelength gives rise to new physical phenomena that cannot be described in terms of the effective-medium approximation. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Nanocomposite materials open new unique opportunities in laser physics and optical technologies. Nanopowder materials, in particular, are at the heart of lasers and optical frequency converters of a new architecture [1–3]. Nano- and mesoporous semiconductor materials can radically enhance nonlinear-optical interactions of laser pulses [4–6]. Photonic-crystal structures are intensively used to control radiative processes and are considered as promising candidates for the creation of a new generation of components and devices for telecommunications, laser physics, and nonlinear optics [7]. Microstructure and photonic-crystal fibers [8, 9] provide unprecedentedly high efficiencies of supercontinuum generation [7, 10, 11] and nonlinear-optical spectral transformation of low-power ultrashort laser pulses [7]. Fibers of this new type are widely used in optical metrology [12, 13], ultrafast nonlinear optics [14], optical coherence tomography [15], nonlinear spectroscopy [16], and photochemistry [17].

Many remarkable properties of nonlinear-optical interactions in nanocomposite materials are due to the local-field enhancement [18] and the possibility of tuning the effective refractive index of such materials by changing the dielectric properties, the content, and the sizes of nanoinclusions [5, 19]. Porous materials, in particular, offer interesting solutions to the phase-matching problem [20] and allow the creation of optically transparent solid-state hosts for composite systems with high quadratic and cubic optical nonlinearities [21].

Recent experiments demonstrate that nanocomposite materials hold much promise for nonlinear-optical frequency conversion of ultrashort pulses [22]. Tunability of the group velocity is one of the key problems in femtosecond nanophotonics—the area of research that deals with the development of highly efficient and compact photonic devices for the control of ultrashort laser pulses based on nanocomposite materials. This work suggests strategies to solve this problem. We will derive and analyze mixing rules for group velocities in nanocomposite materials with different architecture, including lamellar nanotextures, Maxwell Garnett structures, and one-dimensional photonic crystals. We will explore the possibility of controlling the group velocity in nanostructures by engineering their architecture, changing nanocrystal sizes, and varying the dielectric properties and the volume-filling fractions of the constituent materials. It will be shown that the interference of scattered waves in structures with a spatial scale of optical inhomogeneities comparable to the radiation wavelength gives rise to new physical phenomena that cannot be described in terms of the effective-medium approximation.

2. A LAMELLAR-INHOMOGENEOUS NANOTEXTURE

A lamellar-inhomogeneous nanotexture (Fig. 1a) is a classical example of a nanocomposite medium. Nanocomposites of such an architecture consist of parallel nanosheets of material with a dielectric constant $\varepsilon_1 = n_1^2$ embedded in a medium with a dielectric constant $\varepsilon_2 = n_2^2$ (n_1 and n_2 are the refractive indices of constituent materials). Optical properties of such a structure have been analyzed in the literature starting from the 19th century. Lord Rayleigh, in particular, highlighted [23]

1063-7761/04/9901-0037$26.00 © 2004 MAIK "Nauka/Interperiodica"

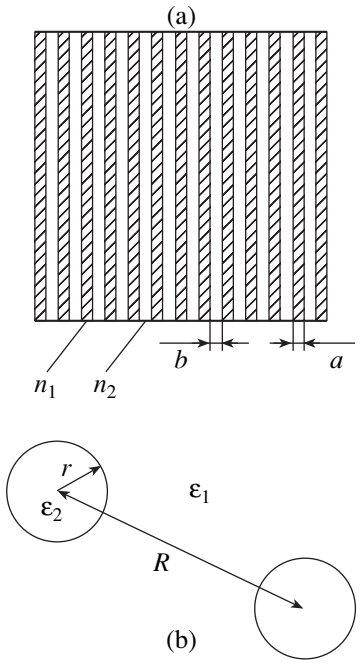


Fig. 1. Nanocomposite materials of different architecture: (a) a lamellar-inhomogeneous nanotexture and (b) a Maxwell Garnett nanocomposite.

a form birefringence of such a layered structure. This signature of anisotropy in nanostructured materials was later studied in seminal papers by Wiener [24] and Ryzhov [25]. As understood later, form birefringence can be used to radically enhance nonlinear-optical processes in nanostructured materials and multilayer waveguides due to improved phase matching between the pump field and the nonlinear signal [5, 19, 20].

Assuming that

$$\omega n_1 a/c, \quad \omega n_2 b/c \ll 1,$$

where ω is the radiation frequency and a and b are the thicknesses of alternating layers of isotropic materials with refractive indices n_1 and n_2, respectively, we can represent the effective refractive indices for ordinary and extraordinary waves in a periodic layered nanotexture (shown in Fig. 1a) as [26]

$$n_o^2 = \frac{a}{d}n_1^2 + \frac{b}{d}n_2^2, \tag{1}$$

$$\frac{1}{n_e^2} = \frac{a}{d}\frac{1}{n_1^2} + \frac{b}{d}\frac{1}{n_2^2}, \tag{2}$$

where $d = a + b$.

Formulas (1) and (2) define elementary mixing rules for the refractive indices of the constituent materials in the case of ordinary and extraordinary waves in lamellar-inhomogeneous nanocomposites. Setting for definiteness $n_1 > n_2$, we find from Eqs. (1) and (2) that the effective refractive indices of a lamellar-inhomogeneous material reach their upper and lower bounds with $b = 0$ and $a = 0$, respectively. These trivial limiting situations correspond to uniform materials with refractive indices n_1 and n_2, respectively.

To find the corresponding mixing rules for the group velocities of light pulses in such a nanostructure, we differentiate Eqs. (1) and (2) in frequency. This operation yields

$$\frac{1}{v_o} = \frac{1}{d^{1/2}(an_1^2 + bn_2^2)^{1/2}}\left(\frac{an_1}{v_1} + \frac{bn_2}{v_2}\right), \tag{3}$$

$$\frac{1}{v_e} = \frac{d^{1/2}}{\left(\frac{a}{n_1^2} + \frac{b}{n_2^2}\right)^{3/2}}\left(\frac{a}{n_1^3}\frac{1}{v_1} + \frac{b}{n_2^3}\frac{1}{v_2}\right), \tag{4}$$

where $v_i = (\partial k_i/\partial\omega)^{-1}$ are the group velocities in the constituent materials ($i = 1, 2$).

In the limiting case of $n_1 = n_2$ and $v_1 = v_2 = v$, Eqs. (3) and (4) are reduced to the expressions for the group velocity of light in an isotropic medium:

$$v_o = v_e = v.$$

Generally, the group-velocity mixing rules for ordinary and extraordinary waves in a lamellar-inhomogeneous medium, as it follows from Eqs. (3) and (4), can be written as

$$\frac{1}{v_{o,e}} = \frac{g_1^{o,e}}{v_1} + \frac{g_2^{o,e}}{v_2}. \tag{5}$$

The weight factors $g_i^{o,e}$ ($i = 1, 2$) in Eq. (5) depend only on the ratio of layer thicknesses, b/a, and the ratio of the refractive indices of constituent materials, n_2/n_1. Figures 2 and 3 display the dependences of the weight factors $g_i^{o,e}$ for the ordinary and extraordinary waves in a lamellar-inhomogeneous material on the ratio of layer thicknesses b/a. For the ordinary wave (Fig. 2), the weight factors g_i^o are monotonic functions of the ratio b/a. The contribution of each of the materials to the group velocity of light in a nanocomposite monotonically decreases as the content of this material is reduced. In the case of the extraordinary wave (Fig. 3), the weight factors g_i^e may nonmonotonically depend on the ratio b/a. Analysis of Eqs. (4) and (5) shows that the factor g_i^e is a monotonic function of b/a for $n_2^2/n_1^2 < 3$ (curve *1* in Fig. 3). With $n_2^2/n_1^2 > 3$, this fac-

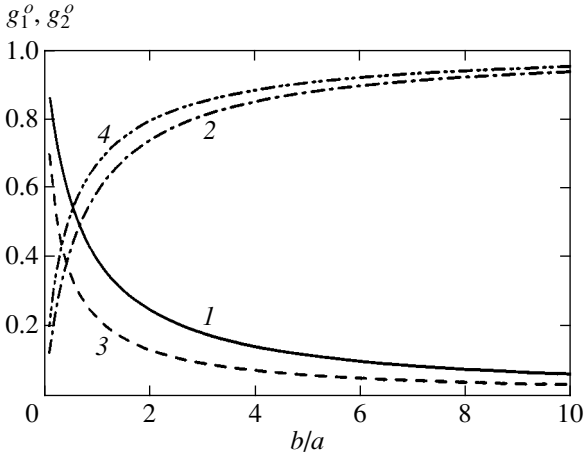


Fig. 2. The weight factors g_1^o (curves *1* and *3*) and g_2^o (curves *2* and *4*) for an ordinary wave in a lamellar-inhomogeneous nanocomposite material as functions of the ratio of layer thicknesses *b*/*a*. The ratio of the refractive indices of the layers forming the structure is n_2/n_1 = 1.5 (*1*, *2*) and 3.0 (*3*, *4*).

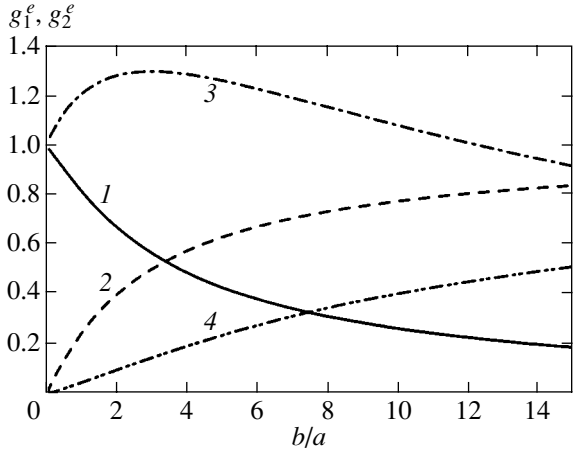


Fig. 3. The weight factors g_1^e (curves *1* and *3*) and g_2^e (curves *2* and *4*) for an extraordinary wave in a lamellar-inhomogeneous nanocomposite material as functions of the ratio of layer thicknesses *b*/*a*. The ratio of the refractive indices of the layers forming the structure is n_2/n_1 = 1.5 (*1*, *2*) and 3.0 (*3*, *4*).

tor reaches its maximum (curve *3* in Fig. 3) for the following value of the ratio *b*/*a*:

$$\left(\frac{b}{a}\right)_{\max} = \frac{1}{2}\frac{n_2^2}{n_1^2} - \frac{3}{2}. \tag{6}$$

Now, we use Eqs. (3)–(5) to find the ratio of the weight factors controlling the contribution of each of the materials to the group velocity of light in a lamellar-inhomogeneous nanotexture:

$$\frac{g_2^{o,e}}{g_1^{o,e}} = F_{o,e}\left(\frac{b}{a}, \frac{n_2}{n_1}\right),$$

$$F_o = \frac{b}{a}\frac{n_2}{n_1}, \tag{7}$$

$$F_e = \frac{b}{a}\left(\frac{n_1}{n_2}\right)^3. \tag{8}$$

Figures 4 and 5 display the factors $F_{o,e}$ as functions of the ratios *b*/*a* and n_2/n_1. Formulas (3)–(8) suggest the possibility of controlling the group velocities of the ordinary and extraordinary modes in lamellar-inhomogeneous nanocomposites by changing the content and the optical properties of the constituent materials. As can be seen from Eqs. (3)–(8) and Fig. 5, the growth in the ratio of the refractive indices n_2/n_1 increases the relative contribution of the material with the refractive index n_2 for the ordinary mode, but decreases the contribution of this material in the case of the extraordinary mode. For high ratios n_2/n_1, the group velocity of the extraordinary mode in a lamellar-inhomogeneous nanotexture, as can be seen from Eq. (4), tends to

$$\nabla_e = (a/d)^{1/2} v_1.$$

Thus, we found that the inverse effective group velocity in a nanotextured material can be represented as a sum of inverse group velocities in each of the materials taken with weight factors, controlled by the optical properties of materials forming the nanostructure and by the nanocomposite architecture. Such a mixing rule

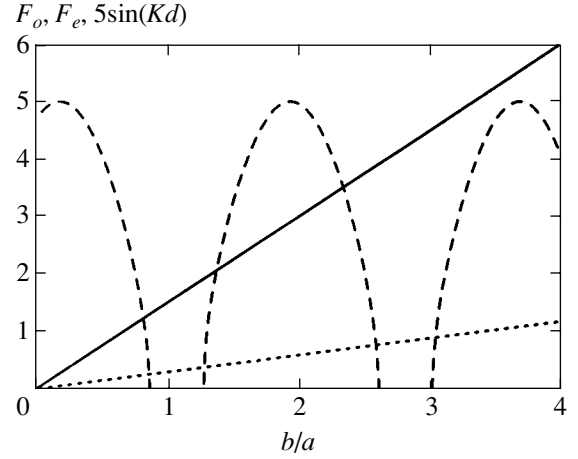


Fig. 4. The factors F_o (solid line) and F_e (dotted line) for the ordinary and extraordinary waves in a lamellar-inhomogeneous nanotexture with a ratio of the refractive indices of the materials forming the structure n_2/n_1 = 1.5 calculated as functions of the ratio of layer thicknesses *b*/*a*. The dashed line shows the parameter $5\sin(Kd)$ as a function of *b*/*a* for a periodic multilayer structure with n_1 = 1.5, n_2 = 2.25, and $\omega n_1 a/c$ = 1.2.

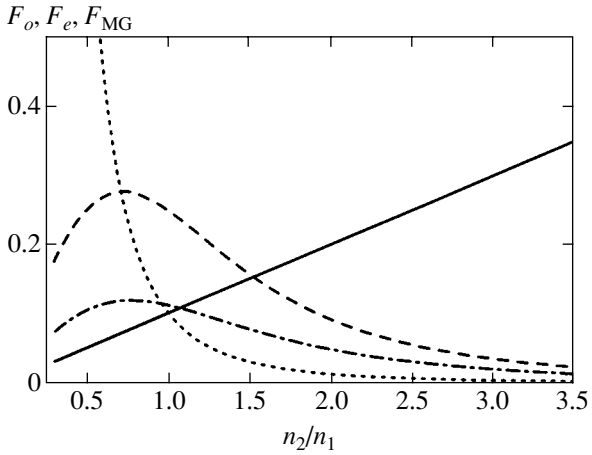


Fig. 5. The factors F_o (solid line) and F_e (dotted line) for the ordinary and extraordinary waves in a lamellar-inhomogeneous nanotexture with a ratio of layer thicknesses $b/a = 0.1$ and the factor F_{MG} for a Maxwell Garnett nanocomposite with $p = 0.1$ (dash–dotted line) and 0.2 (dashed line) calculated as functions of the ratio n_2/n_1 of the refractive indices of the materials forming the nanostructure.

for group velocities is characteristic of effective-medium models. Strong coupling and interference of scattered waves, as will be shown in Section 4, give rise to deviations from this mixing rule.

3. THE MAXWELL GARNETT NANOCOMPOSITE

The Maxwell Garnett model [27] gives an effective dielectric constant of a nanocomposite structure (Fig. 1b) consisting of a material with a dielectric constant $\varepsilon_1 = n_1^2$ with nanoscale spherical inclusions of material with a dielectric constant $\varepsilon_2 = n_2^2$ and a volume filling fraction p. This model assumes that the distance R between the nanoinclusions is much larger than their radius r, $R \gg r$ (see Fig. 1b). The Maxwell Garnett mixing rule for the dielectric constants in a nanocomposite material is written as

$$\varepsilon_{MG} = \varepsilon_1 + 3p\varepsilon_1\frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_2 + 2\varepsilon_1}. \quad (9)$$

Differentiating Eq. (9) in frequency, we arrive at the following mixing rule for the group velocities in the Maxwell Garnett model:

$$v_{MG}^{-1} = n^{-1}(n_2^2 + 2n_1^2)^{-2} \times\left\{[2(2-3p)n_1^4 + (1+3p)n_2^4 + 2(2-3p)n_1^2n_2^2] \times\frac{n_1}{v_1} + 9pn_1^4\frac{n_2}{v_2}\right\}. \quad (10)$$

Thus, similar to the case of a lamellar-inhomogeneous nanotexture, the inverse effective group velocity can be represented within the framework of the Maxwell Garnett model as a sum of weighted inverse group velocities of individual materials forming a nanostructure:

$$\frac{1}{v_{MG}} = \frac{g_1^{MG}}{v_1} + \frac{g_2^{MG}}{v_2}. \quad (11)$$

The ratio of the weight factors in Eq. (11),

$$F_{MG}\left(p, \frac{n_2}{n_1}\right) = \frac{g_2^{MG}}{g_1^{MG}}$$

is written as

$$F_{MG}\left(p, \frac{n_2}{n_1}\right) = \frac{9p}{2(2-3p) + (1+3p)\left(\frac{n_2}{n_1}\right)^4 + 2(2-3p)\left(\frac{n_2}{n_1}\right)^2}\frac{n_2}{n_1}. \quad (12)$$

Since p is small, the dependence of the factor F_{MG} on the volume filling fraction of nanoinclusions is close to a linear function. The dependence of F_{MG} on the ratio of refractive indices n_2/n_1 is nonmonotonic (Fig. 5). For small n_2/n_1, the factor F_{MG} linearly increases with the growth in the ratio of the refractive indices. The maximum value of F_{MG} is achieved, independently of p, with $n_2/n_1 = (2/3)^{1/2}$. For large ratios n_2/n_1, the refractive index and the group velocity in a nanocomposite material of this class tend to the following limiting values:

$$n = n_1(1+3p)^{1/2}, \quad v = v_1(1+3p)^{-1/2}.$$

4. ONE-DIMENSIONAL PHOTONIC CRYSTAL

Photonic crystals [28–30] are materials where the refractive index is periodically modulated in one, two, or three dimensions on a characteristic spatial scale providing a strong coupling of scattered electromagnetic waves. The field of electromagnetic radiation has a form of Bloch waves in such structures. In the elementary case of an ideal one-dimensional photonic crystal, equivalent to an infinite periodic stack of layers with refractive indices n_1 and n_2 and thicknesses a and b (Fig. 1a), the Bloch wave number K meets the following dispersion relation:

$$\cos(Kd) = \cos\left(\frac{\omega}{c}n_1a\right)\cos\left(\frac{\omega}{c}n_2b\right) - \Delta\sin\left(\frac{\omega}{c}n_1a\right)\sin\left(\frac{\omega}{c}n_2b\right), \quad (13)$$

where

$$\Delta = (n_1^2 + n_2^2)/2n_1n_2.$$

Differentiating Eq. (13) in frequency, we derive the mixing rule for group velocities in a one-dimensional photonic crystal:

$$\frac{1}{v_{\mathrm{PBG}}} = \frac{g_1^{\mathrm{PBG}}}{v_1} + \frac{g_2^{\mathrm{PBG}}}{v_2}, \tag{14}$$

where

$$g_1^{\mathrm{PBG}} = \frac{a}{d|\sin(Kd)|}\left[\Delta\sin\left(\frac{\omega}{c}n_1a\right)\cos\left(\frac{\omega}{c}n_2b\right) + \cos\left(\frac{\omega}{c}n_1a\right)\sin\left(\frac{\omega}{c}n_2b\right)\right], \tag{15}$$

$$g_2^{\mathrm{PBG}} = \frac{b}{d|\sin(Kd)|}\left[\sin\left(\frac{\omega}{c}n_1a\right)\cos\left(\frac{\omega}{c}n_2b\right) + \Delta\cos\left(\frac{\omega}{c}n_1a\right)\sin\left(\frac{\omega}{c}n_2b\right)\right]. \tag{16}$$

In the limiting case of

$$\omega n_1 a/c, \omega n_2 b/c \ll 1$$

Eqs. (14)–(16) yield formulas (1) and (2) for the effective group velocity in a lamellar-inhomogeneous medium. Generally, however, the mixing rule for group velocities in a photonic crystal, as it follows from the comparison of Eqs. (1), (2) and (14)–(16), qualitatively differs from the mixing rules for group velocities in a lamellar-inhomogeneous nanotexture or a Maxwell Garnett nanocomposite (Fig. 4). In contrast to effective-medium models, corresponding to the electrostatic approximation, the mixing rule for group velocities in a one-dimensional photonic crystal explicitly involves the dependence on the radiation frequency. The group velocity in a photonic crystal is controlled not only by the dispersion properties of each of the materials constituting the photonic crystal and the morphology of the nanostructure, but also the dispersion of Bloch waves (the inset to Fig. 6), which may radically differ from the dispersion of each of the constituent materials.

Physically, the difference between the mixing rules for group velocities in nanocomposite materials described in terms of the electrostatic effective-medium approximation and the mixing rules for photonic crystals originates from the interference of light waves scattered from refractive-index inhomogeneities. Such interference phenomena are especially pronounced around photonic band gaps of periodic structures, i.e., in the frequency ranges where the scattered waves are strongly coupled with the forward wave (the inset in Fig. 6). One of the most prominent effects observed

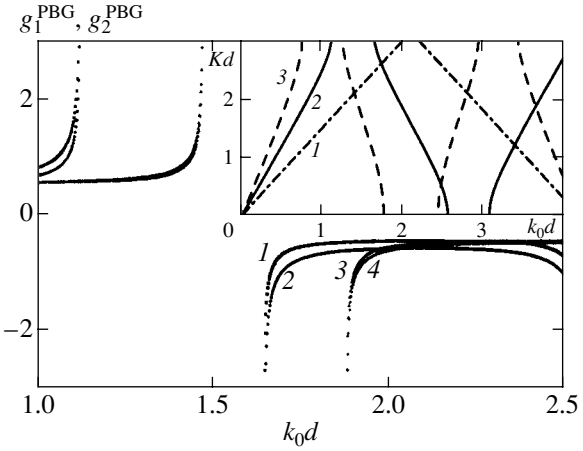


Fig. 6. The weight factors g_1^{PBG} (curves *1* and *3*) and g_2^{PBG} (curves *2* and *4*) as functions of the dimensionless frequency k_0d ($k_0 = \omega/c$) in a one-dimensional photonic crystal with layer thicknesses $a = b = 100$ nm and refractive indices $n_1 = 1.5$ and $n_2 = 2$ (*1*, *2*) and 1.5 (*3*, *4*). The inset shows the dispersion of a one-dimensional photonic crystal consisting of periodically alternating layers with equal thicknesses, $a = b$, and the ratio of refractive indices $n_2/n_1 = 1$ (*1*), 2 (*2*), and 3 (*3*).

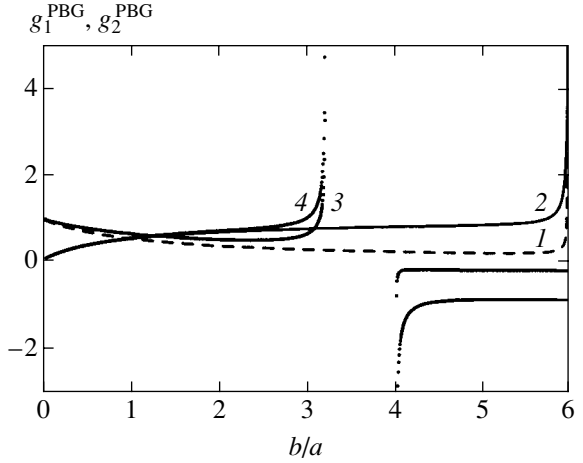


Fig. 7. The weight factors g_1^{PBG} (curves *1* and *3*) and g_2^{PBG} (curves *2* and *4*) in a one-dimensional photonic crystal as functions of the ratio of layer thicknesses b/a for a periodic multilayer structure with $a = 100$ nm, $n_1 = 1.5$, and $n_2 = 2.25$ (*1*, *2*) and 3.375 (*3*, *4*).

under these conditions is the lowering of the group velocity. Near the edges of photonic band gaps, with $Kd \longrightarrow \pi$, we have $v_{\mathrm{PBG}} \longrightarrow 0$ (Figs. 6, 7). The group velocity in this regime is much lower than the group velocities in each of the materials. Such phenomena are never observed in lamellar-inhomogeneous media where the spatial scale of refractive-index modulation is much less than the radiation wavelength (Fig. 4). The role of interference phenomena is insignificant in such materials, and their dispersion can be adequately

described within the framework of effective-medium models [Eqs. (1), (2)]. Inside photonic band gaps, the definition of group velocity should be carefully modified. A comprehensive, methodologically consistent analysis of this issue is provided by Yariv and Yeh [26].

5. CONCLUSIONS

We have derived and analyzed the mixing rules for group velocities in nanocomposite materials with different architecture, including lamellar-inhomogeneous nanotextures, Maxwell Garnett structures, and one-dimensional photonic crystals. We have shown that the group velocity can be controlled for such nanocomposite structures by changing nanocrystal sizes and varying the dielectric properties and the content of the constituent materials. Within the framework of the effective-medium approximation, the inverse effective group velocity in a nanocomposite material can be represented as a sum of inverse group velocities in each of the materials taken with some weight factors, controlled by the optical properties of the materials forming the nanostructure and by the nanocomposite architecture. Strong coupling and interference of scattered waves in structures with a spatial scale of optical inhomogeneities comparable to the radiation wavelength give rise to new physical phenomena that cannot be described in terms of the effective-medium approximation. In particular, energy exchange between strongly coupled forward and reflected waves in photonic crystals may substantially lower the group velocity of electromagnetic radiation, giving rise to a strong local-field enhancement of nonlinear-optical interactions.

ACKNOWLEDGMENTS

This study was supported in part by the President of Russian Federation Grant MD-42.2003.02, the Russian Foundation for Basic Research (projects nos. 03-02-16929 and 02-02-17098), Award no. RP2-2558 of the U.S. Civilian Research and Development Foundation for the Independent States of the Former Soviet Union (CRDF), and the European Research Office of the US Army (Contract no. 62558-03-M-0033).

REFERENCES

1. N. M. Lawand, R. M. Balachandran, A. S. L. Gomes, and E. Sauvain, Nature **368**, 436 (1994).
2. W. Sha, C.-H. Liu, and R. Alfano, Opt. Lett. **19**, 1922 (1994).
3. H. Cao, Y. G. Zhao, S. T. Ho, *et al.*, Phys. Rev. Lett. **82**, 2278 (1999).
4. L. A. Golovan, V. Yu. Timoshenko, A. B. Fedotov, *et al.*, Appl. Phys. B **73**, 31 (2001).
5. P. K. Kashkarov, L. A. Golovan, A. B. Fedotov, *et al.*, J. Opt. Soc. Am. B **19**, 2273 (2002).
6. L. A. Golovan, L. P. Kuznetsova, A. B. Fedotov, *et al.*, Appl. Phys. B **76**, 429 (2003).
7. *Nonlinear Optics of Photonic Crystals*, Ed. by C. M. Bowden and A. M. Zheltikov, Feature Issue of J. Opt. Soc. Am. B **19** (9) (2002).
8. J. C. Knight, T. A. Birks, P. St. J. Russell, and D. M. Atkin, Opt. Lett. **21**, 1547 (1996).
9. J. C. Knight, J. Broeng, T. A. Birks, and P. St. J. Russell, Science **282**, 1476 (1998).
10. J. K. Ranka, R. S. Windeler, and A. J. Stentz, Opt. Lett. **25**, 25 (2000).
11. *Supercontinuum Generation*, Ed. by A. M. Zheltikov, Special Issue of Appl. Phys. B **77** (1/2) (2003).
12. D. J. Jones, S. A. Diddams, J. K. Ranka, *et al.*, Science **288**, 635 (2000).
13. R. Holzwarth, T. Udem, T. W. Hänsch, *et al.*, Phys. Rev. Lett. **85**, 2264 (2000).
14. A. Baltuska, T. Fuji, and T. Kobayashi, Opt. Lett. **27**, 1241 (2002).
15. I. Hartl, X. D. Li, C. Chudoba, *et al.*, Opt. Lett. **26**, 608 (2001).
16. A. B. Fedotov, Ping Zhou, A. P. Tarasevitch, *et al.*, J. Raman Spectrosc. **33**, 888 (2002).
17. S. O. Konorov and A. M. Zheltikov, Opt. Express **11**, 2440 (2003).
18. J. E. Sipe and R. W. Boyd, Phys. Rev. A **46**, 1614 (1992).
19. A. Fiore, V. Berger, E. Rosencher, *et al.*, Nature **391**, 463 (1998).
20. L. A. Golovan', A. M. Zheltikov, P. K. Kashkarov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 274 (1999) [JETP Lett. **69**, 300 (1999)].
21. S. V. Zabotnov, A. B. Fedotov, S. O. Konorov, *et al.*, Opt. Commun. **224**, 309 (2003).
22. S. O. Konorov, D. A. Sidorov-Biryukov, I. Bugar, *et al.*, Appl. Phys. B **78**, 73 (2004).
23. Lord Rayleigh, Philos. Mag. **34**, 481 (1892).
24. O. Wiener, Abh. Sachs. Ges. Akad. Wiss. Math.–Phys. Kl. **32**, 575 (1912).
25. S. M. Ryzhov, Zh. Éksp. Teor. Fiz. **29**, 11 (1955).
26. A. Yariv and P. Yeh, *Optical Waves in Crystals: Propagation and Control of Laser Radiation* (Wiley, New York, 1984; Mir, Moscow, 1987).
27. J. C. Maxwell Garnett, Philos. Trans. R. Soc. London **203**, 385 (1904).
28. J. D. Joannopoulos, R. D. Meade, and J. N. Winn, *Photonic Crystals: Molding the Flow of Light* (Princeton Univ. Press, Princeton, 1995).
29. K. Sakoda, *Optical Properties of Photonic Crystals* (Springer, Berlin, 2001).
30. S. G. Johnson and J. D. Joannopoulos, *Photonic Crystals: The Road from Theory to Practice* (Kluwer, Boston, 2002).

Translated by A. Zheltikov

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 43–50.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 54–62.

ATOMS, SPECTRA, RADIATION

Amplification of Light in a Dilute-Gas Bose–Einstein Condensate

E. D. Trifonov* and N. I. Shamrov

Herzen State Pedagogical University of Russia, St. Petersburg, 191186 Russia

**e-mail: thphys@herzen.spb.ru*

Received July 28, 2003

Abstract—A semiclassical theory is proposed to describe the light amplification in a Bose–Einstein condensate of a dilute atomic gas observed in [1]. Atomic states with well-defined momenta are taken as a basis of wave functions to derive the Maxwell–Bloch equations for the model of a dilute gas of atoms interacting with electromagnetic field. The evolution of light intensity and populations of coherent atomic states with different recoil momenta is described by solving these equations. The solutions obtained are used to demonstrate theoretically the effects observed in [1]. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Production of Bose–Einstein condensates (BECs) of dilute vapors of alkali metals by laser or evaporative cooling is a topic of great current interest [1–6]. Extensive reviews of theoretical and experimental progress in this area can be found in [7, 8].

In experiments on light amplification [1], a vapor condensate consisting of several million sodium atoms confined in a trap 20 μm in diameter and 200 μm long was illuminated by two laser beams, "dressing" and "probe." The dressing beam was red-detuned by 1.7 GHz from the $3S_{1/2}$, $F = 1 \longrightarrow 3P_{3/2}$, $F = 0, 1, 2$ transition. The probe beam was red-detuned from the dressing beam by 91 kHz to ensure resonant two-photon transition to the recoil state of atoms. The beams were in the plane perpendicular to the long axis of the condensate, and they intersected at an angle of 135°. The principal result of the experiment was that the probe beam was amplified in the presence of the dressing beam. When the probe intensity was low, no optical gain was observed after the dressing beam was switched off. However, when the dressing-beam intensity was high, residual probe output was observed even after the dressing beam was switched off. This could be interpreted as a holographic effect: probe output was obtained as a result of diffraction of the dressing beam by a wave grating consisting of atoms in the recoil state. Moreover, the observed delay of probe-pulse output was interpreted as a reduction of the group velocity of light to 1 m/s.

Interaction between light and a degenerate Bose gas of atoms was considered in [9–12]. Semiclassical treatments generally rely on the model of a "dressed" atom, in which excited electronic states are eliminated and analysis is restricted to a basis consisting of ground electronic states with different translational momenta [13, 14]. The Maxwell–Bloch equations written in this basis include phenomenological terms representing relaxation and outgoing radiation. In our studies [15–17] dealing with superradiant Rayleigh scattering [5], we applied a semiclassical theory with the wave functions of excited electronic states retained in the basis and found an explicit solution to the problem of light propagation through the system.

The groundwork for semiclassical analysis of light amplification was laid in [16]. In particular, according to the estimate given in [16] for the chemical potential of a sodium BEC (under the conditions of the experiments reported in [1, 4]) and a comparison of its value with the recoil energy transferred to an atom absorbing or emitting a photon, the ideal-gas model can be used to analyze the interaction of a BEC with electromagnetic field.

In this paper, we refine the BEC model (in particular, we take into account relaxation of the BEC) and develop a more efficient method for solving the Maxwell–Bloch equation that leads to closer agreement with experimental results.

2. MODEL OF A DRESSED ATOM

A qualitative explanation of the effect under consideration is based on the model of a "dressed" atom. Since the incident laser beams are quasi-resonant with an atomic transition, each atom is treated as a two-level electron system. Assuming that the dressing beam is much more intense than the probe, we first solve the problem of interaction between a two-level atom and a dressing field. Denote by ψ_a and ψ_b the ground and excited atomic states with energies E_a and E_b, respectively:

$$\psi_{a,b} = \psi^0_{a,b}\exp\left(-i\frac{E_{a,b}}{\hbar}t\right).$$

1063-7761/04/9901-0043$26.00 © 2004 MAIK "Nauka/Interperiodica"

We represent the wave function as

$$\psi = C_a\psi_a + C_b\psi_b.$$

The Schrödinger equation for a two-level system interacting with a monochromatic field $E\exp(-i\omega_0 t)$ (with a relaxation term added in the rotating-wave approximation) can be written as

$$\begin{aligned}\dot{C}_a &= i\Omega_R\exp(-i\Delta\omega t)C_b,\\ \dot{C}_b &= -\frac{\gamma}{2}C_b + i\Omega_R\exp(i\Delta\omega t)C_a,\end{aligned} \tag{1}$$

where γ is the emission rate constant for the excited state,

$$\Delta\omega = \frac{E_b - E_a}{\hbar} - \omega_0$$

is the detuning, and $\Omega_R = Ed/\hbar$ is the Rabi frequency of the dressing field (dipole transition matrix element d is assumed to be real here). If the atom is initially in the state described by ψ_a, then the wave function is

$$\begin{aligned}\psi &= \exp\left(-i\frac{\varepsilon}{2}t\right)\left[\cos\Omega t + \frac{i\varepsilon}{2\Omega}\sin\Omega t\right]\psi_a\\ &+ i\frac{\Omega_R}{\Omega}\exp\left(i\frac{\varepsilon}{2}t\right)\sin\Omega t\exp\left(-\frac{\gamma}{2}\tau\right)\psi_b,\end{aligned} \tag{2}$$

where

$$\varepsilon = \Delta\omega - i\frac{\gamma}{2}, \quad \Omega = \sqrt{\frac{\varepsilon^2}{4} + \Omega_R^2}.$$

Suppose that $|\varepsilon| \gg \Omega_R$ in accordance with the experimental conditions in [1]. Then,

$$\Omega \approx \frac{\varepsilon}{2} + \frac{\Omega_R^2}{\varepsilon}$$

and

$$\begin{aligned}\psi &= \left[\left(1 - \frac{\Omega_R^2}{\Delta\omega^2}\right)\exp\left(-i\left(\frac{E_a}{\hbar} - \frac{\Omega_R^2}{\Delta\omega}\right)t\right)\right.\\ &+ \left.\frac{\Omega_R^2}{\Delta\omega^2}\exp\left(-i\left(\frac{E_a}{\hbar} + \Delta\omega - i\frac{\gamma}{2} + \frac{\Omega_R^2}{\Delta\omega}\right)t\right)\right]\psi_a^0\\ &+ \frac{\Omega_R}{\Delta\omega}\left[-\exp\left(-i\left(\frac{E_b}{\hbar} + \frac{\Omega_R^2}{\Delta\omega} - i\frac{\gamma}{2}\right)t\right)\right.\\ &+ \left.\exp\left(-i\left(\frac{E_a}{\hbar} + \omega_0 - \frac{\Omega_R^2}{\Delta\omega} - i\frac{\gamma\Omega_R^2}{2\Delta\omega^2}\right)t\right)\right]\psi_b^0,\end{aligned} \tag{3}$$

if it is additionally assumed that $\gamma \ll |\Delta\omega|$.

It is obvious that the first summand in this expression is similar to the ground-state wave function. The second summand consists of a term representing a small admixture of the excited state and a term representing a virtual state whose energy (measured in frequency units) is close to the dressing-field frequency. The virtual state is characterized by weak decay, with the rate constant

$$\tilde{\gamma} = \frac{\Omega_R^2}{\Delta\omega^2}\gamma,$$

and a small shift,

$$\tilde{\Delta}\omega = \frac{\Omega_R^2}{\Delta\omega^2}\Delta\omega.$$

Since the contribution of the virtual state is multiplied by the weighting factor $\Omega_R/\Delta\omega$, the dipole moment of the transition from this state to the ground state is

$$\tilde{d} = \frac{\Omega_R}{\Delta\omega}d.$$

One may say that the dressing beam creates a virtual excited state ψ_d with a population equal to that of the ground state ψ_a and a smaller dipole transition moment (or with equal dipole transition moment and lower population). The probe field can stimulate transition to the ground state, but the resulting translational momentum must be equal to the difference between the momenta of dressing and probe photons by virtue of momentum conservation. Initially, the state ψ_a' is not populated. Since phase memory is preserved, the process can be interpreted as stimulated (or triggered) superradiance [18, 19], and the corresponding time scale is

$$\tilde{\tau}_R = \tilde{\gamma}N_0\lambda^2 L,$$

where N_0 is the concentration of atoms, λ is the radiation wavelength, and L is the sample length in the direction of emission.

Let us consider the steady-state linear polarization induced by a monochromatic probe field of frequency ω via the $\psi_a' \longleftrightarrow \psi_d$ transition:

$$P(\omega) = \frac{d}{i\hbar}N_0\int_0^\infty \langle\psi_d\hat{d}\psi_a'\rangle E(\omega)\exp(-i\omega d)dt,$$

where N_0 is the concentration of atoms and $\hat{d}$ is the dipole moment operator. Substituting the expression for ψ_d, we obtain

$$P(\omega) = \frac{N_0\Omega_R d^2 E(\omega)}{\hbar\Delta\omega}$$

$$\times\left[\frac{-(\omega-\omega_{ba})+i\gamma/2}{(\omega-\omega_{ba})^2+\gamma^2/4}+\frac{(\omega-\omega_0)-i\tilde{\gamma}/2}{(\omega-\omega_0)^2+\tilde{\gamma}^2/4}\right] \quad (4)$$

$$= \alpha(\omega)E(\omega),$$

where $\alpha(\omega)$ is polarizability.

Since the complex refractive index is

$$n(\omega) = \sqrt{1+4\pi\alpha(\omega)},$$

we can assume that $4\pi\alpha(\omega) \ll 1$ to obtain

$$n(\omega) = n_1 - in_2 = 1 + 2\pi\mathrm{Re}\alpha(\omega) - 2\pi i\mathrm{Im}\alpha(\omega),$$

where n_1 is refractive index and n_2 is absorption index. Using (4), we obtain

$$n_1 = 1 + A\left[-\frac{(\omega-\omega_{ba})}{(\omega-\omega_{ba})^2+\gamma^2/4}+\frac{(\omega-\omega_0)}{(\omega-\omega_0)^2+\tilde{\gamma}^2/4}\right],$$
$$n_2 = A\left[-\frac{\gamma/2}{(\omega-\omega_{ba})^2+\gamma^2/4}+\frac{\tilde{\gamma}/2}{(\omega-\omega_0)^2+\tilde{\gamma}^2/4}\right], \quad (5)$$

where

$$A = \frac{2\pi N_0\Omega_R d^2}{\hbar\Delta\omega}.$$

The expressions in (5) are derived under the condition that the atom is at rest. Since the recoil momentum

$$\mathbf{p} = \hbar(\mathbf{k}_0 - \mathbf{k})$$

is transferred to an atom interacting with the dressing and probe fields, these expressions can be rewritten as

$$n_1 = 1 + A\left[-\frac{(\omega-\omega_{ba})}{(\omega-\omega_{ba})^2+\gamma^2/4}+\frac{(\omega-\omega_0)}{(\omega-\omega_0+\omega_g)^2+\tilde{\gamma}^2/4}\right],$$
$$n_2 = \left[-\frac{\gamma/2}{(\omega-\omega_{ba})^2+\gamma^2/4}+\frac{\tilde{\gamma}/2}{(\omega-\omega_0+\omega_g)^2+\tilde{\gamma}^2/4}\right], \quad (5a)$$

where $\hbar\mathbf{k}_0$ and $\hbar\mathbf{k}$ denote the momenta of the dressing and probe photons and $\hbar\omega_g = p^2/2M$ is the recoil kinetic energy for an atom of mass M.

Thus, the absorption index is the sum of two Lorentzians, with minus and plus signs corresponding to absorption and amplification, respectively. In other words, we have a superposition of an absorption spectrum centered at ω_{ba} and an amplification spectrum centered at the frequency $\omega_0 - \omega_g$ corresponding to the exact two-photon resonance condition. The absorption and amplification rates at $\omega_0 - \omega_g$ are proportional, respectively, to $\gamma/\Delta\omega^2$ and

$$\frac{1}{\tilde{\gamma}} = \left(\frac{\Delta\omega}{\Omega_R}\right)^2\frac{1}{\gamma}.$$

It is obvious that the amplification rate substantially exceeds the absorption rate, but only in a very narrow frequency range. Actually, the amplification is limited by the pulse duration and the relaxation time of the coherent atomic state. At frequencies in the vicinity of the exact two-photon resonance, the refractive index is characterized by linear normal dispersion, which obviously leads to a decrease in the group velocity of the amplified pulse.

When applied to a BEC, the results presented above should be interpreted as qualitative estimates, primarily because the BEC atoms are delocalized over the entire condensate volume, and nonlocal dependence of polarization on the probe field strength implies that neither the refractive index nor the absorption index can be well defined.

3. NONLINEAR SCHRÖDINGER EQUATION

Hereinafter, an atom is characterized not only by its electronic state, but also by the state of translational motion with a well-defined momentum. Suppose that all atoms initially are in the ground electronic state with a nearly zero momentum. Momentum conservation in atom–field interactions is formally ensured by selection rules. The matrix element corresponding to interaction between an atom and an electromagnetic wave with a wave vector $\mathbf{k}$ is zero for the atomic state with momentum differing from the initial momentum by $\hbar\mathbf{k}$. The ground and excited states φ_a and φ_b of an atom are denoted by subscripts a and b, respectively, and its translational motion is described by a momentum eigenfunction. For example, the wave-function basis chosen for analyzing a two-level atom is

$$|a;\mathbf{k}\rangle = \frac{1}{\sqrt{V}}\exp(i\mathbf{k}\cdot\mathbf{r})\varphi_a,$$
$$|b;\mathbf{k}'\rangle = \frac{1}{\sqrt{V}}\exp(i\mathbf{k}'\cdot\mathbf{r})\varphi_b\exp(-i\omega_d t), \quad (6)$$

where ω_d is the dressing-field frequency and V is the volume of the system. Note that the functions in (6) differ from the wave functions of a free atom by a phase factor in which the atom energy eigenvalue (including the kinetic energy of translational motion) is replaced by the frequency of a dressing laser field. Representation in this basis is different from the generally used

interaction representation, and the Schrödinger equation is modified accordingly.

The general expression for an atomic wave function can be written as

$$\Psi = \sum_{\substack{\alpha = a, b \\ \mathbf{k}}} C_{\alpha, \mathbf{k}} |\alpha; \mathbf{k}\rangle, \tag{7}$$

where $C_{\alpha, \mathbf{k}}$ are time-dependent coefficients.

Consider the interaction between BEC atoms and the two electromagnetic plane-wave modes E_j ($j = d, p$) with wave vectors $\mathbf{k}_j$:

$$E = i\sum_j E_j \exp(-i\omega_d t + i\mathbf{k}_j \cdot \mathbf{r}) + \text{c.c.}, \tag{8}$$

with

$$k_d = \omega_d/c, \quad k_p = \omega_p/c,$$

where ω_p is the probe frequency and c is the speed of light in free space. It is obvious that the states $|a, \mathbf{k}\rangle$ and $|b, \mathbf{k}'\rangle$ are coupled via interaction with a mode E_j only if $\mathbf{k}' - \mathbf{k} = \pm\mathbf{k}_j$.

Each mode creates a polarization of the condensate, which in turn generates a secondary field E_j' in the same mode. In the one-dimensional case, the secondary field is expressed in terms of polarization as follows:

$$E_j'(y_j, t) = \frac{2\pi i\omega_d}{c}\exp(-i\omega_d t)$$
$$\times \int_0^{D_j} P_j\left(y', t - \frac{|y_j - y'|}{c}\right) \tag{9}$$
$$\times \exp(ik_j|y_j - y'|)dy' + \text{c.c.},$$

where $P_j(y', t)$ denotes the complex amplitudes of polarization, y_j is the coordinate corresponding to a particular mode, and D_j is the condensate length along the coordinate. The polarization amplitudes are expressed in terms of expansion coefficients as the quantum average values of the corresponding polarization operators:

$$P_j(\mathbf{r}) = \frac{1}{V}\sum_i^N \sum_{\alpha\beta} \bar{C}_\alpha^{(i)} C_\beta^{(j)} \langle\alpha|\hat{d}_i \delta(\mathbf{r} - \mathbf{r}_i)|\beta\rangle, \tag{10}$$

where the bar denotes a complex conjugate, N is the total number of atoms in the condensate, $\hat{d}_i$ is the dipole moment operator for the ith atom, and $C_\alpha^{(i)}$ and $C_\beta^{(i)}$ are the expansion coefficients for the ith atom. The subscripts α and β correspond to a, $\mathbf{k}'$ and b, $\mathbf{k}$. The summation is performed only over α and β such that $\mathbf{k} - \mathbf{k}' = \mathbf{k}_j$. Since the expansion coefficients are equal for every atom in the condensate, the polarization complex amplitudes corresponding to the two modes are finally expressed as

$$P_j(y', t) = N_0 d \exp(ik_j y') \sum_{\mathbf{k} - \mathbf{k}' = \mathbf{k}_j} C_{b, \mathbf{k}} \bar{C}_{a, \mathbf{k}'}, \tag{11}$$

where N_0 is the concentration of atoms in the condensate and d is the dipole transition matrix element.

Neglecting delay in the polarization amplitude and contribution of fast-oscillating terms in the integral with respect to y', we obtain

$$E_j' = \frac{2\pi i d\omega_d N_0 y_j}{c}$$
$$\times \exp(-i\omega_d t + ik_j y_j) \sum_{\mathbf{k}' - \mathbf{k} = \mathbf{k}_j} C_{b, \mathbf{k}'} \bar{C}_{a, \mathbf{k}}. \tag{12}$$

Thus, the field generated by a uniform polarization wave is linear in y_j. Now, we replace these variables with their average values, setting $y_d = y_p = D/2$, where D is the transverse size of the condensate.

Representing the atom–field interaction operator as

$$\hat{H}_{\text{int}} = -\hat{d}E, \tag{13}$$

we write the Schrödinger equation for a single atom. In the rotating-wave approximation, we have

$$\dot{C}_{a, \mathbf{k}} = \frac{d}{\hbar}\sum_j \bar{E}_j C_{b, \mathbf{k} + \mathbf{k}_j} - i\frac{W_k}{\hbar} C_{a, \mathbf{k}},$$
$$\dot{C}_{b, \mathbf{k}} = -\frac{d}{\hbar}\sum_j E_j C_{a, \mathbf{k} - \mathbf{k}_j} \tag{14}$$
$$-\left(\frac{\gamma}{2} + i\frac{W_k}{\hbar} + i\Delta\omega\right) C_{b, \mathbf{k}},$$

with

$$E_j = E_j^0 + \frac{\hbar}{d\tau_R} \sum_{\mathbf{k}' - \mathbf{k} = \mathbf{k}_j} C_{b, \mathbf{k}'} \bar{C}_{a, \mathbf{k}}, \tag{15}$$

where

$$\tau_R = c\hbar/\pi\omega_d d^2 N_0 D,$$

$\Delta\omega = \omega_{ba} - \omega_d$ is detuning, W_k is the kinetic energy of an atom with momentum $\mathbf{k}$, γ is the emission rate constant for the excited state (characterizing spontaneous light scattering in any direction), and E_j^0 denotes the incident-wave amplitudes.

Equations (14) and (15) can be substantially simplified when applied to the experimental conditions of [1]. Indeed, since $\Delta\omega$ is the largest parameter in this case, it can be assumed (see [20]) that

$$|\dot{C}_{b,\mathbf{k}}| \ll \Delta\omega |C_{b,\mathbf{k}}|.$$

Under this condition, we can drop the derivative on the left-hand side of (14) and write

$$C_{b,\mathbf{k}} = \frac{d(i-\gamma')}{\hbar\Delta\omega}\sum_j E_j C_{a,\mathbf{k}-\mathbf{k}_j},$$

using the relations

$$W_k \ll \hbar\Delta\omega, \quad \gamma' = \gamma/2\Delta\omega.$$

After $C_{b,\mathbf{k}}$ is eliminated, Eqs. (14) and (15) become

$$\dot{C}_{a,\mathbf{k}} = \frac{d^2(i-\gamma')}{\hbar^2\Delta\omega} \times \sum_{j,j'} \bar{E}_j E_{j'} C_{a,\mathbf{k}+\mathbf{k}_j-\mathbf{k}_{j'}} - i\frac{W_k}{\hbar}C_{a,\mathbf{k}}, \tag{16}$$

$$E_j = E_j^0 + \frac{i-\gamma'}{\Delta\omega\tau_R}\sum_{j'} E_{j'} \sum_{\mathbf{k}'-\mathbf{k}=\mathbf{k}_j} C_{a,\mathbf{k}'-\mathbf{k}_{j'}}\bar{C}_{a,\mathbf{k}}. \tag{17}$$

Since the field amplitudes are expressed in terms of the coefficients of wave-function expansion in terms of the basis chosen here, Eq. (16) is a nonlinear Schrödinger equation.

4. AMPLIFICATION OF COHERENT LIGHT

In an experiment with a finite-size trap and a relatively strong nonradiative coupling between atomic states, the amplification of the probe beam involves relaxation of atomic states. To allow for the relaxation, we should change from the model based on the Schrödinger equation to the Maxwell–Bloch equations. Relaxation of coherent atomic states is another reason for neglecting secondary excitations, and we restrict our analysis to first-order excitations. Figure 1 schematizes the corresponding levels and transitions. Here, the ground and excited states are labeled by odd and even numbers, respectively.

Henceforth, we use the "true" amplitude of the probe field:

$$E_p \longrightarrow E_p \exp[i(\omega_d - \omega_p)t].$$

Furthermore, we redefine the expansion coefficients,

$$C_3 \longrightarrow C_3 \exp[-i(\omega_d - \omega_p)t],$$

$$C_5 \longrightarrow C_5 \exp[i(\omega_d - \omega_p)t]$$

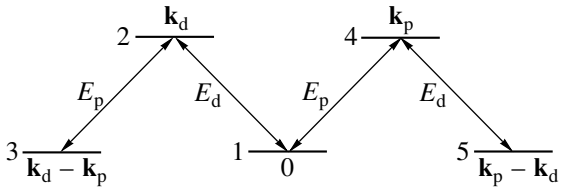


Fig. 1. Atomic states and transitions involved in probe-pulse amplification in the presence of a dressing pulse.

and change to the dimensionless time, field amplitude, and energy

$$\tau = \frac{t}{\tau_s}, \quad \varepsilon_d = \frac{E_d}{E_m}, \quad \varepsilon_p = \frac{E_p}{E_m}, \quad w_k = \frac{W_k}{\hbar\tau_s^{-1}}, \tag{18}$$

where the characteristic time scale is

$$\tau_s = \frac{\Delta\omega}{\Omega_R^2}, \quad \Omega_R = \frac{dE_m}{\hbar} \tag{19}$$

(E_m is the highest field amplitude attained in the experiment). Then, we combine Eqs. (16) and (17) with a phenomenological model of phase relaxation to write the equations for the density matrix

$$R_{\mathbf{k},\mathbf{k}'} = C_{a,\mathbf{k}}\bar{C}_{a,\mathbf{k}'},$$

as

$$\begin{aligned}
\dot{R}_{13} &= -\gamma'[(2|\varepsilon_p|^2 + |\varepsilon_d|^2)R_{13} \\
&+ \varepsilon_p\bar{\varepsilon}_d(R_{33} + R_{11}) + \varepsilon_d\bar{\varepsilon}_p\bar{R}_{35}] \\
&+ i[|\varepsilon_d|^2 R_{13} + \varepsilon_p\bar{\varepsilon}_d(R_{33} - R_{11}) \\
&+ \varepsilon_d\bar{\varepsilon}_p\bar{R}_{35}] - iq_3R_{13} - \gamma_{13}R_{13},
\end{aligned}$$

$$\begin{aligned}
\dot{R}_{15} &= -\gamma'[(2|\varepsilon_d|^2 + |\varepsilon_p|^2)R_{15} \\
&+ \varepsilon_d\bar{\varepsilon}_p(R_{55} + R_{11}) + \varepsilon_p\bar{\varepsilon}_d R_{35}] \\
&+ i[|\varepsilon_p|^2 R_{15} + \varepsilon_d\bar{\varepsilon}_p(R_{55} - R_{11}) \\
&+ \varepsilon_p\bar{\varepsilon}_d R_{35}] - iq_5R_{15} - \gamma_{15}R_{15},
\end{aligned}$$

$$\begin{aligned}
\dot{R}_{35} &= -\gamma'[(|\varepsilon_p|^2 + |\varepsilon_d|^2)R_{35} + \varepsilon_d\bar{\varepsilon}_p(R_{15} + \bar{R}_{13})] \\
&+ i[(|\varepsilon_p|^2 - |\varepsilon_d|^2)R_{35} + \varepsilon_d\bar{\varepsilon}_p(R_{15} - \bar{R}_{13})] \\
&+ i(q_3 - q_5)R_{35} - \gamma_{35}R_{35},
\end{aligned} \tag{20}$$

$$\dot{R}_{11} = -\gamma'[2|\varepsilon_d|^2 + |\varepsilon_p|^2 R_{11}$$

$$+ \varepsilon_d\bar{\varepsilon}_p(R_{13} + \bar{R}_{15}) + \varepsilon_p\bar{\varepsilon}_d(R_{15} + \bar{R}_{13})]$$
$$+ i[\varepsilon_d\bar{\varepsilon}_p(\bar{R}_{15} - R_{13}) + \varepsilon_p\bar{\varepsilon}_d(\bar{R}_{13} - R_{15})],$$

$$\dot{R}_{33} = -\gamma'[2|\varepsilon_p|^2 R_{33} + \varepsilon_d\bar{\varepsilon}_p R_{13} + \varepsilon_p\bar{\varepsilon}_d\bar{R}_{13}]$$
$$+ i(\varepsilon_d\bar{\varepsilon}_p R_{13} - \varepsilon_p\bar{\varepsilon}_d)\bar{R}_{13},$$

$$\dot{R}_{55} = -\gamma'[2|\varepsilon_d|^2 R_{55} + \varepsilon_p\bar{\varepsilon}_d R_{15} + \varepsilon_d\bar{\varepsilon}_p\bar{R}_{15}]$$
$$+ i(\varepsilon_p\bar{\varepsilon}_d R_{15} - \varepsilon_d\bar{\varepsilon}_p\bar{R}_{15}).$$

The field amplitudes ε_p and ε_d can be found by solving the system of linear equations

$$\varepsilon_p = \varepsilon_p^0 + \frac{i-\gamma'}{\Delta}[\varepsilon_p(R_{11} + R_{33}) + \varepsilon_d(R_{13} + \bar{R}_{15})],$$
$$\varepsilon_d = \varepsilon_d^0 + \frac{i-\gamma'}{\Delta}[\varepsilon_d(R_{11} + R_{55}) + \varepsilon_p(\bar{R}_{13} + R_{15})], \quad (21)$$

with

$$\gamma' = \gamma/2\Delta\omega, \quad q_3 = (\omega_d - \omega_p - W_3/\hbar)\tau_s,$$

$$q_5 = (\omega_p - \omega_d - W_5/\hbar)\tau_s, \quad \gamma_{ik} = \Gamma_{ik}\tau_s, \quad \Delta = \Delta\omega\tau_R,$$

where Γ_{ik} $(i \neq k = 1, 3, 5)$ denotes off-diagonal transverse relaxation rates. Experimental estimates for $\Gamma_{ik} \approx \Gamma$ are on the order of 10^4 s^{-1} [5]. The only nonzero initial condition is

$$R_{11}(0) = 1.$$

Note that Eqs. (21) determine fields averaged over the sample, since the secondary fields are linear in coordinate according to (12). Equations (21) can also be used to find the output values of the fields by introducing a factor of 2 into the second summands.

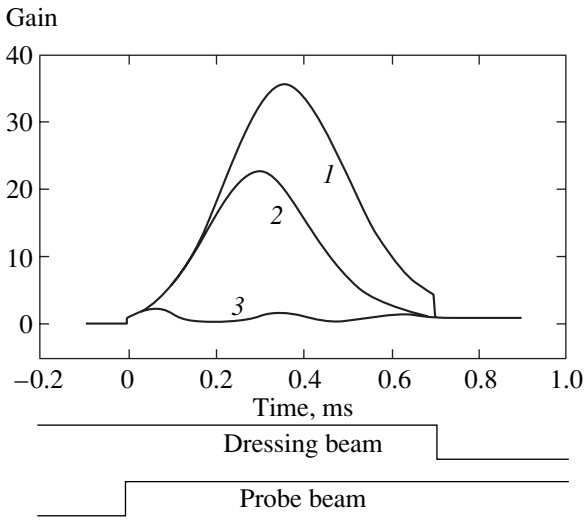


Fig. 2. Probe-pulse output from a dressed condensate for probe light intensities of (*1*) 0.0025, (*2*) 0.005, and (*3*) 0.5 mW/cm^2 and a dressing-beam intensity of 5 mW/cm^2. The curves are normalized to the incident probe intensity. Curve *3* is scaled up by a factor of 2. The probe- and dressing-pulse input profiles are shown at the bottom.

Let us estimate the quantities contained in Eqs. (20) and (21). Suppose that $E_m = 1.45 \times 10^{-2}$ CGS units, which corresponds to laser intensity $I = 100$ mW/cm^2. Since $d = 0.64 \times 10^{-17}$ CGS units and $\Delta\omega = 1.07 \times 10^{10}$ s^{-1}, the reference time and energy are $\tau_s = 1.37 \times 10^{-6}$ s and $\hbar\tau_s^{-1} = 0.77 \times 10^{-21}$ ergs, respectively. Set $N = 5 \times 10^6$, which corresponds to $N_0 = 0.8 \times 10^{14}$ cm^{-3} for the condensate dimensions indicated above. Since $\omega_d = 3.2 \times 10^{15}$ s^{-1} and $D = 20$ μm, we obtain $\tau_R = 0.48 \times 10^{-9}$ s and $\gamma = 0.63 \times 10^8$ s^{-1}. The recoil kinetic energy of a sodium atom that has just emitted or absorbed a photon with $\lambda = 589$ nm is $W = 1.64 \times 10^{-22}$ ergs. Thus, the characteristic pulse width under experimental conditions (500 μs) is about $365\tau_s$, $\gamma' = 0.30 \times 10^{-2}$, and $\Delta = 5.2$. The corresponding energy values are $w_3 = w_5 = 3.42W/\hbar\tau_s^{-1} = 0.73$.

In what follows, we treat the dressing field as prescribed: $\varepsilon_d = \varepsilon_d^0$. Computations based on Eqs. (20) and (21) have shown that level 5 remains unpopulated under experimental conditions even if $\varepsilon_p \sim \varepsilon_d$. According to the equation for R_{15} in (20), this is explained by a strong disparity between the 1–5 transition frequency and the two-photon transition frequency $\omega_d - \omega_p$. For this reason, we henceforth assume that the level 5 population is negligible. Then, Eqs. (20) and (21) reduce to

$$\dot{R}_{13} = -\gamma'[(|\varepsilon_p|^2 + |\varepsilon_d|^2)R_{13} + \varepsilon_p\bar{\varepsilon}_d(R_{33} + R_{11})]$$
$$+ i\varepsilon_p\bar{\varepsilon}_d(R_{33} - R_{11}) - i[q_3 - |\varepsilon_d|^2 + |\varepsilon_p|^2 + \gamma_{13}]R_{13},$$

$$\dot{R}_{11} = -\gamma'[2|\varepsilon_d|^2 R_{11} + \varepsilon_d\bar{\varepsilon}_p R_{13} + \varepsilon_p\bar{\varepsilon}_d\bar{R}_{13}]$$
$$+ i(\varepsilon_p\bar{\varepsilon}_d\bar{R}_{13} - \varepsilon_d\bar{\varepsilon}_p R_{13}), \quad (22)$$

$$\dot{R}_{33} = -\gamma'[2|\varepsilon_p|^2 R_{33} + \varepsilon_d\bar{\varepsilon}_p R_{13} + \varepsilon_p\bar{\varepsilon}_d\bar{R}_{13}]$$
$$+ i(\varepsilon_d\bar{\varepsilon}_p R_{13} - \varepsilon_p\bar{\varepsilon}_d\bar{R}_{13}),$$

$$\varepsilon_p = \frac{\varepsilon_p^0 + ((i-\gamma')/\Delta)\varepsilon_d R_{13}}{1 - ((i-\gamma')/\Delta)R_{33}}. \quad (23)$$

Now, let us discuss the principal effects revealed by analyzing the proposed model. As in experiments, these effects are manifested in computations only in the presence of the dressing beam. Moreover, these manifestations strongly depend on the shape and intensity of the probe pulse.

First, let us assume that the probe pulse incident on a dressed condensate is a step function. In agreement with experiment, the probe pulse is amplified as it propagates into the condensate. The corresponding gain

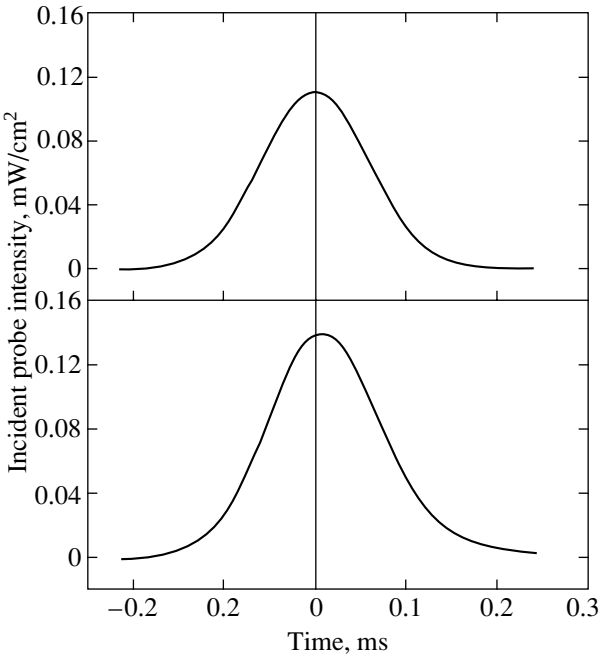


Fig. 3. Pulse delay due to light amplification in a dressed condensate for a Gaussian-shaped incident pulse with duration of 0.14 ms and peak intensity of 0.11 mW/cm^2 in the presence of a dressing beam with an intensity of 5.0 mW/cm^2.

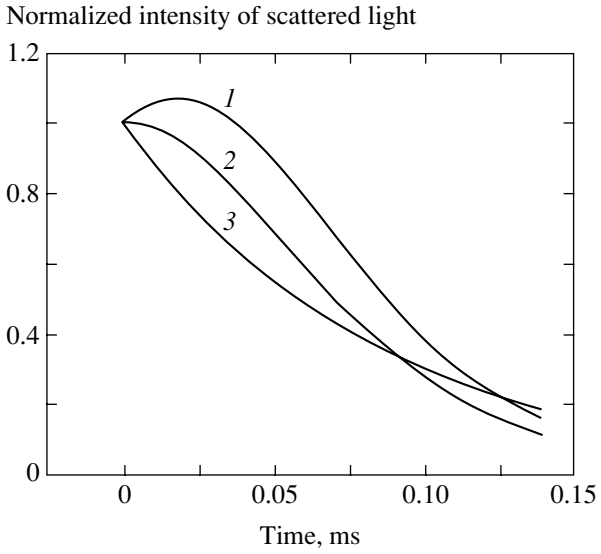


Fig. 4. Intensity of light diffracted by the grating of recoiling atoms created in the dressed condensate with a rectangular Bragg pulse with duration of 0.1 ms and intensity of 1 mW/cm^2 for dressing-beam intensities of (*1*) 2.9, (*2*) 11.5, and (*3*) 13 mW/cm^2. The curves are normalized to the intensity of diffracted light at $t = 0$ (when the pulse was switched off).

strongly depends on the detuning q_3 from the two-photon resonance. According to the equation for R_{13} in (22), an exact two-photon resonance can be ensured only by making use of the fact that the 1–3 transition frequency is shifted by interaction with the fields ε_p and ε_d. Thus, exact two-photon resonance requires that

$$\omega_d - \omega_p = (w_3 + |\varepsilon_d|^2)\tau_s^{-1} = (0.73 + 0.05)\tau_s^{-1} = 0.78\tau_s^{-1}.$$

This value is exactly equal to the frequency difference set in the experiment ($I_d = 5$ mW/cm^2). The gain for the probe pulse depends not only on the detuning, but also on the probe input intensity (see Fig. 2). The highest gain is obtained for low E_p^0, when the concentration of excited atoms is small. With increasing probe intensity, the gain decreases and the level 3 population increases. When $E_p^0 \sim E_d^0$, both gain and concentration of excited atoms are determined by the Rabi beat associated with the 1–3 transition.

The probe pulse propagating through a dressed condensate is not only amplified, but also slowed down. This effect is most clearly manifested when the probe pulse has a Gaussian profile (see Fig. 3). The slowing-down increases with decreasing transverse relaxation rate Γ. When $\Gamma = 0$ and 10^4 s^{-1}, the group velocity of a probe pulse propagating in a dressed condensate is 1 and 2 m/s, respectively. A probe pulse propagating through a dressed condensate is not only amplified and slowed down, but also substantially broadened. These results are in complete agreement with experiment.

Let us turn back to a probe pulse with a steep leading front. In the presence of a high-intensity dressing beam, a considerable number of atoms are excited to level 3. The rate of this process increases with probe front steepness, and a grating of recoiling atoms will remain after the probe beam is switched off. The dressing beam still incident on the condensate will be diffracted by the grating, and probe output will be produced (see Fig. 4). This effect becomes more pronounced as the incident probe intensity increases. This "holographic" effect was observed experimentally.

5. CONCLUSIONS

We emphasize the role played by the Bose–Einstein condensate in the phenomenon discussed here. The system is a dilute gas of relatively low density 10^{13}–10^{14} cm^{-3} at an ultralow temperature $T \sim 10^{-7}$ K. According to an estimate obtained in [16], the chemical potential corresponding to the experimental conditions is lower than the recoil energy by two orders of magnitude. Therefore, collisional interactions can be neglected, and the condensate can be treated as an ideal gas. In the experiment, the amplified pulses had a relatively long duration (~500 μs) and resonance effects were revealed with a resolution finer than the recoil kinetic energy. The existence of these resonances was

also manifested in our computations. The absence of Doppler broadening for a pulse interacting with the initial state of the condensate is essential for these observations, because the Doppler width for a "classical" ideal gas at the phase-transition temperature ($\sim 10^{-7}$ K) would be comparable to the recoil energy. The influence of collisional interaction was taken into account by introducing a phenomenological transverse-relaxation constant Γ. According to our computations, the gain for the probe pulse is lower by an order of magnitude as compared to the case of $\Gamma = 0$ even if the constant is relatively small ($\Gamma \sim 10^4$ s^{-1}). It was noted above that relaxation contributes to the slowing-down of the amplified pulse. The holographic effect described above is obviously affected by transverse relaxation. Our computations were performed for $\Gamma = 3 \times 10^3$ s^{-1}. When this value was increased by an order of magnitude, no residual probe output was observed.

In this study, the states of translational motion of an atom were described by using momentum-operator eigenfunctions with eigenvalues determined by selection rules for an atom interacting with dressing and probe electromagnetic plane waves. This made it possible to assume that the amplitudes of the corresponding polarization waves are uniform in space and are determined by the fields averaged over the sample volume. In further work, the basis of atomic states representing translational motion will be extended in order to abandon the mean-field approximation and thus estimate its influence on the effects analyzed above.

ACKNOWLEDGMENTS

We thank W. Ketterle and V.I. Perel for their interest in this study and to A.S. Troshin for a detailed discussion and helpful remarks. This work was supported by the Ministry of Education of the Russian Federation and by INTAS, project no. 01-855.

REFERENCES

1. S. Inouye, R. F. Low, S. Gupta, *et al.*, Phys. Rev. Lett. **85**, 4225 (2000).
2. M. H. Anderson, J. R. Ensher, M. R. Matthews, *et al.*, Science **269**, 198 (1995).
3. D. S. Hall, M. R. Matthews, C. E. Wieman, and E. A. Cornell, Phys. Rev. Lett. **81**, 1543 (1998).
4. J. Stenger, S. Inouye, D. M. Stamper-Kurn, *et al.*, Phys. Rev. Lett. **82**, 4569 (1999).
5. S. Inouye, A. P. Chikkatur, D. M. Stamper-Kurn, *et al.*, Science **285**, 571 (1999).
6. L. V. Hau, S. E. Harris, Z. Dutton, and C. H. Behoroozi, Nature **397**, 594 (1999).
7. F. Dalfovo, S. Giorgini, L. P. Pitaevskii, and S. Stringari, Rev. Mod. Phys. **71**, 463 (1999).
8. Ph. W. Courteille, V. S. Bagnato, and V. I. Yukalov, Laser Phys. **11**, 659 (2001).
9. G. Lenz, P. Meystre, and E. M. Wright, Phys. Rev. A **50**, 1681 (1994).
10. Y. Castin and K. Mölmer, Phys. Rev. A **51**, R3426 (1995).
11. J. Javanainen, Phys. Rev. Lett. **75**, 1927 (1995).
12. M. G. Moore and P. Meystre, Phys. Rev. Lett. **83**, 5202 (1999).
13. O. E. Mustecaplioglu and L. You, Phys. Rev. A **62**, 063615 (2000).
14. N. Piovella, M. Gatelli, and R. Bonifacio, Opt. Commun. **194**, 167 (2001).
15. E. D. Trifonov, Zh. Éksp. Teor. Fiz. **120**, 1117 (2001) [JETP **93**, 969 (2001)].
16. E. D. Trifonov, Laser Phys. **12**, 211 (2002).
17. E. D. Trifonov, Opt. Spektrosk. **92**, 631 (2002) [Opt. Spectrosc. **92**, 577 (2002)].
18. R. H. Dicke, Phys. Rev. **93**, 99 (1954).
19. M. G. Benedict, A. M. Ermolaev, V. A. Malyshev, I. V. Sokolov, and E. D. Trifonov, *Superradiance: Multiatomic Coherent Emission* (IOP, Bristol and Philadelphia, 1996).
20. E. D. Trifonov and N. I. Shamrov, Opt. Spektrosk. **96**, 294 (2004) [Opt. Spectrosc. **96**, 258 (2004)].

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 51–60.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 63–74.

ATOMS, SPECTRA, RADIATION

The Method of Polarization Tomography of Radiation in Quantum Optics

V. P. Karassiov and A. V. Masalov

Lebedev Physical Institute, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 119991 Russia

e-mail: masalov@mail1.lebedev.ru

Received November 25, 2003

Abstract—The essence and basic structural elements of the method of polarization tomography of quantum radiation (the reconstruction of the polarization states of a field solely from polarization measurements) are described. The essential features of the method are discussed in comparison with usual field and spin tomographies. A general scheme for experimentally implementing this method is suggested and its use in polarization tomography of biphotonic radiation with hidden polarization is briefly analyzed. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In recent years, a description of the quantum properties of light in the language of quasi-probabilistic distributions, in addition to the language of wave functions and density matrices, has become increasingly attractive because of its clearness and closeness to classical intuition [1–4]. There is a strong demand for a clear description of the properties of quantum objects because of extensive studies of nonclassical states and quantum light effects (e.g., see [1, 4–6] and references therein) and rapid development of quantum computation ideas, which make quantum concepts increasingly popular among not only physicists but also engineers [7–10]. Quasi-probability distributions are similar to classical joint probability distributions of several field variables. As density operator projections in the phase space representation [3], they, however, give a rigorous description of the quantum state of an object. If the quasi-probability function is selected appropriately, its "slices" determine truly probabilistic (so-called marginal) distributions of some measurable quantum observable. For example, the Wigner function, which gives a description closest to classical intuition but retains quantum information [1, 2], can be such a selection when light field quadratures are described. Quantum tomography relates quasi-probability distributions to measurement results that characterize the quantum state of an object. At the same time, the procedure for the reproduction of quasi-probability functions from experimentally recorded distributions of observables is a fairly complex problem both mathematically and experimentally [11–17].

Currently, the most elaborate and best implemented experimentally tomography method in quantum optics is field tomography of single-mode radiation [11, 12, 14]. It is based on the tomographic representation of the density operator in the form of an integral expansion in some basis operators (tomographic kernels) [14]. The coefficients of this expansion are the probability distributions of field observables (quadratures) obtained in homodyne measurements [12]. Single-mode radiation models, however, appear to be inadequate in optics. This creates a need for quasi-classical analysis and quantum tomography of multimode fields, for which not only a complete description of the state but also partial characterization of states in terms of certain collective variables is of interest [4, 6, 15–18]. Accordingly, partial tomography of multimode fields, including the experimental reconstruction of their reduced states, also makes sense [15].

This paper is concerned with quantum tomography of the polarization state of light described by quasi-probability distributions in the three-dimensional space of variables $p_{i=1,2,3}$ that are associated with the Stokes vector parameters [15, 16]. Polarization quasi-probability functions are of interest because they are an effective tool for studying many nonclassical states and polarization optics phenomena (see, e.g., [4, 6] and references therein). Moreover, several recent experiments on nonclassical polarized light states [19–21] revealed gaps in theory and the lack of rigorous means for obtaining quasi-classical images of the quantum state of generated radiation.

The general scheme for quantum tomography of polarization light states is similar to that of field tomography of single-mode radiation. Its particular implementations, however, have important special features. Indeed, the polarization variables p_i (and the corresponding $\hat{P}_i$ operators) are functions of the quadratures of several (at least two) radiation modes and, in addition, do not constitute a complete set of field variables [15]. A

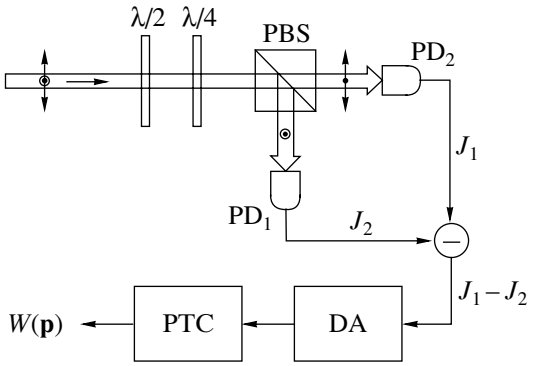


Fig. 1. Scheme of measuring polarization observables of type (8): PBS is the polarization beam splitter, PD_1 and PD_2 are the photodetectors, DA is the analyzer of the distribution of the amplitudes of the difference photocurrent, PTC is the polarization tomography converter that transforms the set of measurement data into the quasi-probability distribution $W(\mathbf{p})$ or $W(\mathbf{s})$.

theoretical apparatus of polarization tomography can be consistently constructed by generalizing single-mode field tomography to the multimode case [13]. We must then reparametrize the quadratures to distinguish between polarization (p_i) and nonpolarization variables and perform the reduction of the latter [15, 17]. This approach is obviously computationally redundant, and its experimental analog is complex to implement (because of the use of multimode homodynation technique).

The alternative approach [15], which can be called the method of polarization tomography of radiation proper, removes these shortcomings. This is the approach on which the present work relies. It involves the direct determination of polarization quasi-probability functions with the use of the *P*-quasi-spin formalism [22, 23] and their experimental reconstruction from probability distributions obtained in polarization measurements [16]. The purpose of this work is to clarify the physical nature and basic equations of the method of polarization tomography of radiation,[1] briefly discuss its special features, and give an example of its experimental implementation. An important problem of polarization tomography is the selection of the polarization quasi-probability function. As in field tomography, we can introduce many polarization quasi-probability functions, each with the necessary mathematical properties. However the best combination of visual informativeness and closeness to classical intuition is only provided by some of them, in field tomography, by the Wigner function. In this work, we selected one of the possible quasi-probability functions from considerations of its mathematical simplicity and the amount of information it contains.

[1] Bearing in mind that, in [15], attention was largely given to some mathematical aspects of the method.

2. THEORETICAL FOUNDATIONS OF POLARIZATION TOMOGRAPHY

The main object of quantum polarization tomography is the quasi-probability distribution of p_i variables, which are associated with the Stokes s_i parameters known in traditional classical optics, in the three-dimensional space. As has been mentioned, such distributions visualize intuitive ideas of quantum fluctuations of radiation and can be used to obtain qualitative and quantitative estimates of quantum fluctuation manifestations in polarization measurements [4, 6]. Note that three measurements are necessary for representing the polarization properties of radiation. The three-dimensional nature of the polarization space of states signifies the transition beyond the two-dimensional Poincaré sphere, which cannot be used to describe the polarization properties of radiation of an indefinite intensity or the properties of nonpolarized light. In the language of observables, data for constructing quasi-probability distributions are extracted from the probability distributions of the eigenvalues S_i of quantum polarization observables $\hat{S}_i = 2\hat{P}_i$ and their linear combinations. The variables S_i, which are proportional to the Stokes parameters s_i for plane-wave monochromatic radiation,[2] are measured using a polarization beam splitter introduced into the light beam to be studied and two ideal photodetectors, whose currents are subtracted from each other. The diversity of variables is determined by a phase plate in front of the beam splitter and the orientation of the beam splitter itself (see Fig. 1). In a narrow sense, polarization tomography is the reproduction of polarization quasi-probability from the specified simplest polarization measurements. There is an obvious analogy between polarization tomography and quadrature (field) tomography developed to study the nonclassical properties of squeezed and other light states. Quadrature tomography also requires recording distributions of many quadratures, and its result is the reproduction of the quasi-probability distribution of light field.

Let us consider the basic elements of the mathematical foundations of polarization tomography. The quantum polarization characteristic function should be mentioned first. It is generated by the shift operator in the space of polarization variables. On the one hand, this characteristic function provides information about the probability distribution of an arbitrary polarization observable that depends on $\hat{S}_i$. On the other, its three-dimensional integral transformation (with some kernel) is a quasi-probability distribution. It follows that the polarization characteristic function relates observable values to quasi-probability. The introduction of the characteristic function, however, does not solve two important problems. The first one is the selection of the

[2] For simplicity, we also call them Stokes variables in the general case when there is no such relation [22].

type of the quasi-probability function, which reduces to the selection of the type of the transformation of the characteristic function. In practice, this selection is made by seeking the quasi-probability function that provides an optimum balance between the amount of quantum information and the classical character of polarization representation. Next, there is another (probably more important) problem, that of the "classical" images of the operators of polarization observables (analogs of Weyl symbols), which are necessary for obtaining the classical form of averaging these observables with the use of quasi-probability functions [24].

Recall that the construction of such quasi-classical images in quadrature tomography can be reduced to the expansion of either the density operator or the operators of arbitrary field observables in the continuous basis of Wigner operators $\hat{W}(\beta)$ with projection onto this basis [1]. The Wigner operators are given by the Fourier transforms of the field shift operator $\hat{D}(\beta) = \exp(\beta\hat{a}^{\dagger} - \beta^{*}\hat{a})$ (which in turn determines the field characteristic function $\chi(\beta;\ \hat{\rho}) = \langle \hat{D}^{\dagger}(\beta)\rangle$ [11]). The continuous family of Wigner operators plays a role similar to that of the family of Glauber coherent states in quasi-classical representations of pure states of the Hilbert electromagnetic field space, initially given in the discrete Fock basis. Note that this expansion of the field density operator and the expansion specified in the introduction ("tomographic" expansion) are key elements of the mathematical apparatus of usual field tomography [14]. Similarly, the second basic element of polarization tomography is the determination of the polarization density operator, which characterizes polarization [only (!)] degrees of freedom of light fields, and the construction of its tomographic and quasi-classical expansions (analogs of the corresponding expansions for the complete field density operator) [15]. A key role is played by the tomographic representation of the polarization density operator, which contains probability distributions of the set of Stokes variables as expansion coefficients. This representation is central to polarization tomography, which is broadly understood as the reproduction of the polarization density operator from measurement data. However note that, because of the special features of commutation relations for the three $\hat{P}_i$ operators, we cannot merely transfer all formal constructions of quadrature tomography to polarization tomography. This in particular refers to the determination of the quasi-classical operator basis that allows not only a suitable quasi-probability function to be selected but also an adequate classical image of any polarization operator to be constructed. Below, we define the specified basic elements of the mathematical foundations of polarization tomography and clarify their physical meaning.

An adequate description of the polarization structure of quantum light fields and the corresponding operators of the values that specify this structure and observables in polarization experiments is of key importance for our purposes. One would think that the role of such basis observables in polarization quantum optics can naturally be played by three Stokes vector operators that correspond to observables in standard polarization experiments of classical optics,

$$\hat{E}_x^{(-)}\hat{E}_x^{(+)} - \hat{E}_y^{(-)}\hat{E}_y^{(+)}, \quad \hat{E}_x^{(-)}\hat{E}_{x'}^{(+)} - \hat{E}_{y'}^{(-)}\hat{E}_{y'}^{(+)},$$

$$\hat{E}_+^{(-)}\hat{E}_+^{(+)} - \hat{E}_-^{(-)}\hat{E}_-^{(+)}.$$

Here, $\hat{E}_\alpha^{(\pm)}$ are the operators of positive-frequency (negative-frequency) field components with polarization α and the indices x, y, x', y', $-$, and + number pairs of orthogonal modes (connected by unitary transformations) in two types of linear and circular radiation polarizations. Like their classical prototypes, these operators are, however, determined correctly only for plane-wave radiation. In addition, they depend on time for multifrequency fields and describe "beats" of frequency modes, which is not directly related to polarization. The description of the polarization structure of quantum radiation suggested by one of these authors [22] is free of these shortcomings. This description uses the $\hat{P}_j$ components of the P-quasi-spin operator $\hat{\mathbf{P}}$ rather than Stokes operators. In the general scheme of quantum electrodynamics, these components replace usual spin components (the usual spin loses its physical meaning in massless fields because of the absence of gauge invariance). For monochromatic plane-wave radiation (a single space-time mode), when there are only two polarization radiation modes with equal wave vectors $\mathbf{k}$, these operators are determined via the differences of the photon number operators [6],

$$\begin{aligned} 2\hat{P}_1 &= \hat{n}_x - \hat{n}_y, \quad 2\hat{P}_2 = \hat{n}_{x'} - \hat{n}_{y'}, \\ 2\hat{P}_3 &= \hat{n}_+ - \hat{n}_-, \quad \hat{n}_\alpha = \hat{a}_\alpha^{\dagger}\hat{a}_\alpha. \end{aligned} \tag{1}$$

Note that the $2\hat{P}_i$ operators then coincide (to within an unimportant factor) with the Stokes operators, whose averages equal the Stokes vector parameters. In addition, the $2\hat{P}_3$ operators coincide with the helicity operator. The components of the P-quasi-spin operator for several spacetime modes are the sums of the corresponding operators over all modes, but the simple relation with Stokes operators mentioned above is then absent even for quasi-monochromatic radiation.

The $\hat{P}_i$ operators satisfy the commutation relations $[\hat{P}_a,\ \hat{P}_b] = i\epsilon_{abc}\hat{P}_c$ of the $su(2)$ algebra [22], which allows us to effectively use this algebra for the polarization characterization of the quantum states of Hilbert (Fock) light field spaces and the derivation of the basic equations of polarization tomography [15]. In particu-

lar, the diversity of the polarization properties of the quantum states of radiation is described using the $\{|P, \mu\rangle\}$ basis of the eigenstates of the operators $\hat{\mathbf{P}}^2 \equiv \hat{P}_1^2 + \hat{P}_2^2 + \hat{P}_3^2 \equiv \hat{P}^2 + \hat{P}$ and $\hat{P}_3$. The basis of the polarization states of radiation $\{|P, \mu\rangle\}$ in the Hilbert space of Fock states is numbered by the quantum numbers P and μ, which determine the P-quasi-spin value and the value of its projection onto the radiation direction (half the total helicity of the field), respectively,

$$\hat{P}|P, \mu\rangle = P|P, \mu\rangle,$$
$$\hat{\mathbf{P}}^2 \equiv \hat{P}_1^2 + \hat{P}_2^2 + \hat{P}_3^2 \equiv \hat{P}^2 + \hat{P}, \quad (2)$$
$$\hat{P}_3|P, \mu\rangle = \mu|P, \mu\rangle.$$

This fixes the electromagnetic field polarization structure [22].

In the simplest case of a single space-time mode when the relation

$$\hat{\mathbf{P}}^2 = \hat{P}^2 + \hat{P} = \hat{N}^2/4 + \hat{N}/2 \longleftrightarrow \hat{P} = \frac{\hat{N}}{2} \quad (3)$$

between the P-quasi-spin operators and the total photon number operator

$$\hat{N} = \hat{n}_x + \hat{n}_y = \hat{n}_{x'} + \hat{n}_{y'} = \hat{n}_+ + \hat{n}_-$$

holds, setting the polarization basis $\{|P, \mu\rangle\}$ in the Hilbert space $L_F(1) = \mathrm{Span}\{|n_+, n_-\rangle\}$ reduces to renumbering Fock states with circular polarization,

$$|P, \mu\rangle \equiv |n_+ = P + \mu, n_- = P - \mu\rangle$$
$$= \frac{(\hat{a}_+^\dagger)^{P+\mu}(\hat{a}_-^\dagger)^{P-\mu}}{\sqrt{(P+\mu)!(P-\mu)!}}|0, 0\rangle, \quad (4)$$
$$2P = 0, 1, 2, \ldots, \infty, \quad |\mu| \le P.$$

It follows from (4) and (2) for one-photon radiation states that the P-quasi-spin values are 1/2 rather than 1 (as for the usual spin), and the $\hat{P}_i$ quasi-spin operators in (1) are given by Pauli matrices in basis (4). Indeed, the P-quasi-spin concept is associated with $SU(2)$ symmetry of free electromagnetic field, which is complementary to the relativistic invariance described by the Poincaré group.[3] The presence of such invariance was mentioned in the literature as early as the beginning of the 20th century and can be deduced from the bivector formulation of the Maxwell equations given by Majorana. More recently, such symmetry was explicitly although sporadically formulated in works on field theory,[4] but systematic use of the P-quasi-spin concept for correctly characterizing the polarization structure of quantum light fields was only suggested in [22]. Generally, when there is an arbitrary number of spacetime modes, the $|P, \mu\rangle$ vectors are given by linear combinations of Fock states that correspond to the Clebsch–Gordan expansion of the tensor products $D^{P_1 = N_1/2} \otimes D^{P_2 = N_2/2} \otimes \ldots$ of the $SU(2)$ group irreducible representations [with bases (4)] in the irreducible components with bases (2) [22].

Setting the polarization basis (2) allows us to introduce the polarization density operator

$$\hat{R} = \sum_{P, \mu, \mu'} R^P_{\mu', \mu}|P, \mu'\rangle\langle P, \mu|, \quad (5)$$
$$R^P_{\mu, \mu'} \equiv \langle P, \mu|\hat{\rho}|P, \mu'\rangle,$$

which is the reduction of the total radiation density operator $R^P_{\mu, \mu'} = \langle P, \mu|\hat{\rho}|P, \mu'\rangle$. The polarization density operator characterizes the quantum properties of light in polarization experiments and completely describes the polarization [only (!)] field state [15]. By way of example, let us write down the matrix elements

$$R^P_{\mu', \mu}(N) = \delta_{P, N/2} c_{P+\mu, P-\mu} c^*_{P+\mu', P-\mu'}, \quad (6a)$$

$$R^P_{\mu', \mu}(\alpha_+, \alpha_-) = \exp(-|\alpha_+|^2 - |\alpha_-|^2)$$
$$\times \frac{(\alpha_+)^{P+\mu}(\alpha_-)^{P-\mu}(\alpha_+^*)^{P+\mu'}(\alpha_-^*)^{P-\mu'}}{\sqrt{(P+\mu)!(P-\mu)!(P-\mu')!(P-\mu')!}} \quad (6b)$$

for quantum-optical radiation states of two polarization modes with circular polarization and a fixed total number of photons in both modes,

$$|N\rangle = \sum_{n_+ + n_- = N} c_{n_+, n_-}|n_+, n_-\rangle, \quad \hat{\rho}(N) = |N\rangle\langle N|,$$

and for two-mode Glauber coherent states,

$$|\alpha_+, \alpha_-\rangle = \exp(-|\alpha_+|^2/2 - |\alpha_-|^2/2)$$
$$\times \sum_{n_+, n_-} \frac{(\alpha_+)^{n_+}(\alpha_-)^{n_-}}{\sqrt{n_+! n_-!}}|n_+, n_-\rangle,$$
$$\hat{\rho} = |\alpha_+, \alpha_-\rangle\langle\alpha_+, \alpha_-|.$$

[3] The term "reduced spin" sometimes used in the literature with reference to the P-quasi-spin (in view of the relation between $2\hat{P}_3$ and the helicity operator mentioned above) is therefore not quite appropriate because it masks the physical nature of this quantity.

[4] Brief historical comments on this point can be found in [23].

Note that the $\hat{R}$ operator is determined by a finite number $[(N+1)^2]$ of the $R^P_{\mu',\mu}$ matrix elements in the first example and by an infinite number of such elements in the second.

Polarization density operator (5) is introduced from the following considerations. For an arbitrary polarization variable (that is, a variable whose operator depends only on bilinear combinations of the $\hat{a}_\alpha$ and $\hat{a}^\dagger_\alpha$ field operators of form (1) and therefore only on $\hat{S}_j$ or $\hat{P}_j$), the matrix elements of its operator $\hat{A}(\hat{\mathbf{P}})$ in the $\{|P, \mu\rangle\}$ basis are block-diagonal (bock with respect to the μ index and diagonal with respect to P),

$$\langle P', \mu'|\hat{A}(\hat{\mathbf{P}})|P, \mu\rangle = \delta_{P',P}A^P_{\mu',\mu}. \quad (7)$$

The quantum mean $\langle\hat{A}(\hat{\mathbf{P}})\rangle = \mathrm{Tr}[\hat{\rho}\hat{A}(\hat{\mathbf{P}})]$ is then expressed via the matrix elements of the polarization density operator as

$$\langle\hat{A}(\hat{\mathbf{P}})\rangle = \sum_{2P=0}^{\infty}\sum_{\mu=-P}^{P}\sum_{\mu'=-P}^{P} A^P_{\mu,\mu'}R^P_{\mu',\mu} \equiv \mathrm{Tr}[\hat{R}\hat{A}(\hat{\mathbf{P}})], \quad (8)$$

rather than via the elements of the total density operator. Note that (8) is similar to the averaging formula for spin systems [25], but, unlike the latter, generally contains infinite sums over quasi-spin P. Indeed, it follows from (4) and example (6b) that light fields generated by real sources are systems with a variable P-quasi-spin [22]. Only the states with a fixed total number of photons N in both polarization modes (6a) are formally equivalent to the states of particles with spin $N/2$.

The above description of the polarization structure of light fields is easy to generalize to an arbitrary number m of spacetime modes by defining the total observables $\hat{P}_i$ as the sums over $\mathbf{k}$ of the "partial" components $\hat{P}_{i\mathbf{k}}$ given by (1) [6, 15]. Equalities (3) and (4) then no longer hold, the P-quasi-spin becomes an independent (of $\hat{N}$) observable, and the polarization basis $|P, \mu; \lambda\rangle \in L_F(m)$ defined by (2) contains additional (nonpolarization) quantum numbers λ. At the same time, (8) remains valid, and the polarization density operator $\hat{R}$ retains its physical meaning, although its formal definition (5) slightly changes: namely, the $|P, \mu'\rangle$ basis vectors in (5) and the trace in the last equality in (8) are determined in the factor-space $L_P(m) \sim L_F(1)$ of the total field space $L_F(m)$ rather than in $L_F(m)$ itself. Accordingly, the $R^P_{\mu,\mu'}$ coefficients are determined by summing the $\langle P, \mu; \lambda|\hat{\rho}|P, \mu'; \lambda\rangle$ elements of the field density matrix over nonpolarization indices λ [15].

As in field tomography, which requires recording the probability distributions of a set of "rotated" quadratures to study the state of a field, in polarization tomography we must deal with a set of rotated Stokes polarization observables [15, 16] of the form

$$\begin{aligned}\hat{S}(\theta, \phi) &= (\hat{S}_1\sin\theta\cos\phi + \hat{S}_2\sin\theta\sin\phi + \hat{S}_3\cos\theta)\\ &= 2\hat{T}(\mathbf{n})\hat{P}_3\hat{T}^\dagger(\mathbf{n}).\end{aligned} \quad (9)$$

Here, $\mathbf{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ corresponds to the radius vector on the Poincaré polarization sphere, and the unitary operator

$$\hat{T}(\mathbf{n}) = \exp[-i\theta(-\hat{P}_1\sin\phi + \hat{P}_2\cos\phi)] \quad (10)$$

describes rotations on this sphere. Every $\hat{S}(\theta, \phi)$ observable has a discrete spectrum ($S = 0, \pm1, \pm2, \ldots$) with degenerate eigenstates

$$\begin{aligned}|P, \mu; \theta, \phi\rangle &= \hat{T}(\mathbf{n})|P, \mu\rangle\\ &= \sum_{\mu'} D^P_{\mu',\mu}(\phi, \theta, -\phi)|P, \mu'\rangle,\end{aligned} \quad (11)$$

$$\hat{S}(\theta, \phi)|P, \mu; \theta, \phi\rangle = 2\mu|P, \mu; \theta, \phi\rangle,$$

which are simultaneously the generalized coherent states of the $SU(2)$ group of polarization invariance of light fields $\{\exp[i(\sum_j u_j P_j)]\}$ [15]. Here, $D^P_{\mu',\mu}(\phi, \theta, -\phi)$ is the Wigner D function of the $SU(2)$ group [25]. The probability distributions $p(S; \theta, \phi)$ of the observables $\hat{S}(\theta, \phi)$ in the state $\hat{R}$ are determined via the averaging

$$p(S; \theta, \phi) = \sum_{P \geq |S/2|}^{\infty}\left\langle P, \frac{S}{2}; \theta, \phi\middle|\hat{R}\middle|P, \frac{S}{2}; \theta, \phi\right\rangle \quad (12)$$

of the polarization density operator in basis (11). Note that, by virtue of (11) and (12), the $p(S; \theta, \phi)$ density distributions are expressed as series in the products of the $R^P_{\mu',\mu}$ elements of the polarization density operator and the $Y_{lm}(\theta, \phi)$ spherical functions,

$$\begin{aligned}p(S; \theta, \phi) &= \sum_{P \geq \left|\frac{S}{2}\right|}^{\infty}\sum_{|\mu| \leq P,\, |\mu'| \leq P} R^P_{\mu',\mu}\\ &\times \sum_{l=0}^{2P}\sum_{m=-l}^{l}(-1)^{(S/2)-\mu}\ldots Y_{lm}(\theta, \phi),\end{aligned} \quad (13)$$

where the multipliers "…" are the products of the Clebsch–Gordan coefficients of the $SU(2)$ group [25].

Another important element of the mathematical apparatus of polarization tomography of radiation is the characteristic polarization function

$$\chi(\omega, \mathbf{n}) = \langle \hat{U}^{\dagger}(\omega, \mathbf{n}) \rangle = \sum_{P, \mu, \mu'} R^{P}_{\mu, \mu'} U^{P}_{\mu', \mu}(-\omega, \mathbf{n}) = \sum_{S=-\infty}^{\infty} p(S; \theta, \phi) \exp\left(i\omega\frac{S}{2}\right), \tag{14}$$

which is determined using the unitary shift operator

$$\begin{aligned} \hat{U}(\omega, \mathbf{n}) &= \exp(-i\omega \mathbf{n} \cdot \hat{\mathbf{P}}) \\ &= \hat{T}(\mathbf{n}) \exp(-i\omega \hat{P}_3) \hat{T}^{\dagger}(\mathbf{n}) \\ &= \sum_{P, \mu, \mu'} U^{P}_{\mu', \mu}(\omega, \mathbf{n}) |P, \mu'\rangle\langle P, \mu| \\ &= \sum_{P, \mu} |P, \mu; \theta, \phi\rangle\langle P, \mu; \theta, \phi| \exp(-i\omega\mu) \end{aligned} \tag{15}$$

of the $SU(2)$ polarization group mentioned above [15].

The elements of the description of quantum light polarization described above and their mathematical properties [15, 25] can be used to obtain the tomographic representation of the polarization density operator

$$\begin{aligned} \hat{R} &= \int_{S^2_+} d^2\mathbf{n} \sum_{S=-\infty}^{\infty} p(S; \theta, \phi) \hat{K}(S; \mathbf{n}) \\ &\equiv \int_0^{\pi/2} \sin\theta d\theta \int_0^{2\pi} d\phi \sum_{S=-\infty}^{\infty} p(S; \theta, \phi) \hat{K}(S; \mathbf{n}), \end{aligned} \tag{16}$$

which is central to the polarization tomography of radiation. Here, the coefficients are the probability distributions $p(S; \theta, \phi)$, that is, polarization tomograms determined experimentally. The tomography operator kernel has the form

$$\begin{aligned} \hat{K}(S; \mathbf{n}) &= \hat{T}(\mathbf{n}) \hat{K}_0(S) \hat{T}^{\dagger}(\mathbf{n}), \\ \hat{K}_0(S) &= \frac{2\hat{P} + 1}{2\pi^2} \int_0^{2\pi} \cos\left[\omega\left(\frac{S}{2} - \hat{P}_3\right)\right] \sin^2\frac{\omega}{2} d\omega, \end{aligned} \tag{17}$$

valid for an arbitrary number of spacetime modes (taking into account the above comments concerning the determination of the total observables $\hat{P}_i$). Note that, as distinct from homodyne field tomography, the $\hat{K}(S; \mathbf{n})$ kernel in (16) is not singular, as follows from the spectral decomposition of $\hat{K}_0(S)$ in the polarization basis $\{|P, \mu\rangle\}$ [15]. The representation of the polarization density operator in the form of expansion (16) in the partially continuous basis (17) is the first principal result of this work.

Basic relation (16) is inverse to (12) and closes the equation that relates the polarization density operator to the set of distribution data extracted from experimental measurements. Both these relations constitute the mathematical foundation of polarization quantum tomography in the broad sense. Note that, unlike discrete representation (5), the expansion (16) of the $\hat{R}$ operator in the "partially continuous" operator basis $\{\hat{K}(S; \mathbf{n})\}$ has the disadvantage of extraordinary redundancy of measurement data because the θ and ϕ variables are continuous, as is the array of polarization tomograms $p(S; \theta, \phi)$ necessary for the exact reconstruction of the polarization density operator. This redundancy is caused by the overcompleteness of the set of basis operators compared with the complete system of the countable basis of $|P, \mu'\rangle\langle P, \mu|$ dyadics in definition (5). It can be lessened by "reducing" the basis of $\hat{K}(S; \mathbf{n})$ operators and the set of the corresponding tomograms $p(S; \theta, \phi)$ to countable subsets in the angular variables of the $\mathbf{n}(\theta, \phi)$ vector. These subsets are complete, as in spin tomography [26]. In addition, such a "size reduction" of the array of measurement data can be strengthened if a priori information about the character of radiation under study is available. For instance, it follows from (6) that multipole expansions (13) for the probabilities $p(S; \theta, \phi)$ contain finite numbers of terms for N-photon states of monochromatic plane-wave radiation with polarization density operator (6a). For Glauber coherent states with polarization density operator (6b), such expansions, although infinite, determine the probabilities $p(S; \theta, \phi)$ at arbitrary S as a family of functions on a sphere parametrized by the α_+ and α_- values. It can therefore be expected that qualitative tomography can be obtained from no more than $(N + 1)^2$ measurements in the first case, whereas the number of measurements necessary in the second case is determined by statistically estimating the parameters [27].

To complete the scheme of tomography, it is necessary to calculate the matrix elements of the polarization density operator $\hat{R}$ in some basis by (16). This problem reduces to calculating the matrix elements of tomographic kernels $\hat{K}(S; \mathbf{n})$ in the same basis, which effects the transition from (16) to its projection onto the corresponding c-number version. For instance, such a projection of equality (16) on the basis of $|P, \mu'\rangle\langle P, \mu|$ dyadics was performed in [15], which led to the c-number version of this equality for the reproduction of the $R^{P}_{\mu', \mu}$ matrix elements from measurement data. To complete the tomography scheme in a narrow sense, we

must, however, augment it by quasi-probability functions as the coordinates of the density operator in some basis of quasi-classical (numbered by phase space points) Hermitian operators [24]. Note that, with respect to the basis of dyadics, such a quasi-classical operator basis plays a role similar to that of Glauber coherent states with respect to a discrete Fock basis when the states of the Hilbert electromagnetic field space are described. A suitable basis for determining a correct quasi-probability function is selected from considerations of the simplicity and clearness of the relations that arise in the corresponding quasi-classical analysis of quantum problems. For instance, in field tomography with the Wigner function as quasi-probability, such a basis is formed by Wigner operators $\hat{W}(\beta)$ determined by the Fourier transform of the field shift operator $\hat{D}(\beta)$ [1].

In polarization tomography, it is also expedient to select some integral transformation of the shift operator (15) of the $SU(2)$ group into the three-dimensional polarization phase space $R^3 = \{\mathbf{p} = (p_1, p_2, p_3)\}$ as a quasi-classical operator basis. This space is the factor space of the complete field phase space ($R^3 \equiv C^2/S^1$ for one spacetime mode) and has the coordinates $p_{i=1,2,3}$ associated with the Stokes vector parameters [6, 15]. The simplest operator basis of this kind is given by the Fourier transforms of the shift operator $\hat{U}(\omega, \mathbf{n})$ numbered by the phase points $\mathbf{p} \in R^3$ [6],

$$\hat{W}(\mathbf{p}) = \int_{R^3} \frac{d\omega d\mathbf{n}}{(2\pi)^3} \hat{U}(\omega, \mathbf{n}) e^{i\omega \mathbf{n}\cdot\mathbf{p}} = \int_{R^3} \frac{d\omega d\mathbf{n}}{(2\pi)^3} \sum_{P,\mu} |P, \mu; \theta, \phi\rangle\langle P, \mu; \theta, \phi| e^{i\omega(\mathbf{n}\cdot\mathbf{p} - \mu)}. \quad (18)$$

As distinct from the Wigner operators of field tomography, operators (18) are, however, singular and, in addition, are not mutually orthogonal, that is,

$$\mathrm{Tr}[\hat{W}(\mathbf{p})\hat{W}(\mathbf{p}')] \neq \delta(\mathbf{p} - \mathbf{p}'). \quad (19)$$

This prevents us from obtaining a complete quasi-classical description of radiation polarization, in particular, the quasi-classical form of the quantum mean (8). This shortcoming can be compensated for by the introduction of one more operator basis $\{\hat{V}(\mathbf{p})\}$ with the property of being biorthogonal to $\hat{W}(\mathbf{p})$,

$$\mathrm{Tr}[\hat{W}(\mathbf{p})\hat{V}(\mathbf{p}')] = \delta(\mathbf{p} - \mathbf{p}'). \quad (20)$$

This basis is formed by differently transforming the shift operator $\hat{U}(\omega, \mathbf{n})$ [6, 15, 28]:

$$\hat{V}(\mathbf{p}) = \frac{2\hat{P} + 1}{4\pi^2} \int_0^{2\pi} \sin^2\frac{\omega}{2} d\omega \int_0^{\pi} d\cos\theta \times \int_0^{2\pi} d\phi \hat{U}(\omega, \mathbf{n}) e^{i\omega \mathbf{n}\cdot\mathbf{p}} = \int_0^{2\pi} \sin^2\frac{\omega}{2} d\omega \int_0^{\pi} d\cos\theta \int_0^{2\pi} d\phi \times \sum_{P,\mu} \frac{2P + 1}{4\pi^2} |P, \mu; \theta, \phi\rangle\langle P, \mu; \theta, \phi| e^{i\omega(\mathbf{n}\cdot\mathbf{p} - \mu)}. \quad (21)$$

Because the operators $\hat{W}(\mathbf{p})$ and $\hat{V}(\mathbf{p})$ are biorthogonal [property (20)], they can conveniently be used in quantum calculations in pairs, which is similar to the use of P- and Q-representation pairs in quasi-classical calculations of quantum optics values. For instance, setting the expansion of the polarization density operator in the $\hat{V}(\mathbf{p})$ basis in the form

$$\hat{R} = \int_{R^3} W(\mathbf{p}) \hat{V}(\mathbf{p}) d\mathbf{p}, \quad (22)$$

we obtain the polarization quasi-probability function $W(\mathbf{p})$ as the projection of the density operator onto the $\hat{W}(\mathbf{p})$ basis,

$$W(\mathbf{p}) = \mathrm{Tr}[\hat{R}\hat{W}(\mathbf{p})]. \quad (23)$$

The quantum mean in (8) can then be represented in the classical form

$$\langle \hat{A}(\hat{\mathbf{P}}) \rangle = \int_{R^3} V(\mathbf{p}; \hat{A}) W(\mathbf{p}) d\mathbf{p}, \quad (24)$$

where the quasi-classical partner of the quasi-probability distribution $W(\mathbf{p})$ is the $V(\mathbf{p}; \hat{A})$ projection (the Weyl symbol) of the $\hat{A}(\hat{\mathbf{P}})$ quantum operator onto the $\hat{V}(\mathbf{p})$ basis,

$$V(\mathbf{p}; \hat{A}) = \mathrm{Tr}[\hat{A}\hat{V}(\mathbf{p})]. \quad (25)$$

It follows that the pair of biorthogonal operator bases $\hat{W}(\mathbf{p})$ and $\hat{V}(\mathbf{p})$ closes the mathematical scheme of quantum polarization tomography.

In [4, 6], the projection of the $\hat{R}$ operator onto basis (18) was taken as the polarization quasi-probability function. Among various biorthogonal pairs, bases (18) and (21) have the advantage of coincidence of the Weyl

symbol of the shift operator with the classical exponential function,

$$V(\mathbf{p};\ \hat{U}^{\dagger}(\omega,\mathbf{n})) = \mathrm{Tr}[\hat{U}^{\dagger}(\omega,\mathbf{n})\hat{V}(\mathbf{p})] = \exp(i\omega\mathbf{n}\cdot\mathbf{p}). \quad (26)$$

In addition, the Weyl symbol of an $f(\mathbf{n}\cdot\hat{\mathbf{P}})$-type operator is merely $f(\mathbf{n}\cdot\mathbf{p})$. What is more, by virtue of (22)–(26), the quasi-probability $W(\mathbf{p})$ can then be determined as the Fourier transform of the characteristic function $\chi(\omega, \mathbf{n})$,

$$W(\mathbf{p}) = \frac{1}{(2\pi)^3}\int_{R^3}\chi(\omega,\mathbf{n})\exp(-i\omega\mathbf{n}\cdot\mathbf{p})d\omega d\mathbf{n}, \quad (27)$$

which corresponds with the classical definition of the joint probability distribution. The selection of the quasi-probability distribution according to (27) [this selection is equivalent to setting bases in forms (18) and (21)] is our second principal result. The adequacy of such a quasi-probabilistic distribution to the problem of the characterization of the polarization state of radiation was verified by calculating several quasi-probability distribution examples [4]. It was found that distributions of this type were fairly sensitive to the quantum special features of the polarization states of light, in particular, to the interference of the amplitudes of the probability of superposition states [6].

Lastly, the projection of (16) onto the operator basis (18) with application of the formalism described in [25] and the theory of generalized functions (distribution theory) yields the relation between the set of experimental distributions (polarization tomograms) $p(S; \theta, \phi)$ and the quasi-probability function, namely,

$$W(\mathbf{p}) = \int_{S_+^2} d\mathbf{n}\sum_{S=-\infty}^{\infty} p(S;\theta,\phi)K(S,\mathbf{n};\mathbf{p}),$$

$$K(S,\mathbf{n};\mathbf{p}) = \mathrm{Tr}[\hat{K}(S;\mathbf{n})\hat{W}(\mathbf{p})] = -\frac{2}{\pi^2}\delta^{(2)}(S-2\mathbf{n}\cdot\mathbf{p}). \quad (28)$$

This equation is most important for practice. Together with (16) and (17), it fully solves the problem of ideal (100% photodetector efficiency) polarization tomography of radiation for an arbitrary number of spacetime modes [15]. Equation (28) for the reproduction of the quasi-probability function from experimental measurements is the third principal result of this work. However in practice, as in field tomography [14], this ideal scheme should be corrected for nonunit quantum efficiency of photodetectors by the introduction of "smoothing" factors into basis equations (16) and (28).

Consider some special features of the suggested method of polarization tomography of radiation. First, the singularity of the tomographic number kernel $K(S, \mathbf{n}; \mathbf{p})$ in (28) is caused by the singular character of the quasi-classical basis operators (18) [by virtue of the periodicity properties of $\hat{U}(\omega, \mathbf{n})$] rather than the tomographic operator kernel $\hat{K}(S; \mathbf{n})$ in (17), which is here quite regular, as distinct from the case of homodyne field tomography [11, 14]. For this reason, we can use modifications of the scheme of polarization tomography of light described above by applying "smoother" polarization quasi-probability functions, for instance, those obtained by the reduction of multimode Wigner field functions [15]. Next, the basic relations (16) and (28) of polarization tomography can also be used in spin tomography [26, 29, 30] because the $\hat{P}_i$ operators obey the *su*(2) algebra. The variable character of the *P*-quasi-spin of light fields [with the summation over the infinite spectrum $0 \le P \le \infty$ in (8)] entails conceptual differences between the two methods (in arbitrary versions) [15]. In particular, since the spin density matrix is finite-dimensional, a finite (!) number of spin tomograms [26] given on a spherical lattice, which is two-dimensional because the total spin is fixed, in the "phase space" of spin systems is sufficient for performing tomography of spin systems. At the same time, because of the presence of infinite sums over *S* in (28) and the summation over *P* in the definition of $p(S; \theta, \phi)$ in (12), two-dimensional spin tomograms [29, 30] can only be used to reconstruct particular polarization state forms, namely, those with fixed *P*-quasi-spin values, for which the corresponding polarization quasi-probability functions are effectively determined on the two-dimensional sphere [4]. Examples of such states are *N*-photon states of type (6a). To summarize, the method of polarization tomography of radiation formally includes spin tomography, but not vice versa. Conceptually, these differences arise because spin wave functions in spin tomography are separated a priori from coordinate wave functions, whereas polarization variables are extracted as collective variables from the complete field description.

3. A PRACTICAL IMPLEMENTATION OF THE POLARIZATION TOMOGRAPHY METHOD

The experimental implementation of polarization tomography of radiation was described for the first time in [16]. The source of radiation was an optical parametric oscillator, which was a potassium titanyl phosphate (KTP) crystal in a cavity with pumping by radiation from a continuous argon laser. Ideally, radiation from the parametric oscillator should be a flux of photon pairs, one photon of each pair belonging to the *x* and the other, to the *y* polarization mode. Such radiation is called *XY*-biphoton light in theory [4]; this is one of spectacular examples of nonclassical nonpolarized

light varieties (for which $\langle \hat{S}_{i=1,2,3} \rangle = 0$) that in addition possesses hidden radiation polarization. Since its discovery [31], *XY*-biphoton light has been extensively studied both theoretically and experimentally [4, 32, 33]. The quantum state of ideal biphotonic light is described by the equation

$$\begin{aligned} |XY\rangle &= \exp(z\hat{a}_x^\dagger \hat{a}_y^\dagger - z^* \hat{a}_x \hat{a}_y)|0, 0\rangle \\ &= \frac{1}{\cosh r}\exp(e^{i\vartheta}\hat{a}_x^\dagger \hat{a}_y^\dagger \tanh r)|0, 0\rangle, \quad z = re^{i\vartheta}; \end{aligned} \tag{29}$$

it is the eigenstate of the $\hat{S}_1$ operator with zero eigenvalue. Its polarization quasi-probability function $W(\mathbf{p})$ is calculated as the Fourier transform (27) of the characteristic polarization function (14). In the approximation of the suppression of characteristic function values outside the main period limits (the approximation of a large number of photons $\langle n \rangle \gg 1$), the $W(\mathbf{p})$ function is given by the equation [4]

$$\begin{aligned} &W(s_1, s_2, s_3) \equiv W(\mathbf{p} = (s_1/2, s_2/2, s_3/2)) \\ &\approx \frac{1}{2\pi}\frac{\delta(s_1)}{\sqrt{\langle n\rangle(\langle n\rangle + 2)}}\frac{1}{\sqrt{s_2^2 + s_3^2}}\exp\left(-\frac{\sqrt{s_2^2 + s_3^2}}{\sqrt{\langle n\rangle(\langle n\rangle + 2)}}\right), \end{aligned} \tag{30}$$

where $\langle n \rangle \equiv \langle XY|\hat{N}|XY\rangle = 2\sinh^2 r$ is the total number of photons. The quasi-probability density is concentrated in the (s_2, s_3) plane. Clearly, this corresponds to the absence of noise when the Stokes parameter s_1 is recorded. Unlike the exact equation, approximation (30) is analytically simple, which substantially facilitates comparison with experiment.

Radiation polarization state measurements proper described in [16] were performed according to the scheme shown in Fig. 1. In this scheme, a family of polarization observables $\hat{S}(\theta, \phi)$ is obtained by rotating the $\lambda/2$ and $\lambda/4$ phase plates. At a fixed arrangement of the polarization elements, the polarization observable values are proportional to the difference of the J_1 and J_2 photocurrents from photodiodes PD_1 and PD_2. The resulting set of difference current values is transformed into the probability distribution $p(J_1 - J_2; \theta, \phi)$. Thanks to the use of photodiodes with a high quantum efficiency and calibration, the $p(J_1 - J_2; \theta, \phi)$ distributions correspond to the $p(S; \theta, \phi)$ distributions. The distributions of the polarization variables were recorded for a set of the θ and ϕ values on a grid in steps of $\pi/10$. The recorded distributions were close to Gaussian with zero means; their variances depended on the θ and ϕ angles. The variances for the polarization observables s_2 and s_3 coincided with the Poisson level, that is, with the standard quantum level, to within measurement errors. Noise, however, decreased below the Poisson level by 24% as the polarization observable approached s_1. The

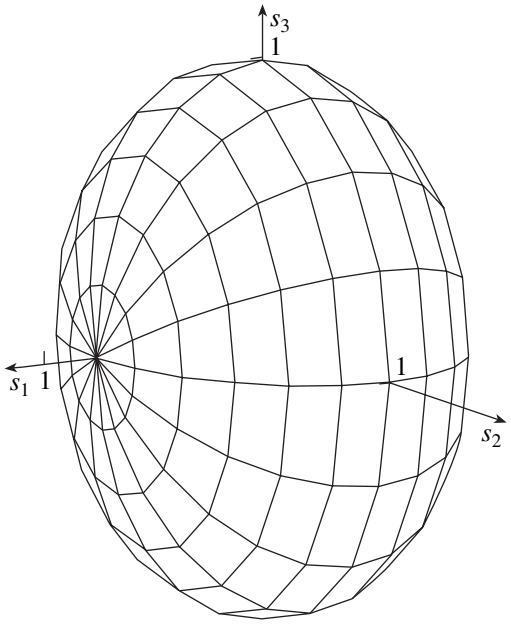


Fig. 2. Level $1/\sqrt{e}$ surface (from the maximum) of the quasi-probability function $W(\mathbf{s})$ reproduced from the experimental data reported in [17].

array of data obtained on the $p(S; \theta_i, \phi_j)$ lattice was extrapolated to the continuous set $p(S; \theta, \phi)$. Substituting this set into (28) yields a quasi-probability distribution in the form of the three-dimensional Gaussian function (Fig. 2)

$$\begin{aligned} W(s_1, s_2, s_3) &= \frac{1}{\sqrt{0.76(2\pi\langle n\rangle)^3}} \\ &\times \exp\left(-\frac{s_1^2}{0.76 \times 2\langle n\rangle} - \frac{s_2^2 + s_3^2}{2\langle n\rangle}\right). \end{aligned} \tag{31}$$

For estimating the quality of state reconstruction by the method of polarization tomography of radiation, it is expedient to compare distribution (31) with the distribution for the polarization state of *XY*-biphoton radiation generated during spontaneous parametric light scattering. Qualitatively, distribution (31) is similar to (30) (both are squeezed in the s_1 direction). Distribution (31), however, differs from (30) by its finite width in the s_1 direction and by the character of its decrease in the (s_2, s_3) plane. This difference is primarily caused by the formation of the experimentally studied radiation under generation rather than spontaneous scattering conditions. Its state therefore has a more complex structure than state (29). Its correct description requires using a more exact model.

4. CONCLUSIONS

To summarize, the method of polarization tomography of radiation that we developed and tested is a fairly

simple and very effective tool for studying the polarization properties of light. This method gives information about the quantum state of radiation polarization sufficient for predicting the results of all possible polarization measurements. Central to the method is equation (28) for the reproduction of the quasi-probability function from measurement data. This equation is in turn based on the general tomographic representation (16) of the polarization density operator and the correct selection (27) of the type of quasi-probability distribution (23). It appears that the most important area in applying polarization tomography is analysis of complex multiphoton light states, when a priori information on the quantum structure of polarization modes is lacking.

In conclusion, let us turn to practical aspects of polarization tomography application. It follows from (16) and (28) that the exact reconstruction of the quantum polarization state requires the use of a lattice-continuous array $\{p(S; \theta, \phi)\}$ of experimental data (polarization tomograms), which is impracticable. This raises the problem of the "correct discretization" of these equations, which sharply decreases the amount of measurements and replaces the integration in the classical variables θ and ϕ by finite summation over them with retention of the main characteristics of the quasi-probability functions $W(\mathbf{p})$ [34]. This problem can be solved by reducing the overcompleteness of tomographic kernels $\hat{K}(S; \mathbf{n})$ and $K(S, \mathbf{n}; \mathbf{p})$ (by analogy with tomography of spin systems [26]) and the "resolving power" of quasi-probability retrieval. However on the whole, the problem of optimizing practical polarization tomography procedures requires further study.

ACKNOWLEDGMENTS

This work was financially supported by INTAS (grant no. 01-2122) and the Russian Foundation for Basic Research (project no. 02-02-17498).

REFERENCES

1. V. I. Tatarskiĭ, Usp. Fiz. Nauk **139**, 587 (1983) [Sov. Phys. Usp. **26**, 311 (1983)].
2. M. Hillery, R. F. O'Connell, M. O. Scully, and E. P. Wigner, Phys. Rep. **106**, 121 (1984).
3. H.-W. Lee, Phys. Rep. **259**, 147 (1995).
4. V. P. Karassiov and A. V. Masalov, Laser Phys. **12**, 948 (2002).
5. V. Buzek and P. L. Knight, in *Progress in Optics*, Ed. by E. Wolf (North-Holland, Amsterdam, 1995), Vol. 34, p. 1.
6. V. P. Karassiov and A. V. Masalov, J. Opt. B: Quantum Semiclassic. Opt. **4**, S366 (2002).
7. B. B. Kadomtsev, Usp. Fiz. Nauk **164**, 449 (1994) [Phys. Usp. **37**, 425 (1994)].
8. A. Steane, Rep. Prog. Phys. **61**, 117 (1998).
9. *The Physics of Quantum Information: Quantum Cryptography, Quantum Teleportation, Quantum Computation*, Ed. by D. Bouwmeester, A. K. Ekert, and A. Zeilinger (Springer, Berlin, 2000; Postmarket, Moscow, 2002).
10. B. A. Grishanin and V. N. Zadkov, Radiotekh. Élektron. (Moscow) **47**, 1029 (2002).
11. K. Vogel and H. Risken, Phys. Rev. A **40**, 2847 (1989).
12. D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, Phys. Rev. Lett. **70**, 1244 (1993).
13. H. Kühn, D.-G. Welsh, and W. Vogel, Phys. Rev. A **51**, 4240 (1995).
14. U. Leonhardt, H. Paul, and G. M. D'Ariano, Phys. Rev. A **52**, 4899 (1995).
15. V. P. Karasev, Kratk. Soobshch. Fiz., No. 9, 34 (1999).
16. P. A. Bushev, V. P. Karasev, A. V. Masalov, and A. A. Putilin, Opt. Spektrosk. **91**, 558 (2001) [Opt. Spectrosc. **91**, 526 (2001)].
17. M. G. Raymer, A. C. Funk, and D. F. McAlister, in *Quantum Communications, Computing and Measurement 2*, Ed. by P. Kumar, G. M. D'Ariano, and O. Hirota (Kluwer Academic/Plenum, New York, 2000), p. 147.
18. G. S. Agarwal, J. Lehner, and H. Paul, Opt. Commun. **129**, 369 (1996).
19. N. Korolkova, G. Leuchs, R. Loudon, *et al.*, Phys. Rev. A **65**, 052306 (2002).
20. R. Schnabel, W. P. Bowen, N. Treps, *et al.*, Phys. Rev. A **67**, 012316 (2003).
21. J. Heersink, T. Gaber, S. Lorenz, *et al.*, Phys. Rev. A **68**, 013815 (2003).
22. V. P. Karassiov, J. Phys. A **26**, 4345 (1993).
23. V. P. Karassiov, J. Russ. Laser Res. **15**, 391 (1994); J. Russ. Laser Res. **21**, 370 (2000).
24. R. L. Stratonovich, Zh. Éksp. Teor. Fiz. **31**, 1012 (1956) [Sov. Phys. JETP **4**, 891 (1956)].
25. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Sci., Singapore, 1988).
26. S. Weigert, Phys. Rev. Lett. **84**, 802 (2000).
27. D. Hudson, *Statistics. Lectures on Elementary Statistics and Probability* (Geneva, 1964; Mir, Moscow, 1970).
28. S. M. Chumakov, A. B. Klimov, and K. B. Wolf, Phys. Rev. A **61**, 034101 (2000).
29. V. A. Andreev and V. I. Man'ko, Zh. Éksp. Teor. Fiz. **114**, 437 (1998) [JETP **87**, 239 (1998)].
30. A. B. Klimov, O. V. Man'ko, V. I. Man'ko, *et al.*, J. Phys. A **35**, 6101 (2002).
31. V. P. Karassiov and V. I. Puzyrevskii, J. Sov. Laser Res. **10**, 229 (1989).
32. D. N. Klyshko, Zh. Éksp. Teor. Fiz. **111**, 1955 (1997) [JETP **84**, 1065 (1997)].
33. A. V. Burlakov, S. P. Kulik, G. O. Rytikov, and M. V. Chekhova, Zh. Éksp. Teor. Fiz. **122**, 738 (2002) [JETP **95**, 639 (2002)].
34. F. Natterer, *The Mathematics of Computerized Tomography* (Wiley, Chichester, 1986; Mir, Moscow, 1990).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 61–72.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 75–88.

PLASMA, GASES

Dusty Photoresonant Plasma with Coulomb Collisions

A. G. Leonov[a], A. F. Pal'[b], A. N. Starostin[b], and A. V. Filippov[b,*]

[a]*Moscow Institute of Physics and Technology, Dolgoprudnyĭ, Moscow oblast, 141700 Russia*
[b]*State Science Center of the Russian Federation, Troitsk Institute for Innovation and Fusion Research, Troitsk, Moscow oblast, 142190 Russia*
**e-mail: fav@triniti.ru*

Received January 14, 2004

Abstract—We have studied the charging of dust particles in a dense photoresonant sodium plasma with electron and ion densities as high as 10^{16} cm^{-3} produced by laser pumping of the resonance level of Na, which was a small admixture (up to 1%) in an argon buffer gas. We show that the charge of dust particles with a radius of 10 μm at maximum reaches 3×10^5 electron charges and that the potential of the dust particles at a low electron bulk loss rate agrees well with the orbital motion limited (OML) model data. The behavior of the electric field near a dust particle was found to be nonmonotonic. We established that the distribution of the potential near a solitary charged dust particle agrees well with the Debye one, but the screening length proves to be much larger than even the electron Debye length; the discrepancies are largest at the afterglow stage of the photoresonant plasma, when the sodium ion with a low recombination coefficient is the main plasma ion. We determined the domain of parameters for a dense plasma where an ensemble of dust particles can crystallize. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Today, increasing attention is being paid to the study of a heterogeneous plasma with a condensed disperse phase (a plasma with dust particles) [1–8] due to its unique properties. In recent years, it has been called a complex plasma increasingly often, by analogy with complex liquids, to emphasize the variety of fundamental differences between such a plasma and an ordinary gas plasma [6–8]. In a complex plasma, an ordered structure, a Coulomb crystal, is formed under certain conditions; studying this crystal is of great interest from both fundamental and applicative points of view. For example, ordered plasma structures from radioactive dust can be used in radioisotopic photovoltaic power supplies and nuclear batteries [9].

In an ideal plasma, the interaction between charged particles is described by the Debye screened potential (which is occasionally also called the Yukawa potential)

$$U(r) = \frac{e^2q^2}{r}\exp(-r/R_\mathrm{D}), \tag{1}$$

where e is the absolute value of the electron charge, q is the particle charge in elementary charges, and R_D is the Debye length. Based on Monte-Carlo numerical simulations, the authors of [10–14] showed that the crystallization condition for particles that interact according to law (1) is specified by the following two parameters: the nonideality parameter

$$\Gamma_S = \frac{e^2q^2}{aT_d}\exp\left(-\frac{a}{R_\mathrm{D}}\right) \tag{2}$$

and the structural parameter

$$\lambda = \frac{a}{R_\mathrm{D}}. \tag{3}$$

Here, T_d is the dust particle temperature in energy units, $a = \sqrt[3]{1/n_d}$ is the mean interparticle distance, and n_d is the dust particle number density. The classical plasma Debye length (R_D) and the Debye lengths of the electron ($R_{\mathrm{D},e}$) and ion ($R_{\mathrm{D},i}$) components are defined by

$$R_\mathrm{D} = \left(\frac{1}{R_{\mathrm{D},e}^2} + \frac{1}{R_{\mathrm{D},i}^2}\right)^{-1/2},$$

$$R_{\mathrm{D},e} = \sqrt{\frac{T_e}{4\pi e^2 n_{e\infty}}}, \tag{4}$$

$$R_{\mathrm{D},i} = \sqrt{\frac{T_i}{4\pi e^2 n_{i\infty}}},$$

where T_e and T_i are the electron and ion temperatures in energy units, and $n_{e\infty}$ and $n_{i\infty}$ are their densities in the unperturbed plasma far from the dust particle.

The dust component in a plasma can crystallize when the nonideality parameter exceeds a critical value [10–14]:

$$\Gamma_S \geq \Gamma_{SC}(\lambda). \tag{5}$$

The critical values of the nonideality parameter for the range of structural parameters $\lambda > 1$ that is most characteristic of dust plasma experiments are well described by the dependence [15]

$$\Gamma_{SC} \approx 54/\lambda^{1.38}. \quad (6)$$

Note that almost all studies of a complex plasma are carried out by introducing a fine-dispersed component into a weakly ionized low-temperature plasma with an electron density n_e no higher than 10^9–10^{12} cm^{-3} by various means (see [1–8]). The Debye length at low electron and ion concentrations is large, while the dust particle charge weakly depends on n_e (according to analytical theories, it does not depend on n_e at all for a solitary dust particle). All of these factors allow criterion (5) to be satisfied at reasonable experimental parameters. According to (4), the Debye length decreases at higher plasma charged particle densities. This causes the structural parameter to increase, which seems to suggest that no ordered structures can be formed in this case. However, it was found both in experiments and in numerical calculations that the screening length for the charge of a dust particle, which characterizes the spatial dependence of its potential, could differ significantly from its classical value under actual low-temperature plasma conditions (for a discussion, see [15]). As the electron and ion densities increase, the screening length begins to exceed even the electron Debye length by several factors [15]. This is because the linearized Debye–Hückel theory is inapplicable to strongly nonideal open systems, an example of which is a complex plasma. Since dust particles continuously absorb electrons and ions, such a plasma can exist only when energy is continuously supplied to generate new charge carriers in place of absorbed ones. Since the degree of plasma perturbation near a dust particle increases with electron density, the deviations from the Debye–Hückel theory also increase with n_e. Therefore, nothing can be said in advance about the Debye length and, accordingly, the structural parameter in a dense plasma.

In addition, the dust particle charge in a dense plasma can be expected to depend on the electron and ion densities, because the electron and, possibly, ion drift and diffusion, which determine the charge, will be largely governed by Coulomb collisions rather than collisions with neutral particles that are mainly responsible for the transport processes in a weakly ionized plasma.

Therefore, it seems interesting to study the charging of dust particles and the possibilities for their Coulomb crystallization in a dense plasma in which the electron and ion densities are several orders of magnitude higher than those reached, for example, in a glow discharge. The only example of a dense plasma in which the temperature of the heavy particles is low enough to be able to disregard the evaporation and thermal decomposition of dust particles appears to be a photoresonant plasma [16, 17]. Such a plasma is produced when a gaseous medium is exposed to laser radiation whose photon energy is close to the energy of the resonance atomic or molecular transition. Alkali and alkali-earth metals in mixtures with a buffer gas are commonly used as such a medium. The photoresonant plasma has a high concentration, $n_e \approx 10^{15}$–10^{17} cm^{-3}, at a relatively low electron temperature, $T_e \approx 0.2$–0.5 eV, and is a fairly unique physical object, because this state cannot be obtained by other methods (see [18]).

In this paper, we study a dense photoresonant plasma with an electron density $n_e \approx 10^{16}$ cm^{-3} with injected micrometer-size dust particles. A physical model for the charging of dust particles in the hydrodynamic approximation is described in the second section. The charging of dust particles in a dense argon plasma is investigated in the third section. In the fourth section, we present the results of our study of the charging of dust particles as applied to the actual parameters of the photoresonant sodium plasma corresponding to the experimental conditions [19]. In the Conclusions, we formulate our main results.

2. A PHYSICAL MODEL FOR THE CHARGING OF DUST PARTICLES IN A PHOTORESONANT PLASMA

The charging of dust particles in the hydrodynamic (HD) approximation is described by the following self-consistent system of equations [20, 21]:

$$\begin{aligned}
\frac{\partial n_e}{\partial t} + \nabla \mathbf{j}_e &= Q_{\text{ion}} - \beta_{ei} n_e n_i - \beta_3 n_e^2 n_i, \\
\frac{\partial n_i}{\partial t} + \nabla \mathbf{j}_i &= Q_{\text{ion}} - \beta_{ei} n_e n_i - \beta_3 n_e^2 n_i, \\
\nabla \mathbf{E} &= 4\pi e (n_i - n_e), \quad (7) \\
\mathbf{j}_e &= -\nabla (D_e n_e) - k_e n_e \mathbf{E}, \\
\mathbf{j}_i &= n_i k_i \mathbf{E} - D_i \nabla n_i.
\end{aligned}$$

Here, $n_{e(i)}$, $k_{e(i)}$, and $D_{e(i)}$ are the electron (ion) density, mobility, and diffusion coefficient, respectively; $\mathbf{j}_{e(i)}$ is the electron (ion) flux density; Q_{ion} is the gas ionization rate; β_{ei} is the electron–ion dissociative or photoradiative recombination coefficient; β_3 is the three-body recombination coefficient; and $\mathbf{E}$ is the strength of the self-consistent electric field.

We will consider the charging of a spherical dust particle with a radius r_0 placed at the center of a spherically symmetric Seitz–Wigner cell with a radius a_d:

$$a_d = \left(\frac{4}{3}\pi n_d\right)^{-1/3}. \quad (8)$$

We assume that there is one dominant type of positive ions in the plasma, which, as will be seen below, is the case in a photoresonant plasma.

The validity conditions for the HD approximation are

$$l_e \ll r_0 + d, \quad l_i \ll r_0 + d, \tag{9}$$

where l_e and l_i are the electron and ion mean free paths, respectively; r_0 is the dust particle radius; and d is the characteristic size of the region where the plasma quasi-neutrality breaks down. In the case considered here, as will be seen below, conditions (9) are always satisfied for ions. In contrast, these conditions are violated for electrons, with l_e and d being comparable. Therefore, the orbital motion limited (OML) model is also unsuitable for describing the electron transport. Below, we will show that using effective boundary conditions allows the electron transport to be properly described in this regime as well; in the collisionless limit, the electron flux on the dust particle is equal to the OML flux.

As was shown in [22], if the condition

$$l_u \gg r_0 + d \tag{10}$$

is satisfied, then the electron temperature remains virtually constant in the region of plasma perturbation by the dust particle. Therefore, the transport coefficients and the rate constants of the electron production and loss processes may be assumed to be constant and can be determined from the plasma parameters far from the dust particle. In inequality (10), l_u is the electron energy relaxation length:

$$l_u = (N\sqrt{\sigma_m \sigma_u})^{-1}, \tag{11}$$

where N is the density of the plasma component that determines the electron energy losses, σ_m is the transport cross section, and σ_u is the effective cross section for inelastic collisions.

Thus, the complete system of equations to describe the charging of dust particles in the one-dimensional approximation in a spherical coordinate system with the origin at the center of the dust particle takes the form

$$\begin{aligned}
&\frac{\partial n_e}{\partial t} - \frac{1}{r^2}\frac{\partial}{\partial t}\left[r^2\left(D_e\frac{\partial n_e}{\partial r} + k_e n_e E\right)\right] \\
&\quad = Q_{\text{ion}} - \beta_{ei} n_e n_i - \beta_3 n_e^2 n_i, \\
&\frac{\partial n_i}{\partial t} + \frac{1}{r^2}\frac{\partial}{\partial t}\left[r^2\left(-D_i\frac{\partial n_i}{\partial r} + k_i n_i E\right)\right] \\
&\quad = Q_{\text{ion}} - \beta_{ei} n_e n_i - \beta_3 n_e^2 n_i, \\
&\frac{1}{r^2}\frac{\partial (r^2 E)}{\partial r} = 4\pi e(n_i - n_e).
\end{aligned} \tag{12}$$

For the electron and ion balance equations, we used effective boundary conditions on the left (on the particle surface) [21]:

$$\begin{aligned}
&\left[n_e - \gamma_e \frac{l_e}{r}\frac{\partial (n_e r)}{\partial r}\right]_{r = r_0} = 0, \\
&\left[n_i - \gamma_i \frac{l_i}{r}\frac{\partial (n_i r)}{\partial r}\right]_{r = r_0} = 0,
\end{aligned} \tag{13}$$

where γ_e and γ_i are constants that weakly depend on the ratios of the mean free paths to the dust particle radius and zero boundary conditions on the right (at the boundary of the Seitz–Wigner cell) for the electron and ion fluxes and for the electric field:

$$\mathbf{j}_e\big|_{r = a_d} = 0, \quad \mathbf{j}_i\big|_{r = a_d} = 0, \quad \mathbf{E}\big|_{r = a_d} = 0. \tag{14}$$

The following dependences of γ_e and γ_i on the ratios l_e/r_0 and l_i/r_0 were used [22]:

$$\begin{aligned}
\gamma_{e(i)} &= \frac{2}{3} + \frac{1}{6}\lambda_{e(i)}^{5/2}, \quad \lambda_{e(i)} \le 1, \\
\gamma_{e(i)} &= \frac{4}{3} - \frac{1}{2}\lambda_{e(i)}^{-1}, \quad \lambda_{e(i)} \ge 1,
\end{aligned} \tag{15}$$

where $\lambda_{e(i)} = l_{e(i)}/r_0$. These formulas also yield correct fluxes in the collisionless limit. Indeed, conditions (13) for electrons specify the flux density on the dust particle defined by

$$j_{e,0} = -\frac{1}{4} n_e\big|_{r = r_0} v_{T,e} - \alpha \frac{D_e}{r_0}\left[\frac{\partial (n_e r)}{\partial r}\right]_{r = r_0}, \tag{16}$$

where α is related to the parameter γ_e by

$$\gamma_e = \frac{4}{3}(1 - \alpha), \tag{17}$$

$v_{T,e} = \sqrt{8T_e/\pi m_e}$ is the thermal electron velocity. In the hydrodynamic limit, $\alpha = 1/2$ and, accordingly, $\gamma_e = 2/3$; in the collisionless limit, $\alpha = 0$ and $\gamma_e = 4/3$.

The boundary conditions (13) were widely used in the theory of neutron and electromagnetic radiation transport [23, 24]. The reason why these remain the same when passing to the case with drift motion is as follows. The anisotropy in the velocity distribution function for electrons and ions caused by the electric field is much smaller than the anisotropy due to the absorption of charged particles by the dust particle. Let us show this for the electron component as an example (the situation for ions is analogous). The stationary electron distribution function may be represented as

$$f(r, \mathbf{v}) \approx f_0(r, v) + f_1(r, \mathbf{v}) + f_2(r, \mathbf{v}), \tag{18}$$

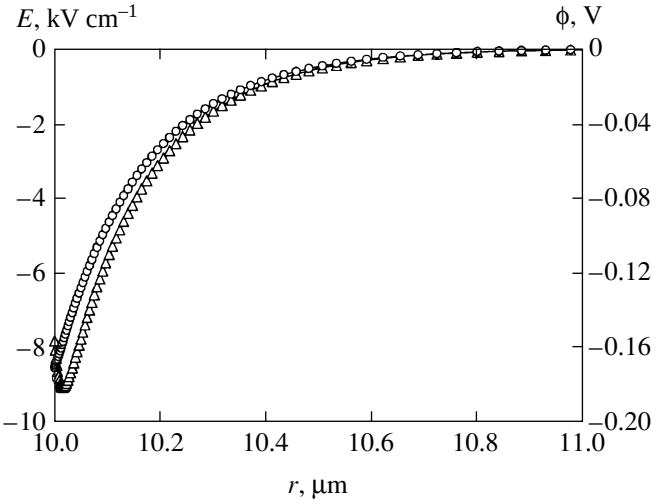


Fig. 1. Radial distributions of the electric field (△) and potential (○) of a solitary dust particle in an argon plasma: $T_e = 0.4$ eV, $r_0 = 10$ μm, $n_e = 10^{16}$ cm^{-3}, and $n_d = 10^8$ cm^{-3} (the Seitz–Wigner cell radius is $a_d = 13.4$ μm).

where f_0 is the spherically symmetric part; f_1 is the anisotropic part attributable to the electric field; f_2 is the anisotropic part attributable to the absorption of electrons by the dust grain; and $\mathbf{v}$ and v are the vector and magnitude of the electron velocity, respectively. If the dust grain does not absorb charged particles, then, in the first approximation, $f_2 \approx 0$. In this case, the anisotropy in the distribution function is caused only by the electric field; therefore (see, e.g., [25]),

$$\frac{f_1}{f_0} \approx \frac{v_{\text{drift}}}{v_{T,e}}, \tag{19}$$

where v_{drift} is the electron drift velocity. In moderate electric fields, this ratio is always small.

For an absolutely absorbing dust grain, the boundary condition for the distribution function is

$$f(r_0, \mathbf{v}) = 0, \quad \theta \le \frac{\pi}{2}, \tag{20}$$

where θ is the angle between the normal to the dust particle surface and the electron velocity direction. Since f_1 is small, we find from (20) that the following relation must hold at the boundary:

$$f_2(r_0, \mathbf{v}) = -f_0(r_0, v), \quad \theta \le \frac{\pi}{2}. \tag{21}$$

This relation indicates that f_0 and f_2 are comparable near the dust grain, suggesting that the drift motion near the dust particle may be disregarded in moderate electric fields [21]. Therefore, the boundary conditions in the presence of drift are the same as those in the absence of drift.

Coulomb collisions become important at high charged particle densities; in this case, electron–electron collisions will cause the electron distribution function to become Maxwellian, which significantly increases the reliability of the HD results. The Coulomb collisions were taken into account as follows. First, we determined the electron mean free path for *e*–*e* and *e*–*i* collisions [25]:

$$l_{ee} = l_{ei} = \frac{9}{4}\frac{T_e^2}{e^4 n_e \ln\Lambda}, \tag{22}$$

where $\ln\Lambda$ is the Coulomb logarithm defined by the relation [25]

$$\ln\Lambda = \ln\left(\frac{3}{2\sqrt{\pi}}\frac{T_e^{3/2}}{e^3 n_e^{1/2}}\right). \tag{23}$$

Then, we determined the total electron mean free path

$$\frac{1}{l} = \frac{1}{l_{en}} + \frac{1}{l_{ei}}, \tag{24}$$

where l_{en} is the electron mean free path without Coulomb collisions that was calculated using the transport cross sections from [26]. Finally, we determined the electron diffusion coefficient from the classical formula

$$D_e = \frac{1}{3} l v_{T,e}. \tag{25}$$

The effective boundary conditions (13) incorporate the effective mean free path that also includes the contribution from *e*–*e* collisions:

$$l_e = \left(\frac{1}{l_{en}} + \frac{1}{l_{ee}} + \frac{1}{l_{ei}}\right)^{-1} = (l_{en}^{-1} + 2l_{ei}^{-1})^{-1}. \tag{26}$$

The ionization rate was specified to ensure the required electron density $n_{e\infty}$ far from the dust particle:

$$Q_{\text{ion}} = (\beta_{ei} + \beta_3 n_{e\infty}) n_{e\infty}^2. \tag{27}$$

3. THE CHARGING OF DUST PARTICLES IN ARGON

The distributions of the electric field strength and potential obtained by numerically simulating the charging of a solitary dust particle with the radius $r_0 = 10$ μm in argon at room temperature and the number density $N = 2.5 \times 10^{19}$ cm^{-3} are shown in Fig. 1. At high pressures, the main ion in an argon plasma is Ar_2^+; its dissociative recombination coefficient is given by $\beta_{ei} = 0.85 \times 10^{-6}(0.026/T_e)^{0.67}$ cm^3 s^{-1} [27], and its mobility k_i is equal to 2.1 cm^2 V^{-1} s^{-1} [28]. As will be seen from the

subsequent analysis, the rapid dissociative recombination of the molecular ion plays a major role in the charging kinetics of dust particles. The electron temperature was assumed to be $T_e = 0.4$ eV, and the electron density far from the particle was varied within the range $n_{e\infty} = 10^{14}$–10^{16} cm^{-3}. In argon, the three-body recombination may be disregarded for the parameters under consideration.

As we see from Fig. 1, the radial dependence of the electric field passes through a minimum. Such a dependence was also observed at lower densities $n_{e\infty}$ and in our calculations with the main plasma ion Na^+ early in the development of a photoresonant plasma (see below). The radial behavior of the potential proves to be fairly regular and monotonic. Since the nonmonotonic behavior of the electric field near the dust particle is quite unexpected, let us consider this effect in more detail.

The Ar_2^+ mean free path under normal conditions is $l_i \approx 0.05$ μm, and the region where the plasma quasineutrality breaks down is $d \approx 1$ μm in size; therefore, the hydrodynamic description of the ion transport is correct. The electron mean free path is $l \approx 1.5$ μm and $l_e \approx 1.1$ μm. Since d and l_e are comparable, the hydrodynamic approach for the electron component is incorrect. (Note that the electron mean free path l_e and the distance from the point of minimum to the dust particle surface are different in scale). Our numerical simulations indicate that the electrons have a Boltzmann distribution with a high accuracy (see Fig. 2). In this case, the electron flux on the dust particle in the collisionless limit is equal to the OML flux. Indeed, in the collisionless limit, it follows from the effective boundary conditions (16) that the electron flux on the dust particle in the stationary case is Maxwellian:

$$j_e\big|_{r=r_0} \approx \frac{1}{4} n_e\Big|_{r=r_0} v_{T,e}. \tag{28}$$

Given the Boltzmann electron distribution, this leads to the following relation for the total electron flux:

$$\begin{aligned} J_e\big|_{r=r_0} &\approx 4\pi r_0^2 j_e\big|_{r=r_0} \\ &= \pi r_0^2 v_{T,e} n_{e,\infty} \exp(e\phi|_{r=r_0}/T_e), \end{aligned} \tag{29}$$

which is identical to the relation for the OML electron flux [29].

Figure 2 shows the electron and ion distributions in the immediate vicinity of the dust particle, from which we see why the electric field is nonmonotonic. According to the Boltzmann distribution, the electron density near the dust particle decreases to a value (determined by the electric field potential and the electron density and temperature far from it) that significantly exceeds the ion density (in the hydrodynamic limit $l_i/r_0 \longrightarrow 0$, the ion density at the boundary tends to zero). Therefore, a region of negative space charge emerges near the dust particle; this region is responsible for the nonmonotonic behavior of the electric field. Actually, however, the electron density at the boundary near a catalytic wall (absorbing all electrons) must also be equal to zero. Therefore, the Boltzmann law for the electron distribution will be violated near the dust particle. Since there is an electron flux on the dust particle that is not balanced by the reverse flux, there is no thermodynamic equilibrium; the closer the dust particle surface, the lower the reverse flux, and the larger the deviation from equilibrium. Using the effective boundary conditions allows us to correctly determine the fluxes rather than the densities of the plasma particles near the catalytic wall. An analysis on a kinetic level is required to correctly determine the electron and ion densities near the dust grain; i.e., the inhomogeneous and nonstationary Boltzmann equations for the distribution functions with collisions should be solved together with the Poisson equation for the self-consistent electric field, which currently seems to be an intractable problem. Calculations with a varied electron mean free path at constant electron temperature indicate that the distributions of all quantities after the minimum of the electric field weakly depend on l_e. Therefore, it may be assumed that we correctly determine all physical quantities after the minimum of the electric field.

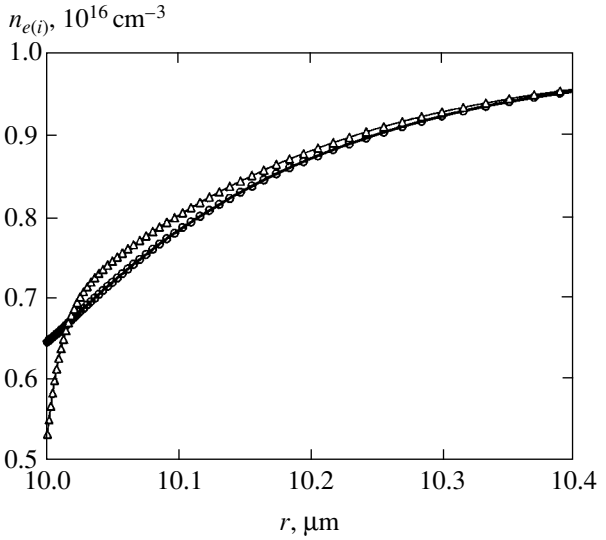


Fig. 2. Radial distributions of the electron (○) and ion (△) densities in the immediate vicinity of a solitary dust particle in argon at $T_e = 0.4$ eV, $r_0 = 10$ μm, $n_e = 10^{16}$ cm^{-3}, and $n_d = 10^8$ cm^{-3} ($a_d = 13.4$ μm). The thick curve represents the Boltzmann distribution for electrons.

Since, as we mentioned above, the charge of a solitary dust particle in standard charging theories [21, 29] does not depend on the electron density far from it, $n_{e\infty}$, studying the influence of $n_{e\infty}$ on the charge is of interest. In this paper, we checked whether the dust particle was solitary. To this end, we checked whether the elec-

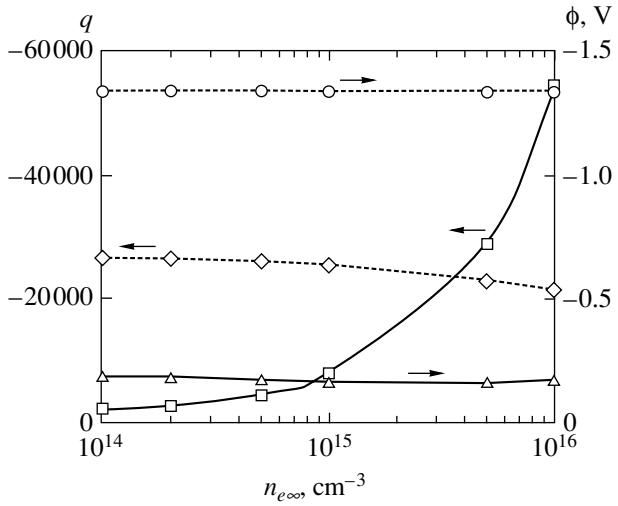


Fig. 3. Charge and potential of a solitary dust particle in an argon plasma with Coulomb collisions versus electron density far from it at $T_e = 0.4$ eV and $r_0 = 10$ μm: the calculated charge (□); the analytical charging theory (◇) [4, 9]; the calculated potential (△); and the OML potential (○) [29].

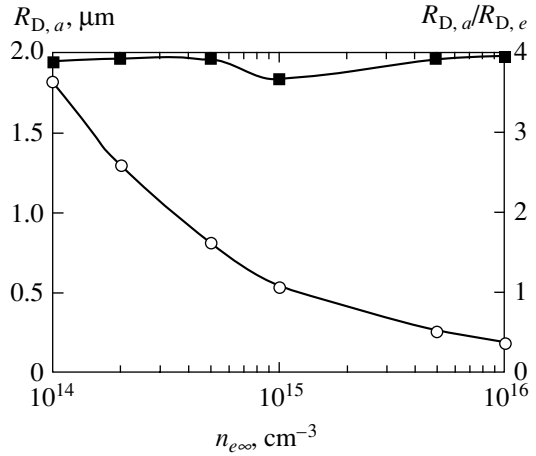


Fig. 4. Screening length for the field of a solitary dust particle obtained by fitting the calculated potential by the Debye potential in an argon plasma versus electron density far from it at $T_e = 0.4$ eV and $r_0 = 10$ μm (○) and reduced to the classical electron Debye length (■).

tron and ion densities at the right boundary (at the boundary of the Seitz–Wigner cell) were equal to the given one:

$$n_e\big|_{r = a_d} = n_i\big|_{r = a_d} = n_{e\infty}. \qquad (30)$$

The Seitz–Wigner radius was chosen in such a way that condition (30) was satisfied. Naturally, a_d increased with decreasing $n_{e\infty}$.

Figure 3 shows the charge and potential of dust particles in an argon plasma at various electron densities far from them. In contrast to the case of a weakly ionized plasma, the dust particle charge strongly depends on the electron density and, as our calculations indicate, reaches a stationary level at the values of $n_{e\infty}$ under consideration in fractions of a nanosecond. We clearly see from Fig. 3 that the dust particle potential is virtually independent of $n_{e\infty}$ and is much lower than the calculated OML potential. These peculiarities will be discussed below. As we see from Fig. 3, the analytical charging theory [4, 9] yields a weakly falling dependence of the charge on $n_{e\infty}$ (because the electron mobility decreases due to the increase in *e–i* collision frequency) that is at variance with the steeply rising dependence observed in our calculations.

We fitted the calculated potential by the Debye potential

$$\phi(r) = \frac{eq_a}{r}\exp(-r/R_{\mathrm{D},a}), \qquad (31)$$

where r is the distance from the center of the dust particle to the point under consideration; q_a and $R_{\mathrm{D},a}$ are the parameters of the fit—the charge and the screening length. The screening lengths obtained by fitting are compared with the plasma and electron Debye lengths in Fig. 4. We see that $R_{\mathrm{D},a}$ is appreciably larger than even $R_{\mathrm{D},e}$, but the ratio of $R_{\mathrm{D},a}$ to $R_{\mathrm{D},e}$ weakly depends on the plasma density, being close to 4 in the entire $n_{e\infty}$ range under study.

The fitting results indicate that the calculated potential is satisfactorily represented by the Debye potential. Given the finite size of the dust particles, the Debye–Hückel theory yields the following screened potential, called the DLVO (Derjagin–Landau–Verwey–Overbeek) potential:

$$\phi = \frac{eq_{\mathrm{eff}}}{1 + r_0/R_{\mathrm{D},a}}\frac{1}{r}\exp\left(-\frac{r - r_0}{R_{\mathrm{D},a}}\right). \qquad (32)$$

In this case, the electric field is defined by

$$E = \frac{eq_{\mathrm{eff}}}{r^2}\frac{1 + r/R_{\mathrm{D},a}}{1 + r_0/R_{\mathrm{D},a}}\exp\left(-\frac{r - r_0}{R_{\mathrm{D},a}}\right) \qquad (33)$$

and the field at the grain–plasma boundary is a Coulomb one:

$$E\big|_{r = r_0} = \frac{eq_{\mathrm{eff}}}{r_0^2}. \qquad (34)$$

One of the methods for determining the charge q of dust particles in our hydrodynamic calculations involved precisely the boundary condition (34), which is a corollary of the Gauss–Ostrogratski theorem and expresses the law of charge conservation in the spherically symmetric problem. The second method involved the integration of the difference between the ion and electron fluxes on the dust particle. The degree of coincidence

between the charges determined by the two methods served as evidence of their accurate determination; the accuracy generally exceeded five decimal digits. The charge q_{eff} is related to the fitted charge q_a by

$$q_a = \frac{q_{\text{eff}}}{1 + r_0/R_{\text{D},a}} \exp\left(\frac{r_0}{R_{\text{D},a}}\right). \quad (35)$$

Clearly, if $r_0 \ll R_{\text{D},a}$, then both charges, q_{eff} and q_a, will virtually coincide.

Fitting the potential by formula (32) allows us to determine whether a Coulomb crystal of dust particles can be formed using criterion (5) for the dense-plasma parameters considered here. Since, as we see from Fig. 4, a dust particle in argon perturbs the plasma only at small distances, the interaction energy of the dust particles will be comparable to their thermal energy only at high dust number densities exceeding 10^7 cm^{-3}. In this case, $\lambda \approx 18$ at $n_{e\infty} = 10^{14}$ cm^{-3}, and the structural parameter becomes even larger as the electron density increases. There are no calculations of the crystallization conditions at such large λ in the literature. However, based on simple estimates, we may conclude that the interaction between the dust particles in argon will be weak at reasonable dust particle number densities (10^6–10^7 cm^{-3} or lower).

4. PHOTORESONANT PLASMA

According to [16–19, 30, 31], at a high density of the medium and upon exposure to pulsed laser radiation, a photoresonant plasma is produced as follows. At the first stage of the breakdown, the resonance transition is saturated by laser radiation and initial ionization emerges in the processes involving excited atoms—associative ionization, multiphoton (usually two-photon) ionization of the resonance level—in laser-induced collisions. The seed electrons then rapidly gain energy in hyperelastic collisions with resonantly excited atoms of the medium, intensively populating and ionizing the higher lying levels. In this case, laser radiation again plays a certain subsidiary role, speeding up the growth of the electron density through the one-photon photoionization of highly excited levels. In the long run, the development of an electron avalanche leads to almost complete ionization of the medium in a narrow channel along the laser beam. It thus follows that the resonant breakdown of the medium requires that the intensity of the laser radiation be only on the order of the saturation intensity. The latter is fairly low (10–1000 W cm^{-2}) in metal vapors due to the large dipole moment and the small width of the resonance transition (of course, in actual experiments, much higher intensities, ~10^5–10^8 W cm^{-2}, at a pulse duration of 10^{-8}–10^{-6} s are required to produce a plasma channel ~1–10 cm in length in vapors with concentrations of 10^{15}–10^{17} cm^{-3} due to strong absorption and finite development time of the electron avalanche).

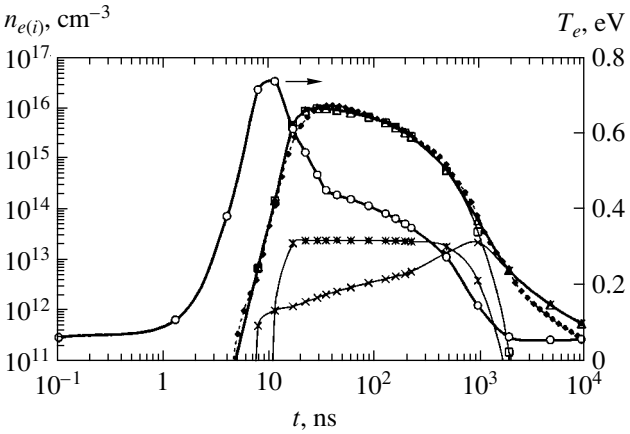


Fig. 5. The pattern of electron temperature and density variations in a photoresonant sodium plasma (based on the data of [19]): the electron (△), Na^+ (□), Na_2^+ (×), $NaAr^+$ (∗) densities, the electron density calculated here (◆), and the electron temperature (○).

The properties of a photoresonant sodium plasma in the presence of a buffer (argon) gas were studied in detail previously [19]; we also performed numerical simulations of the plasma parameters that showed good agreement with experimental data. For this reason, when simulating the charging of dust particles in a photoresonant plasma, we relied on our previous results [19]. All of the experiments in this work were carried out using a heated cell at a temperature of the sodium vapors in its hot zone up to $T \approx 700$ K and their concentration $N_{\text{Na}} = 1.8 \times 10^{16}$ cm^{-3}. The argon concentration in the hot zone of the cell was higher by two orders of magnitude, $N_{\text{Ar}} = 2.6 \times 10^{18}$ cm^{-3}. The vapors were ionized by the radiation of a dye laser with an intensity up to 10^8 W cm^{-2} and a pulse duration at half maximum of about 16 ns that was tuned to the wavelength of the resonance transition in the sodium atom, $3^2S_{1/2}$–$3^2P^0_{3/2}$ (589.995 nm). Figure 5 shows the pattern of electron temperature and density variations in a photoresonant sodium plasma numerically simulated in [19]. These dependences, which agree with experimental data, were used here as the input parameters of the problem.

As we see from Fig. 5, Na^+ is the main ion of the photoresonant plasma until 1 μs; subsequently, the diatomic ion Na_2^+ gradually becomes the main ion. $NaAr^+$ accounts for less than 1% of the density of the main ion. Note that $NaAr^+$ ions play a crucial role in the drop in electron temperature after the completion of the laser pulse. For dissociative recombination of these ions, the ionization energy goes into breaking the bond and into the kinetic energy of the dissociation products. In contrast, for three-body Na^+ recombination, a highly excited sodium atom with a long radiative lifetime is formed; subsequently, it is involved in the electron

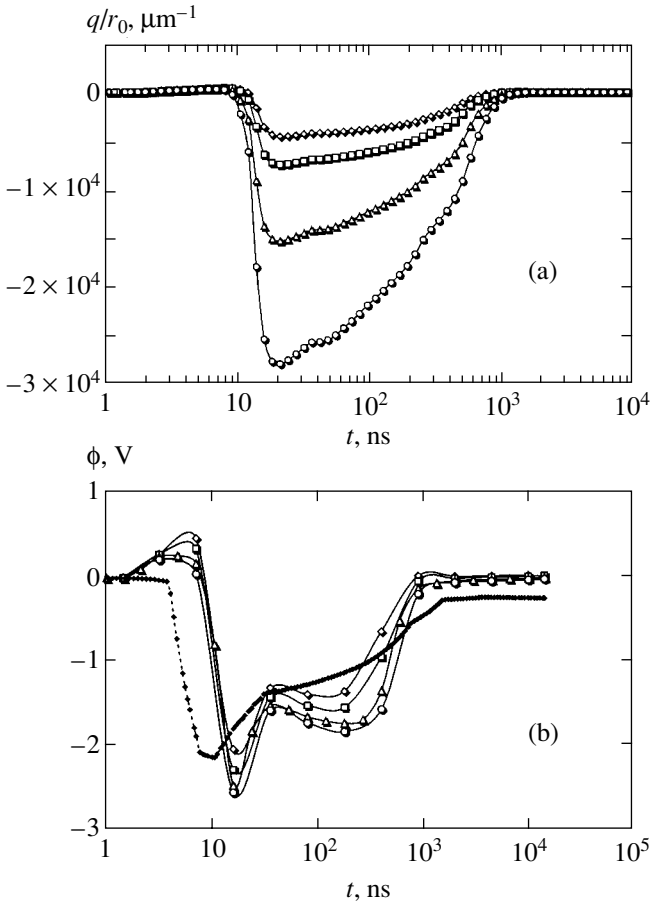


Fig. 6. Evolution of the charge (a) and potential (b) of a solitary dust particle in a photoresonant sodium plasma at $n_d = 2 \times 10^5$ cm^{-3} ($a_d = 106$ μm). The dust particle radius is $r = 1$ (◇), 2 (□), 5 (△), and 10 (○) μm; the calculated OML potential (◆).

heating in impacts of the second kind or can be ionized. Therefore, energy is found to be removed from a photoresonant plasma without $NaAr^+$ ions mainly in elastic collisions of electrons with argon atoms. Despite their low density, the $NaAr^+$ ions lead to the removal of a substantial amount of energy from this system, causing the electrons to cool down. There were no atomic and molecular argon ions due to the low plasma temperature and the high ionization potential of the argon atom. Here, we solved a nonstationary problem with a time-varying ionization rate chosen in such a way that the electron and ion densities in the unperturbed plasma at each instant of time during our calculations were close to those in [19] (the triangles in Fig. 5). The ionization in a photoresonant plasma is a complex dynamical phenomenon that involves excited sodium atoms, electrons heated in impacts of the second kind, and, at the initial stage, laser radiation, with the ionization being clearly seen until 10^3 ns. Three-body recombination with a coefficient $\beta_3 = 8.75 \times 10^{-27} T_e^{-9/2}$ cm^6 s^{-1} [25] (the temperature is in eV) and dissociative recombination with a coefficient $\beta_{ei} = 3 \times 10^{-7}(0.026/T_e)^{1/2}$ cm^3 s^{-1} [19] were assumed to be the main Na^+ and Na_2^+ loss processes, respectively. In our calculations, we considered only one ion with its mobility and diffusion and recombination coefficients recalculated to plasma ion composition [19]. The recalculation was performed in a simple way—the mobility was determined as the sum of the Na^+ and Na_2^+ mobilities multiplied by the fraction of the corresponding ion [19]:

$$k_i = k_{i1}\frac{n_{i1}}{n_i} + k_{i2}\frac{n_{i2}}{n_i},$$

where the quantities with subscripts 1 and 2 refer to monoatomic and diatomic sodium ions, respectively; $n_i = n_{i1} + n_{i2}$. The dissociative recombination coefficient of Na_2^+ was multiplied by the fraction of this ion, and the diffusion coefficient was determined from the Einstein relation

$$D_i = D_{i1}\frac{n_{i1}}{n_i} + D_{i2}\frac{n_{i2}}{n_i} \equiv \frac{k_i T_i}{e}.$$

The mobilities of the monoatomic and diatomic sodium ions reduced to normal conditions were determined using the Langevin formula [28]:

$$k_{i1(2)} = 36/\sqrt{\alpha\mu_{1(2)}}\ \text{cm}^2/\text{V s}, \qquad (36)$$

where $\mu_{1(2)}$ is the reduced mass of the monoatomic or diatomic sodium ion and the neutral (argon) particle, and $\alpha = 11.08$ is the polarizability of the argon atom in atomic units.

In our calculations, the possible perturbation of the photoresonant plasma by dust particles could be disregarded, because even at their concentration of 10^7 cm^{-3} and at a maximum particle charge q no larger than 3×10^5 (see below), the total density of the charge concentrated in the particles was less than 3×10^{12} cm^{-3}, which is more than three orders of magnitude lower than the density of the plasma itself (see Fig. 5).

The evolution curves for the charge of dust particles with various radii in a photoresonant plasma are shown in Fig. 6a. The calculations were performed with a dust particle density of 2×10^5 cm^{-3}. We see that the dust particle charge reaches 3×10^5 electron charges. The dust particle charge increased by almost an order of magnitude compared to a purely argon plasma, because the recombination coefficient of Na^+ is lower than that of Ar_2^+. At the initial stage of photoresonant plasma ignition until 10 ns, when the electron density is not yet high enough, the behavior of the electric field was nonmonotonic, while the dust particle charge and potential were even positive (see Fig. 6b). However, as we noted above, this is because the hydrodynamic approach near

the dust particle at the early stage is incorrect. The dust particle charge proves to be significant until 10^3 ns. Subsequently, the main plasma ion is the diatomic sodium ion (see Fig. 5) with a large dissociative recombination coefficient; therefore, the plasma particle density and the dust particle charge decrease sharply. Figure 6b presents the evolution curves for the surface potential of dust particles that are seen to be unexpectedly close to one another at various dust particle radii and to the calculated OML potential. In these OML calculations [29], the electron and ion (two types) fluxes on the dust particle were specified by using the plasma parameters from [19]. The pattern of charge variations was determined by time integration using a semi-explicit scheme of the difference between the ion and electron fluxes. (In the OML model, the potential of a dust particle does not depend on its radius.) A certain lead in the growth of the OML potential curve stems from the fact that the fluxes were calculated from the current parameters of the unperturbed plasma far from the dust particle.

To determine whether a Coulomb crystal can be formed in a photoresonant plasma, we fitted the calculated potential by the Debye potential. Figure 7 shows the screening length for the dust particle field obtained by fitting and compares their values with the electron Debye length determined from the classical formula (4). We see that the screening length obtained by fitting is larger than the electron Debye length, with $R_{D,a}$ (at approximately the same electron density) in argon being appreciably smaller (see Fig. 4) than that in a photoresonant sodium plasma. This is attributable to the different electron–ion recombination rates, because the size of the region of plasma perturbation by the dust particle depends on the bulk loss rate of the charged plasma particles. The size of this region R can be estimated by equating the recombination and diffusion loss (due to ambipolar escape to dust particles) rates of the electrons and ions.

We obtain

$$R = \pi\sqrt{D_a/n_{i\infty}(\beta_{ei} + \beta_3 n_{e\infty})}, \quad (37)$$

where D_a is the ambipolar diffusion coefficient defined by the relation [25]

$$D_a = D_i(1 + T_e/T_i). \quad (38)$$

When passing to a photoresonant sodium plasma, the recombination coefficient decreases by two orders of magnitude compared to the case of particle charging in pure argon considered in the third section, which causes the region of plasma perturbation by the dust particle to increase in size by an order of magnitude. As a result, the screening length also increases. As we see from Fig. 7b, after $t = 1$ μs, the ratio $R_{D,a}/R_{D,e}$ decreases sharply, because the type of the main plasma ion changes. The fact that the Na^+ recombination coefficient is small leads to approximately equal potentials of the dust particle in the HD and OML model approximations, because the bulk destruction of charged particles near the dust particle decreases in importance. When Na_2^+ ions dominate, as we see from Fig. 6b, the HD and OML data strongly disagree, as in an argon plasma.

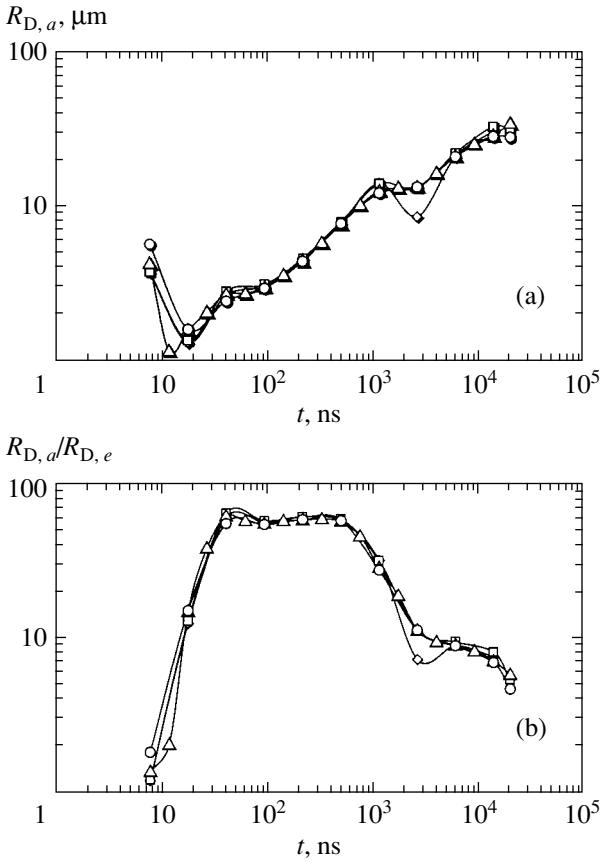


Fig. 7. Screening length for the dust particle charge in μm (a) and reduced to the electron Debye length (b) in a photoresonant sodium plasma versus time. The dust particle radius is $r = 1$ (◇), 2 (□), 5 (△), and 10 (○) μm.

The dust particle density at which condition (5) is satisfied can be estimated from the fitting results based on relation (2). These data are presented in Fig. 8. Beyond (before and after) these curves, the crystallization condition could not be satisfied at all dust particle densities. In the dense photoresonant sodium plasma under consideration, the crystallization condition for particles with a radius $r_0 = 10$ μm can be satisfied at dust particle densities $n_d < 10^7$ cm^{-3}.

In conclusion, note that the data in Figs. 3 and 6 clearly indicate that there is no vacuum relationship between the charge and the potential in a dense plasma. In general, the relationship between the charge of a dust particle and its potential depends on the spatial plasma charge distribution near the dust particle [22]. In the

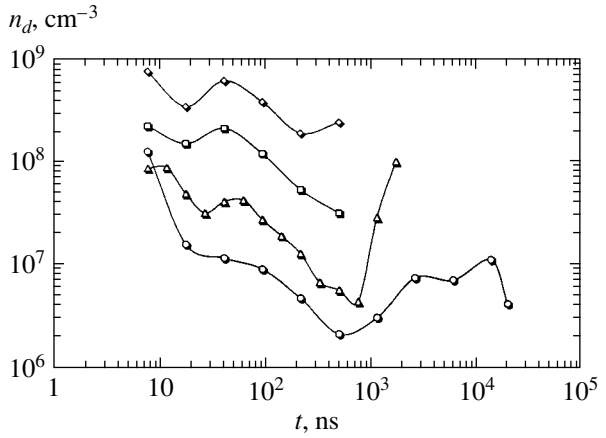


Fig. 8. Dust particle density above which $\Gamma_S > \Gamma_{SC}$ in a photoresonant sodium plasma versus time. The dust particle radius is r = 1 (◇), 2 (□), 5 (△), and 10 (○) μm.

spherically symmetric case, this relationship is given by the boundary condition on the dust particle

$$\phi|_{r = r_0} = \frac{eq}{r_0} + 4\pi e\int_{r_0}^{\infty}(n_i - n_e)rdr. \qquad (39)$$

The derived expression shows that, due to the appearance of space charge, the dust particle potential will be lower in absolute value than that in a vacuum for the so-called vacuum charge–potential relationship. If the screening of the dust particle charge can be described by the screened Debye potential, then, as follows from (32), the charge–potential relationship is given by [22]

$$\phi|_{r = r_0} = \frac{eq}{1 + r_0/R_{D,a}}\frac{1}{r_0}. \qquad (40)$$

Therefore, the potential proves to be much lower than the vacuum potential for $R_{D,a} \ll r_0$. Note also that the OML model yields the surface potential of the dust particle rather than its charge.

We see from Figs. 3 and 6 that the OML potential is higher than or comparable to the HD potential. Filippov *et al.* [22] found that the OML potential in He at atmospheric pressure, when condition (9) is also satisfied for electrons, is lower than the calculated HD potential. As was pointed out in [32, 33], when passing from collisionless to hydrodynamic electron and ion transport, the dust particle charge passes through a minimum as the pressure rises. The different behavior of the HD potential compared to the OML potential in this study and in [22] is a reflection of precisely this fact.

Let us now estimate the contribution of photoelectron emission under the effect of laser radiation. The photon energy in the experiments was 2.1 eV, and the work function of the ordinary materials except Cs was higher than this value. For example, the work function for sodium is 2.35 eV [34]. Given the Fowler dependence below the threshold, the quantum efficiency for sodium (any dust particle in a medium of sodium vapors will be coated with a layer of this metal) will not exceed 10^{-6}. Consequently, the flux density of the one-photon photoemission electrons can be estimated as $j_{1ph} \le 0.8 \times 10^{20}$ cm^{-2} s^{-1} at a laser intensity of 100 MW cm^{-2}. A high laser intensity can lead to a two-photon photoeffect, but, according to the experiments [35, 36], the efficiency of this process will not exceed $(1–3) \times 10^{-7}$ in our conditions. Consequently, the flux density of the two-photon photoelectrons can be estimated as $j_{2ph} = (1–3) \times 10^{19}$ cm^{-2} s^{-1}. As our calculations show, in the stationary case, the ion flux (which is equal to the electron flux) is 20–50% of the Langevin flux, which is defined by the relation [21]

$$J_L = 4\pi e k_i q n_{i\infty} = q\beta_L n_{i\infty}, \qquad (41)$$

where β_L is the Langevin recombination coefficient [28]. Let us estimate the Langevin electron flux density with typical parameters at maximum of $q \approx -2.8 \times 10^5$, $n_{e\infty} \approx 10^{16}$ cm^{-3}, r_0 = 10 μm, and k_i = 28.3 cm^2 V^{-1} s^{-1}. We obtain

$$j_L = J_L/4\pi r_0^2 \approx 1.1 \times 10^{22} \text{ cm}^{-2} \text{ s}^{-1},$$

which appreciably exceeds the photoelectron flux density. It should also be noted that photoemission affects the charging of dust particles only during a short laser pulse.

5. CONCLUSIONS

Our results have shown that, as expected, as one passes from a weakly ionized plasma to a fairly dense plasma and as the electron density increases up to ~10^{16} cm^{-3}, the dust particle charge begins to significantly depend on the plasma density (see Figs. 3 and 6), because Coulomb collisions are switched on. As n_e increases, such collisions begin to crucially affect the drift and diffusion velocities. Our numerical simulations also showed that the calculated potential of a solitary dust particle in a dense plasma decreases with increasing distance from it much more slowly than predicted by the classical Debye–Hückel theory; fitting the calculated data yields a screening length that is much (up to almost two orders of magnitude) larger than the classical electron Debye length (see Figs. 4 and 7). Once the dust particle charge has passed through its maximum at about 20 ns, it begins to decrease (see Fig. 6). Nonetheless, the dust particle density at which the photoresonant plasma crystallizes also gradually decreases, and this tendency is preserved until 10^3 ns (see Fig. 8). The reason is that the structural parameter λ decreases simultaneously, because the screening length $R_{D,a}$ increases (Fig. 7). In contrast, the exponen-

tial dependence of the nonideality parameter Γ_S on λ at $a > R_{D,a}$ defined by formula (2) is much sharper than the dependence of Γ_S on the particle charge and the dependence of the critical parameter Γ_{SC} on λ (see (6)). However, after the change of the type of main plasma ions, the decrease in the critical dust particle density changes to its slow increase.

Let us consider the possibilities for the formation of ordered structures in a photoresonant plasma in actual experiments. As a rough estimate of the upper limit for the formation time, we may use the time it takes for a macroparticle with thermal energy to shift by the mean interparticle distance, which is several fractions of a second for particles 10 μm in diameter in the photoresonant sodium plasma considered above in a mixture with an argon buffer gas. Therefore, for a plasma–dust structure to be formed, the photoresonant plasma must be maintained over this time interval, which can be done under pulse-periodic laser radiation with a fairly short pulse repetition period. Copper vapor lasers that are used to pump dye lasers can operate with pulse repetition frequencies as high as 100 kHz [37], which, in principle, makes the formation of ordered structures possible during the pulse-periodic formation of a photoresonant plasma. In addition, we can easily vary the photoresonant plasma concentration within the range 10^{15}–10^{17} cm^{-3} (see, e.g., [38]) and increase the buffer gas density by an order of magnitude and, thus, reduce the diffusion coefficient (at an argon pressure of the order of the atmospheric pressure and even at a plasma concentration of ~10^{16} cm^{-3}, the electron Coulomb collisions significantly dominate over the collisions with neutral atoms, see Section 2).

The question as to the possible size of the region where dust structures exist is yet to be solved, because laser radiation will rapidly scatter in the dust plasma at the high dust densities considered here. Indeed, the extinction cross section for large dust particles in the visible spectral range for an arbitrary refractive index of the particle material is ~$2\pi r_0^2$ [39]. In our case, the extinction coefficient can reach ~12 cm^{-1} for a particle concentration $n_d \approx 2 \times 10^6$ cm^{-3} and radius $r_0 \approx 10$ μm (i.e., laser radiation will be able to penetrate into the plasma to a depth of only about one millimeter). However, for nonabsorbing (transparent or totally reflecting) dust particles, laser radiation will only scatter, illuminating a region with the characteristic size determined by the small extinction coefficient in the resonant medium (~0.01–1 cm^{-1}).

Note that the effect of laser radiation with a pulse duration of ~10 ns and a peak intensity up to 10^8 W cm^{-2} on the plasma dust heating will be negligible when the particles are made of a transparent material (e.g., alumina powder) or a material that effectively reflects laser radiation (e.g., made of pure aluminum with a reflectance of above 90% for photons with a wavelength of 590 nm). Accordingly, the impact of the radiometric force on the levitation of dust particles is also negligible due to the short heating pulse duration, although this force at maximum will exceed gravity by an order of magnitude for a short time. The light pressure force under the action of a short laser pulse is appreciably weaker than the radiometric force. Estimates also indicate that particle heating by a spherically symmetric ion stream may also be disregarded for the plasma parameters under consideration.

In the experiment [40], melamineformaldehyde dust particles with a radius of 2 μm in an inductively coupled neon plasma at a pressure of 40–50 Pa began to "feel" laser radiation with a wavelength of 534 nm and with a duration at half-maximum of 7–11 ns only at a pulse energy of 15–20 mJ. The original laser beam about 1 cm in diameter was focused by a spherical lens with a focal length of 7.5 cm into a spot less than 100 μm in diameter. Consequently, we have 10^{10} W cm^{-2} as a rough estimate of the laser energy flux density that did not perturb the dusty component of the plasma in the experiment [40]. This value is two orders of magnitude higher than the intensity of the laser radiation required to produce a photoresonant plasma.

Finally, our study leads us to conclude that it is quite realistic to produce a photoresonant plasma and to investigate its properties. We see the following way of improving the crystallization conditions for the dust component in such a plasma. As we noted above, the decrease in dust particle charge, screening length, and, accordingly, nonideality parameter late in the decay of a photoresonant plasma is attributable to the recovery of diatomic sodium ions with a high recombination coefficient. To solve this problem, we may decrease the sodium atomic density to the minimum possible level, reduce the electron density by an order of magnitude, and replace argon with neon as the buffer gas, which will allow us to significantly reduce the formation of molecular ions in the photoresonant plasma and to decrease the electron energy losses in elastic collisions. This will also lead to a reduction in the laser intensity required to produce a photoresonant plasma, which will lower the thermal load on dust particles.

ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research (project no. 04-02-16775a) and a grant from the President of the Russian Federation for State Support of Leading Scientific Schools (no. NSh-1257.2003.2).

REFERENCES

1. *Dusty Plasmas in the New Millennium*, Ed. by R. Bharuthram, M. A. Hellberg, P. K. Shukla, and F. Verheest (AIP, Melvill, New York, 2002), AIP Conf. Proc., Vol. 649.
2. V. N. Tsytovich, Usp. Fiz. Nauk **167**, 57 (1997) [Phys. Usp. **40**, 53 (1997)].

3. A. P. Nefedov, O. F. Petrov, and V. E. Fortov, Usp. Fiz. Nauk **167**, 1215 (1997) [Phys. Usp. **40**, 1163 (1997)].
4. B. M. Smirnov, Usp. Fiz. Nauk **170**, 495 (2000) [Phys. Usp. **43**, 453 (2000)].
5. P. M. Shukla, Phys. Plasmas **8**, 1791 (2001).
6. A. Piel and A. Melzer, Plasma Phys. Controlled Fusion **44**, R1 (2002).
7. Ch. Hollenstein, Plasma Phys. Controlled Fusion **42**, R93 (2000).
8. V. N. Tsytovich, G. E. Morfill, and H. Thomas, Fiz. Plazmy **28**, 675 (2002) [Plasma Phys. Rep. **28**, 623 (2002)].
9. V. Yu. Baranov, I. A. Belov, A. V. Dem'yanov, *et al.*, in *Isotopes*, Ed. by V. Yu. Baranov (IzdAT, Moscow, 2000), p. 626.
10. M. O. Robbins, K. Kremer, and G. S. Grest, J. Chem. Phys. **88**, 3286 (1988).
11. M. J. Stevens and M. O. Robbins, J. Chem. Phys. **98**, 2319 (1993).
12. E. J. Meijer and D. Frenkel, J. Chem. Phys. **94**, 2269 (1991).
13. R. T. Farouki and S. Hamaguchi, Appl. Phys. Lett. **61**, 2973 (1992).
14. S. Hamaguchi, R. T. Farouki, and D. H. E. Dubin, Phys. Rev. E **56**, 4671 (1997).
15. A. F. Pal', D. V. Sivokhin, A. N. Starostin, *et al.*, Fiz. Plazmy **28**, 32 (2002) [Plasma Phys. Rep. **28**, 28 (2002)].
16. V. A. Kas'yanov and A. N. Starostin, in *Plasma Chemistry*, Ed. by B. M. Smirnov (Énergoatomizdat, Moscow, 1990), No. 16 [in Russian].
17. I. M. Beterov, A. V. Eletskiĭ, and B. M. Smirnov, Usp. Fiz. Nauk **155**, 265 (1988) [Sov. Phys. Usp. **31**, 535 (1988)].
18. A. V. Eletskiĭ, Yu. N. Zaĭtsev, and S. V. Fomichev, Zh. Éksp. Teor. Fiz. **94** (5), 98 (1988) [Sov. Phys. JETP **67**, 920 (1988)].
19. A. G. Leonov, A. N. Starostin, and D. I. Chekhov, Zh. Éksp. Teor. Fiz. **111**, 1274 (1997) [JETP **84**, 703 (1997)].
20. A. F. Pal', A. N. Starostin, and A. V. Filippov, Fiz. Plazmy **27**, 155 (2001) [Plasma Phys. Rep. **27**, 143 (2001)].
21. A. F. Pal', A. O. Serov, A. N. Starostin, *et al.*, Zh. Éksp. Teor. Fiz. **119**, 272 (2001) [JETP **92**, 235 (2001)].
22. A. V. Filippov, N. A. Dyatko, A. F. Pal', and A. N. Starostin, Fiz. Plazmy **29**, 214 (2003) [Plasma Phys. Rep. **29**, 190 (2003)].
23. B. Davison, *Neutron Transport Theory* (Clarendon, Oxford, 1957; Atomizdat, Moscow, 1960).
24. G. I. Marchuk, *Methods for Nuclear Reactor Calculations* (Atomizdat, Moscow, 1961) [in Russian].
25. Yu. P. Raĭzer, *The Physics of Gas Discharge* (Nauka, Moscow, 1987) [in Russian].
26. J. L. Pack, R. E. Voshall, A. V. Phelps, and L. E. Kline, J. Appl. Phys. **71**, 5363 (1992).
27. V. A. Ivanov, Usp. Fiz. Nauk **162**, 35 (1992) [Sov. Phys. Usp. **35**, 17 (1992)].
28. B. M. Smirnov, *Ions and Excited Atoms in Plasma* (Atomizdat, Moscow, 1974) [in Russian].
29. M. S. Barnes, J. H. Keller, J. C. Forster, *et al.*, Phys. Rev. Lett. **68**, 313 (1992).
30. R. M. Measures, J. Quant. Spectrosc. Radiat. Transf. **10**, 107 (1970).
31. R. M. Measures and P. G. Cardinal, Phys. Rev. A **23**, 804 (1981).
32. A. V. Zobnin, A. P. Nefedov, V. A. Sinel'shchikov, and V. E. Fortov, Zh. Éksp. Teor. Fiz. **118**, 554 (2000) [JETP **91**, 483 (2000)].
33. Yu. A. Mankelevich, M. A. Olevanov, and T. V. Rakhimova, Zh. Éksp. Teor. Fiz. **121**, 1288 (2002) [JETP **94**, 1106 (2002)].
34. *Physical Quantities. Handbook*, Ed. by I. S. Grigor'ev and E. Z. Meĭlikhov (Énergoatomizdat, Moscow, 1991) [in Russian].
35. M. C. Teich and G. J. Wolga, Phys. Rev. **171**, 809 (1968).
36. E. M. Logophetis and P. L. Hartman, Phys. Rev. Lett. **18**, 581 (1967).
37. M. A. Alaev, A. I. Baranov, N. M. Vereshchagin, *et al.*, Kvantovaya Élektron. (Moscow) **3**, 1134 (1976).
38. O. L. Landen, R. J. Winfield, D. D. Burgess, *et al.*, Phys. Rev. A **32**, 2963 (1985).
39. M. Born and E. Wolf, *Principles of Optics*, 4th ed. (Pergamon, Oxford, 1969; Nauka, Moscow, 1970).
40. A. D. Usachev, private communication (2003).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 73–82.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 89–98.

FLUIDS

Dynamics of Jumpwise Temperature Pitch Variations in Planar Cholesteric Layers for a Finite Strength of Surface Anchoring¶

V. A. Belyakov[a], I. W. Stewart[b], and M. A. Osipov[b]

[a]*Landau Institute for Theoretical Physics, Moscow, 117334 Russia*
[b]*Department of Mathematics, University of Strathclyde, Glasgow, G1 1XH, Great Britain*
e-mail: bel@landau.ac.ru
Received September 15, 2003

Abstract—The dynamics of pitch jumps in cholesteric layers with a finite surface anchoring strength under variations in temperature is investigated theoretically. General expressions are presented that connect the dynamics of pitch jumps with the parameters that determine the process, such as the viscosity, the specific form of the anchoring potential, and the dimensionless parameter $S_d = K_{22}/Wd$, where W is the depth of the anchoring potential, K_{22} is the twist elastic modulus, and d is the layer thickness. It is found that the shape of the anchoring potential significantly influences the temporal behavior of the cholesteric helix in the process of a pitch jump. To illustrate this revealed dependence of the pitch jump dynamics on the shape and strength of the anchoring potential, the problem was investigated for two different models of the surface anchoring potential for a jump mechanism in connection with the director at the surface slipping over the barrier of the anchoring potential. Calculations for the unwinding (winding) of the helix in the process of the jump were performed to investigate the case of infinitely strong anchoring on one surface and finite anchoring on the other, which is important in applications. The results show that an experimental investigation of the dynamics of the pitch jumps will make it possible to distinguish different shapes of the finite strength anchoring potential and, in particular, it will provide a means for determining whether the well-known Rapini–Papoular anchoring potential is the best suited potential relevant to the dynamics of pitch jumps in cholesteric layers with a finite surface anchoring strength. The optimal conditions for experimental observation of these phenomena are briefly considered. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Recent investigations of temperature pitch variations in planar cholesteric layers and of the influence of finite surface anchoring and thermodynamic fluctuations on these variations [1, 2] have revealed some novel effects that are interesting in the general context of the physics and practical applications of liquid crystals. Similar investigations of pitch variation under the influence of applied external fields have also been carried out [3, 4]. Some experimental and application aspects of pitch variations in external fields were considered in [5]. It has been known for quite some time that the temperature evolution of a cholesteric liquid crystal (CLC) structure [6, 7] in samples with a finite surface anchoring energy may be continuous in some temperature ranges with jumpwise changes at definite temperature points, with a strong hysteresis effect occurring when the direction of temperature variations is changed [8–10]. This problem was recently investigated in [1, 2]. However, only thermodynamic equilibrium states of cholesteric layers were studied in these theoretical papers, and the problem of the dynamics of jumpwise transitions was not considered. Nevertheless, the dynamics of liquid crystals in restricted geometries are of general physical interest and are, therefore, of especially great concern in liquid crystal applications [6]. For example, the switching time between two bistable states studied experimentally in [11, 12], as was shown theoretically [13], is directly connected to the liquid crystal dynamics in the corresponding restricted geometry. In this article, the dynamics of the aforementioned jumpwise transitions are studied. We begin with the simplest cases that reveal jumpwise transitions, namely, theoretical investigations of the dynamics of temperature-induced jumpwise variations of the cholesteric pitch and director distribution in a planar layer of CLC with infinitely strong anchoring at one of its boundary surfaces and anchoring of a finite strength at its other boundary surface. We note that the dynamics of such jumpwise transitions is directly dependent on both the viscosity properties of liquid crystals and the characteristics of the surface anchoring potential. This is the motivation for introducing and considering different models for the surface anchoring potential. In addition to the well-known Rapini–Papoular anchoring potential, another possible model for the anchoring potential,

¶This article was submitted by the authors in English.

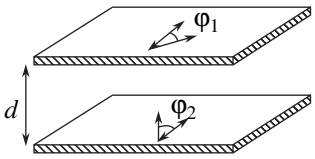


Fig. 1. The case of nonidentical anchoring at the surfaces of a cholesteric layer (for infinitely strong anchoring at the bottom surface $\varphi_2 = 0$).

the *B*-potential, is used in the calculations presented below.

2. SOME RESULTS OF EQUILIBRIUM INVESTIGATIONS

We present some results on the temperature behavior of the cholesteric helix in a planar cholesteric layer of finite thickness having a finite anchoring strength at one of its surfaces and infinite anchoring at the other, as depicted in Fig. 1. We first restrict the analysis of the temperature variations of the director configuration in the layer by assuming that the pitch-jump mechanism is connected with the overcoming of the anchoring barrier by the director at the surface and, moreover, that any liquid crystal (LC) thermal fluctuations may be neglected.

Below, we follow the approach and investigations reported earlier [1, 2] and concentrate on the transitions between N and $N + 1$ half-turns of the director in the layer, which proceed without strong local disturbances of the director configuration. Such a transition is unique if the finite anchoring at the second surface is sufficiently weak, i.e., the dimensionless parameter $S_d = K_{22}/Wd > 1/2\pi$, where W is the depth of the anchoring potential, K_{22} is the twist elastic constant, and d is the layer thickness. The case of small $S_d < 1/2\pi$ was studied numerically in [9].

We start by finding the temperature behavior of φ and the angle of deviation of the director from the alignment direction orientation at the surface of the cholesteric layer with finite anchoring (see Fig. 1). Following [6], we can write the free energy of the layer as

$$F(T) = W_s(\varphi) + \frac{1}{2}K_{22}d\left[\frac{2\pi}{p_d(T)} - \frac{2\pi}{p(T)}\right]^2, \qquad (1)$$

where $W_s(\varphi)$ is the surface anchoring potential, K_{22} is the elastic twist modulus, d is the layer thickness, $p(T)$ is the natural pitch value in an infinite sample of the cholesteric liquid crystal, and $p_d(T)$ is the corresponding pitch value in the layer.

Because the pitch value in the layer $p_d(T)$ is determined by the angle φ and the natural pitch $p(T)$ may be expressed via the angle $\varphi_0(T)$, the angle of the director deviation from the alignment direction at the surfaces with finite anchoring in the absence of any anchoring forces, expression (1) for the free energy is then readily represented as a function of these angles, namely,

$$F(T) = W_s(\varphi) + \frac{1}{2}\frac{K_{22}}{d}[\varphi - \varphi_0(T)]^2. \qquad (2)$$

We note that Eq. (2) is obtained from Eq. (1) using a simple change of variables $\varphi = qd$, where $q = 2\pi/p$ and $z = 0$ at the surface with infinite anchoring. The angle φ can be found from the condition for a minimum of the free energy in Eq. (2). Consequently, φ must satisfy the equation

$$\frac{\partial W_s}{\partial \varphi} + \frac{K_{22}}{d}[\varphi - \varphi_0(T)] = 0. \qquad (3)$$

An analysis of Eq. (3) reveals that a smooth change in the director deviation angle φ is possible when φ is less than some critical angle φ_c. As φ achieves the critical value φ_c, a jumplike change of the pitch occurs. For $S_d > 1/2\pi$, the transition to the unique new configuration of the helix occurs that differs by one in the number of half-turns N. In this case, it is possible to restrict the range of values of φ to the interval $[-\pi/2, \pi/2]$ using the formula $\varphi = N\pi + \varphi'$, where the integer $N = \mathrm{Int}[\varphi/\pi]$ is the number of half-turns within the layer thickness. In the case where $S_d > 1/2\pi$, all solutions for φ' fit into the domain $[-\pi/2, \pi/2]$. In the rest of the paper, we only use the variable φ', with the prime dropped for simplicity. The critical value of the director deviation angle φ_c corresponds to the configuration with N director half-turns in the layer when it is at an instability point.

The pitch in the layer just before the jump and the corresponding natural pitch are expressed via φ_c as

$$p_d(T_c) = \frac{2d}{N + \varphi_c/\pi}, \quad p(T_c) = \frac{2d}{N + \varphi_0(T_c)/\pi}, \qquad (4)$$

where T_c is the jump temperature. The angle $\varphi_0(T_c)$ (the natural one at the jump point temperature) is given by

$$\varphi_0(T_c) = \varphi_c + \frac{d}{K_{22}}\left[\frac{\partial W_s(\varphi)}{\partial \varphi}\right]_{\varphi = \varphi_c}. \qquad (5)$$

The value of φ after the jump, denoted by φ_j, which is basically connected to the pitch $p_{dj}(T_c)$ in the layer after the jump, is determined by the solution of the equation

$$\frac{\partial W_s(\varphi)}{\partial \varphi} + \frac{K_{22}}{d}[\varphi - \varphi_0(T_c) + \pi] = 0, \qquad (6)$$

where $\varphi_0(T_c)$ is determined by Eq. (5). As has been determined previously [1, 2], the variations of the pitch in the layer and, in particular, the hysteresis, are determined by the dimensionless parameter $S_d = K_{22}/Wd$, where W is the depth of the anchoring potential; these

are rather universal phenomena because they are not directly dependent on the sample thickness. This means that, for every specific form of the anchoring potential, expressions (3)–(6) can be transformed to forms that include the parameters of the problem, namely, d, K_{22}, and W, which only appear in combinations that reduce to the dimensionless parameter S_d.

3. MODEL ANCHORING POTENTIALS

To obtain some quantitative predictions, we must assume some specific form of the anchoring potential. We use the widely known Rapini–Papoular (RP) anchoring model potential [6, 14]

$$W_s(\varphi) = -\frac{W}{2}\cos^2\varphi. \quad (7)$$

For this potential, the critical angle at which the anchoring is identical at both surfaces of the layer is $\varphi_c = \pi/4$. The analysis in [1, 2] demonstrated that the essential features of the director temperature variations are directly dependent on the particular shape of the anchoring potential. It is therefore quite natural that we perform similar calculations for a potential that differs from the RP model anchoring potential. The second model potential to be investigated, which we call the *B*-potential, is given by the expression (see Fig. 2)

$$W_s(\varphi) = -W\left(\cos^2\frac{\varphi}{2} - \frac{1}{2}\right), \quad -\frac{\pi}{2} < \varphi < \frac{\pi}{2}, \quad (8)$$

which is continued periodically for $|\varphi| > \pi/2$ in accordance with the relation $W_s(\varphi) = W_s(\varphi - \pi)$. The behavior of potential (8) is similar to the case of the RP potential for small φ. However, it differs substantially from the RP model when φ is close to $\pi/2$. In particular, for identical anchoring at both surfaces, the critical angle φ_c for potential (8) is independent of the strength of the anchoring (via the parameter S_d) and is equal to $\pi/2$. Because there are too many parameters in the general case for different anchoring at the two surfaces, we consider a specific case in detail, namely, the case in which there is infinitely strong anchoring on one surface of a layer and finite anchoring on the other.

4. INFINITELY STRONG ANCHORING AT ONE SURFACE

We now apply the above general expressions to the specific case where infinitely strong anchoring is assumed at one of the layer surfaces and finite anchoring, described by the potential $W_s(\varphi)$, at the other,

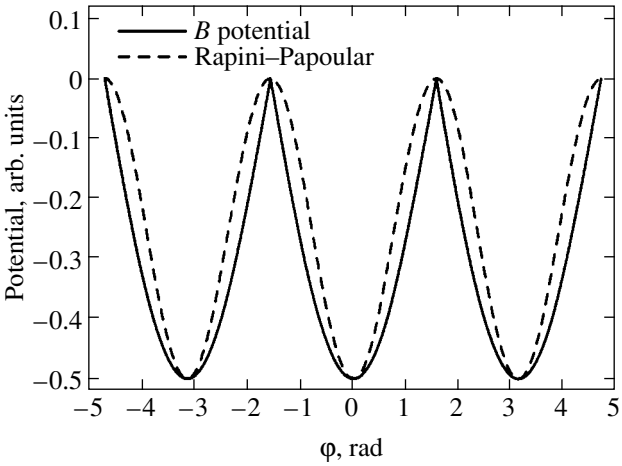


Fig. 2. Qualitative plots of RP potential (7) and *B* potential (8).

where $W_s(\varphi)$ is either of the two finite anchoring potentials mentioned above.

4.1. The Rapini–Papoular Potential

For the RP potential (7), the free energy given by Eq. (2) becomes

$$\frac{F(T)}{W} = \frac{1}{2}[-\cos^2\varphi + S_d(\varphi - \varphi_0(T))^2]. \quad (9)$$

The angle of the director deviation at the surface, φ, is determined by

$$\sin(2\varphi) + 2S_d(\varphi - \varphi_0(T)) = 0, \quad (10)$$

while the critical angle φ_c is determined by the relationship $\cos(2\varphi_c) + S_d = 0$, that is,

$$\varphi_c = \frac{1}{2}\arccos(-S_d). \quad (11)$$

Equation (11) shows that the critical angle φ_c depends on the parameter S_d, in contrast to the case where there the anchoring is identical at both surfaces [1, 2]. It changes from $\pi/4$ at $S_d = 0$ to $\pi/2$ at $S_d = 1$.

The value of the pitch for a bulk sample, or the alternative description involving the free rotation angle $\varphi_0(T_c)$ corresponding to the jump point, is determined from Eq. (10) as

$$\varphi_0(T_c) = \varphi_c + \frac{1}{2S_d}\sin(2\varphi_c). \quad (12)$$

The solution of Eq. (11) exists only for $0 < S_d < 1$. This means that, for weak anchoring (or thin layers), jumpwise changes of the director configuration in the layer may be absent. However, it should be mentioned that,

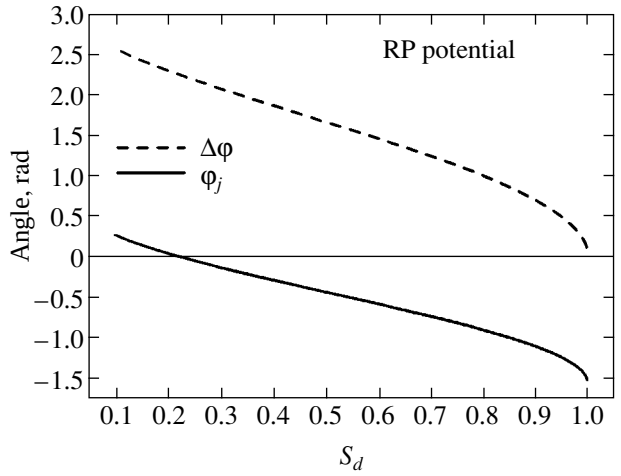


Fig. 3. The dependence of the postjump angle φ_j and the angular width $\Delta\varphi$ of the jump on the dimensionless parameter S_d for the RP potential.

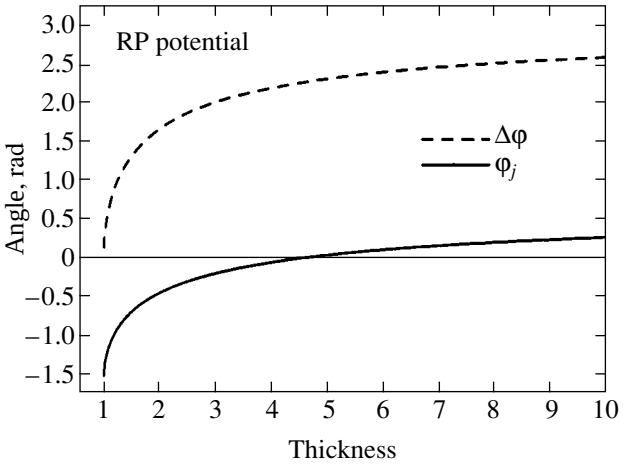


Fig. 4. The dependence of the postjump angle φ_j and the angular width $\Delta\varphi$ of the jump on the layer thickness (in units of the penetration length K_{22}/W for fixed values of K_{22} and W) calculated for the RP potential.

because Eq. (11) for the critical angle was obtained for the RP anchoring model potential, the previous statement is model-dependent. Therefore, experimental investigations of jumpwise changes of the director configuration in a layer may be used for determining the shape of the actual anchoring potential and its deviations from the RP anchoring model potential. The value of φ_j, or its equivalent in terms of the pitch $p_{dj}(T_c)$ in a layer after the jump, is determined by solving

$$\sin(2\varphi_j) + 2S_d\left\{\varphi_j - \frac{1}{2}[\arccos(-S_d)] - \frac{\sin(2\varphi_c)}{2S_d} + \pi\right\} = 0. \quad (13)$$

4.2. The B-potential

When the finite anchoring potential is taken as the *B*-potential in Eq. (8), the expression for the free energy given by Eq. (2) becomes

$$\frac{F(T)}{W} = \frac{1}{2}\left[-2\cos^2\frac{\varphi}{2} + 1 + S_d(\varphi - \varphi_0(T))^2\right]. \quad (14)$$

The angle φ of the director deviation is determined from the relation

$$\sin\varphi + 2S_d(\varphi - \varphi_0(T)) = 0. \quad (15)$$

The critical angle φ_c for the *B*-potential is given by $\pi/2$ for any value of S_d (see Fig. 2), as in the case for identical anchoring at both surfaces. The value of the pitch for a bulk sample, or its equivalent in terms of the free rotation angle $\varphi_0(T_c)$ corresponding to the jump point, is determined from Eq. (15) as

$$\varphi_0(T_c) = \frac{\pi}{2} + \frac{1}{2S_d}. \quad (16)$$

The value of φ_j, equivalent to the knowledge of the pitch in a layer after the jump at $p_{dj}(T_c)$, is determined by solving the equation

$$\sin\varphi_j + 2S_d\left(\varphi_j - \frac{1}{2S_d} + \frac{\pi}{2}\right) = 0. \quad (17)$$

The values given above for the angles of the director deviation just before and after the jump, equivalent to knowing the values of the pitches in the layer and the corresponding value of the pitch in a bulk sample, completely determine the initial and final states of the dynamical problem to be solved. As an example, φ_j and the angular width of the jump, i.e., $\Delta\varphi = \varphi_j - \varphi_c$, calculated versus S_d and the layer thickness d for the RP and *B* potentials are presented in Figs. 3 and 4 and Figs. 5 and 6, respectively.

5. JUMP DYNAMICS

The plane geometry of the problem under consideration, and symmetry arguments, allow us to suppose, in a first approximation to the problem, that the hydrodynamical flow during the pitch jump motion in a planar cholesteric layer is of minor significance. We therefore initiate a study of the above problem by neglecting hydrodynamical flow. Moreover, we also neglect the fluid and director inertial terms; this approach is usually adopted in the theory of liquid crystals.

In our approach, the director configuration is specified by one time-dependent variable $\varphi(t)$, i.e., by the director orientation at the surface, which, unlike in the

previous sections, is now time-dependent. The dynamics of φ is in general described by the Landau–Khalatnikov equation [15]

$$\gamma\frac{d\varphi}{dt} = -\frac{\partial F}{\partial\varphi}, \tag{18}$$

where $F = F(\varphi)$ is the total free energy and γ is the kinetic coefficient. This equation can also be derived using the general continuum theory of liquid crystals, which enables one to determine the parameter γ. In the absence of flow and general external forces, it is known from the general continuum theory of liquid crystals [16] that

$$\frac{d}{dt}F = -\int_{\Omega}\mathcal{D}\,d\Omega, \tag{19}$$

where $\mathcal{D}$ is the Rayleigh dissipation function [17] and Ω is the sample volume.

5.1. Simplified Dynamical Solutions

For the further simplification of the problem, we assume that the director distribution in the layer is quasistatic. This means that the director orientation angle $\phi(t, z)$ in the bulk of the sample can be easily related to the above time-dependent orientation angle $\phi(t)$ through the relation

$$\frac{\partial\phi}{\partial t} = \frac{z}{d}\frac{d\varphi}{dt}, \tag{20}$$

a form that is motivated by the well-known twist solution that occurs in the isothermal situation of cholesterics. This is a consequence of assuming that the director configuration inside the layer at any given time t is quasistatic; that is, the helical structure within the layer remains undistorted and corresponds to some value of the pitch, which changes with time; infinitely strong anchoring at one boundary then justifies the assumption made in Eq. (20). It follows from Eqs. (18) and (19) that, for a sample of unit dimensions in the x and y directions,

$$\frac{d}{dt}F(\varphi(t)) = -\int_{0}^{d}\mathcal{D}\,dz. \tag{21}$$

With the above assumptions on flow, the dissipation function in this case is simply (cf. [18, pp. 11–13; 19])

$$\mathcal{D} = \gamma_1\left(\frac{\partial\phi}{\partial t}\right)^2, \tag{22}$$

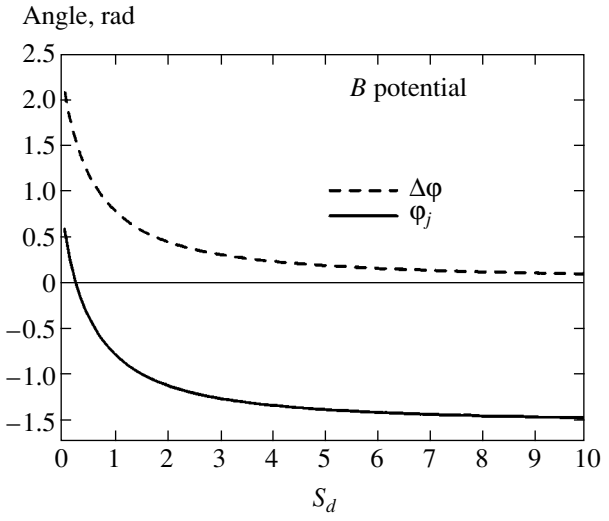


Fig. 5. The dependence of the postjump angle φ_j and the angular width Δφ of the jump on the dimensionless parameter S_d for the B potential.

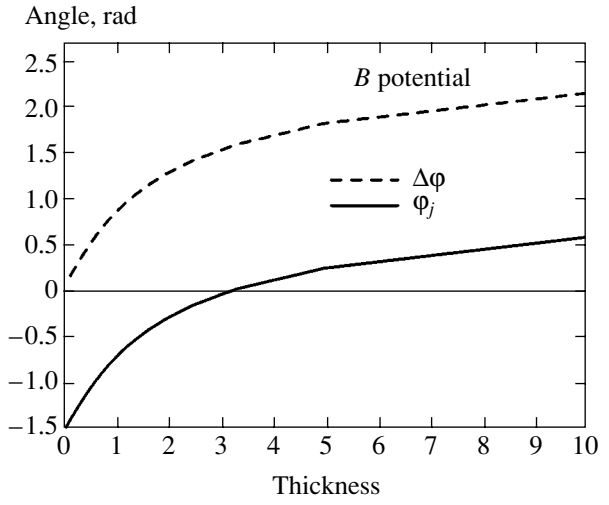


Fig. 6. The dependence of the postjump angle φ_j and the angular width Δφ of the jump on the layer thickness (in units of the penetration length K_{22}/W for fixed values of K_{22} and W) calculated for the B potential.

where $\gamma_1 > 0$ is the twist viscosity. Because

$$\frac{dF}{dt} = \frac{d\varphi}{dt}\frac{dF}{d\varphi}, \tag{23}$$

we can use relations (20) and (23) and insert $\mathcal{D}$ defined by Eq. (22) in Eq. (21) to find that

$$\frac{d\varphi}{dt} = -\frac{1}{\gamma_1}\frac{3}{d}\frac{dF}{d\varphi}. \tag{24}$$

This expression for the dynamics in the layer allows us to derive the solution for φ implicitly by integration of (24) from time $t = 0$, where the director deviation

angle at the surface is equal to the critical value, i.e., $\varphi(0) = \varphi_c$, up to time t, where $\varphi = \varphi(t)$. The resulting solution for φ is

$$t = -\gamma_1 \frac{d}{3} \int_{\varphi_c}^{\varphi} \left(\frac{dF}{d\hat{\varphi}} \right)^{-1} d\hat{\varphi}. \quad (25)$$

The solution φ must, of course, lie within the range $\varphi_c \le \varphi < \varphi_j$, that is, between the initial and final states identified above in Section 4. The values of φ_c and φ_j were discussed in Section 4 and are clearly dependent on the specific form of the anchoring potential and may be dependent on the value of the parameter S_d. The complete duration τ of the pitch jump is found by replacing the upper limit in the integral in Eq. (25) by φ_j.

The solution in Eq. (25) also allows us to define the relaxation time for the jump process. For example, the relaxation time t_r may be defined as the derivative $dt/d\varphi$ evaluated at the middle-during-the-jump-value of $\varphi = (\varphi_c + \varphi_j)/2$. Explicitly, we have from (25) that

$$t_r = -\gamma_1 \frac{d}{3} \left(\frac{dF}{d\varphi} \right)^{-1}_{\varphi = (\varphi_c + \varphi_j)/2}. \quad (26)$$

We now discuss the solutions for the RP and *B* potentials separately.

5.2. Rapini–Papoular Potential

In the case of the RP-potential, the pitch jump only occurs when $0 < S_d < 1$ (sufficiently strong anchoring). The explicit form for the integral appearing in solution (25) can be obtained via Eqs. (2) and (7) to find that the solution can be written as

$$t = -S_d \frac{2\gamma_1 d^2}{3K_{22}} \int_{\varphi_c}^{\varphi} \left[\sin(2\hat{\varphi}) + 2S_d \left(\hat{\varphi} + \pi - \varphi_c - \frac{\sin(2\varphi_c)}{2S_d} \right) \right]^{-1} d\hat{\varphi}, \quad (27)$$

where the critical angle φ_c is given by Eq. (11). The maximum possible upper limit of integration is φ_j, which is determined from Eq. (13) for a given value of S_d.

In the simple special case where $S_d = 1/2$, $\varphi_c = \pi/3$ and the upper limit φ_j is determined from the expression

$$\sin(2\varphi_j) + \varphi_j + \frac{\pi}{3} = 0, \quad (28)$$

obtained from Eq. (13).

5.3. The B-potential

For the *B*-potential, the pitch jump occurs for $0 < S_d < \infty$ (i.e., at any anchoring strength), and the explicit expression for the implicit solution is given via Eqs. (2), (8), and (25). The solution is given by

$$t = -S_d \frac{2\gamma_1 d^2}{3K_{22}} \times \int_{-\pi/2}^{\varphi} \left[\sin\hat{\varphi} - 1 + 2S_d \left(\hat{\varphi} + \frac{\pi}{2} \right) \right]^{-1} d\hat{\varphi}, \quad (29)$$

where we recall that φ_c is always $-\pi/2$ for the *B* potential. The upper limit of integration φ_j is determined from Eq. (17) for any given value of S_d. We note that, in this result, the form of the integrand and integration limits are given for the director configuration after the jump, that is, when N has changed by 1, and, consequently, $\varphi_0(T_c)$ is replaced by $\varphi_0(T_c) - \pi$.

As a simple example, we consider the case where $S_d = 1/\pi$. The implicit solution (29) for φ is then given by

$$t = -\frac{1}{\pi} \frac{2\gamma_1 d^2}{3K_{22}} \int_{-\pi/2}^{\varphi} \left[\sin\hat{\varphi} + 2\frac{\hat{\varphi}}{\pi} \right]^{-1} d\hat{\varphi}, \quad (30)$$

where the upper integration limit in the right-hand side is formally zero, because φ_j must be zero at $S_d = 1/\pi$ by relation (17). We note that the integral in (30) diverges logarithmically as φ approaches $\varphi_j = 0$, which formally indicates that the time taken to complete the jump is infinite. Nevertheless, there are physical mechanisms that ensure a cutoff to this limit such that this divergence does not occur in reality. For example, the cutoff may be due to thermal fluctuations within the cholesteric layer, and a nonzero upper limit may then be determined by the temperature.

6. GENERAL RESULTS

We now use solutions (27) and (29) for the RP and *B* potentials, respectively. To simplify the results and give qualitative plots for data, we introduce the timescale [6, p. 226]

$$\bar{t} = t \frac{\pi^2 K_{22}}{\gamma_1 d^2}, \quad (31)$$

which is a typical type of scaling that occurs in liquid crystal problems [16, 19].

6.1. Rapini–Papoular Potential

The solution (27) for the RP potential is given by

$$\bar{t} = -S_d\pi^2\frac{2}{3}\int_{\varphi_c}^{\varphi}\left[\sin(2\hat{\varphi}) + 2S_d\left(\hat{\varphi} + \pi - \varphi_c - \frac{\sin(2\varphi_c)}{2S_d}\right)\right]^{-1} d\hat{\varphi}, \quad (32)$$

where, by (11), $\varphi_c = \arccos(-S_d)/2$ and the relaxation time is given, via (26), by

$$\bar{t}_r = -S_d\pi^2\frac{2}{3}\left[\sin(2\varphi) + 2S_d\left(\varphi + \pi - \varphi_c - \frac{\sin(2\varphi_c)}{2S_d}\right)\right]^{-1}_{\varphi = \frac{\varphi_c - \pi + \varphi_j}{2}}, \quad (33)$$

with φ_j obtained from Eq. (13). Figure 3 shows the dependence of φ_j on S_d. It is also possible to define the switching time. We define the switching time $\bar{t}_s$ as the time taken for the orientation angle φ to change by a half of the jump angle width, i.e., to change from φ_c to $(\varphi_c - \pi + \varphi_j)/2$. It is given by

$$\bar{t}_s = -S_d\pi^2\frac{2}{3}\int_{\varphi_c}^{\frac{\varphi_c - \pi + \varphi_j}{2}}\left[\sin(2\hat{\varphi}) + 2S_d\left(\hat{\varphi} + \pi - \varphi_c - \frac{\sin(2\varphi_c)}{2S_d}\right)\right]^{-1} d\hat{\varphi}. \quad (34)$$

The solution φ, the relaxation time $\bar{t}_r$, and dependence of the switching time $\bar{t}_s$ on S_d for the RP potential are shown in Figs. 7–9.

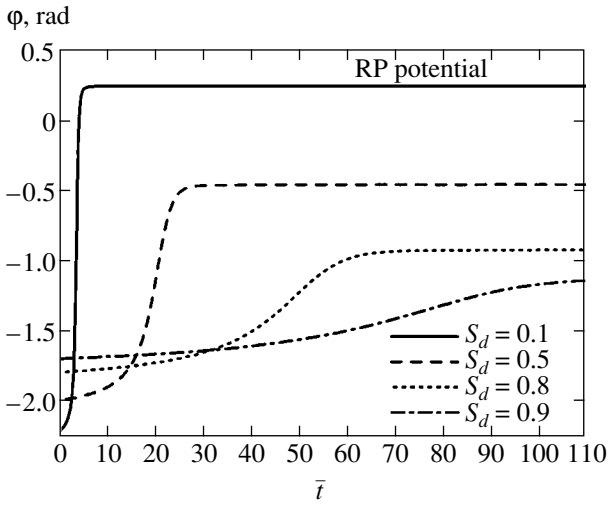


Fig. 7. Temporal behavior of the director orientation angle φ at the surface during a jump for the RP potential at the indicated values of S_d.

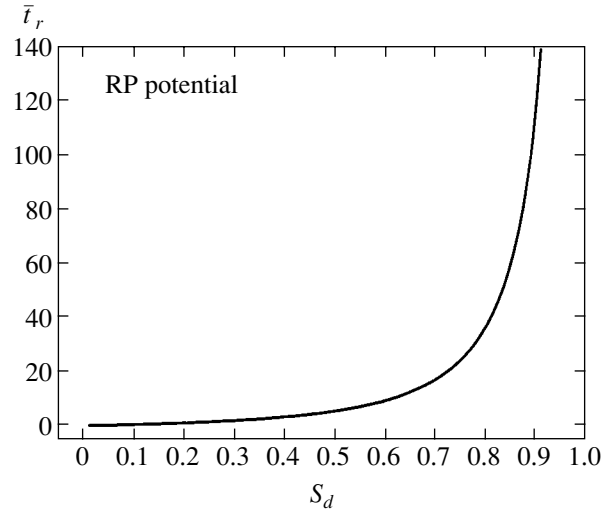


Fig. 8. The dependence of the relaxation time $\bar{t}_r$ defined by (33) on S_d for the RP potential.

6.2. The B potential

The solution (29) for the *B* potential is given by

$$\bar{t} = -S_d\pi^2\frac{2}{3}\int_{-\frac{\pi}{2}}^{\varphi}\left[\sin\hat{\varphi} - 1 + 2S_d\left(\hat{\varphi} + \frac{\pi}{2}\right)\right]^{-1} d\hat{\varphi}, \quad (35)$$

and the relaxation time is

$$\bar{t}_r = -S_d\pi^2\frac{2}{3}\left[\sin\varphi - 1 + 2S_d\left(\varphi + \frac{\pi}{2}\right)\right]^{-1}_{\varphi = \frac{\varphi_j - \pi/2}{2}}, \quad (36)$$

where φ_j is calculated from Eq. (17). It is also possible to define a switching time analogous to that for the RP potential by the relation

$$\bar{t}_s = -S_d\pi^2 \times \frac{2}{3}\int_{-\frac{\pi}{2}}^{\frac{\varphi_j - \pi/2}{2}}\left[\sin\hat{\varphi} - 1 + 2S_d\left(\hat{\varphi} + \frac{\pi}{2}\right)\right]^{-1} d\hat{\varphi}. \quad (37)$$

The solution φ and the relaxation time $\bar{t}_r$ for the *B*-potential are shown in Figs. 10 and 11, while the dependence of the switching time $\bar{t}_s$ on S_d is shown in Fig. 12.

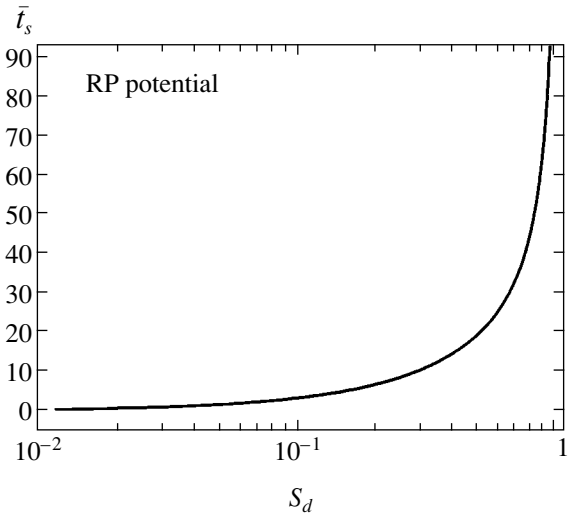


Fig. 9. The dependence of the switching time $\bar{t}_s$ defined by (34) on S_d calculated for the RP potential.

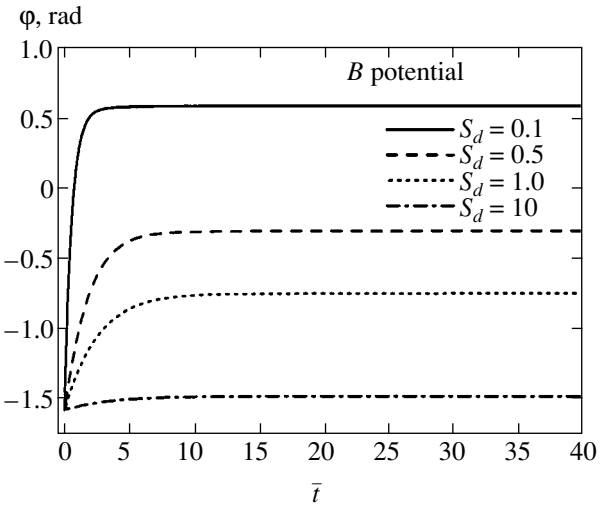


Fig. 10. Temporal behavior of the director orientation angle φ at the surface during a jump for the *B* potential at the indicated values of S_d.

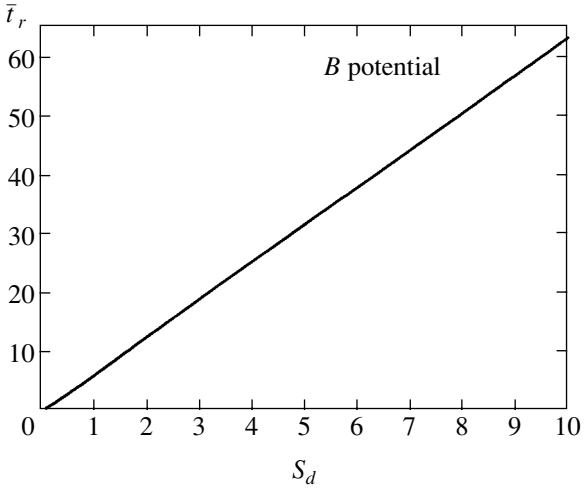


Fig. 11. The dependence of the relaxation time $\bar{t}_r$ defined by (33) on S_d for the *B* potential.

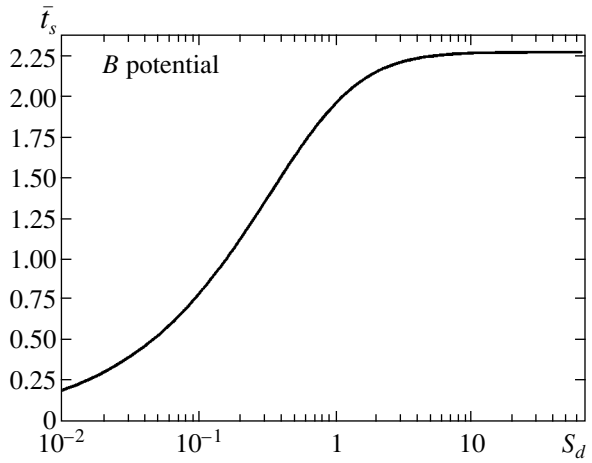


Fig. 12. The dependence of the switching time $\bar{t}_s$ defined by (37) on S_d calculated for the *B* potential.

We note that, at the initial stage of the jump, the surface viscosity (see, e.g., the introduction of the surface viscosity discussed in [13]) may restrict the velocity of director rotation at the layer surface. It may be taken into account by adding the surface viscosity term $\gamma_s(d\varphi/dt)^2$ to Eq. (21). The corresponding additional term may play a role in very thin layers and may in principle be detected by experiment.

7. CONCLUSIONS

The results obtained above reveal the qualitatively important physical properties of jump dynamics. For instance, such dynamics have a direct dependence on the strength and shape of the anchoring potential. Although the specific calculations of the jump dynamics were performed under the simplifying assumptions mentioned above, there is no doubt that the qualitative features of the jump dynamics remain valid in general for the phenomenon as a whole. It is therefore interesting to discuss under which circumstances the solutions described above are quantitatively valid, and what modifications to these solutions would be required under other conditions in order to obtain a further quantitative description of pitch jump dynamics.

We consider the quasistatic approximation to the dynamics of the pitch jump. It may work quantitatively if the time of propagation of the disturbance between the surfaces of the layer is smaller than the characteristic time of the jump. Estimating the velocity of perturbation propagation in the cholesteric as $v_p = K_{22}/\gamma_1 p$, where p is the cholesteric pitch, we find that the pertur-

bation propagation time is p/d, in the de Gennes units of time used in the calculations (cf. Eq. (31)), and is therefore small if p/d is small. Examination of Figs. 7–12 allows us to determine in which range of the parameter S_d and for which potential the accepted approximations are valid. In any case, it is clear that, in general, the jump time (or the corresponding relaxation time) is shorter for the B potential compared to the RP potential, and therefore, as the values of S_d decrease, the approximation becomes invalid for the B potential at larger values of S_d than for the RP potential. If we assume that $p/d < 0.1$, then the calculations presented above show that the obtained results are of quantitative meaning for $S_d > 0.1$.

As a first step in overcoming the quasistatic approximation made here, we may regard solutions of the equations in the previous section with a time-dependent space scale and limit the integration in Eq. (19) to the time t by the distance relation $d(t) = v_p t$ if $d(t) < d$. Within this approach, a quantitative description of the pitch jump dynamics may be obtained for values of the parameter S_d not limited by the above condition $S_d > 0.1$. However, one must bear in mind that, for smaller values of S_d, other pitch-jump mechanisms of may be at work. This is connected with the fact that, for values of S_d smaller than its critical value $1/2\pi$ [1], along with the jumps corresponding to a change in the number N of half-turns at the layer thickness by one (which corresponds to the transitions of the director configuration not currently in the ground state), jumps greater than an increment of one to the value of N are possible. The calculated results for $S_d < 1/2\pi$, therefore, require a special discussion to determine the range of applicability to the pitch jump dynamics in the framework of the mechanism accepted here.

Comparison of Figs. 7–12 (see also Figs. 3 and 5) shows that experimental measurements make it possible to obtain a qualitative conclusion about the applicability of the RP potential to describing jumps (at $S_d > 1$, the jumps are absent at all for RP potential). It is commonly accepted that, for small angular deviations of the director from the alignment direction, the surface anchoring potential is quadratic in the deviation angle, as is the case with the RP potential. However, for large deviation angles (which are essential for the occurrence of pitch jumps), the question about the shape of the associated anchoring potential remains open. Therefore, the results presented above show that experimental investigations of the pitch jump dynamics provide a unique opportunity to study the actual shape of the anchoring potential at large angular deviations of the director from the alignment direction. Other approaches employed so far in this area (see, e.g., [20, 21]) mainly enable one to determine the anchoring strength that characterizes the anchoring energy at small deviations from the alignment direction.

Concerning the experimental observation of the dynamics of pitch jumps, it should be kept in mind that, without special precautions, it is quite improbable that the pitch jump in the cell occurs in the whole cell simultaneously and that it most probably occurs in limited areas of its surface, whereas the formulas presented above assume that the jump process occurs throughout the whole cell simultaneously. To ensure that the process occurs in the whole cell, it is possible to perform an experiment stabilizing the director configuration in the layer by applying a rather weak external field and turning the field off just as the director at the surface achieves the critical angle. This approach seems to be very similar to the one applied for studying the dynamics of the Frederiks transition [6, 22]. The same effect of a jump in a whole cell may be achieved by mechanical rotation of the layer surface by a small angle in the director configuration state of the layer close to the jump.

When the homogeneity of the jump over the surface of the cell is ensured, the dynamics of the jump may be followed by the conventional approach of measuring the time-dependent transmission [8], reflection spectra [23], or rotation of the plane of light polarization [10].

ACKNOWLEDGMENTS

The authors are grateful for the UK EPSRC (grant no. GR/S34311/01) that enabled this work to be undertaken. The partial funding of this work by the Russian Foundation for Basic Research (project no. 03-02-16173) is greatly appreciated by V.A.B.

REFERENCES

1. V. A. Belyakov and E. I. Kats, Zh. Éksp. Teor. Fiz. **118**, 560 (2000) [JETP **91**, 488 (2000)].
2. V. A. Belyakov, P. Oswald, and E. I. Kats, Zh. Éksp. Teor. Fiz. **123**, 1040 (2003) [JETP **96**, 915 (2003)].
3. V. A. Belyakov and E. I. Kats, Zh. Éksp. Teor. Fiz. **120**, 430 (2001) [JETP **93**, 380 (2001)].
4. V. A. Belyakov, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 99 (2002) [JETP Lett. **76**, 88 (2002)].
5. G. Chilaya, G. Hauck, H. D. Koswig, *et al.*, Cryst. Res. Technol. **32**, 401 (1997).
6. P. G. de Gennes and J. Prost, *The Physics of Liquid Crystals*, 2nd. ed. (Mir, Moscow, 1982; Clarendon, Oxford, 1993).
7. P. Oswald and P. Pieranski, *Les Cristaux Liquides: Concepts et Propriétés Physiques Illustrées par des Expériences* (Gordon and Breach, Paris, 2000).
8. H. Zink and V. A. Belyakov, Mol. Cryst. Liq. Cryst. **265**, 445 (1995); Pis'ma Zh. Éksp. Teor. Fiz. **63**, 37 (1996) [JETP Lett. **63**, 43 (1996)].
9. S. P. Palto, Zh. Éksp. Teor. Fiz. **121**, 308 (2002) [JETP **94**, 260 (2002)].

10. W. Kuczynski, private communication.
11. R. Barberi and G. Durand, Appl. Phys. Lett. **58**, 2907 (1991).
12. R. Barberi, M. Giocondo, and G. Durand, Appl. Phys. Lett. **60**, 1085 (1992).
13. P. J. Kedney and F. M. Leslie, Liq. Cryst. **24**, 613 (1998).
14. L. M. Blinov, E. I. Kats, and A. A. Sonin, Usp. Fiz. Nauk **152**, 449 (1987) [Sov. Phys. Usp. **30**, 604 (1987)].
15. E. M. Lifshitz and L. P. Pitaevskiĭ, *Course of Theoretical Physics*, Vol. 10: *Physical Kinetics* (Nauka, Moscow, 1979; Pergamon, Oxford, 1981).
16. G. Vertogen and W. H. de Jeu, *Thermotropic Liquid Crystals, Fundamentals* (Springer, Berlin, 1988).
17. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics*, 3rd ed. (Pergamon, London, 1959; Nauka, Moscow, 1986).
18. F. M. Leslie, Adv. Liq. Cryst. **4**, 1 (1979).
19. I. W. Stewart, *The Static and Dynamic Continuum Theory of Liquid Crystals* (Taylor and Francis, London) (in press).
20. B. Jerome, Rep. Prog. Phys. **54**, 391 (1991).
21. H. Yokoyama, Mol. Cryst. Liq. Cryst. **165**, 265 (1988).
22. F. Brochard, P. Pieranski, and E. Guyon, Phys. Rev. Lett. **28**, 1681 (1972).
23. H. F. Gleeson, private communication.

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 83–94.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 99–112.

FLUIDS

Switching Time Dispersion and Retention of Bistable States in Langmuir–Blodgett Ferroelectric Films

A. R. Geivandov[a], S. P. Palto[a], S. G. Yudin[a], and L. M. Blinov[a,b]

[a]*Institute of Crystallography, Russian Academy of Sciences, Moscow, 119333 Russia*
[b]*INFM, Physical Department, University of Calabria, I-87046 Italy*
e-mail: ageivandov@yandex.ru

Received December 1, 2003

Abstract—Polarization switching and retention of each of the two polarized states in Langmuir–Blodgett (LB) ferroelectric films are studied using nonlinear dielectric spectroscopy. It is found that polarized states can be preserved for a long time, but the polarization dynamics in 10–40-nm-thick LB films is characterized by a considerable switching-time dispersion. In addition, ferroelectric LB films exhibit clearly manifested asymmetry of switching to states with opposite directions of polarization. To explain the experimental results, a polarization-switching model is proposed that takes into account the energy of interaction of a ferroelectric polymer with boundary surfaces. The effect of inhomogeneity of the LB film structure on the ferroelectric switching dynamics is also discussed. © *2004 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

Langmuir–Blodgett (LB) films are prepared by successive transfer of organic monomolecular layers from the interface between two subphases (normally, water–air) to a solid substrate. Thus, LB films are stacks of monolayers with controllable chemical composition and molecular orientation [1, 2]. This makes it possible to produce model samples for basic research.

The discovery of ferroelectric LB films [3] was a considerable achievement. The first ferroelectric LB films were prepared from the copolymer vinylidene fluoride and trifluoroethylene (PVDF/TrFE). Copolymers PVDF/TrFE have been studied in detail owing to their ferroelectric properties [4] and applications as piezoelectric transducers [5]. Earlier, PVDF/TrFE films were mainly prepared by spin-coating technique or by deposition from a solution. These methods do not provide continuous films of a thickness exceeding 50 nm. The application of the LB technique opened the way for studying ferroelectricity in ultrathin films [3] on the order of 1 nm in thickness, which in fact led to the discovery of 2D ferroelectricity. Among other things, it was proved that the LB films prepared from PVDF/TrFE are highly crystalline and clearly manifest ferroelectric properties such as the first-order phase transition as well as pyro- and piezoelectric effects [6, 7]. It is the absence of a critical thickness below which ferroelectricity vanishes that served as a proof of the 2D nature of ferroelectricity in LB films made of PVDF/TrFE [3]. The analysis of these films has not covered the details of polarization switching and the longevity of polarized states so far. Here, we report on the first results of such studies obtained with the help of nonlinear dielectric spectroscopy (NDS). The NDS method was used by us earlier to measure the Landau–Ginzburg coefficients in the model of the ferroelectric phase transition in LB films [7].

The article consists of two main parts. In Section 2, the experimental results obtained with the help of the NDS method are described. The polarization switching models and the results of numerical simulation of switching dynamics are considered in Section 3. The observed specific features of switching require the inclusion of the surface interaction and structural inhomogeneity of the films. Allowance for spatial inhomogeneity of the films also makes it possible to answer some questions arising when the homogeneous Landau–Khalatnikov model [8] is used for describing the switching dynamics in LB films.

2. EXPERIMENT

2.1. Sample Preparation

We studied LB films of the PVDF/TrFE copolymer of composition 70/30, consisting of 20 monolayers.[1] Monomolecular layers were successively deposited from the water surface onto the surface of a glass substrate with a sputtered aluminum electrode at room temperature (20–22°C) under a surface pressure of 3 mN/m.

[1] We assume that only one monolayer is formed on the surface of water although local formations of several monolayers cannot be ruled out in view of weak amphiphility.

1063-7761/04/9901-0083$26.00

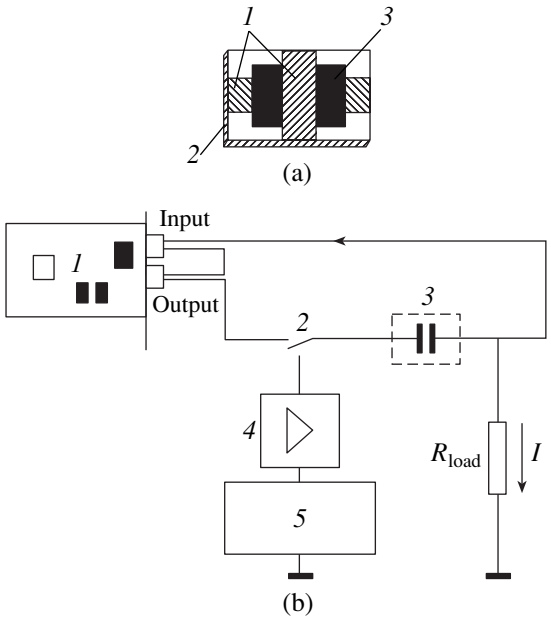


Fig. 1. (a) Sample structure: *1*—aluminum electrodes, *2*—glass substrate, *3*—LB film. (b) Circuit diagram for measurements by the NDS method: *1*—sound card of the computer and PhysLab program as lock-in amplifier and generator; *2*—key for switching between the polarization and measurement circuits; *3*—sample; *4*—amplifier; *5*—pulse voltage generator; R_{load}—load resistor.

In this study, instead of the horizontal deposition method, in which the substrate plane is parallel to the water surface during the transfer of a monolayer, we used the classical Langmuir–Blodgett method and the setup described in [9]. The transfer of monolayers was carried out by pulling the substrate from water (*Z*-type transfer) so that the normal to the substrate surface was oriented at approximately 45° to the water surface. This improved the homogeneity of the films. A second aluminum electrode was deposited onto the LB polymer film. The total area of electrode overlapping was 0.01 ± 0.0005 cm^2. The permittivity was set at 10 (this value was found for bulk samples); the film thickness measured by a capacitive method was 30 nm. The structure of the sample is shown in Fig. 1a.

2.2. Experimental Setup

The electric measuring circuit is shown in Fig. 1b. The total current in the circuit is defined by the relation

$$I = I_R + I_C = \frac{U}{R} + \frac{d(CU)}{dt}, \tag{1}$$

where I_R and I_C are the active and reactive current components, respectively; R is the film resistance; and U is the voltage applied across the film. In expression (1), capacitance C is a function of the applied voltage and the reactive current component can be represented by the formula

$$\begin{aligned} I_C &= C(U)\frac{dU}{dt} + U(t)\frac{dC(U)}{dt} \\ &\approx \left(C_0 + U(t)\frac{dC}{dU}\right)\frac{dU}{dt} \\ &+ U(t)\frac{dC}{dU}\frac{dU}{dt} = C_0\frac{dU}{dt} + 2U(t)\frac{dC}{dU}\frac{dU}{dt}, \end{aligned} \tag{2}$$

where $C_0 = C(U = 0)$. In formula (2), we confine ourselves only to the first-order terms in the expansion of the field dependence of capacitance.

When a sinusoidal voltage $U(t) = U_0\sin(\omega t)$ is applied to the sample (U_0 is the amplitude and ω is the circular frequency), relation (1) combined with (2) assumes the form

$$\begin{aligned} I &= \frac{U_0}{R}\sin(\omega t) + \omega U_0 C_0\cos(\omega t) \\ &+ \omega U_0^2\sin(2\omega t)\frac{dC}{dU}. \end{aligned} \tag{3}$$

It can easily be seen that, using phase-sensitive current detection at fundamental frequency ω, we can independently measure the first two terms in relation (3), which are proportional to the ohmic losses and the capacitance of the sample. These components have a mutual phase shift of 90° and can easily be separated using phase-sensitive registration. The third term in relation (3) reflects the nonlinear contribution and is observed at higher harmonics. It is well known that, in the general case, the components of the electric displacement vector in a nonlinear dielectric can be written in the form

$$D_i = P_{s,i} + \varepsilon_{ij}E_j + \chi_{ijk}E_jE_k + \ldots, \tag{4}$$

where ε_{ij} and χ_{ijk} are the tensors of linear permittivity and second-order susceptibility, respectively. Here and below, we assume summation over repeating indices. To allow for the nonlinearity of the capacitance, it is expedient to introduce the differential permittivity tensor

$$\varepsilon_{ij}^* = \frac{\partial D_i}{\partial E_j} \approx \varepsilon_{ij} + \chi_{ijk}E_k. \tag{5}$$

In our geometry, where the z axis is normal to the substrate and corresponds to values of indices equal to 3, the derivative of the capacitance in relation (3) can be written in the form

$$\frac{dC}{dU} \propto \frac{\partial\varepsilon_{33}^*}{\partial E_3} \approx \chi_{333}. \tag{6}$$

Thus, taking into account this relation, we can write the last nonlinear term in relation (3) in the form

$$I_{2\omega} \approx \frac{\omega S U_0^2}{d^2} \varepsilon_0 \chi_{333} \sin(2\omega t), \tag{7}$$

where S is the area of electrode overlapping, d is the film thickness, and $\varepsilon_0 \approx 8.85 \times 10^{-12}$ F/m.

Expression (7) reflects the fundamental property of the NDS method. Like the second optical harmonic generation, the NDS makes it possible to measure the second-order permittivity typical of noncentrosymmetric media. It should be emphasized that, in the case of inhomogeneous films with a domain structure, the symmetry of the entire system after averaging can be higher than that of individual domains. For this reason, the observation at the second harmonics requires ordered or polarized samples.

The NDS method is based on phase-sensitive measurements of the first and higher harmonics. In analyzing ferroelectricity, it is convenient to introduce the quantities defined in terms of the ratios of the Fourier components measured with a lock-in amplifier [7]:

$$A_2 = \frac{-\sqrt{2}\Phi_{2x}}{2\left(\sqrt{2}\Phi_{1y} - \omega\varepsilon_0 \frac{U_0 S}{d}\right)^3} \frac{U_0}{d} (\omega S)^2, \tag{8}$$

$$A_3 = \frac{\sqrt{2}\Phi_{3y}}{3\left(\sqrt{2}\Phi_{1y} - \omega\varepsilon_0 \frac{U_0 S}{d}\right)^4} \frac{U_0}{d} (\omega S)^3, \tag{9}$$

$$A_4 = \frac{\sqrt{2}\Phi_{4x}}{15\left(\sqrt{2}\Phi_{1y} - \omega\varepsilon_0 \frac{U_0 S}{d}\right)^5} \frac{U_0}{d} (\omega S)^4, \tag{10}$$

$$A_5 = \frac{-2\sqrt{2}\Phi_{5y}}{5\left(\sqrt{2}\Phi_{1y} - \omega\varepsilon_0 \frac{U_0 S}{d}\right)^6} \frac{U_0}{d} (\omega S)^5. \tag{11}$$

Here, S is the area of electrode overlapping; d is the film thickness; $\varepsilon_0 \approx 8.85 \times 10^{-12}$ F/m; and Φ_{1y}, Φ_{2x}, Φ_{3y}, Φ_{4x}, and Φ_{5y} are the effective values of the x and y Fourier components detected by a lock-in amplifier from the first to the fifth harmonic, respectively. It should be recalled that the x component corresponds to a inphase reference signal (proportional to $\sin(k\omega t)$, where k is the number of a harmonic), while the y component is proportional to $\cos(k\omega t)$, i.e., having a phase shift of 90° relative to the reference signal.

The convenience of using the quantities defined by relations (8)–(11) becomes obvious if we consider ferroelectricity in the framework of the Landau–Ginzburg model. In accordance with this model, the contribution to the free energy density of a ferroelectric, which is associated with polarization P, can be written in the form

$$F = \frac{1}{2}\alpha P^2 + \frac{1}{4}\beta P^4 + \frac{1}{6}\gamma P^6 - EP, \qquad \alpha = \alpha_0(T - T_0), \tag{12}$$

where α_0, β, and γ are the temperature-independent Landau–Ginzburg coefficients; T_0 is the Curie temperature; and E is the electric field magnitude. Thus, as was demonstrated in [7], the quantities defined by relations (8)–(11) for a uniformly polarized ferroelectric can be written in the form

$$A_2 = P_s(3\beta + 10\gamma P_s^2), \tag{13}$$

$$A_3 = \beta + 10\gamma P_s^2, \tag{14}$$

$$A_4 = \gamma P_s, \tag{15}$$

$$A_5 = \gamma, \tag{16}$$

where P_s is the spontaneous polarization.

Here, we will be interested primarily in the value of quantity A_2, which reflects the sign and magnitude of the spontaneous polarization component along the normal to the film. Indeed, since the value of P_s^2 is independent of the orientation of the polar axis, factor $3\beta + 10\gamma P_s^2$ in relation (13) at a fixed temperature can be treated as a constant. The normal component of P_s for a homogeneous film depends on the orientation of the polar axis relative to the film normal. Real films are inhomogeneous since they consist of a large number of ferroelectric domains with polar axes randomly distributed over the film plane. The transverse component of the total polarization of the film vanishes due to axial symmetry and the resultant (remnant) polarization vector can be directed only along the normal to the film. Thus, the value of A_2 being measured is proportional to the value of remnant polarization P_r, which is the result of macroscopic averaging of the polarizations of all ferroelectric domains in the sample.

In addition to the NDS method, we also used the well-known Merz method [10] for measuring the polarization current to directly obtain the total switched polarization. This enabled us to carry out approximate calibration of the NDS method, i.e., to establish a relation between the effective values of the response being measured at the second harmonic and the absolute value of the remnant polarization. In the Merz method, we used a triangular waveform of frequency 80 Hz. The current response was measured as the voltage across a 1.2-kΩ load resistor connected in series with the sample. The polarization hysteresis loops were obtained

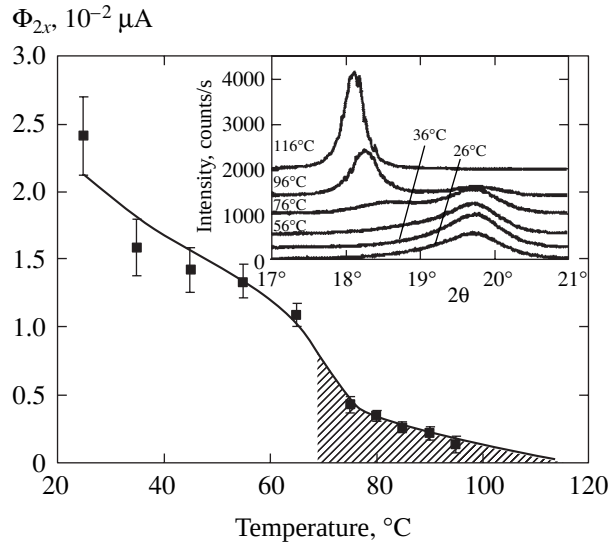


Fig. 2. Temperature dependence of polarization being switched during heating of LB films prepared from the VDF/TrFE copolymer with a composition of 70/30, measured by the NDS method. The polarizing pulse duration is 50 ms and amplitude is 20 V. The inset shows X-ray diffraction at different temperatures for a sample consisting of 150 monolayers [12], demonstrating the region of coexistence of the ferroelectric and paraelectric phases.

after integrating the current response and subtracting the linear contributions from the capacitance and resistance of the sample.

After determining the breakdown voltage of the film (for our samples containing 20 monolayers, it was approximately equal to 25 V), we chose several regimes for a detailed analysis of polarization switching and relaxation. The samples were polarized by a single dc voltage pulse (see Fig. 1b). The pulse was formed by generator *5* (Tektronix AFG-320) and was amplified by amplifier *4* to the required amplitude. Key *2* made it possible to switch circuits for polarization and measurement of sample *3*. The NDS measurements were carried out using the PhysLab program operating in the Windows 98 operating system. This software makes it possible to generate a sinusoidal voltage in a frequency range of 10 Hz–20 kHz with an amplitude from 1 to 2500 mV and with a very low coefficient of harmonics. The PhysLab lock-in amplifier makes it possible not only to simultaneously detect the real and imaginary current components at the fundamental frequency, but also to measure the corresponding components at other harmonics. The total coefficient of intrinsic nonlinear distortions of the generator and the detector system of the PhysLab using the Creative Technologies sound card SB PCI 128 amounts approximately to 0.005%, which makes it possible to study the nonlinear properties of our samples by the NDS method with a considerable margin. We used a sinusoidal voltage at a frequency of 1 kHz and an amplitude from 0.5 to 1.5 V, which is much lower that the value of the coercive field corresponding to an applied voltage from 8 to 10 V for samples with a thickness of 20 monolayers. Switching between different polarization states was carried out by using a single rectangular pulse of the corresponding sign with one of the amplitude values 10, 15, or 20 V. We will denote the polarization states after switching by a positive or negative pulse as $P^{(+)}$ and $P^{(-)}$, respectively (the pulse potential was measured relative to the electrode on the substrate). It is important to emphasize that the sign of quantities A_2 and Φ_2 should not necessarily coincide with the sign of polarization. It will be demonstrated in the experimental part of this paper that the negative sign of signal Φ_2 corresponds to the positive sign of polarization (state $P^{(+)}$). From the standpoint of the Landau–Ginzburg theory, this is possible if factor $3\beta + 10\gamma P_s^2$ in relation (13) is negative. In our case, PVDF/TrFE is characterized by the first-order ferroelectric phase transitions and $\beta < 0$; consequently, the sign of the factor can indeed be negative.

2.3. Experimental Results

Figure 2 shows the temperature dependence of the current response at the second harmonic after the sample polarization by a single rectangular pulse. This dependence illustrates the ferroelectric phase transition in PVDF/TrFE (70/30). The response of the second harmonic decreases at 70°C, but remains finite up to 110°C. This demonstrates considerable blurring of the phase transition and confirms the well-known phase coexistence in a temperature range of 70–110°C, where both the ferroelectric and the paraelectric phase exist in accordance with X-ray diffraction data (see the inset to Fig. 2). Thus, we can switch polarization even in the region of phase coexistence and the NDS method makes it possible to observe this switching.

Figure 3 shows the polarization hysteresis loops obtained with the help of the Merz method at room temperature. The value of remnant polarization P_r (at $U = 0$) increases with the amplitude of the applied triangular voltage and attains values on the order of 20 mC/m^2, which is approximately one-fourth of the maximal value approximately equal to 80 mC/m^2 [4]. The lower value of P_r can be explained by partial switching of ferroelectric domains at a frequency of 80 Hz, at which the loops were measured. This hypothesis is also confirmed by the NDS data shown in Figs. 4–6, which were recorded for various values of the amplitude and duration of the polarizing pulse.

In spite of the fact that the hysteresis loops recorded at 80 Hz are highly symmetric (see Fig. 3), the results of experiments with pulsed switching presented in Fig. 4 demonstrate a considerable asymmetry of the opposite polarization states. For example, state $P^{(-)}$ attained after the application of a negative pulse is less advantageous as compared to the opposite state $P^{(+)}$ (positive signal Φ_2 corresponds to the $P^{(-)}$ state). If the

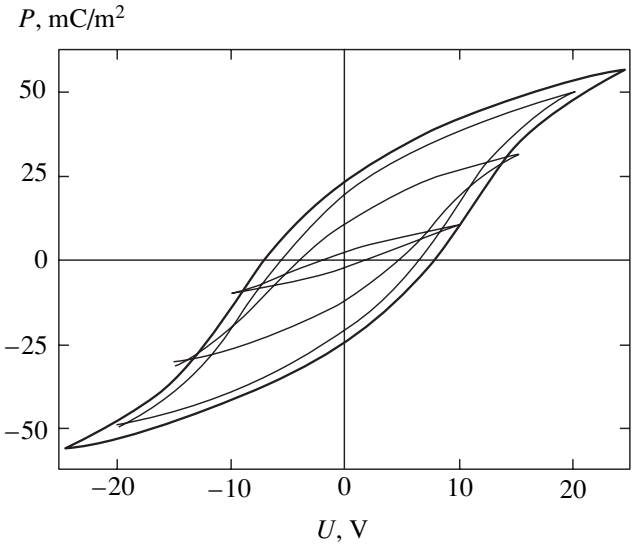


Fig. 3. Polarization hysteresis loops in LB films consisting of 20 monolayers of copolymer PVDF/TrFE, measured at a frequency of 80 Hz.

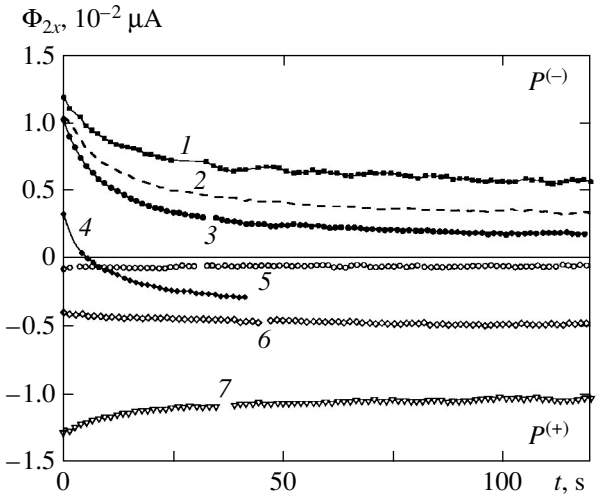


Fig. 4. Relaxation of the second harmonic of the current response after polarization of the LB films by a single pulse of various duration. The pulse amplitude $U = 10$ V for polarization in the $P^{(+)}$ state (negative values of Φ_2), $U = -10$ V for polarizing the $P^{(-)}$ state (positive value of Φ_2). The curves correspond to a pulse duration of 50 s (*1*), 25 s (*2*), 10 s (*3*), 5 s (*4*), 5 ms (*5*), 10 ms (*6*), and 500 ms (*7*).

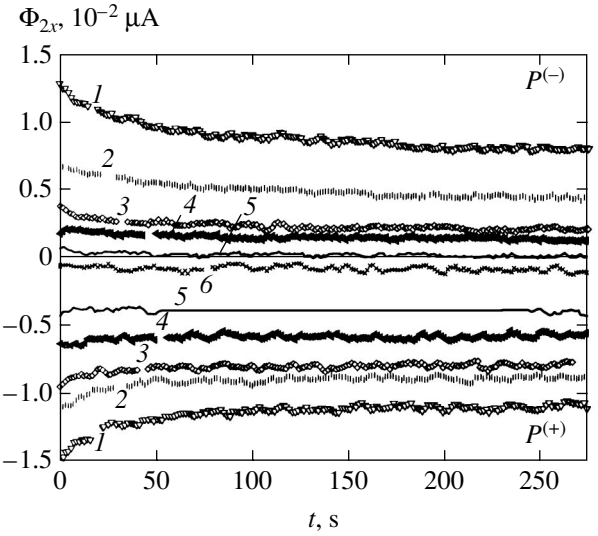


Fig. 5. Relaxation of the second harmonic of the current response after the polarization of the LB films by a single pulse of various duration. The pulse amplitude $U = 15$ V for polarization in the $P^{(+)}$ state (negative values of Φ_2), $U = -15$ V for polarizing the $P^{(-)}$ state (positive value of Φ_2). The curves correspond to a pulse duration of 10 s (*1*), 500 ms (*2*), 50 ms (*3*), 5 ms (*4*), 1 ms (*5*), and 0.2 ms (6).

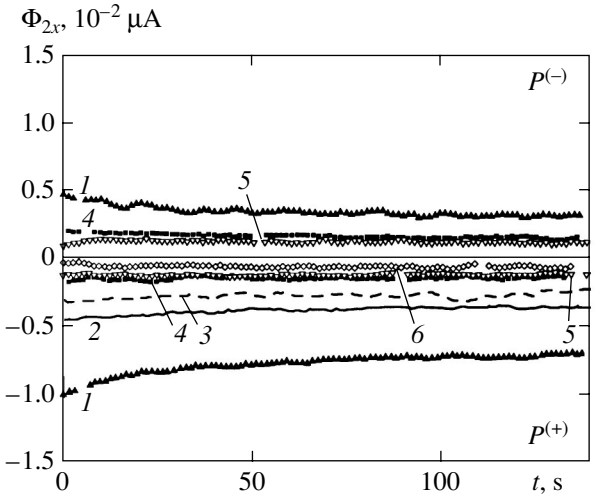


Fig. 6. Relaxation of the second harmonic of the current response after the polarization of the LB films by a single pulse of various duration. The pulse amplitude $U = 20$ V for polarization in the $P^{(+)}$ state (negative values of Φ_2), $U = -20$ V for polarizing the $P^{(-)}$ state (positive value of Φ_2). The curves correspond to a pulse duration of 100 ms (*1*), 10 ms (*2*), 5 ms (*3*), 1 ms (*4*), 0.5 ms (*5*), and 0.2 ms (6).

duration of the applied pulse with an amplitude of 10 V is smaller than 5 s, the system cannot be switched from state $P^{(+)}$ to state $P^{(-)}$ for a long time (the lifetime of the $P^{(-)}$ state amounts to about 10 s after the polarization; see Fig. 4). On the other hand, the switching by a positive pulse from state $P^{(-)}$ to state $P^{(+)}$ is possible even for a pulse duration of 5 ms, although the remnant polarization in this case is much smaller than the maximal value attained after polarization by long pulses (with a duration of at least 500 ms; see the negative Φ_2 curves in Fig. 4). The duration of the switching (which amounts to 50 s for the $P^{(-)}$ state) is explained by the fact that the amplitude of the applied voltage pulse (equal to –10 V in the present case) is quite close to the coercive voltage

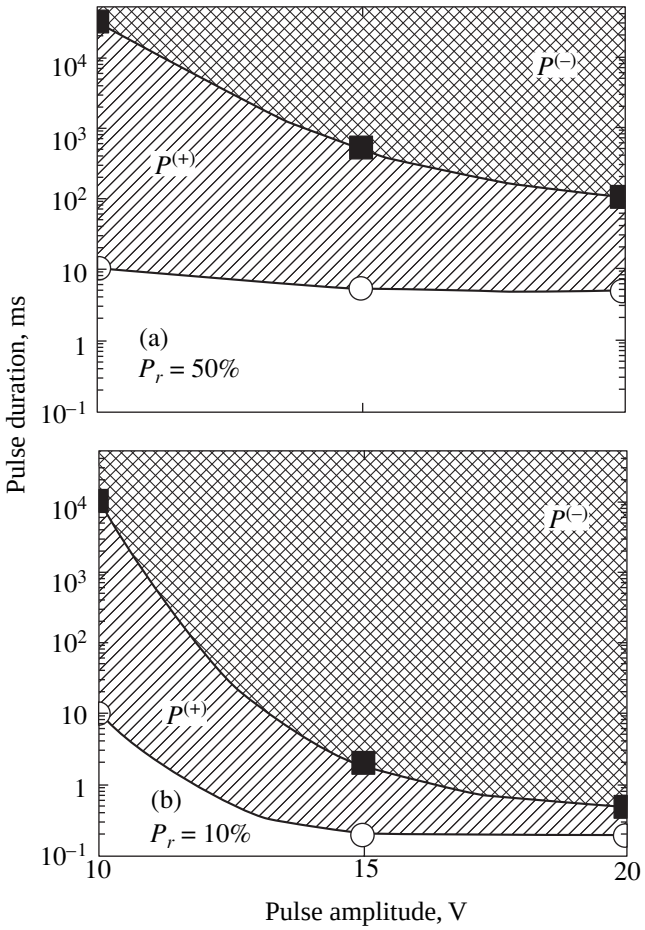


Fig. 7. Switching diagrams for polarization states $P^{(+)}$ and $P^{(-)}$. Shaded areas determine the values of the pulse amplitude and duration required for polarization switching. The area above the curve with circles correspond to switching from state $P^{(-)}$ to state $P^{(+)}$, while the area above the curve with squares corresponds to switching from state $P^{(+)}$ to $P^{(-)}$. The remnant polarization level is (a) 10% and (b) 50% of the attainable remnant polarization (100%) obtained after the application of a 100-ms pulse of an amplitude of 20 V.

equal to approximately 8 V (see Fig. 3). An increase in the amplitude of the polarizing voltage from 15 to 20 V leads to a much faster switching (see the data in Figs. 5 and 6). It should be noted that the polarization being switched for two states for a pulse amplitude of 20 V is still noticeable even for a pulse duration of only 500 μs. A comparison of the data in Figs. 6 and 3 makes it possible to approximately calibrate the value of Φ_2 measured by the NDS method. In accordance with the data presented in Fig. 3, the value of the remnant polarization at a frequency of 80 Hz is approximately equal to 20 mC/m^2, which approximately corresponds to a current response of $\Phi_2 \approx 0.005$ μA after polarization by a 10-ms pulse with an amplitude of 20 V (see Fig. 6).

The data shown in Fig. 6 can be used for estimating the value of remnant polarization attained in the $P^{(+)}$ state at about 50 mC/m^2 ($\Phi_2 \approx 0.012$ μA), which is 30% smaller than the maximal theoretical value of P_s for the PVDF/TrFE material with a composition of 70/30. It can readily be seen from the data presented in Figs. 4–6 that polarization relaxation virtually ceases 100 s after the application of a polarizing pulse. The stability of the NDS method has made it possible to check the polarization state a week after the experiment and to make sure that the remnant polarization has changed insignificantly.

Figure 7 shows the data summarizing the results obtained from an analysis of switching with different pulses and illustrates the conditions imposed on the amplitude and duration of the polarizing pulse, which are necessary for switching between the two polarization states. The data presented in Figs. 7a and 7b correspond to different levels of remnant polarization relative to its maximal value. It can be clearly seen from these figures that the states $P^{(+)}$ and $P^{(-)}$ are not equivalent. Switching to the $P^{(+)}$ state is more efficient (i.e., it occurs at smaller amplitudes and durations of the polarizing pulse as compared to switching to the analogous state $P^{(-)}$). Indirect indications of such an asymmetry also follow from the observed asymmetry of the pyroelectric hysteresis loops recorded for LB films with a thickness of 8–15 monolayers [11]. The asymmetry in polarization switching can be explained by the influence of the substrate. Although both electrodes are made of the same metal, the presence of the glass substrate and the method of successive deposition of the layers may break the symmetry of the system as a whole. In other words, the lower interface between aluminum and the LB film nearest to the substrate may differ from the upper interface (it should be noted that not only the material of the substrate, but also its morphology, roughness, degree of purity, etc., also play a significant role). It is interesting that the switching asymmetry is not reflected in the hysteresis loops depicted in Fig. 3, which were recorded at a frequency of 80 Hz. This can be explained by the effects emerging due to the injection of a free charge from the electrodes to the film. The charge injection apparently occurs too slowly to affect the hysteresis loops measured at a comparatively high frequency. The sample asymmetry can be associated with an "built-in" electric field, which is screened by the injected charges only partially. As a result, the effective coercive field expressed in terms of the external voltage applied to the electrodes becomes different for states $P^{(+)}$ and $P^{(-)}$. In our case, these results can be explained by a higher effective value of the coercive field required for switching to the $P^{(-)}$ state. This question will be discussed in greater detail below in the analysis of the switching model.

The results shown in Fig. 7 show that the switching rate depends on the pulse amplitude. If the amplitude of the applied voltage is only slightly higher than the corresponding amplitude of the coercive field, the switching time becomes very long. Switching by short pulses,

which is especially important for practical applications, requires comparatively large voltage amplitudes. In addition, in the latter case we are dealing with only partial switching. The obtained results show that only about 10% of the maximal polarization is switched when pulses with a duration of about 200 μs and only with an amplitude close to the breakdown voltage of the film (25 V) are used. These experimental results will be explained below in greater detail using numerical simulation.

3. SIMULATION

3.1. General Approach

In the general case, the free energy density of an inhomogeneous intrinsic ferroelectric can be treated as a function of fifteen variables:

$$F = F(r_i, P_i, P'_{ij}). \quad (17)$$

Here, P_i are the polarization vector components serving as the order parameter; r_i denote the components of the radius vector (x, y, z) of a spatial point in the ferroelectric; indices i and j correspond to the x, y, and z directions; and P'_{ij} are the derivative of the polarization vector components,

$$P'_{ij} = \frac{\partial P_i}{\partial r_j}, \quad i, j \in \{x, y, z\}. \quad (18)$$

In this case, the system of three dynamic equations can be written in the form

$$I\frac{d^2P_i}{dt^2} + \xi_{ij}\frac{dP_i}{dt} = -\frac{\partial F}{\partial P_i} + \frac{\partial}{\partial r_j}\left(\frac{\partial F}{\partial P'_{ij}}\right). \quad (19)$$

Here and below, we assume summation over repeating subscripts. On the left-hand side of Eqs. (19), we have the sum of the terms reflecting inertia and friction (viscosity), which are balanced on the right-hand side by the forces or torques associated with the reaction of the medium and the contribution from the external field (all quantities correspond to a unit volume). Equations (19) are the result of direct application of the well-known Euler–Lagrange variational principles for determining the driving forces that bring the system to an equilibrium state corresponding to the minimal total free energy. The inertial contributions and dissipation processes are determined by the mass or the moment of inertia tensor I and the viscosity (frictional) tensor ξ, respectively. In our opinion, the first inertial term in Eqs. (19) can be discarded in the case of polymers. It is well known that this term can be disregarded for nematic liquid crystals that have a lower viscosity as compared to more "crystalline" ferroelectric liquid crystals [13]. Thus, we can rightfully assume that the term associated with viscosity has a larger weight than the inertial contribution for polymers as well.

Equations (19) are not rigorous in the sense that the viscosity tensor is actually a rank four tensor and not a rank two tensor as follows from Eqs. (19). In these equations, we confine ourselves to either purely translational, or purely rotational motion. In the latter case, we are dealing with rotational viscosity, which is a rank two tensor [13], while components P_i are not independent. For example, in the case of a purely rotational mechanism, the additional constraint $P_x^2 + P_y^2 + P_z^2 = \text{const}$ is imposed on the polarization vector and Eqs. (19) should be supplemented with a term proportional to λP_i, where λ is a Lagrange multiplier and quantity I is treated as components of rank two tensor I_{ij}.

To solve Eqs. (19), it is important to specify boundary conditions as well as the initial conditions. Otherwise, the mathematical formulation of the problem becomes incorrect. We can avoid solving the boundary value problem only in special cases (e.g., for an infinitely large homogeneous medium for which the free energy density is independent of the spatial coordinates and can be referred to an arbitrary unit of volume). To solve the boundary value problem, we must take into account the interaction with the boundary surfaces. We propose that the representation

$$W_s = W_1(P_i n_{e,i}) + \frac{1}{2}W_2(P_i n_{e,i})^2 \quad (20)$$

be used for the surface energy, where summation is carried out over repeating indices i, and $n_{e,i}$ are the components of the unit vector $\mathbf{n}_e$ along the preferred ("easy") direction on the surface. We introduce the concept of a preferred direction (the easy axis) by analogy with the description of the surface interaction (anchoring) in the theory of liquid crystals, where this concept is generally accepted [14]. The easy axis corresponds to extremal values of the surface interaction. We can give a clear physical interpretation of the terms in Eq. (20). Indeed, an electric field is always present at the surface, its spatial distribution being determined by the symmetry properties of this surface. Consequently, the first and second terms can be treated as reflecting the polar and nonpolar interactions of this field with the polarization vector, respectively. If the substrate surface is isotropic, we have only one preferred direction coinciding with normal $\mathbf{n}_s$ to the surface ($\mathbf{n}_e \equiv \mathbf{n}_s$). If two boundary electrodes are present in experiment, the polar contributions from the upper and lower surfaces of the electrodes to the surface energy have opposite signs due to inversion of the normal to the surface ($\mathbf{n}_{s1} = -\mathbf{n}_{s2}$) directed into the ferroelectric.

Using Eqs. (19) and (20), we can construct an infinitely large number of various models depending on the representation of the free energy density. Let us con-

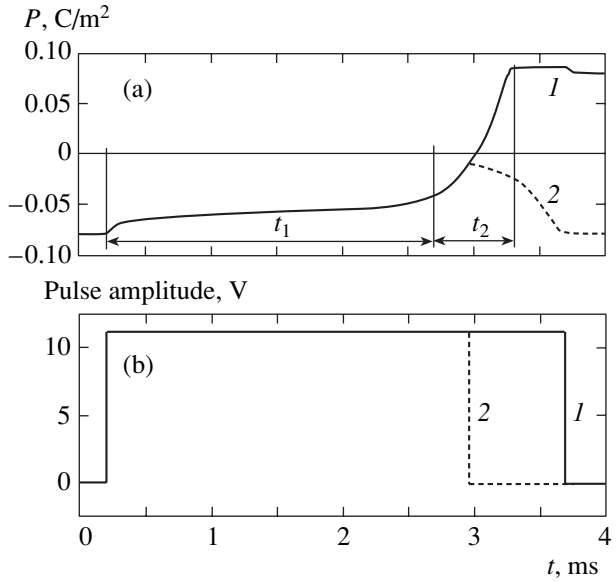


Fig. 8. Calculated response of polarization switching (a) to applied rectangular voltage pulse (b). Curve *2* corresponds to the situation when the voltage pulse duration is insufficient for polarization switching ($\tau < t_1 + t_2/2$; see text).

sider the simplest cases taking into account specific properties of ferroelectric LB films.

3.2. Quasi-homogeneous Model Taking into Account Polar Surface Energy

It is well known that thick polymer ferroelectric films are inhomogeneous in all three directions. In the case of ferroelectric LB films, the situation radically changes in view of their extremely small thickness. The LB films can be treated as monodomain objects in the *z* direction (along the normal to the film), although these films remain inhomogeneous in the *xy* plane (in transverse directions). If the transverse domain size is larger than the film thickness, we can assume in the first approximation that the ferroelectric domains are independent (we disregard the elastic energy concentrated near the boundaries between domains). However, the surface energy of the interaction may be different for different domains since the surface is nonuniform in the general case (it may be rough, may contain defects, etc.).

It was established earlier that many ferroelectric properties of the LB films are correctly described by the Landau–Ginzburg theory [15]. Even the dynamics simulated using the Landau–Khalatnikov homogeneous model (disregarding the surface interaction) leads to reasonable results [8]. In this case, the free energy of a homogeneous ferroelectric domain taking into account the polar interaction at the surface can be expressed as

$$F = \frac{1}{2}\alpha P_z^2 + \frac{1}{4}\beta P_z^4 + \frac{1}{6}\gamma P_z^6 - EP_z + W_{1,s1}P_z\delta(0) - W_{1,s2}P_z\delta(z-d), \quad (21)$$

where $\delta(z)$ is the Dirac δ-function reflecting localization of the surface interaction and subscripts $s1$ and $s2$ correspond to the first and second surfaces, respectively. Allowing for the small thickness of the LB films, we can "delocalize" the contribution of the surface energy over the film thickness d by replacing $(W_{1,s1}P_z\delta(0), W_{1,s2}P_z\delta(z-d))$ by $(W_{1,s1}P_z/d, W_{1,s2}P_z/d)$, thus simplifying the boundary problem. The abovementioned replacement is possible due to the integral equivalence of the two representations. Indeed, the contribution to the total free energy per unit area, for example, from the first surface is

$$\int_0^d W_{1,s1}P_z\delta(0)dz = W_{1,s1}P_z = \int_0^d \frac{W_{1,s1}P_z}{d}dz. \quad (22)$$

In this relation, we proceed from the independence of P_z of the coordinates; for this reason, quantity $W_{1,s1}P_z$ can be taken out of the integral. In fact, the proposed substitution is an approximation that allows us to remain in the framework of the homogeneous Landau–Ginzburg model, in which P_z is independent of the coordinates. This is the condition of the identical effect of the surface on the ferroelectric irrespective of the spatial coordinates of an arbitrary point in the bulk.

Taking into account relations (21), (22), and simplified equations (19), we can write the dynamic equation for the *z* component of polarization in the form

$$\xi\frac{dP_z}{dt} = -(\alpha P_z + \beta P_z^3 + \gamma P_z^5) + \frac{W_1 + U}{d}, \quad (23)$$

where $W_1 = W_{1,s2} - W_{1,s1}$ is the difference in the amplitudes of the polar surface energies for two surfaces and U is the applied voltage. In relation (23), we discarded the inertial term and assumed that the viscosity tensor is isotropic. It can be seen that the difference in the polar interaction with the surfaces is equivalent to the presence of an "built-in" electric field breaking the symmetry of ferroelectric switching since the sign of W_1 is preserved irrespective of the sign of U. The physical reason for this difference can be, for example, the difference in the work functions for charges passing from the upper and lower electrodes into the polymer [15].

Figure 8 shows the results of numerical solution of Eq. (23) for a rectangular voltage pulse applied to the sample. We can note two characteristic times t_1 and t_2 of the switching process. The value of t_1 reflects the critical nature of switching in the sense that it tends to infinity when the electric field attains the coercive value [8]. The value of t_2 can be estimated using expression (23)

in the range of small polarization P_z, when higher order terms in relation (23) can be discarded:

$$\xi\frac{dP_z}{dt} \approx -\alpha P_z + \frac{W_1 + U}{d},$$
$$\frac{1}{t_2} \approx \frac{1}{\xi}\left(-\alpha + \frac{W_1 + U}{dP_s}\right). \quad (24)$$

If the duration τ of a polarizing voltage pulse is smaller than $t_1 + t_2/2$, no polarization switching takes place (see curve *2* in Fig. 8). However, this statement is valid only in the monodomain approximation or in the case when both the film and the boundary conditions are homogeneous. If, however, the film or the surface is inhomogeneous in the *xy* plane, domains with different surface energies are present. Such a system as a whole exhibits dispersion of the coercive field and, hence, dispersion in switching times. It should be noted that the polarizing voltage is fixed in experiments; for this reason, local variations of the film thickness also make an additional contribution to the dispersion of the coercive field.

Figure 9 shows the dependences of switching time τ on the amplitude of a polarizing voltage pulse, which are calculated for different values of the difference between the amplitudes of polar surface energies W_1. The calculations were made for the Landau–Ginzburg coefficients measured experimentally in [7, 16] ($\alpha_0 = 1.4\times10^8$ J m C^{-2} K^{-1}, $T_0 = 74$°C, $\beta \approx -3.9\times10^{12}$ J C^{-4} m^5, and $\gamma \approx 7.9\times10^{14}$ J C^{-6} m^9). The same figure contains experimental points (for the data presented in Fig. 7). Curves *1* and *1** in Fig. 9 correspond to zero value of the difference between the polar surface energies and reflect symmetric switching. However, the figure demonstrates a clearly manifested asymmetry which can be satisfactorily described taking into account the difference in values of W_1 approximately equal to 1 V (solid curves in Fig. 9). It was found that the calculated curve will correspond to experimental points corresponding to a larger fraction of polarization being switched if the viscosity is considerably increased. In other words, quantitative agreement can be reached by varying the viscosity ξ and the difference in polar energies W_1 only for a specific set of experimental data (e.g., for the 10% level of polarization being switched). However, the experimental points corresponding to 50% of the polarization being switched cannot be described using the same value of viscosity. The dispersion of polar surface interaction, which leads only to a shift in the curves along the abscissa axis in Fig. 9, is insufficient for a comprehensive description of switching. This means that other important factors determining the dispersion of switching time also exist; the most important of these factors will be considered below.

3.3. The Role of the Quadratic Term in the Surface Energy

In the above-mentioned quasi-homogeneous model, we discarded the second term in expression (20) for surface energy. However, this term can be important. In the framework of the homogeneous approach described above, this term appears in the free energy density:

$$(W_{2,s1} + W_{2,s2})\frac{P_z^2}{2d} \equiv \frac{W_2 P_z^2}{2d}. \quad (25)$$

In other words, this term effectively "redefines" coefficient α of the first term on expression (21) and dynamic equation (23) assumes the form

$$\xi\frac{dP_z}{dt} = -\left(\left(\alpha + \frac{W_2}{d}\right)P_z + \beta P_z^3 + \gamma P_z^5\right) + \frac{W_1 + U}{d}. \quad (26)$$

The sign of W_2 may be positive or negative depending on whether the normal or planar orientation of the polarizability vector is preferred for a specific surface. If W_2/d assumes different values for different domains in the plane, this also leads to a certain contribution to the switching time dispersion. In contrast to the polar term affecting the switching time only in terms of the coercive field, the quadratic term produces a more direct effect.

Obviously, the contribution (25) appearing in coefficient α may suppress the temperature-induced divergence in the permittivity in the vicinity of the phase transition. This suppression is indeed observed (to verify this, it is sufficient to compare the temperature

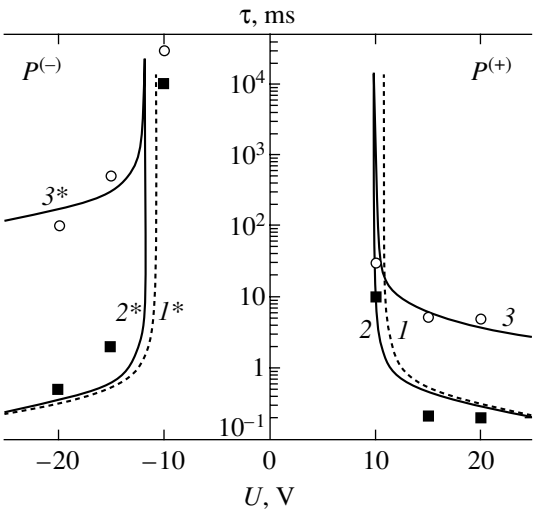


Fig. 9. Calculated dependences of the minimal pulse duration (switching time) on the amplitude of the applied voltage, required for polarization switching and corresponding experimental points: W_1 = 0 (*1*, *1**), 1 (*2*, *2**, *3*, *3**) V, $\xi = 3\times10^6$ (*1*, *1**, *2*, *2**), 4×10^7 (*3*), and 1.5×10^9 (*3**) V s/(m C). The symbols correspond to experimental data for 10 (■) and 50% (○) of switched polarization.

dependences of the permittivity for bulk samples and ultrathin films [3, 4, 15]). The measurement of the dependence of the first Landau coefficient α on the film thickness could provide information on the quadratic surface interaction.

3.4. Inhomogeneous Model

In spite of the fact that the proposed quasi-homogeneous model appears very attractive, its applicability is limited. Even for very thin films, our assumption concerning the statistical independence of domains in the *xy* plane is not always valid. For example, different values of viscosity required for describing switching of different fractions of polarization (see Fig. 9) cannot be described using the quasi-homogeneous model. For this reason we propose a more general form of representation of free energy density, which may become an object for subsequent investigations.

In our opinion, while constructing an inhomogeneous model, we must primarily recognize the importance of the vectorial nature of the order parameter. Polarization can obviously be presented as the product of the scalar part of P_s, which reflects the amplitude of polarization vector $|\mathbf{P}|$ (scalar order parameter) and the unit vector $\mathbf{p}$ directed along the local resultant polarization vector:

$$P_i = |\mathbf{P}| p_i \equiv P_s p_i. \tag{27}$$

In spite of its triviality, this expression indicates the existence of two different mechanisms of ferroelectric polarization switching, one of which is connected with a change in the scalar order parameter and the other with the collective variation of the polarization direction or the rotation of unit vector $\mathbf{p}$.

In the case of inhomogeneous ferroelectric films, we can associate elastic stresses (e.g., near domain walls) not only with the deformation of the lattice involving the variation of its translational parameters, but also with the deformation of the vector polarization field in the sense of variation of the direction of unit vector $\mathbf{p}$. The corresponding theory of elasticity for a uniaxial vector field was developed by Oseen and Frank in the first half of the last century [17, 18]. The application of the Frank–Oseen approach to liquid crystals (LCs) has made it possible not only to grasp the switching of LCs in electric and magnetic fields as well as complex physical processes of formation of LC domains and various types of instability, but also to describe all these processes quantitatively, predicting the behavior of LC systems to a high degree of accuracy. The Frank–Oseen approach is also applicable for a quantitative description of ferroelectric LCs [19]. Like LCs, polymer systems can be attributed to objects known as soft matter. In our opinion, the theory of elasticity developed for LCs can be adapted for describing ferroelectric polymers as polar nematics. The Frank–Oseen theory is of general significance. Its main postulate is the invariance of the elastic energy density per unit volume relative to the field vector inversion ($F_{\text{elastic}}(\mathbf{p}) = F_{\text{elastic}}(-\mathbf{p})$). Similar invariance is also postulated for free energy in zero external fields in the homogeneous Landau–Ginzburg model for ferroelectrics. In accordance with the Frank–Oseen approach, the free elastic energy density per unit volume of a polymer ferroelectric, which is associated with deformation of the vector polarization field, must contain the following linearly independent terms invariant to the inversion of vector $\mathbf{p}$: $A_1(\text{div}\mathbf{p})^2$, $A_2(\mathbf{p} \cdot \text{curl}\mathbf{p})^2$, and $A_3[\mathbf{p} \times \text{curl}\mathbf{p}]^2$. These contributions correspond to splay, twist, and bend of the vector field, respectively. We can also introduce a term proportional to $(\mathbf{p} \cdot \text{curl}\mathbf{p})$, which is also invariant to inversion of $\mathbf{p}$. However, this term is not independent and can be accounted for via the substitution $A_2(\mathbf{p} \cdot \text{curl}\mathbf{p})^2 \longrightarrow A_2(\mathbf{p} \cdot \text{curl}\mathbf{p} + q)^2$. Thus, taking into account expression (20), we can represent the free energy density in our problem in the form

$$\begin{aligned} F &= F_{\text{LG}} + A_1(\text{div}\mathbf{p})^2 + A_2(\mathbf{p} \cdot \text{curl}\mathbf{p} + q)^2 \\ &\quad + A_3[\mathbf{p} \times \text{curl}\mathbf{p}]^2 - P_s(\mathbf{p} \cdot \mathbf{E}) \\ &\quad + \frac{1}{2}[W_{1,s1}P_s(\mathbf{p} \cdot \mathbf{n}_e) + W_{2,s1}P_s(\mathbf{p} \cdot \mathbf{n}_e)^2]\delta(0) \\ &\quad + \frac{1}{2}[W_{1,s2}P_s(\mathbf{p} \cdot \mathbf{n}_e) + W_{2,s2}P_s(\mathbf{p} \cdot \mathbf{n}_e)^2]\delta(z-d), \end{aligned} \tag{28}$$

where F_{LG} is the Landau–Ginzburg free energy density and A_1, A_2, and A_3 are the elastic coefficients for the three types of deformation mentioned above. It should be emphasized that, speaking of the invariance of the elastic part of free energy to the inversion operation, we do not treat $\mathbf{p}$ as a pseudovector as is usually done for nonpolar nematics. In our case, $\mathbf{p}$ is a true vector. Speaking of invariance, we mean that the elastic part of the bulk free energy remains unchanged upon the substitution of $-\mathbf{p}$ for $\mathbf{p}$. It is also clear that no such invariance takes place for the surface contribution when we are considering the polar interaction with the surface (see the first term in expression (20)) as well as for the interaction with an external electric field (term $P_s(\mathbf{p} \cdot \mathbf{E})$ in relation (28)). These are exactly the contributions to free energy that distinguish our "polar nematic" from a "traditional" nonpolar nematic. We do not impose any constraints on the signs of coefficients A_1, A_2, and A_3. Negative values of some coefficients may lead to spontaneous formation of spatially inhomogeneous or domain structures, which is typical of ferroelectrics. Twisted textures can be observed even for positive elastic coefficients if $q \neq 0$.

In the framework of expression (28), switching of a homogeneous ferroelectric is determined only by the Landau–Ginzburg expression for F_{LG} and depends

exclusively on the scalar order parameter $|\mathbf{P}|$. We refer to such switching as scalar switching. The Landau–Ginzburg model describing such a switching is one-dimensional since the switching is completely described only by the spontaneous polarization amplitude. In the Landau–Ginzburg model, no torque is applied to the polarization vector (the resultant polarization vector is always parallel or antiparallel to the electric vector). For this reason, to switch the polarization, we must change the amplitude of the order parameter (its scalar part) rather than the orientation of the polarization vector in the 3D space. A change in the scalar part of the order parameter requires considerable energy, comparable to the energy of intermolecular interactions. In fact, it involves the displacement of charges along the field. This is the reason for very high values of the coercive field in the Landau–Ginzburg model. On the contrary, in the case of an inhomogeneous ferroelectric, switching can be three-dimensional (we define this type of switching as vectorial). Owing to nonzero values of the divergence and curl of the vector field in expression (28), the torque applied to the polarization vector differs from zero and local switching may be a result of collective rotation of the polarization vector without a noticeable change in the scalar order parameter. As a rule, this requires a much lower energy and, hence, is preferable. The coefficients of friction (viscosity) for scalar and vector switching may differ significantly in the general case since these coefficients correspond to different components of the viscosity tensor. Thus, the huge dispersion in our values of viscosity (by three orders of magnitude; see the caption to Fig. 9) can easily be explained in this model. In accordance with relation (28), vector switching may have a threshold only due to the surface interaction or the defects appearing in the bulk. For instance, surface-stabilized ferroelectric LCs are an ideal example of vector switching [13]. The extremely small (~0.1 V/μm) coercive field in these crystals is completely determined by the surface interaction. These simple arguments are in accordance with the fact that the coercive field in thick ferroelectric films is much smaller than the value predicted by the Landau–Ginzburg homogeneous model. In very thin LB films, the role of the surface interaction increases to such an extent that it may suppress the vector mechanism of switching. In our opinion, this is the main reason for the good agreement attained for the first time between the values of the coercive field for very thin films and the predictions of the Landau–Ginzburg theory [15].

A detailed analysis of expression (28) is beyond the scope of this paper. Further study of Eqs. (28) and (19) will clarify the emergence of the vector and scalar mechanism of switching provided that experimental data for films of various thickness are available. In this sense, ferroelectric LB films are quite interesting objects which make it possible to study these two basically different aspects of ferroelectric switching.

4. CONCLUSIONS

Using nonlinear dielectric spectroscopy, the switching and retention of remnant polarization of LB films prepared from the copolymer PVDF/TrFE with a composition of 70/30 have been studied for the first time. It is shown that ferroelectric LB films can be switched to states with opposite polarizations, which are preserved for a long time. The LB films exhibit a great dispersion of the switching time as well as clearly manifested asymmetry in polarization switching. Such parameters as the switching time and polarizing voltage, which are required to attain an equivalent level of remnant polarization, are different for states with opposite directions of polarization. In our opinion, the reason for the asymmetry of switching lies in different intensities of the polar interaction of the ferroelectric polymer with the boundary surfaces (electrodes). The results of numerical simulation of dynamic switching in the Landau–Ginzburg model taking into account the surface interaction successfully describe the experimental data only for extremely strong dispersion of viscosity coefficients for different groups of ferroelectric domains. Thus, a comprehensive description of switching in the framework of the Landau–Ginzburg homogeneous model alone can hardly be obtained. We propose an inhomogeneous switching model including, in addition to the scalar Landau–Ginzburg mechanism, the vectorial switching mechanism based on the Frank–Oseen theory of elasticity for a uniaxial vector field. The combination of two switching mechanisms makes it possible to explain the observed switching time dispersion and peculiarities of switching in ferroelectric LB films.

ACKNOWLEDGMENTS

We are grateful to S. Ducharme, V.M. Fridkin, M.I. Barnik, and V.V. Lazarev for fruitful discussions, as well as to J. Choi *et al.* [12] for permission to demonstrate their results in the inset to Fig. 2.

This study was supported by the Russian Foundation for Basic Research (project nos. 03-02-17288 and 04-02-16466) and by the program "New Materials and Structures" of the Physics Departments, Russian Academy of Sciences. S.P. Palto received a grant from the Russian Science Support Foundation.

REFERENCES

1. L. M. Blinov, Usp. Fiz. Nauk **155**, 443 (1988) [Sov. Phys. Usp. **31**, 623 (1988)].
2. M. C. Petty, *Langmuir–Blodgett Films: An Introduction* (Cambridge Univ. Press, Cambridge, 1996).
3. A. V. Bune, V. M. Fridkin, S. Ducharme, *et al.*, Nature **391**, 874 (1998).
4. T. Furukawa, Phase Transit. **18**, 143 (1989).
5. T. T. Wang, J. M. Herbert, and A. M. Glass, *The Applications of Ferroelectric Polymers* (Chapman and Hall, New York, 1988).

6. L. M. Blinov, V. M. Fridkin, S. P. Palto, *et al.*, Usp. Fiz. Nauk **170**, 247 (2000) [Phys. Usp. **43**, 243 (2000)].
7. S. P. Palto, G. N. Andreev, N. N. Petukhova, *et al.*, Zh. Éksp. Teor. Fiz. **117**, 1003 (2000) [JETP **90**, 872 (2000)].
8. G. Vizdrik, S. Ducharme, V. M. Fridkin, and S. G. Yudin, Phys. Rev. B **68**, 094113 (2003).
9. S. G. Yudin, S. P. Palto, V. A. Khavrichev, *et al.*, Thin Solid Films **210**–**211**, 46 (1992).
10. W. J. Merz, J. Appl. Phys. **27**, 938 (1956).
11. L. M. Blinov, V. M. Fridkin, S. P. Palto, *et al.*, Thin Solid Films **284**–**285**, 469 (1996).
12. J. Choi, C. N. Borca, P. A. Dowben, *et al.*, Phys. Rev. B **61**, 5760 (2000).
13. S. T. Lagerwall, in *Handbook of Liquid Crystals*, Ed. by D. Demus, J. Goodby, G. W. Gray, H.-W. Spiess, and V. Vill (Wiley–VCH, Weinheim, 1998), Vol. 2B, p. 515.
14. R. Barberi, I. Dozov, M. Giocondo, *et al.*, Eur. Phys. J. B **6**, 83 (1998).
15. S. Ducharme, S. P. Palto, and V. M. Fridkin, in *Handbook of Thin Film Materials*, Ed. by H. S. Nalwa (Academic, New York, 2002), Vol. 3, Chap. 11.
16. S. P. Palto, A. M. Lotonov, K. A. Verkhovskaya, *et al.*, Zh. Éksp. Teor. Fiz. **117**, 342 (2000) [JETP **90**, 301 (2000)].
17. C. W. Oseen, Ark. Mat., Astron. Fys. A **19**, 1 (1925).
18. F. C. Frank, Discuss. Faraday Soc. **25**, 19 (1958).
19. S. P. Palto, Kristallografiya **48**, 130 (2003) [Crystallogr. Rep. **48**, 124 (2003)].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 9–18.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 14–25.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Modulational Instabilities in Neutrino–Antineutrino Interactions¶

M. Marklund[a], P. K. Shukla[b,c], G. Betschart[a,d], L. Stenflo[c], D. Anderson[a], and M. Lisak[a]

[a]*Department of Electromagnetics, Chalmers University of Technology SE-412 96, Göteborg, Sweden*
[b]*Institut für Theoretische Physik IV, Fakultät für Physik und Astronomie, Ruhr-Universität Bochum D-44780, Bochum, Germany*
[c]*Department of Plasma Physics, Umeå University SE-901 87, Umeå, Sweden*
[d]*Department of Mathematics and Applied Mathematics, University of Cape Town 7701, Rondebosch, Cape Town, South Africa*
e-mail: marklund@elmagn.chalmers.se

Received January 13, 2004

Abstract—Using a semiclassical approach, we analyze the collective behavior of neutrinos and antineutrinos in a dense background. Applying the Wigner transform technique, we show that the interaction can be modeled by a coupled system of nonlinear Vlasov-like equations. From these equations, we derive a dispersion relation for neutrino–antineutrino interactions on a general background. The dispersion relation admits a novel modulational instability. Moreover, we investigate the modifications of the instability due to thermal effects. The results are examined, together with a numerical example, and we discuss the induced density inhomogeneities using parameters relevant to the early Universe. © *2004 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

Neutrinos have fascinated people ever since they were first introduced by Pauli in 1931. Since then, neutrinos have gone from hypothetical to an extremely promising tool for analyzing astrophysical events, and neutrino cosmology is now one of the hottest topics in modem times due to the discovery that neutrinos may be massive [1]. Because of their weak interaction with other particles, neutrinos can travel great distances without being affected appreciably by material obstacles. They can therefore give us detailed information about events taking place deep within, e.g., supernovas. Furthermore, because the neutrinos decoupled from matter at a redshift z of the order 10^{10}, as compared to $z \approx 10^3$ for photons, it is possible that neutrinos could give us a detailed understanding of the early Universe, if such a signal could be detected [2]. Massive neutrinos have also been a possible candidate for hot dark matter necessary for explaining certain cosmological observations, such as rotation curves of spiral galaxies [3]. Therefore, massive neutrinos could have a profound influence on the evolution of our Universe. Unfortunately, due to the Tremaine–Gunn bound [4], the necessary mass of the missing particles (if these are fermions) for explaining the formation of dwarf galaxies seems to make neutrinos of any species unlikely single candidates for dark matter. As a remedy to this problem, interacting hot dark matter has been suggested [5, 6], because the interaction prevents free-streaming smoothing of small-scale neutrino inhomogeneities. Thus, dark matter in astrophysics not only is a mystery but also plays an essential role in determining the dynamics of the Universe, its large-scale structures (galaxies and super clusters). However, so far, the suggested "sticky" neutrino models have not been successful in dealing with the dwarf galaxy problem [5].

The first successful indication that neutrinos have a nonzero mass came in 1998 through laboratory experiments of atmospheric neutrinos and their oscillations [7]. Although the allowed neutrino masses encompass a wide range,[1] it is currently believed that neutrinos have masses below 2 eV. This conclusion is further supported by independent cosmological observations (see, e.g., [9]). Thus, the masses of neutrinos are indeed very small, and the classical analysis by Tremaine and Gunn would therefore indicate that neutrinos can in no way be considered a sole candidate for dark matter. This conclusion is reanalyzed in this paper within the electro-weak framework, where neutrino–neutrino interactions occur as a natural consequence of the theory.

We thus consider the nonlinear interaction between neutrinos and antineutrinos in the lepton plasma of the early Universe, adopting a semiclassical model. Neutrinos and antineutrinos interact with dense plasmas through the charged and neutral weak currents arising

¶This article was submitted by the authors in English.

[1] Some estimates even support the notion that neutrinos may contribute up to 20% of the matter density of the Universe [8].

from the Fermi weak nuclear interaction forces. Charged weak currents involve the exchange of the charged vector bosons associated with the processes involving interactions between leptons and neutrinos of the same flavor, while neutrino weak currents involve the exchange of the neutral vector bosons associated with processes involving neutrinos of all types interacting with arbitrary charged and neutral particles. Asymmetric flows of neutrinos and antineutrinos in the early Universe plasma may be created by the ponderomotive force of nonuniform intense photon beams or by shock waves. Here, using an effective field theory approach, a system of coupled Wigner–Moyal equations for nonlinearly interacting neutrinos and antineutrinos is derived, and it is shown that these equations admit a modulational instability. Finally, we discuss the relevance of our results in the context of the dark matter problem, and it is moreover suggested that the nonlinearly excited fluctuations could be used as a starting point for obtaining a better understanding of the process of galaxy formation. It turns out that the short-time evolution of the primordial neutrino plasma medium in the temperature range $1 < T < 10$ MeV is governed by collisionless collective effects involving relativistic neutrinos and antineutrinos.

2. DISPERSION RELATION AND THE MOTION OF NEUTRINO BUNCHES

As a primer, we study the implication of the known dispersion of neutrinos on a thermal neutrino/antineutrino background, using the eikonal representation and the WKBJ approximation.

We suppose that a single neutrino (or antineutrino) moves in a Fermionic sea composed of a neutrino–antineutrino mixture. The energy E of the neutrino (antineutrino) is then given by (see, e.g., [10, 11])

$$E = \sqrt{p^2c^2 + m^2c^4} + V_\pm(\mathbf{r}, t), \tag{1}$$

where $\mathbf{p}$ is the neutrino (antineutrino) momentum; c, the speed of light in vacuum; and m, the neutrino mass. The effective potential for a neutrino moving on a background of its own flavor and in thermal equilibrium is given by [10] (see also [12–15])[2]

$$V_\pm(\mathbf{r}, t) = \pm 2\sqrt{2}G_F(n - \bar{n}), \tag{2a}$$

while the potential for a neutrino moving on a background of a different flavor is

$$V_\pm(\mathbf{r}, t) = \pm\sqrt{2}G_F(n - \bar{n}), \tag{2b}$$

where

$$\frac{G_F}{(\hbar c)^3} \approx 1.2 \times 10^{-5}\ \text{GeV}^{-2},$$

G_F is the Fermi constant, n ($\bar{n}$) is the density of the background neutrinos (antineutrinos), and + (–) represents the propagating neutrino (antineutrino). Expressions (2) are valid in the rest frame of the background. As is seen from (1) and (2), while neutrinos moving in a background of neutrinos and antineutrinos change their energy by an amount $\sim G_F(n - \bar{n})$, the antineutrinos change their energy by $\sim -G_F(n - \bar{n})$ [16]. The extra factor of 2 in (2a) compared to (2b) comes from exchange effects between identical particles [13].

Relation (1) can be interpreted as a dispersion relation for relativistic and nonrelativistic neutrinos, with the identifications $E = \hbar\omega$ and $\mathbf{p} = \hbar\mathbf{k}$, i.e.,

$$\omega = c\sqrt{k^2 + \frac{m^2c^2}{\hbar^2}} + \frac{V_\pm}{\hbar}, \tag{3}$$

where $\hbar$ is the Planck constant divided by 2π. Using Eq. (3), the eikonal representation

$$E \longrightarrow \hbar\omega_0 - \frac{i\hbar\partial}{\partial t}, \quad \mathbf{p} \longrightarrow \hbar\mathbf{k}_0 + i\hbar\nabla$$

and the WKBJ approximation [17, 18]

$$\left|\frac{\partial\Psi}{\partial t}\right| \ll \omega_0|\Psi|, \quad |\nabla\Psi| \ll |\mathbf{k}_0||\Psi|,$$

we obtain a Schrödinger equation for slowly varying neutrino (antineutrino) wave packets $\Psi(\mathbf{r}, t)$ modulated by long-scale density fluctuations (i.e., neutrino bunches)[3]

$$i\left(\frac{\partial}{\partial t} + \mathbf{v}_g \cdot \nabla\right)\Psi + \frac{\hbar c^2}{2E_0}\left[\nabla^2 - \left(1 - \frac{m^2c^4}{E_0^2}\right)(\mathbf{n}_0 \cdot \nabla)^2\right]\Psi - \frac{V_\pm}{\hbar}\Psi = 0, \tag{4}$$

where

$$\mathbf{v}_g = c\mathbf{k}_0(k_0^2 + m^2c^2/\hbar^2)^{-1/2}$$

is the group velocity[4] of relativistic neutrinos and antineutrinos, which have similar energy spectra, E_0 is the neutrino energy in the absence of interactions, $\mathbf{n}_0 = \mathbf{k}_0/|k_0|$, and $\mathbf{k}_0$ is the vacuum wave vector. We now suppose that the neutrino bunches themselves are nearly in thermal equilibrium (to be quantified in the next section). Then, we have the case of self-interacting neutri-

[2] For a more detailed description of the potential, see the next section.

[3] See also [16] for a similar treatment of neutrino–electron interactions.

[4] We note that, when the scale length of the density inhomogeneity is comparable to the wavelength of the modulated neutrino wave packets, we must modify the coupled Schrödinger equations to account for differing group velocities of neutrinos and antineutrinos in a Fermionic sea. We expect a shift in the momentum of Eq. (13) and a slower growth rate of the modulational instability of neutrino quasiparticles involving short-scale density inhomogeneities.

nos and antineutrinos, and the densities in the potential $V_\pm$ are given in terms of the sums

$$\begin{aligned} n &= \sum_{i=1}^{M} n_i = \sum_{i=1}^{M} \left\langle |\Psi_{i+}|^2 \right\rangle, \\ \bar{n} &= \sum_{i=1}^{N} \bar{n}_i = \sum_{i=1}^{N} \left\langle |\Psi_{i-}|^2 \right\rangle, \end{aligned} \tag{5}$$

where Ψ_{i+} and Ψ_{i-} are the neutrino and antineutrino wave packets, respectively (with i numbering the wave packets), and the angular bracket denotes the ensemble average. In this case, the relativistic neutrino and antineutrino wave packets are comoving with the background, and Eq. (4) thus yields

$$i\frac{\partial \Psi_{i\pm}}{\partial t} + \frac{\hbar c^2}{2E_0}\left(\nabla_\perp^2 + \frac{m^2c^4}{E_0^2}\nabla_\parallel^2\right)\Psi_{i\pm} - \frac{V_\pm}{\hbar}\Psi_{i\pm} = 0, \tag{6}$$

where

$$\nabla_\perp^2 = \nabla^2 - (\mathbf{n}_0 \cdot \nabla)^2 \text{ and } \nabla_\parallel^2 = (\mathbf{n}_0 \cdot \nabla)^2.$$

Expressions (2a) and (5) reveal that self-interactions between relativistic neutrinos and antineutrinos produce a nonlinear asymmetric potential in Eq. (6). By further rescaling the coordinate along $\mathbf{n}_0$, Eq. (4) can finally be written as the coupled system

$$i\frac{\partial \Psi_{i\pm}}{\partial t} + \frac{\alpha}{2}\nabla^2\Psi_{i\pm} \mp \beta(n-\bar{n})\Psi_{i\pm} = 0, \tag{7}$$

where

$$\alpha = \frac{\hbar c^2}{E_0}, \quad \beta = \frac{2\sqrt{2}G_F}{\hbar}$$

for neutrinos moving on the same flavor background.

Equation (7) shows that this approach can lead to some interesting effects. The case of a single self-interacting neutrino bunch shows that the formation of dark solitary structures is possible. Furthermore, the slightly more complicated case of two interacting bunches, of either the neutrino–neutrino or neutrino–antineutrino type, can result in splitting and focusing of the wave packets [19].

3. KINETIC DESCRIPTION

In the preceding section, we investigated the case of a neutrino bunch close to thermal equilibrium. In general, this may of course not be the case, and Eq. (2) must be modified. The more precise form of the potential $V_\pm$ for equal species due to neutrino forward scattering is given by [20]

$$\begin{aligned} V_\pm(t, \mathbf{r}, \mathbf{p};\, f_{i\pm}) &= \pm 2\sqrt{2}G_F \int d\mathbf{q}(1 - \hat{\mathbf{p}}\cdot\hat{\mathbf{q}}) \\ &\times \left[\sum_{i=1}^{M} f_{i+}(t, \mathbf{r}, \mathbf{q}) - \sum_{i=1}^{N} f_{i-}(t, \mathbf{r}, \mathbf{q})\right], \end{aligned} \tag{8}$$

where hatted quantities denote the corresponding unit vector and $f_{i+}(t, \mathbf{r}, \mathbf{q})$ ($f_{i-}(t, \mathbf{r}, \mathbf{q})$) is the neutrino (antineutrino) distribution function corresponding to bunch i. The distribution functions are taken to be normalized such that

$$\begin{aligned} n_i(t, \mathbf{r}) &= \int d\mathbf{q}\, f_{i+}(t, \mathbf{r}, \mathbf{q}), \\ \bar{n}_i(t, \mathbf{r}) &= \int d\mathbf{q}\, f_{i-}(t, \mathbf{r}, \mathbf{q}), \end{aligned} \tag{9}$$

where n_i ($\bar{n}_i$) is the number density of the ith neutrino (antineutrino) bunch.

We first note that, when the distribution is thermal, potential (8) reduces exactly to (2a). Second, when the neutrinos have an almost thermal distribution, i.e., the corresponding distribution function can be expressed as (dropping the indices for notational simplicity)

$$f(t, \mathbf{r}, \mathbf{p}) = f_0(p) + \delta f(t, \mathbf{r}, \mathbf{p}), \quad |\delta f| \ll |f_0|,$$

we obtain the following form of the potential:

$$\begin{aligned} V_\pm(t, \mathbf{r}, \mathbf{p};\, f_{i\pm}) &= \pm 2\sqrt{2}G_F \\ \times \left[(n-\bar{n}) - \int d\mathbf{q}(\hat{\mathbf{p}}\cdot\hat{\mathbf{q}})\left(\sum_{i=1}^{M}\delta f_{i+} - \sum_{i=1}^{N}\delta f_{i-}\right)\right]. \end{aligned} \tag{10}$$

The last term is small and may therefore be neglected, and we obtain

$$V_\pm(t, \mathbf{r}) \approx \pm 2\sqrt{2}G_F(n - \bar{n}),$$

in accordance with expressions (2a), thus justifying the equation of motion (7).

Now, we define a distribution function for the collective neutrino states by Fourier transforming the two-point correlation function for $\Psi_\pm$, according to [21]

$$\begin{aligned} f_{i\pm}(t, \mathbf{r}, \mathbf{p}) &= \frac{1}{(2\pi\hbar)^3}\int d\mathbf{y} \exp(i\mathbf{p}\cdot\mathbf{y}/\hbar) \\ &\times \langle \Psi_{i\pm}^*(t, \mathbf{r} + \mathbf{y}/2)\Psi_{i\pm}(t, \mathbf{r} - \mathbf{y}/2)\rangle, \end{aligned} \tag{11}$$

where $\mathbf{p}$ represents the momentum of the neutrino (antineutrino) quasiparticles (we note that the ensemble average was not present in the original definition [21] but has important consequences when the wave packet has a random phase). With definition (11), it follows that

$$\left\langle |\Psi_{i\pm}(t, \mathbf{r})|^2 \right\rangle = \int d\mathbf{p}\, f_{i\pm}(t, \mathbf{r}, \mathbf{p}). \tag{12}$$

Thus, using (11) and (6) together with potential (8), we obtain the generalized Wigner–Moyal equation for $f_{i\pm}$,

$$\frac{\partial f_{i\pm}}{\partial t} + \frac{c^2\mathbf{p}}{E_0}\frac{\partial f_{i\pm}}{\partial \mathbf{r}} - \frac{2V_\pm}{\hbar}\sin\left[\frac{\hbar}{2}\left(\frac{\overleftarrow{\partial}}{\partial \mathbf{r}}\frac{\vec{\partial}}{\partial \mathbf{p}} - \frac{\overleftarrow{\partial}}{\partial \mathbf{p}}\frac{\vec{\partial}}{\partial \mathbf{r}}\right)\right] f_{i\pm} = 0, \tag{13}$$

where the sin operator is defined in terms of its Taylor expansion and the arrows denote the direction of operation. In the case of potential (2a), the last term in the sin operator drops out, and Eq. (13) reduces to the standard Wigner–Moyal equation [21]. Equation (13) was obtained in [20] using a density matrix approach.

Retaining only the lowest-order terms in $\hbar$ (i.e., taking the long-wavelength limit), we obtain the coupled Vlasov equations

$$\left[\frac{\partial}{\partial t} + \left(\frac{c^2\mathbf{p}}{E_0} + \frac{\partial V_\pm}{\partial \mathbf{p}}\right)\frac{\partial}{\partial \mathbf{r}}\right] f_{i\pm} - \frac{\partial V_\pm}{\partial \mathbf{r}}\frac{\partial f_{i\pm}}{\partial \mathbf{p}} = 0. \tag{14}$$

The term $\partial V_\pm/\partial \mathbf{p}$ represents the group velocity. While higher-order group velocity dispersion is present in (13), this is not the case in (14). Thus, information is partially lost by using Eq. (14). Furthermore, while Eq. (14) preserves the number of quasiparticles, Eq. (13) shows that this conclusion is in general not true, i.e., the particle number in a phase-space volume is not constant, and the higher-order terms $\partial^n V_\pm/\partial \mathbf{r}^n$ may moreover contain vital short-wavelength information. Equations similar to (14) have been used to study neutrino–electron interactions in astrophysical contexts [11].

We now suppose that we have small amplitude perturbations on a background of constant neutrino and antineutrino densities $n_i = n_{i0}$ and $\bar{n}_i = \bar{n}_{i0}$, respectively,

$$f_{i\pm}(t, \mathbf{r}, \mathbf{p}) = f_{i0\pm}(\mathbf{p}) + \delta f_{i\pm}(\mathbf{p})\exp[i(\mathbf{K}\cdot\mathbf{r} - \Omega t)], \tag{15}$$

and $|\delta f_{i\pm}| \ll |f_{i0\pm}|$, where $\mathbf{K}$ and Ω are the perturbation wave vector and frequency, respectively. Thus, Eqs. (13) give

$$i\left[\Omega - \frac{c^2\mathbf{p}\cdot\mathbf{K}}{E_0} - \frac{2i}{\hbar}V_{0\pm}\sin\left(-\frac{i\hbar}{2}\frac{\overleftarrow{\partial}}{\partial \mathbf{p}}\cdot\mathbf{K}\right)\right]\delta f_{i\pm} + \frac{2}{\hbar}\delta V_\pm \sin\left(\frac{i\hbar}{2}\mathbf{K}\cdot\frac{\vec{\partial}}{\partial \mathbf{p}}\right) f_{i0\pm} = 0, \tag{16}$$

where $\delta V_\pm = V_\pm(t, r, \mathbf{p}; \delta f_{i\pm})$ and $V_{0\pm} = V_\pm(t, r, \mathbf{p}; \delta f_{i0\pm})$ from Eq. (8). Eliminating $\delta f_{i\pm}$ from (16), using $\delta V_- = -\delta V_+$, we have

$$\delta V_+(\mathbf{p}) = \frac{4\sqrt{2}iG_F}{\hbar}\int d\mathbf{q}(1 - \hat{\mathbf{p}}\cdot\hat{\mathbf{q}})\delta V_+(\mathbf{q}) \times \left[\sum_{i=1}^{M}\frac{\sin\left(\frac{i\hbar}{2}\mathbf{K}\cdot\frac{\vec{\partial}}{\partial \mathbf{q}}\right) f_{i0+}(\mathbf{q})}{\Omega - \frac{c^2\mathbf{q}\cdot\mathbf{K}}{E_0} - \frac{2i}{\hbar}V_{0+}(\mathbf{q})\sin\left(\frac{i\hbar}{2}\frac{\overleftarrow{\partial}}{\partial \mathbf{q}}\cdot\mathbf{K}\right)} + \sum_{i=1}^{N}\frac{\sin\left(\frac{i\hbar}{2}\mathbf{K}\cdot\frac{\vec{\partial}}{\partial \mathbf{q}}\right) f_{i0-}(\mathbf{q})}{\Omega - \frac{c^2\mathbf{q}\cdot\mathbf{K}}{E_0} - \frac{2i}{\hbar}V_{0-}(\mathbf{q})\sin\left(\frac{i\hbar}{2}\frac{\overleftarrow{\partial}}{\partial \mathbf{q}}\cdot\mathbf{K}\right)}\right]. \tag{17}$$

Assuming that $\delta f_{i\pm}(\mathbf{p})$ is a symmetric function of $\mathbf{p}$, which is a reasonable physical restriction, implies that $\delta V_\pm$ is independent of $\mathbf{p}$, and Eq. (17) simplifies to the dispersion relation

$$1 = \frac{4\sqrt{2}iG_F}{\hbar} \times \int d\mathbf{q}\left[\sum_{i=1}^{M}\frac{\sin\left(\frac{i\hbar}{2}\mathbf{K}\cdot\frac{\partial}{\partial \mathbf{q}}\right) f_{i0+}(\mathbf{q})}{\Omega - \frac{c^2\mathbf{q}\cdot\mathbf{K}}{E_0} - \frac{2i}{\hbar}\sin\left(\frac{i\hbar}{2}\mathbf{K}\cdot\frac{\partial}{\partial \mathbf{q}}\right)V_{0+}(\mathbf{q})} + \sum_{i=1}^{N}\frac{\sin\left(\frac{i\hbar}{2}\mathbf{K}\cdot\frac{\partial}{\partial \mathbf{q}}\right) f_{i0-}(\mathbf{q})}{\Omega - \frac{c^2\mathbf{q}\cdot\mathbf{K}}{E_0} - \frac{2i}{\hbar}\sin\left(\frac{i\hbar}{2}\mathbf{K}\cdot\frac{\partial}{\partial \mathbf{q}}\right)V_{0-}(\mathbf{q})}\right], \tag{18}$$

where we have dropped the arrows indicating the direction of operation. We note that, if the background distribution is thermal, $V_{0\pm}$ is independent of $\mathbf{p}$, and the last term in the denominators of Eq. (18) vanishes.

3.1. The One-Dimensional Case

The simplest way to analyze dispersion relation (18) is to reduce the dimensionality of the problem. We therefore first consider the one-dimensional case, where we may use the identity

$$2\sin\left(\frac{i\hbar K}{2}\frac{\partial}{\partial p}\right)h(p) = i\left[h\left(p + \frac{\hbar K}{2}\right) - h\left(p - \frac{\hbar K}{2}\right)\right]$$

in order to rewrite dispersion relation (18) as

$$1 = -\frac{2\sqrt{2}iG_F}{\hbar} \times \int dq \left\{ \sum_{i=1}^{M} \frac{f_{i0+}(q+\hbar K/2) - f_{i0+}(q-\hbar K/2)}{\Omega - c^2 qK/E_0 + \Delta_+(q)} + \sum_{i=1}^{N} \frac{f_{i0-}(q+\hbar K/2) - f_{i0-}(q-\hbar K/2)}{\Omega - c^2 qK/E_0 + \Delta_-(q)} \right\}, \quad (19)$$

where we have introduced

$$\Delta_\pm(q) \equiv \frac{1}{\hbar}\left[V_{0\pm}\left(q + \frac{\hbar K}{2}\right) - V_{0\pm}\left(q - \frac{\hbar K}{2}\right)\right].$$

In the case of monoenergetic beams, i.e.,

$$f_{i0+}(p) = n_{i0}\delta(p - p_{i0}), \quad f_{i0-}(p) = \bar{n}_{i0}\delta(p - \bar{p}_{i0}),$$

Eq. (19) reduces to

$$\begin{aligned} 1 = -\frac{2\sqrt{2}G_F}{\hbar}\Bigg\{ & \sum_{i=1}^{M} n_{i0}\Bigg[-\frac{\hbar c^2 K^2}{E_0} + \Delta_+\left(p_{i0} + \frac{\hbar K}{2}\right) - \Delta_+\left(p_{i0} - \frac{\hbar K}{2}\right)\Bigg]\Bigg[\left(\Omega - \frac{c^2 p_{i0} K}{E_0}\right)^2 - \left(\frac{\hbar c^2 K^2}{2E_0}\right)^2 \\ & + \left(\Omega - \frac{c^2 p_{i0} K}{E_0}\right)\left[\Delta_+\left(p_{i0} + \frac{\hbar K}{2}\right) + \Delta_+\left(p_{i0} - \frac{\hbar K}{2}\right)\right] \\ & + \frac{\hbar c^2 K^2}{2E_0}\left[\Delta_+\left(p_{i0} + \frac{\hbar K}{2}\right) - \Delta_+\left(p_{i0} - \frac{\hbar K}{2}\right)\right] \\ & + \Delta_+\left(p_{i0} + \frac{\hbar K}{2}\right)\Delta_+\left(p_{i0} - \frac{\hbar K}{2}\right)\Bigg]^{-1} \\ & + \sum_{i=1}^{N} \bar{n}_{i0}\Bigg[-\frac{\hbar c^2 K^2}{E_0} + \Delta_-\left(\bar{p}_{i0} + \frac{\hbar K}{2}\right) - \Delta_-\left(\bar{p}_{i0} - \frac{\hbar K}{2}\right)\Bigg] \\ & \times \Bigg[\left(\Omega - \frac{c^2 \bar{p}_{i0} K}{E_0}\right)^2 - \left(\frac{\hbar c^2 K^2}{2E_0}\right)^2 \\ & + \left(\Omega - \frac{c^2 \bar{p}_{i0} K}{E_0}\right)\left[\Delta_-\left(\bar{p}_{i0} + \frac{\hbar K}{2}\right) + \Delta_-\left(\bar{p}_{i0} - \frac{\hbar K}{2}\right)\right] \\ & + \frac{\hbar c^2 K^2}{2E_0}\left[\Delta_-\left(\bar{p}_{i0} + \frac{\hbar K}{2}\right) - \Delta_-\left(\bar{p}_{i0} - \frac{\hbar K}{2}\right)\right] \\ & + \Delta_-\left(\bar{p}_{i0} + \frac{\hbar K}{2}\right)\Delta_-\left(\bar{p}_{i0} - \frac{\hbar K}{2}\right)\Bigg]^{-1}\Bigg\}, \end{aligned} \quad (20)$$

where

$$V_{0\pm}(p) = \pm 2\sqrt{2}G_F\Bigg[(n_0 - \bar{n}_0) - \operatorname{sgn} p\left(\sum_{i=1}^{M} n_{i0}\operatorname{sgn} p_{i0} - \sum_{i=1}^{N} \bar{n}_{i0}\operatorname{sgn}\bar{p}_{i0}\right)\Bigg] \quad (21)$$

by Eq. (8).

We consider the simplest case of interacting neutrinos and antineutrinos with $M = N = 1$. We assume that they have equal densities $n_0 = \bar{n}_0$ and are counter-propagating, i.e., $p_0 = -\bar{p}_0 > 0$. From (21), we then obtain the potential

$$V_{0\pm}(p) = \mp 4\sqrt{2}G_F \operatorname{sgn} p n_0, \quad (22)$$

while Eq. (20) yields

$$\begin{aligned} \sigma\left(\frac{\hbar c^2 K^2}{E_0} - 2\sigma\varepsilon\right)\Bigg\{ & \Bigg[\left(\Omega - \frac{c^2 p_0 K}{E_0}\right)^2 - \left(\frac{\hbar c^2 K^2}{2E_0}\right)^2 \\ & - 2\sigma\varepsilon\left(\left(\Omega - \frac{c^2 p_0 K}{E_0}\right) - \frac{\hbar c^2 K^2}{2E_0}\right)\Bigg]^{-1} \\ & + \Bigg[\left(\Omega + \frac{c^2 p_0 K}{E_0}\right)^2 - \left(\frac{\hbar c^2 K^2}{2E_0}\right)^2 \\ & + 2\sigma\varepsilon\left(\left(\Omega + \frac{c^2 p_0 K}{E_0}\right) + \frac{\hbar c^2 K^2}{2E_0}\right)\Bigg]^{-1}\Bigg\} = 1, \end{aligned} \quad (23)$$

where

$$\sigma = \frac{2\sqrt{2}G_F n_0}{\hbar},$$

$$\varepsilon = 1 - \operatorname{sgn}(p_0 - \hbar K) = \begin{cases} 0, & p_0 > \hbar K \\ 1, & p_0 = \hbar K \\ 2, & p_0 < \hbar K. \end{cases}$$

Thus, for $\varepsilon = 0$, the growth rate is given by (see Figs. 1 and 2)

$$\frac{\Gamma^2}{K^2} = \sqrt{4v^2\left(\frac{\hbar c^2 K}{2E_0}\right)^2 + 4v^2 v_F^2 + v_F^4} - v^2 - v_F^2 - \left(\frac{\hbar c^2 K}{2E_0}\right)^2, \quad (24)$$

where $\Gamma = -i\Omega$ is the instability growth rate and

$$v_F^2 \equiv 2\sqrt{2}G_F n_0 c^2/E_0.$$

We also note that, as expected, the instability disappears in the limit $v \longrightarrow 0$, just stating the well-known

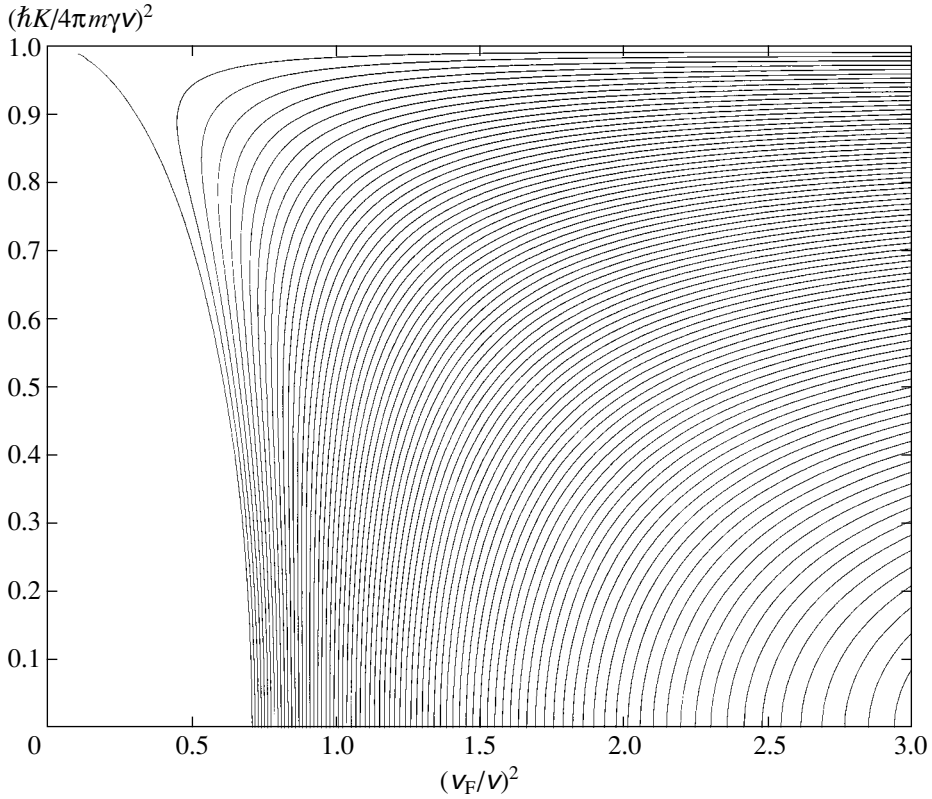


Fig. 1. A contour plot of the values of $(\Gamma/Kv)^2$, when $\varepsilon = 0$, for which the instability occurs. The function $(\Gamma/Kv)^2$ is constant along the contours and is plotted in terms of the variables $(v_F/v)^2$ and $(\hbar K/4\pi m\gamma v)^2$. Outside the contours, $\Gamma^2 < 0$.

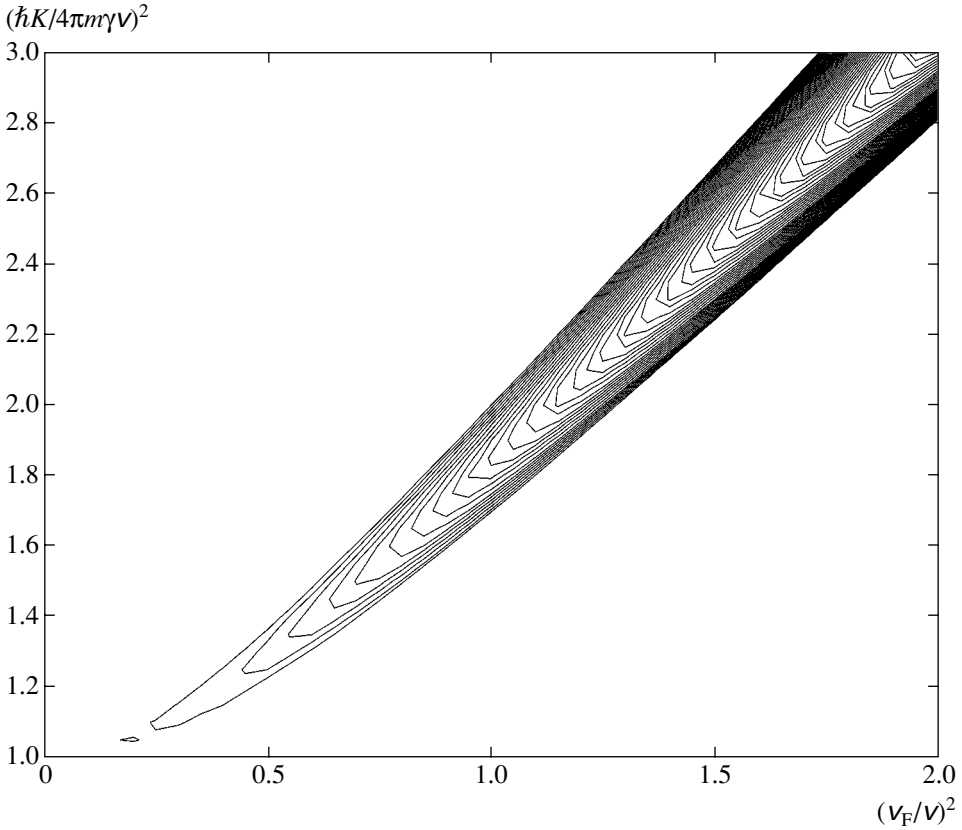


Fig. 2. The same plot as in Fig. 1 but for $\varepsilon = 2$.

fact that there must be a nonzero relative velocity between the beams in order for the instability to occur. Because Γ^2 is positive, we have

$$\left(\frac{v}{v_F}\right)^2 - 2 < \left(\frac{\hbar c^2 K}{2E_0 v_F}\right)^2 < \left(\frac{v}{v_F}\right)^2, \tag{25}$$

i.e.,

$$l_0 < l < l_0(1 - 2v_F^2/v^2)^{-1/2}, \tag{26}$$

where we have introduced the length scales

$$l = \frac{2\pi}{K}, \quad l_0 = \frac{\hbar c^2}{2E_0 v}.$$

Thus, a higher neutrino momentum can retain a smaller instability length scale. It is clear from (24) that (i) the instability remains for arbitrary velocities (see Figs. 1 and 2) and (ii) the higher the neutrino velocity, the smaller the corresponding instability length scale l.

3.2. Partial Incoherence and Thermal Effects

Partial incoherence can in general lead to a lower growth rate, similar to the inverse Landau damping. As an example of the results of stochastic effects, e.g., thermal fluctuations, we consider the following example. Let the indeterminacy of the neutrino collective state manifest itself in a random phase $\varphi(x)$ of the background wave packet, with the width Δp defined according to

$$\begin{aligned}\langle \exp(-i\varphi(x + y/2))\exp(i\varphi(x - y/2))\rangle \\ = \exp(-\Delta p|y|/\hbar).\end{aligned}$$

Due to this random spread, the modulational instability is damped, as we show below. The Wigner function corresponding to the random phase assumption is given by the Lorentz distribution

$$f_0(p) = \frac{n_0}{\pi}\frac{\Delta_p}{(p - p_0)^2 + \Delta p^2}. \tag{27}$$

With this, we obtain Eq. (24) with

$$\Gamma \longrightarrow \Gamma_D + \Delta c^2 pK/E_0,$$

where Γ_D is the reduced growth rate. Thus, we see that the broadening tends to suppress the growth. Moreover, a positive growth rate Γ_D requires that

$$\frac{2v\Delta p}{\hbar\Gamma} < \frac{l}{l_0}, \tag{28}$$

where Γ is given by Eq. (24). Hence, the general property of a spread in momentum space, here exemplified by a random phase, is to put bounds on the modulational instability length scale l.

Incoherent effects among the neutrinos and antineutrinos can also be approached for a background obeying the Fermi–Dirac statistics, i.e.,

$$f_{0\pm}(p) = \frac{cn_0}{\ln 4 k_B T_\pm}\left[1 + \exp\left(\frac{c|p|}{k_B T_\pm}\right)\right]^{-1}, \tag{29}$$

where we set $M = N = 1$ and assume $n_0 = \bar{n}_0$. Here, we have neglected the mass of the neutrinos (which leads to the correct result to the lowest order). For simplicity, we assume that $T_\pm = T$, and therefore dispersion relation (19) takes the form

$$\begin{aligned}1 = -\frac{4\sqrt{2}cG_F n_0}{\ln 4\hbar k_B T}\int_{-\infty}^{\infty} dp\left(\Omega - \frac{c^2 pK}{E_0}\right)^{-1} \\ \times\left[\left(1 + \exp\left(\frac{c|p + \hbar K/2|}{k_B T}\right)\right)^{-1}\right. \\ \left. + \left(1 + \exp\left(\frac{c|p - \hbar K/2|}{k_B T}\right)\right)^{-1}\right].\end{aligned} \tag{30}$$

Dispersion relation (30) cannot be solved analytically, but it can be expressed as

$$1 = -Q[P(I(\Omega_n, K_n)) + i\pi g(\Omega_n, K_n)], \tag{31}$$

where $P(I(\Omega_n, K_n))$ is the principal value of the integral

$$\begin{aligned}I = \int_0^{\infty} dx(1 + e^x)^{-1} \\ \times\left[\frac{\Omega_n + K_n^2}{(\Omega_n + K_n^2)^2 - K_n^2 x^2} + \frac{\Omega_n - K_n^2}{(\Omega_n - K_n^2)^2 - K_n^2 x^2}\right],\end{aligned} \tag{32}$$

and $g = g_+ + g_-$, where

$$\begin{aligned}g_\pm(\Omega_n, K_n) = \frac{\Omega_n \pm K_n^2}{1 + \exp(|\Omega_n \pm K_n^2|/\sqrt{2}K_n)}, \\ Q \equiv \frac{4}{\ln 4}\frac{2\sqrt{2}G_F n_0}{k_B T}\frac{E_0}{k_B T},\end{aligned} \tag{33}$$

and we have introduced the dimensionless variables

$$\Omega_n \equiv \frac{\hbar E_0}{(k_B T)^2}\Omega, \quad K_n \equiv \frac{\hbar c}{\sqrt{2}k_B T}K.$$

The constant Q gives the ratio of the potential energy contribution of the background and the individual neutrino energy to the thermal energy of the background.

Furthermore, $E_0 \approx k_B T$, thus simplifying the expression for Q. The contributions from real and imaginary parts to the dispersion relation are plotted in Figs. 3

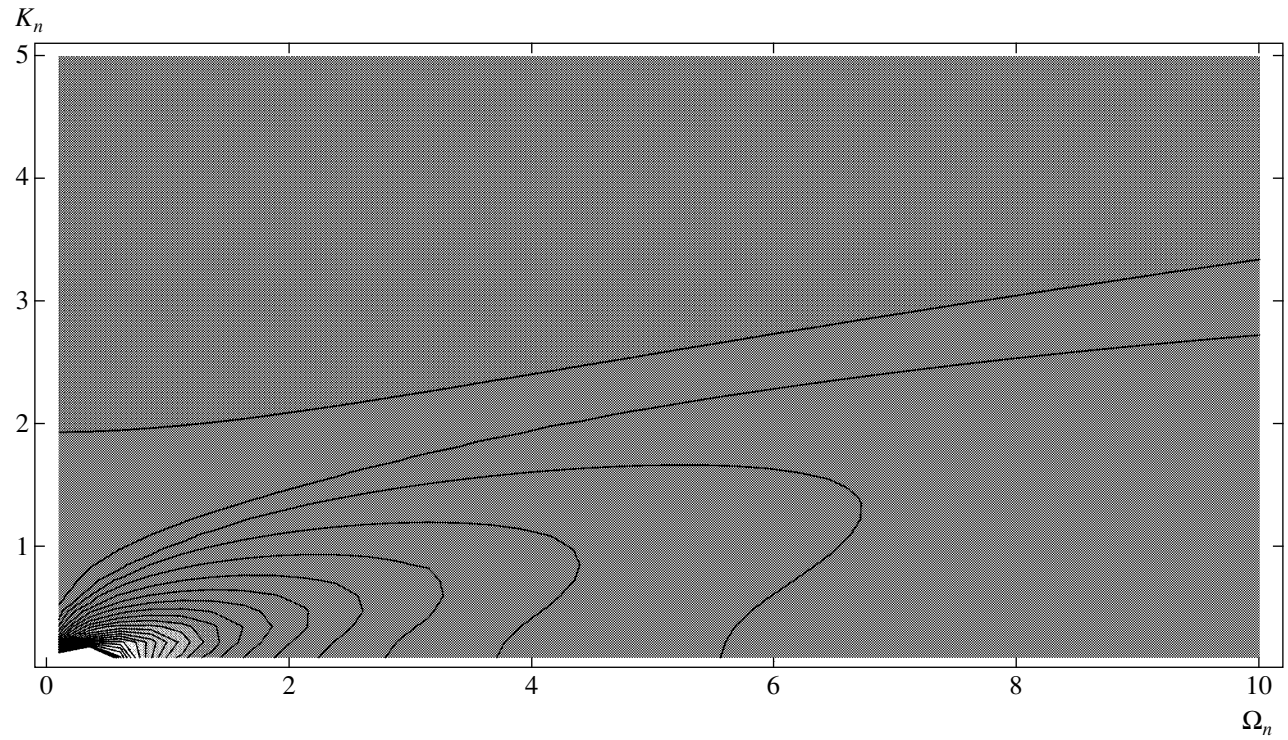


Fig. 3. A contour plot of the Cauchy principal value $P(I(\Omega_n, K_n))$ as a function of the dimensionless variables Ω_n and K_n. We note that $P(I(\Omega_n, K_n)) \geq 0$, being largest for small Ω_n and K_n, and approaching zero at infinity. The uppermost contour has $P(I(\Omega_n, K_n)) = 0$.

and 4, respectively. We note that, for very short length scales, i.e., large K, the quantity $\Omega_n - K_n^2$ becomes negative and the imaginary part in Eq. (31) changes sign, which cannot be seen from long-wavelength limit equation (14). This behavior can in principle lead to growth instead of damping of the perturbations (see [22] for a general discussion of this behavior). We can obtain a quantitative measure of the growth/damping rate as follows. For any fixed K_{n0}, the dimensionless growth/damping rate

$$\Gamma_n = -i\mathrm{Im}\Omega_n$$

can be expressed as (with the value at Ω_{n0} denoted by 0)

$$\Gamma_n = \pi\frac{(Q^{-1} + P(I_0))(\partial g/\partial\Omega_{n0}) - g_0(\partial P(I)/\partial\Omega_{n0})}{\pi^2(\partial g/\partial\Omega_{n0})^2 + (\partial P(I)/\partial\Omega_{n0})^2} \quad (34)$$

to the first order around (Ω_{n0}, K_{n0}). Therefore, $\Gamma_n > 0$ if

$$(Q^{-1} + P(I_0))\left(\frac{\partial \ln g}{\partial\Omega_{n0}}\right) > \frac{\partial P(I)}{\partial\Omega_{n0}}.$$

Moreover, using values given in Section 4, one can show that $Q^{-1} \approx 3 \times 10^9$. Thus, Q^{-1} dominates the contribution to the growth/damping rate over a wide range of (Ω_n, K_n), and a positive growth rate is implied as long as $\partial g/\partial\Omega_{n0} > 0$.

4. APPLICATIONS

As a model for hot dark matter, massive neutrinos have for some time been one of the prime candidates, but as such they have faced the problem of the scale of the inhomogeneities that they can support. Due to the conservation of phase-space density, the Tremaine–Gunn limit constrains the neutrino mass for isothermal spheres of a given size. For dwarf galaxies, for which there is ample evidence of dark matter [23], the necessary mass of the neutrino is uncomfortably large [4, 24]. On the other hand, as was pointed out in [5], interacting dark matter can in principle change this picture. Here, we see from Eq. (26) that, as the neutrino momentum increases, the typical length scale l of the inhomogeneity that can be supported by the modulational instability decreases. From the definition of l_0, we note that, as v tends to c, $l_0 \longrightarrow 0$, and due to Eq. (26), the allowed scale of inhomogeneity becomes squeezed between two small values. On the other hand, if $v \sim v_F$ (a condition stating that the neutrino number density must reach extreme values), the upper inhomogeneity scale limit diverges. A minimum requirement for the effect to be of importance is that the instability growth rate is larger than the Hubble parameter H. An estimate of the growth rate can be obtained as follows. At the onset of "free

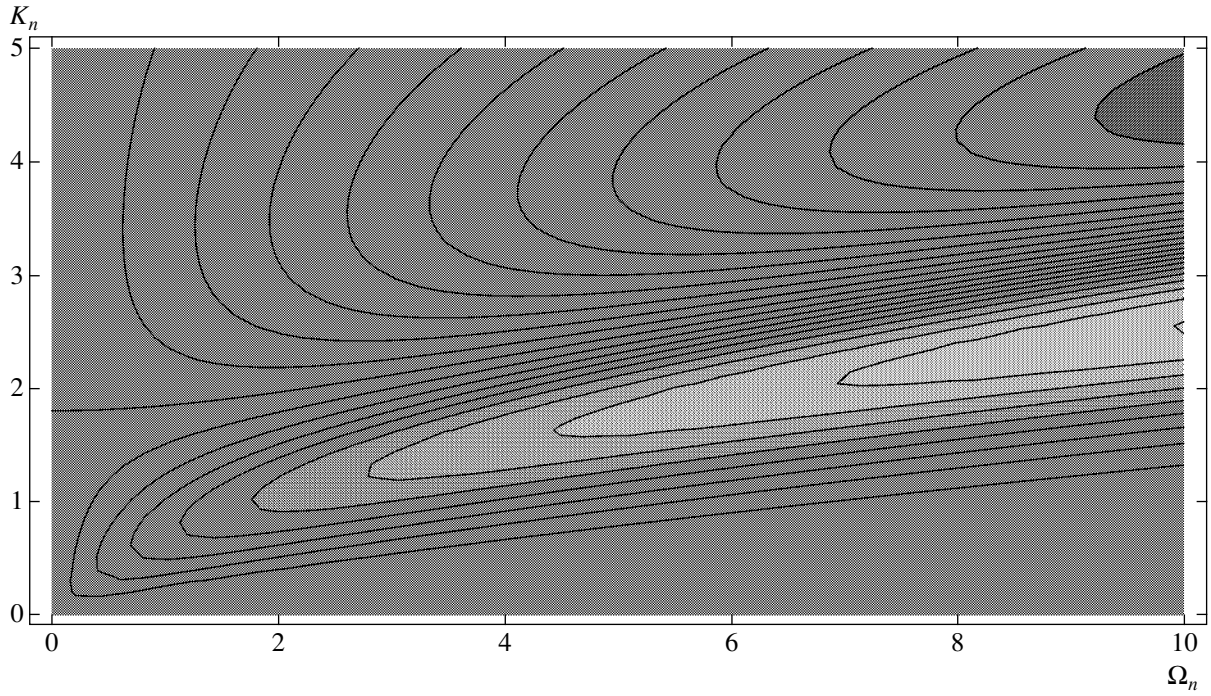


Fig. 4. A contour plot of the contribution $g(\Omega_n, K_n)$ of the poles to integral (30) as a function of the dimensionless variables Ω_n and K_n. The darker areas represent negative values; the lighter, positive values; and g is zero on the contour emanating from $(\Omega_n, K_n) = (0, 1.8)$.

streaming" of neutrinos (i.e., their decoupling from matter and radiation) at $z \sim 10^{10}$, the neutrino number density can be estimated as $n_0 \approx 2.1 \times 10^{38}$ m^{-3} (see, e.g., [25]). Furthermore, we assume that the neutrino mass is in the range $m \sim 1$ eV and find $v_F \approx 9 \times 10^{-4} E_0^{-1/2}$ m/s. The temperature of the neutrinos, given by

$$T_\nu = (4/11)^{1/3} T_0 (1 + z)$$

at neutrino decoupling (with T_0 being the present day CMB temperature) [25], is $T_\nu \approx 2 \times 10^{10}$ K. Thus, the thermal energy is roughly five orders of magnitude greater than the assumed rest mass of the neutrino, and in this sense the neutrinos can be well approximated as ultrarelativistic. In this case, using values of $(\hbar c^2 K/2E_0 v_F)^2$ in the middle range of inequality (25), we find from Eq. (24) that

$$\Gamma \approx \frac{2\sqrt{2} G_F n_0}{\hbar} \sim 16 \times 10^{10} \text{ s}^{-1}$$

for the values specified above. With the a critical density assumed for the Universe, the Hubble time becomes

$$H^{-1} \approx H_0^{-1} (1 + z)^{-3/2} \sim 5 \times 10^2 \text{ s}$$

at the redshift 10^{10}, and therefore $\Gamma/H \gg 1$.

Although the two-stream instability may seem contrived as a cosmological application, the important issue displayed by this example is the nongravitational growth of inhomogeneities, given a small perturbation of a homogeneous, although anisotropic, background. The criticism of neutrinos as dark matter candidates is in particular based on the fact that they are ultrarelativistic for long times, with free-streaming smoothing as a result [4]. According to this, we would have to accept a top-down scenario for structure formation if neutrinos would indeed be the missing dark matter [15]. These arguments are presented with the prerequisite that only gravitational instabilities are of importance after neutrino decoupling. Other instabilities, such as the one presented in this work, could in principle alter this picture, loosening the bounds set by the Tremaine–Gunn limit. The fact that the growth rate exceeds the inverse of the Hubble time by many orders of magnitude makes it clear that the mechanism may be of some importance. Moreover, an analogous estimate for the Fermi–Dirac background, although done in a simplistic manner, indicates that the growth of the large-K perturbations may be of importance. Furthermore, although here we have used parameters relevant to a cosmological setting, it could also be of interest to use the current formalism as a tool to investigate neutrino interactions within supernovas, where the two-stream instability scenario may occur as a more natural ingredient than perhaps within cosmology.

5. CONCLUSIONS

We have considered the nonlinear coupling between neutrinos and antineutrinos in a dense plasma. We have found that their interactions are governed by a system of Wigner–Moyal equations which admit a modulational instability of the neutrino/antineutrino beams against large-scale (in comparison with the neutrino wavelength) density fluctuations. Physically, the instability arises because interpenetrating neutrino and antineutrino beams are like quasiparticles, carrying free energy that can be coupled to inhomogeneities due to a resonant quasiparticle-wave interaction that is similar to the Cherenkov interaction. Nonlinearly excited density fluctuations can be associated with the background inhomogeneity of the early Universe and possibly counteract the free-streaming smoothing of the small-scale primordial fluctuations, thus making massive neutrinos plausible as a candidate for hot dark matter.

ACKNOWLEDGMENTS

This research was partially supported by the Sida/NRF (grant no. SRP-2000-041), the Swedish Research Council through Contract no. 621-2001-2274, and the Deutsche Forschungsgemeinschaft through the Sonderforschungsbereich 591.

REFERENCES

1. Q. R. Ahmad *et al.* (SNO Collab.), Phys. Rev. Lett. **89**, 011301 (2002).
2. A. D. Dolgov, Phys. Rep. **370**, 333 (2002).
3. Y. Sofue and V. C. Rubin, Annu. Rev. Astron. Astrophys. **39**, 137 (2001).
4. S. Tremaine and J. E. Gunn, Phys. Rev. Lett. **42**, 407 (1979).
5. G. G. Raffelt and J. Silk, Phys. Lett. B **192**, 65 (1987).
6. F. Atrio-Barandela and S. Davidson, Phys. Rev. D **55**, 5886 (1997).
7. Y. Fukuda *et al.* (Super-Kamiokande Collab.), Phys. Rev. Lett. **81**, 1562 (1998); Phys. Lett. B **436**, 33 (1998); Phys. Rev. Lett. **82**, 1810 (1999).
8. O. Elgaroy, O. Lahav, W. J. Percival, *et al.*, Phys. Rev. Lett. **89**, 061301 (2002).
9. S. Hannestad, Phys. Rev. D **66**, 125011 (2002); K. N. Abazajian and S. Dodelson, Phys. Rev. Lett. **91**, 041301 (2003).
10. T. K. Kuo and J. Pantaleone, Rev. Mod. Phys. **61**, 937 (1989).
11. L. O. Silva, R. Bingham, J. M. Dawson, *et al.*, Astrophys. J. **127**, 481 (2000).
12. H. A. Weldon, Phys. Rev. D **26**, 2789 (1982).
13. D. Nötzhold and G. Raffelt, Nucl. Phys. B **307**, 924 (1988).
14. H. Nunokawa, V. B. Semikoz, A. Y. Smirnov, and J. W. F. Valle, Nucl. Phys. B **501**, 17 (1997).
15. G. G. Raffelt, *Stars as Laboratories for Fundamental Physics* (Univ. of Chicago Press, Chicago, 1996).
16. N. L. Tsintsadze, J. T. Mendonca, and L. N. Tsintsadze, Phys. Plasmas **5**, 3512 (1998); P. K. Shukla, L. Stenflo, L. N. Tsintsadze, and N. L. Tsintsadze, Phys. Plasmas **9**, 3625 (2002).
17. V. I. Karpman, *Nonlinear Waves in Dispersive Media* (Nauka, Moscow, 1973; Pergamon, Oxford, 1975).
18. A. Hasegawa, *Plasma Instabilities and Nonlinear Effects* (Springer, Berlin, 1975).
19. M. Marklund, P. K. Shukla, and L. Stenflo, Phys. Scr. (2004) (in press).
20. J. Pantaleone, Phys. Lett. B **342**, 250 (1995).
21. E. P. Wigner, Phys. Rev. **40**, 749 (1932); J. E. Moyal, Proc. Cambridge Philos. Soc. **45**, 99 (1949); V. B. Semikoz, Physica A (Amsterdam) **142**, 157 (1987).
22. D. Anderson, L. Helczynski, M. Lisak, and V. Semenov, Phys. Rev. E **69**, 025601 (2004).
23. B. Carr, Annu. Rev. Astron. Astrophys. **32**, 531 (1994).
24. Ya. B. Zel'dovich and R. A. Syunyaev, Sov. Astron. Lett. **6**, 249 (1980); A. G. Doroshkevich, Ya. B. Zel'dovich, R. A. Syunyaev, and M. Yu. Khlopov, Sov. Astron. Lett. **6**, 252 (1980); Sov. Astron. Lett. **6**, 257 (1980); P. J. E. Peebles, Astrophys. J. **258**, 415 (1982); S. D. M. White, C. S. Frenk, and M. Davis, Astrophys. J. Lett. **274**, L1 (1983); G. G. Raffelt, New Astron. Rev. **46**, 699 (2002).
25. P. J. E. Peebles, *Principles of Physical Cosmology* (Princeton Univ. Press, Princeton, 1993).

Journal of Experimental and Theoretical Physics, Vol. 99, No. 1, 2004, pp. 95–100.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 1, 2004, pp. 113–119.

FLUIDS

Wigner Liquid Conductivity in a Parallel Magnetic Field

É. G. Batyev

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

e-mail: batyev@isp.nsc.ru

Received December 2, 2003

Abstract—It is assumed that comparatively low-mobility objects (clusters of a small number of electrons) can appear in a two-dimensional strongly correlated electronic system (Wigner liquid) against the background of mobile Fermi-type carriers. These formations can get "pinned" to inhomogeneities and play the role of additional scatterers. Clusters of two and three electrons are discussed (for a short-range order in the arrangement of electrons, as in a triangular lattice). The number of these clusters depends on both temperature and the parallel magnetic field. This results in the temperature and field dependences of the resistance and magnetization of the system. According to a simple model, resistance increases and the metal–dielectric transition occurs as the parallel magnetic field grows stronger. The model predicts a nonlinear magnetic field dependence of magnetization. © *2004 MAIK "Nauka/Interperiodica".*

One of the interesting features of two-dimensional low-density electronic systems is a strong dependence of resistance on spin polarization induced by a parallel magnetic field (see reviews [1, 2]). Resistance increases severalfold as the magnetic field grows stronger. In addition, judging from the temperature behavior of resistance, the metal–dielectric transition occurs as spin polarization increases. Naturally, the question arises as to the reasons for such behavior.

Generally, various mechanisms of an increase in resistance as spin polarization grows can be suggested. For instance, we would observe such an increase if the effective mass of quasi-particles grew with increasing polarization. Experiment is at variance with this suggestion [2]. Another reason might be an increase in the interaction with impurities. Precisely this mechanism was considered in [3, 4]. A change in the number of scattering centers, as suggested in [5], is also possible. Lastly, the reason can be the formation of a nonuniform (two-phase) state [6].

In this work, we suggest another approach, which follows from the concepts of the properties of a strongly correlated system discussed in [7]. In that work, the exchange interaction of particles in a system of two-dimensional electrons of a low density (and, accordingly, the effective mass of Fermi excitations) was estimated. The model of two particles (nearest neighbors) which were in a common potential well created by the environment and interacted with each other by the Coulomb law was used. Each particle was largely in its own potential minimum, and they rarely switched places. The ensuing exchange splitting of levels $E_A - E_S$, where E_A and E_S are the energies that correspond to antisymmetric Ψ_A and symmetrical Ψ_S coordinate functions, was calculated in [7]. The exchange interaction Hamiltonian H_{ex} responsible for this splitting can be written in the form

$$H_{\text{ex}} = \lambda_S\left\{(\mathbf{S}_1\mathbf{S}_2) + (\mathbf{Q}_1\mathbf{Q}_2) + 4(\mathbf{S}_1\mathbf{S}_2)(\mathbf{Q}_1\mathbf{Q}_2) - \frac{1}{4}\right\}, \quad (1)$$

where indices 1 and 2 number particles, **S** is the spin operator of a particle, **Q** is the quasi-spin operator of a particle (if there are two valleys, as with two-dimensional electrons in the inversion layer of silicon, this operator equals 1/2), and the constant –1/4 in braces is added for convenience. The λ_S value and effective mass m^* are related to the splitting of levels by the equation

$$2\lambda_S = E_A - E_S \sim n/m^*, \quad (2)$$

where n is the concentration of carriers. Note that (1) is symmetrical with respect to the spin and quasi-spin operators. In what follows, we consider the two-valley situation, although (1) can also be used when there is only one valley and no quasi-spin, as for particles within one valley.

Several points should be mentioned in connection with the aforesaid. In a system with a low density of electrons, the hierarchy of energies is as follows:

(1) The highest energy corresponds to Coulomb interaction of electrons and is taken as the energy unit. Precisely Coulomb interaction causes Wigner crystal formation and establishes short-range order in Wigner liquids.

(2) The characteristic energy of plasma oscillations, or the energy of zero-point oscillations, is on the order of $1/\sqrt{r_S}$, where r_S is the dimensionless distance

between electrons ($r_S \gg 1$). Zero-point oscillations prevent electron crystallization up to very large r_S values (according to [8], crystallization occurs at $r_S \approx 37 \pm 5$).

(3) The Fermi energy follows next. This energy is on the order of $1/r_S$, if it is calculated using the unrenormalized (band) mass m as in a gas. In reality, the Fermi energy is still lower, because $m^* > m$. This energy is of the same order as the exchange interaction energy, see (2). This is the energy scale that we consider below.

Initial idea. Seemingly, we must proceed following the Landau theory of Fermi liquids, that is, consider a system of Fermi-type quasi-particles (with a renormalized effective mass) by analogy with the Fermi gas, whose properties, however, depend on the distribution function of the quasi-particles. The question arises if this is sufficient for solving our problem. Indeed, we can imagine a situation when, for instance, the energy (per particle) of a pair of neighboring particles considered in the model mentioned above [7] is lower than the Fermi energy of the quasi-particles. For this reason, a certain number of such pairs could appear against the background of the gas of the quasi-particles. We must then take into account the influence of such pairs on the properties of the system.

We may consider some other units instead of pairs. For short-range order, as in a Wigner crystal (that is, as in a triangular lattice), the three nearest neighbors are the most suitable candidate. Clearly, such formations are more energetically favorable (per particle) than pairs. For this reason, preference should probably be given to sets of three particles.

Considering more complex units in a liquid would hardly make sense, because their formation would require ordered arrangement of not only the nearest but also next-nearest neighbors. For this reason, we will concentrate on pairs and triples of particles.

Clearly, the number of such clusters depends on magnetic field (because they have spins) and temperature. For this reason, the number of mobile Fermi-type carriers (fermions) will also be field- and temperature-dependent. It is natural to assume that such clusters have low mobilities, because the jump amplitude for them is smaller than the jump amplitude for a separate particle (for a set of three particles, this amplitude is exponentially small). This means that they can "stick" even on small inhomogeneities, which prevents them from contributing to the current. They can scatter mobile carriers (fermions), which also influences the conductivity of the system. The description suggested below is based on the picture outlined above.

To a certain degree, all of the above resembles the model used in [5]. But the model used in this work is directly related to the properties of a strongly correlated system and is inherent in it. In this sense, our model is universal and independent of sample properties.

As far as [6] is concerned, note the following. In that work, a nonuniform (two-phase) state of the system was considered and it was mentioned that the corresponding region can be small. Outside this region, there is a uniform state, and the model under consideration can then be applicable (to a liquid). In the two-phase state, our model can be used to describe the liquid phase.

Clusters. Let us begin with a pair. The quantum numbers that correspond to a pair with the lowest energy are ($S = 1$, $Q = 0$) and ($S = 0$, $Q = 1$) (here, S and Q are the total spin and quasi-spin of a system of two particles). Using (1), we can easily show that the lowest pair energy E_2 is

$$E_2/2 = -3\lambda_S/8 \tag{3}$$

(here, the energy is per particle).

Next, consider a cluster of three particles that occupy the vertices of an equilateral triangle. Let each pair of the nearest neighbors interact according to (1). The state with the quantum numbers ($S = 3/2$, $Q = 3/2$) has the highest energy and is of no interest to us (here, S and Q are the total spin and quasi-spin of a system of three particles).

Let us consider states with the quantum numbers ($S = 3/2$, $Q = 1/2$) and ($S = 1/2$, $Q = 3/2$), which have equal energies. This energy will be denoted by $E_3(3/2)$ according to the maximum possible spin. It is given by

$$\frac{1}{3}E_3\left(\frac{3}{2}\right) = -\frac{7}{12}\lambda_S \tag{4}$$

(also per particle). Note that the states with the given quantum numbers S, S_3, Q, and Q_3 are doubly degenerate.

Lastly, consider the state of a cluster of three particles with the quantum numbers ($S = 1/2$, $Q = 1/2$). The energy of this state $E_3(1/2)$ is given by

$$E_3(1/2)/3 = -3\lambda_S/4. \tag{5}$$

This state is not degenerate at the given quantum numbers S, S_3, Q, and Q_3.

Model. We use the simplest model, namely, a gas of mobile carriers (fermions) plus clusters. The energy of the system is

$$E = \sum_{\mathbf{p},\sigma}[\epsilon(\mathbf{p}) + H\sigma]n_\sigma(\mathbf{p}) + \sum_{\Sigma,\nu}\kappa(\Sigma)[\gamma(\Sigma)E_0(\Sigma,\nu) + HS_3]N(\Sigma,\nu). \tag{6}$$

Here, the first sum refers to fermions (the summation is over momentum and spin projection $\sigma = \pm 1/2$, and ϵ and n_σ are the energy and concentration of fermions,

respectively). The second sum is for clusters. We use the notation

$$(S, S_3, Q, Q_3) \longrightarrow \Sigma$$

for the set of quantum numbers. The summation is over these quantum numbers and state numbers ν. The $\gamma(\Sigma)E_0(\Sigma, \nu)$ value is the energy of the corresponding cluster in the absence of a magnetic field (it does not depend on the spin S_3 and quasi-spin Q_3 projections), and $\gamma(\Sigma)$ is the number of particles in a cluster. The $\kappa(\Sigma)$ constant takes into account additional degeneracy of the level [$\kappa = 2$ for states with energy (4), and $\kappa = 1$ for the other states]. Lastly, H is the magnetic field in energy units,

$$g^*\mu_B B \longrightarrow H.$$

Here, g^* is the effective g-factor (it is assumed to be the same for fermions and clusters) and μ_B is the Bohr magneton. Energy (6) can be used to include clusters of various types.

Further, it is assumed that only one cluster can occur in each state ν, and the total number of these states is on the order of the total number of particles. Seemingly, a cluster should be treated as a particle (with a fairly large effective mass) in a liquid and be characterized by a momentum. It, however, makes sense to model a system of clusters by assuming localized states for them because of inhomogeneities.

Equilibrium properties. The characteristics of clusters can be found using the partition function $Z(\nu)$ for a given state,

$$Z(\nu) - 1 = \sum_{\Sigma} \kappa(\Sigma) \exp\left\{-\frac{\gamma(\Sigma)[E_0(\Sigma, \nu) - \mu] + HS_3}{T}\right\}, \quad (7)$$

where μ is the chemical potential of the system. The fermion distribution function is

$$n_\sigma(\mathbf{p}) = \left\{\exp\left[\frac{\epsilon(\mathbf{p}) - \mu_\sigma}{T}\right] + 1\right\}^{-1}, \quad \mu_\sigma \equiv \mu - H\sigma. \quad (8)$$

If the energy of a fermion quadratically depends on its momentum, $\epsilon(\mathbf{p}) = p^2/2m^*$, the concentration of fermions with a specified spin projection in a particular valley is given by

$$n_\sigma = \frac{1}{V}\sum_{\mathbf{p}} n_\sigma(\mathbf{p}) = \frac{m^*T}{2\pi}\ln\left\{1 + \exp\left(\frac{\mu_\sigma}{T}\right)\right\}. \quad (9)$$

As usual, the chemical potential is determined by setting the number of particles (the number of fermions plus the number of electrons in clusters). For two valleys, this condition takes the form

$$n = n_F + n_\Gamma, \quad n_F = 2\sum_\sigma n_\sigma,$$
$$n_\Gamma = \sum_{\Sigma, \nu} \frac{-T}{Z(\nu)}\frac{\delta Z(\nu)}{\delta E_0(\Sigma, \nu)}. \quad (10)$$

Here, n is the concentration of electrons, n_F is the concentration of mobile carriers (fermions), and n_Γ is the concentration of electrons in clusters [expressed the functional derivative of (7)].

Up to this point, we wrote equations in the general form taking into account clusters containing various numbers of electrons. Further, we restrict our consideration to clusters of three electrons, for which lower energies were obtained than for electron pairs [see (3)–(5)]. For these clusters, we will take into account not only states with spin 1/2 [see (5)] but also states (4). Although they are higher in energy in the absence of a magnetic field, the contribution of states with spin 3/2 can predominate in very strong magnetic fields. In addition, we ignore the spread of cluster energies for simplicity; that is, we assume that E_0 is independent of state number ν. Instead of the notation $E_0(\Sigma)$, for which a set of quantum numbers should be specified every time, the notation

$$E_0(1/2, 1/2) \longrightarrow \epsilon_0,$$
$$E_0(1/2, 3/2) = E_0(3/2, 1/2) \longrightarrow \epsilon_1 \quad (11)$$

can more conveniently be used (the arguments correspond to the spin and quasi-spin values). As follows from (4) and (5), $\epsilon_0 < \epsilon_1$.

Bearing this in mind, partition function (7) can be written as

$$Z = 1 + 4\exp\left[-\frac{3(\epsilon_0 - \mu)}{T}\right]\cosh\frac{H}{2T}$$
$$+ 8\exp\left[-\frac{3(\epsilon_1 - \mu)}{T}\right]\left\{\cosh\frac{3H}{2T} + 3\cosh\frac{H}{2T}\right\}. \quad (12)$$

The concentration of electrons in clusters is

$$n_\Gamma = 3n_0\frac{Z-1}{Z}, \quad (13)$$

where n_0 is the concentration of cluster states.

Next, we find the spin moment M as a function of magnetic field. The ratio between M and its maximum

possible value $M_m = n/2$ can be written in the form

$$\frac{M}{M_m} = \eta\frac{2T}{3Z}\left(\frac{\partial Z}{\partial H}\right)_\mu + \frac{T}{2\epsilon_F}\left\{\frac{H}{2T} + \ln\frac{\cosh[(\mu + H/2)/(2T)]}{\cosh[(\mu - H/2)/(2T)]}\right\}. \quad (14)$$

Here, the first term on the right-hand side is the contribution of clusters [see (12)], and the second term is the contribution of fermions. The constant η corresponds to the maximum fraction of electrons which can be placed into cluster states [it appears that η should be assumed to be smaller than (on the order of) one], and ϵ_F is the Fermi energy of the two-valley system at $T = 0$ in the absence of clusters,

$$\eta = \frac{3n_0}{n}, \quad \epsilon_F = \frac{\pi n}{2m^*}. \quad (15)$$

Zero temperature. Suppose that $\epsilon_0 < \epsilon_F$. In a magnetic field at zero temperature, we must only include minimum-energy cluster levels, that is,

$$\epsilon_0 \longrightarrow \epsilon_0 - \frac{H/2}{3}, \quad \epsilon_1 \longrightarrow \epsilon_1 - \frac{H}{2}.$$

The shift of the energy of fermions is $\delta\epsilon_{\uparrow,\downarrow} = \pm H/2$. As a result, the concentration of fermions (with various spin projections) and electrons in clusters (with the most favorable spin projection) in the interval $0 < H < 3\epsilon_0/2$ is given by

$$n_\uparrow = \frac{m^*}{\pi}\left(\epsilon_0 - \frac{2H}{3}\right), \quad n_\downarrow = \frac{m^*}{\pi}\left(\epsilon_0 + \frac{H}{3}\right), \quad n_\Gamma = n - \frac{m^*}{\pi}\left(2\epsilon_0 - \frac{H}{3}\right). \quad (16)$$

In high fields,

$$3\epsilon_0/2 < H < 3(2\epsilon_F - \epsilon_0), \quad (17)$$

cluster levels are depleted even to zero and the $n_\downarrow$ value increases to n (by the linear law). We did not include the ϵ_1 level thus far; that is, we assumed that $\epsilon_1 > 2\epsilon_F$.

Next, suppose that $\epsilon_0 < \epsilon_1 < 2\epsilon_F$. The ϵ_1 level comes into play in magnetic fields $H/3 > \epsilon_1 - \epsilon_0$. Two situations can arise. First, this level can become involved when $n_\uparrow = 0$; that is, when the magnetic field falls within interval (17). Cluster levels are then depleted, but not to zero [they are not at all depleted at the left boundary of interval (17), that is, at $\epsilon_1 = 3\epsilon_0/2$]. Secondly, this level may become involved at $n_\uparrow > 0$ ($\epsilon_1 < 3\epsilon_0/2$). Consider this situation in more detail.

Equation (16) is, as previously, valid in the magnetic field range $0 < H < 3(\epsilon_1 - \epsilon_0)$. At $H/3 = \epsilon_1 - \epsilon_0$, cluster states with the initial energy ϵ_1 rather than ϵ_0 are filled (this occurs in a jump if the spread of cluster energies is ignored). In the magnetic field range $3(\epsilon_1 - \epsilon_0) < H < \epsilon_1$, we have

$$n_\uparrow = \frac{m^*}{\pi}(\epsilon_1 - H), \quad n_\downarrow = \frac{m^*}{\pi}\epsilon_1, \quad n_\Gamma = n - \frac{m^*}{\pi}(2\epsilon_1 - H). \quad (18)$$

This situation persists until the field reaches the value $H = \epsilon_1$, when the state of complete polarization ($n_\uparrow = 0$) is attained and the $n_\downarrow$ and n_Γ values reach saturation. We obtain curves with kinks for all values including resistance, and even a curve with jumps for magnetic moment. Naturally, the picture becomes smoothed as the temperature increases.

Note that the above equations are valid while the number of electrons in clusters that we obtain is smaller than the maximum possible number; that is, $n_\Gamma < 3n_0$. This requirement is satisfied if the condition

$$1 - \frac{\epsilon_1}{2\epsilon_F} < \eta$$

is met. Otherwise, the equations are valid unless the cluster levels are filled completely. For example, if this condition holds for $3(\epsilon_1 - \epsilon_0) < H$, then the corresponding magnetic field is

$$\frac{H_0}{2\epsilon_F} = \eta + \frac{\epsilon_1}{\epsilon_F} - 1.$$

Now, let us turn to the magnetic moment. Consider the case of $\epsilon_0 < \epsilon_F$, $\epsilon_0 < \epsilon_1 < 3\epsilon_0/2$ [see (16) and (18) for two regions of field variations]. We find that

$$\frac{M}{M_m} = \frac{5H}{9\epsilon_F} + \frac{1}{3}\left(1 - \frac{\epsilon_0}{\epsilon_F}\right), \quad 0 < H < 3(\epsilon_1 - \epsilon_0). \quad (19)$$

These are low fields. In high fields up to complete polarization, we have

$$\frac{M}{M_m} = 1 + \frac{H - \epsilon_1}{\epsilon_F}, \quad 3(\epsilon_1 - \epsilon_0) < H < \epsilon_1. \quad (20)$$

Magnetic moment jumps occur at $H = 0$ and $H = 3(\epsilon_1 - \epsilon_0)$. (Clearly, these jumps become smoothed at finite temperatures and because of a probable spread of cluster levels.) The ratio between the susceptibilities in the low and high fields is

$$\frac{\chi_1}{\chi_2} = \frac{5}{9}.$$

Note that complete polarization then occurs in magnetic field $H = \epsilon_1 < 3\epsilon_0/2 < 3\epsilon_F/2$, whereas, in the absence of clusters, complete polarization occurs at $H = 2\epsilon_F$.

In relation to the magnetic moment jump near $H = 0$, note the following. Direct magnetic moment measurements in a parallel magnetic field were performed

in [9]. It was found that magnetization exhibited nonlinear behavior, and the spin susceptibility grew as the magnetic field decreased. This circumstance was interpreted as a contribution of localized spins. Possibly, it was a manifestation of clusters discussed in this work.

Resistance. Specific resistance ρ will be described by the usual equation

$$\rho = \frac{m}{n_F e^2 \tau}, \quad (21)$$

where τ is the relaxation time. The question is what should be taken as τ. If interaction with impurities is weak (it is certainly not weak for clusters) or the amount of impurities is low, the relaxation time is proportional to their concentration; that is, for scattering by clusters, we have $1/\tau \sim n_\Gamma$. This is the main thing for fairly pure samples. Possibly, we must take into account other scatterers whose concentration is n_i, and the above approximate equality can then in the simplest case be replaced by

$$1/\tau \propto n_\Gamma + \alpha n_i, \quad (22)$$

where α takes into account the difference between scatterers. Using (13) and (10) in (21) then yields

$$\rho \propto \frac{\eta(Z-1) + AZ}{(1-\eta)Z + \eta}, \quad A \equiv \alpha \frac{n_i}{n}. \quad (23)$$

The contribution with the coefficient A in (14) is essential in the absence of clusters (if $Z = 1$) and corresponds to what can be called residual resistance. This situation can arise at zero temperature in the absence of a magnetic field if $\epsilon_0 > \epsilon_F$. However, if clusters are always present, it can be assumed that $A \longrightarrow 0$ for fairly pure samples.

Consider the behavior of resistance at low temperatures $T \ll \epsilon_F$. If $H = 0$, (10) and (13) give an equation for determining the chemical potential, namely,

$$n - \frac{2m^*\mu}{\pi} = \frac{3n_0}{(1/4)\exp[3(\epsilon_0 - \mu)/T] + 1} \quad (24)$$

(cluster states with the energy $\epsilon_1 > \epsilon_0$ are ignored).

Let us write the chemical potential in the form

$$\mu = \epsilon_0 + \mu_1.$$

Equation (24) then becomes

$$1 - \frac{\epsilon_0}{\epsilon_F} = \frac{\mu_1}{\epsilon_F} + \frac{\eta}{(1/4)\exp(-3\mu_1/T) + 1}. \quad (25)$$

If cluster states are partially filled at absolute zero ($\epsilon_0 < \epsilon_F$), the chemical potential behaves as $\mu \longrightarrow \epsilon_0$ as T tends to zero. At certain parameter values, (24) has a zero solution, that is,

$$1 - \frac{\epsilon_0}{\epsilon_F} = \frac{4\eta}{5}, \quad \mu_1 = 0. \quad (26)$$

The number of clusters then remains constant. If (22) is valid, the resistance is also constant; that is, this is a condition of a separatrix.

If parameter changes are small, we can ignore the first term on the right-hand side of (25), which is small compared with the left-hand side. This approximation yields

$$\frac{3\mu_1}{T} \approx -\ln\left(5\frac{\beta_R}{\beta_L} - 4\right), \quad (27)$$

where the left- and right-hand sides of (26) are for convenience denoted by β_L and β_R, respectively. It follows that, if $\beta_R/\beta_L > 1$ ($\mu_1 < 0$), the number of fermions decreases as the temperature increases, whereas the number of clusters (scattering centers) grows; that is, the resistance increases. This corresponds to the metallic phase. Otherwise, if $\beta_R/\beta_L < 1$ ($\mu_1 > 0$), the resistance decreases as the temperature increases, which corresponds to the dielectric phase. In both cases, the resistance changes linearly at low temperatures.

The temperature behavior of the resistance depends on the magnetic field value. Consider the magnetic field interval specified for (18). It is then sufficient to include cluster states with spin $S = 3/2$ and quasi-spin $Q = 1/2$. We obtain (24)–(26) with the replacement

$$\epsilon_0 \longrightarrow \epsilon_1 - H/2.$$

For instance, (26) becomes

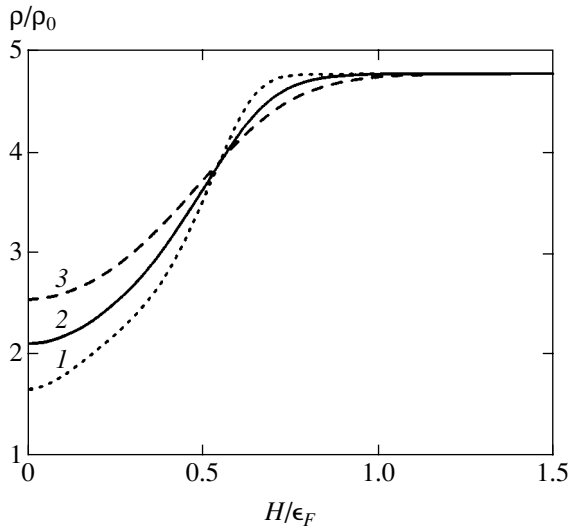
$$1 - \frac{\epsilon_1 - H/2}{\epsilon_F} = \frac{4\eta}{5}, \quad \mu_1 = 0. \quad (28)$$

As $H < \epsilon_1$, the condition for the appearance of a separatrix takes the form

$$\frac{\epsilon_1}{2\epsilon_F} < 1 - \frac{4\eta}{5}.$$

In case of deviations from condition (28), we have an equation for μ_1 that coincides with (27), except that the right- and left-hand sides of (25) should be understood as the corresponding terms of (28). The left-hand side increases as the magnetic field grows, whereas the right-hand side remains constant. It therefore follows from (27) that $\mu_1 < 0$ in lower fields and the number of mobile carriers decreases, whereas the number of scattering centers increases as the temperature rises; that is, the resistance increases with temperature, as is characteristic of metallic phases. In high fields, everything is the other way around; that is, we have a dielectric phase. It follows that, judging from the temperature dependence of the resistance, the metal–dielectric transition can occur as spin polarization increases. Such resistance behavior is in qualitative agreement with experiment.

The reason for the appearance of a separatrix (and the metal–dielectric transition) is well understood within the framework of the model under consider-



Magnetic field dependence of resistance at various temperatures: (*1*) $T/\epsilon_F = 0.05$, (*2*) $T/\epsilon_F = 0.10$, and (*3*) $T/\epsilon_F = 0.15$.

ation. The filling of cluster states increases as the magnetic field grows. If the degree of filling is low at zero temperature, it increases with temperature, and if the degree of filling cluster states is large, it decreases as temperature grows. In a certain intermediate situation (in a certain magnetic field), the filling of cluster states does not change with temperature; that is, the resistance remains constant, as is characteristic of the metal–dielectric transition point.

We did not compare our results with actual experiments. By way of example, consider resistance curves obtained taking into account scattering by clusters only. The magnetic field dependences of the resistance (ρ_0 is the resistance at $T = H = 0$) at various temperatures are shown in the figure. The curves were obtained using (23) and the parameters $\epsilon_0/\epsilon_F = 0.95$, $\epsilon_1/\epsilon_F = 1.1$, $\eta = 0.2$, and $A = 0$. The point at which the curves intersect corresponds to the separatrix. Here, all cluster states can be filled, which occurs in magnetic fields $H > H_0 = 0.6\epsilon_F$ (at $T = 0$), while complete spin polarization is still absent. The resistance, however, already reaches saturation in such fields [the equation for H_0 is given between (18) and (19)]. The maximum resistance value is $(\rho/\rho_0)_{max} = 4.75$.

In the example considered above, the number of mobile carriers (fermions) decreases as the magnetic field increases (by up to 16% at absolute zero and to a lesser degree at $T \neq 0$, for instance, approximately by 8.5% at $T/\epsilon_F = 0.15$). Accordingly, the Hall coefficient should increase. The experiment performed in [10] (also see [11]), however, shows that the Hall coefficient is to a high accuracy constant over the whole range of parallel magnetic field variations even to complete spin polarization. It is unclear thus far whether the constant value of the Hall coefficient is a universal property or the property of the samples studied in [10]. Unfortunately, the data obtained in [10] are insufficient (for instance, it is unknown whether or not a separatrix is observed) to determine what the model gives for this experiment.

Conclusions about an increase in the resistance as the magnetic field increases follow from the model that we use, which includes cluster formation from three electrons (the picture remains qualitatively unchanged also for clusters of two electrons). The validity of (21) and (22) is assumed. These equations are valid if the interaction of mobile carriers with clusters can be considered weak. In reality, the interaction of Fermi quasiparticles with clusters is not weak. It is on the order of one in dimensionless units (the Fermi energy is taken as the energy unit, and the mean distance between the particles, as the length unit). Equation (22) can therefore be written in the limit of a small number of clusters, $n_\Gamma \ll n$; qualitatively, this equation is valid up to $n_\Gamma \sim n$. At certain n_Γ values (on the order of n), this simple picture can, however, fail us, because we must then take into account more complex phenomena, for instance, the localization of fermions.

ACKNOWLEDGMENTS

The author thanks A.V. Chaplik, M.V. Éntin, and Z.D. Kvon for discussions. This work was financially supported by the Russian Foundation for Basic Research (project no. 02-02-16159), INTAS (grant no. 2212), the Council under President of Russian Federation for Support of Scientific Schools, and the program of the Ministry of Industry and Science of Russian Federation.

REFERENCES

1. E. Abrahams, S. V. Kravchenko, and M. P. Sarachik, Rev. Mod. Phys. **73**, 251 (2001).
2. S. V. Kravchenko and M. P. Sarachik, Rep. Prog. Phys. **67**, 1 (2004).
3. V. T. Dolgopolov and A. Gold, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 42 (2000) [JETP Lett. **71**, 27 (2000)].
4. G. Zala, B. N. Narozhny, and I. L. Aleiner, Phys. Rev. B **65**, 020201 (2002).
5. B. L. Altshuler and D. L. Maslov, Phys. Rev. Lett. **82**, 145 (1999).
6. B. Spivak, Phys. Rev. B **67**, 125205 (2003).
7. É. G. Batyev, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 840 (2002) [JETP Lett. **76**, 711 (2002)].
8. B. Tanatar and D. M. Ceperly, Phys. Rev. B **39**, 5005 (1989).
9. O. Prus, Y. Yaish, M. Reznikov, *et al.*, Phys. Rev. B **67**, 205407 (2003).
10. S. A. Vitkalov, H. Zheng, K. M. Mertes, *et al.*, Phys. Rev. B **63**, 193304 (2001).
11. S. A. Vitkalov, Phys. Rev. B **64**, 195336 (2001).

Translated by V. Sipachev