Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 1–9.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 5–15.

ATOMS, SPECTRA, RADIATION

Selective Multiphoton IR Dissociation of SF_6 Molecules under Nonequilibrium Conditions of a Pulsed Gasodynamically Cooled Molecular Flow Interacting with a Solid Surface

G. N. Makarov[a, *] and A. N. Petin[b]

[a]*Institute of Spectroscopy, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia*
[b]*Troitsk Institute of Innovation and Thermonuclear Investigations, Troitsk, Moscow oblast, 142092 Russia*
**e-mail: g.makarov@isan.troitsk.ru*
Received May 16, 2000

Abstract—The process of the isotope-selective multiphoton IR dissociation of SF_6 molecules under the nonequilibrium conditions of a pulsed gasodynamically cooled molecular flow interacting with a solid surface was experimentally studied. The SF_6 molecules dissociate as a result of excitation in a shock wave generated in the flow, in the flow incident onto the sold surface, and in an unperturbed flow (in the absence of the solid). The experiment was based on detecting the luminescence from HF* molecules ($\lambda \approx 2.5$ μm) accompanying the SF_6 dissociation in the presence of H_2 or CH_4, the emission intensity being a measure of the SF_6 dissociation yield. The molecular beam parameters were studied. The time-of-flight spectra of SF_6 in the flow interacting with the surface were measured under various experimental conditions. The spectral and energy characteristics of the SF_6 dissociation process were determined in the flow interacting with the solid surface and in the unperturbed flow. The dissociation product (SF_4) yield was measured and the coefficient of its enrichment with the ^{34}S isotope was determined. It is demonstrated that, using the shock wave formation, it is possible to increase the efficiency of the isotope-selective dissociation of SF_6 molecules. An explanation of the observed results is proposed. The gas density and temperature in the incident flow and in the shock wave were estimated. The results are analyzed and compared to the other published data on the SF_6 dissociation in a molecular beam. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The isotope-selective dissociation of molecules under the action of high-intensity laser radiation has been studied in sufficient detail (see, e.g., [1–4]). The main factors affecting the process selectivity and the dissociation yield are understood. Based on this process, an experimental isotope separation plant was created (Kaliningrad, Russia) that separates carbon isotopes using the method of the selective dissociation of CF_2HCl (Freon-22) molecules.

Experiments on the isotope-selective dissociation of SF_6 molecules were performed for the most part under static conditions in a cell at room temperature [5–7] or at low temperatures ($T \approx 190$ [8], 175 [9], 140 K [7]). The isotope-selective dissociation of SF_6 was also studied under nonequilibrium conditions in a molecular beam [10–12] or in a gasodynamically cooled jet [13] and flow [14]. The stronger gas cooling in the jet or flow in the latter case leads to a significant narrowing of the IR absorption bands of molecules, which increases the selectivity of their excitation and dissociation [13–15]. However, the jets and flows are characterized by low efficiency of the photochemical processes. Because of a small concentration of molecules and the low temperature of the gas, the rates of chemical reactions, including those leading to the target products of interest, are rather small. In some cases, a considerable fraction of radicals formed at a low concentration of molecules in the flow is lost on the walls, not forming products (e.g., during the dissociation of CF_3I [16–18]).

A more favorable situation can be realized during the interaction of a pulsed gasodynamically cooled supersonic molecular flow interacting with a solid surface. A shock wave formed in front of the surface in this system is characterized by nonequilibrium conditions which can be inverse to those in the incident flow: $T_{1,\,\mathrm{tr}} \le T_{1,\,\mathrm{rot}} \le T_{1,\,\mathrm{vib}}$ in the incident flow versus $T_{2,\,\mathrm{tr}} \ge T_{2,\,\mathrm{rot}} \ge T_{2,\,\mathrm{vib}}$ in the shock wave, where $T_{i,\,\mathrm{tr}}$, $T_{i,\,\mathrm{rot}}$, and $T_{i,\,\mathrm{vib}}$ ($i = 1, 2$) are the translational, rotational, and vibrational temperatures of molecules. The shock wave formation is accompanied by a significant increase in the dissociation yield of molecules excited in the incident flow and provides for the modified nonequilibrium conditions of the selective photochemical processes in the shock wave, offering an interesting object for investigation.

Recently [19] we discovered a considerable (severalfold) increase in the yield of dissociation products (at a virtually unchanged selectivity of the process) for the molecules (SF_6, CF_3I) excited in a pulsed flow incident onto a solid surface. In [20, 21] we reported the first

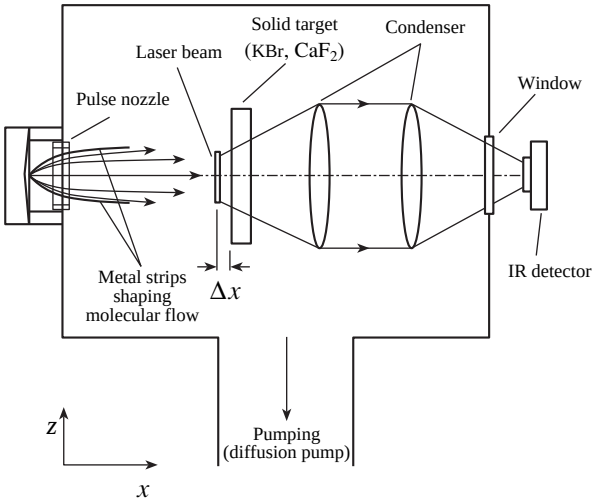

Fig. 1. A schematic diagram of the experimental setup (cross section in the *xz* plane; the laser beam is directed along the *y*-axis).

experimental data on the isotope-selective dissociation of SF_6 in the shock wave formed in this system. This study is devoted to a detailed investigation of the isotope-selective dissociation of SF_6 in a pulsed gasodynamically cooled molecular flow interacting with a solid surface.

2. NONEQUILIBRIUM CONDITIONS IN THE INCIDENT MOLECULAR FLOW AND IN THE SHOCK WAVE

During the rapid cooling of a gas expanded on leaving a nozzle, the gasdynamic equilibrium between various degrees of freedom in the gas molecules is violated. The resulting difference from the state of local equilibrium depends on the number of collisions z_{col} necessary for the relaxation of a given degree of freedom. Polyatomic molecules usually obey the relationship $z_{tr} \le z_{rot} \le z_{vib}$. Therefore, the corresponding effective temperatures in the flow obey the condition [22]

$$T_{1,\,tr} \le T_{1,\,rot} \le T_{1,\,vib}. \quad (1)$$

The interaction of a pulsed molecular flow with a solid surface leads to the formation of a shock wave [23–25] because of a difference between the translational, rotational, and vibrational relaxation rates [26], which may lead to nonequilibrium conditions that are "inverse" relative to those (1) in the incident flow:

$$T_{2,\,tr} \ge T_{2,\,rot} \ge T_{2,\,vib}. \quad (2)$$

In view of a large vibrational-translational relaxation time (e.g., in SF_6 the corresponding relaxation rate constant is $p\tau_{vib-tr} \approx 150$ μs Torr [27]), the vibrational temperature of molecules in the shock wave formed in a pulsed rarefied gas flow may be virtually the same as that in the incident flow ($T_{2,\,vib} \approx T_{1,\,vib}$), while the translational and rotational temperatures are different ($T_{2,\,tr} > T_{1,\,tr}$;$T_{2,\,rot} > T_{1,\,rot}$). Thus, the shock wave features modified nonequilibrium conditions characterized by a vibrational temperature of molecules markedly lower than their translational and rotational temperatures. We have studied the selective dissociation of SF_6 under these conditions.

3. EXPERIMENTAL SETUP AND METHOD

Figure 1 shows a schematic diagram of the experimental setup. The flow of molecules was formed in a pulsed nozzle of the "stream loop" type [28]. The nozzle entrance hole diameter was 0.75 mm and the nozzle open pulse time duration (at halfheight) was 100 μs. The gas pressure at the nozzle entrance could be varied in a range from 0.1 to 3.5 atm. The nozzle output channel had a cone shape with an apex angle of 60° and a height of 15 mm. The total number of molecules N_{fl} in the flow emitted from the nozzle per pulse depended on the input gas pressure and was varied in our experiments from 5×10^{15} to 1.5×10^{17} mol/pulse. The nozzle could operate both in the regime of single pulses and in the repeated pulse mode at a pulse repetition rate of up to 1 Hz.

The molecular flow was formed in a vacuum chamber (with a volume of $V_{ch} \approx 20$ l) evacuated to a pressure of $(1–2) \times 10^{-6}$ Torr by an oil diffusion pump (ensuring a pumping rate of 500 l/s). The molecular flow was formed with the aid of two thin metal strips fixed at the nozzle output cone so as to form a dihedral angle with the edge parallel to the *y* axis and a variable curvature radius in the *xz* plane (Fig. 1).

Spaced by $x \approx 50–150$ mm from the nozzle, a solid target (KBr, CaF_2, or LiF crystal) was placed behind the nozzle and oriented so that the surface was perpendicular to the gas flow. The interaction of the supersonic pulsed molecular flow with the solid surface led to the formation of a shock wave in front of the surface [23–25] with significantly inhomogeneous, nonstationary, and nonequilibrium conditions established in this region. A characteristic size of the shock wave front, equal (by an order of magnitude) to the mean free path of molecules [23, 24], under our experimental conditions was 0.2–5 mm. At a small distance Δx between the solid surface and the laser beam (excitation zone), the molecules could be excited (depending on the time delay between the nozzle open pulse and the exciting laser pulse) either in the incident flow or in the shock wave (see, e.g., Fig. 3).

The molecules were excited by the radiation of a tunable TEA CO_2 laser. The laser radiation pulse represented a leading front peak (with an ~100 ns width at halfheight) and a trailing front of an ~0.5 μs duration, the pulse energy being approximately equally distributed between the leading and trailing parts. The total pulse energy reached up to 3 J. The molecules were excited at a distance of $\Delta x = 1.5–8$ mm from the solid surface. The laser radiation was focused into this region

with the aid of a cylindrical lens with a focal distance of 12 cm. The lens axis was parallel to the solid surface. The laser beam cross section at the focal spot was 0.18×12.5 mm^2.

The nozzle, the CO_2 laser, and the lock-in registration system including an HF* emission detector (or a pyroelectric detector with amplifier) and an S9-8 oscillograph were switched by a delay pulse generator based on a GI-1 oscillator. The synchronism between the laser pulses and the pulsed molecular flow was monitored with the aid of the pyroelectric detector measuring a signal induced by the vibrationally excited SF_6 molecules [29, 30] or by measuring the HF* luminescence signal.

The dissociation of molecules in the gas flow was studied by detecting the luminescence from HF* molecules ($\lambda \approx 2.5$ μm). The vibrationally excited HF* molecules are formed in the reaction between fluorine atoms (the primary product of the dissociation of the SF_6 molecule) and hydrogen or methane [31]. The HF* luminescence intensity is well correlated with the SF_6 dissociation yield [13, 32]. The luminescence was measured using a PbS-based IR detector with a working sensor area of 1×1 cm^2. The transmission bandwidth of the receiver with amplifier (×100) was about 16 kHz. The spectral composition of the IR luminescence was determined with the aid of color light filters. A weak IR emission in the region of 2.5 μm was also observed from SF_6 molecules excited in the absence of H_2 or CH_4. This is probably related to the presence of a small amount of residual hydrocarbon impurities in the initial gaseous SF_6, which was purified by condensation and pumping prior to experiments. Upon adding H_2 or CH_4, the luminescence intensity increased more than ten times. Most of the data presented below refer to SF_6 excited in a mixture with H_2 or CH_4. In additional experiments, however, we have studied the dissociation of SF_6 in the flow containing either a little amount of hydrogen or methane or none at all. These experiments were performed in order to compare the results with other data reported for the dissociation of SF_6 in a molecular beam without a carrier.

We have also measured the yield of SF_4 (another dissociation product) and the coefficient of its enrichment with ^{34}S isotope. A procedure used for collecting the dissociation products and for the IR analysis of these products and the residual gas phase upon laser excitation of the gasdynamic flow was described in detail elsewhere [14, 33]. The coefficient of the SF_4 enrichment with the ^{34}S isotope was determined as

$$K_{34}^{\text{prod}} = \frac{1}{\zeta}\frac{[^{34}SF_4]}{[^{32}SF_4]}, \qquad (3)$$

where $[^{34}SF_4]/[^{32}SF_4]$ is the ratio of the concentrations of the SF_4 molecules with the corresponding sulfur isotopes and $\zeta = {}^{34}S/{}^{32}S \approx 0.044$ is the sulfur isotope ratio in the initial SF_6 gas. The ratio of the concentrations of the $^{34}SF_4$ and $^{32}SF_4$ molecules in the products was determined using the IR absorption intensity measured in the region of ν_6 vibrations ($\nu_6 \approx 728$ cm^{-1} for $^{32}SF_4$ [34]) for which the isotope shift between $^{34}SF_4$ and $^{32}SF_4$ amounts to approximately 12.3 cm^{-1} [35].

4. RESULTS AND DISCUSSION

4.1. Determination of the Molecular Flow Parameters

The major parameters of a pulsed molecular flow, affecting the formation and characteristics of the shock wave, are the flow pulse duration and velocity, the gas density (concentration), and the translational, rotational, and vibrational temperatures of molecules in the flow.

The pulse duration and average velocity of the flow of SF_6 molecules were determined with the aid of a pyroelectric detector using a time-of-flight (TOF) technique described in [29, 30]. The average velocity of the flow of SF_6 molecules in the excitation zone (i.e., at a distance of $x \geq 50$ mm from the nozzle exit) was $v_x = 420 \pm 20$ m/s. The TOF spectra of molecules were also studied by detecting the HF* luminescence (see Section 4.3).

The temperature of molecules in the flow was not studied in these experiments. However, since the flow was created by means of a pulsed nozzle similar to that described in [30], we may expect that the translational, rotational, and vibrational temperatures of molecules in the flow at the same pressures are comparable to those reported in the work cited.

The concentration N_1 of SF_6 molecules in the flow was estimated using the total number of molecules N_{fl} emitted from the nozzle per pulse and the calculated flow volume V_{fl} ($N_1 \approx N_{\text{fl}}/V_{\text{fl}}$). The N_{fl} value was determined from the pressure increment Δp in the vacuum chamber measured upon emitting n gas pulses from the nozzle in the absence of pumping ($T \approx 300$ K):

$$N_{\text{fl}} = \frac{(p_{\text{fin}} - p_{\text{in}})V_{\text{ch}}}{nkT} = \frac{\Delta p V_{\text{ch}}}{nkT}. \qquad (4)$$

Alternatively, the N_{fl} value was determined from the IR absorption spectrum of SF_6 molecules collected from the vacuum chamber into a cell upon emitting n gas pulses from the nozzle. The estimated and measured N_{fl} values showed good agreement. The flow volume V_{fl} was estimated from its geometry, The flow cross section at a distance of $x \approx 50$ mm from the nozzle was $S_{\text{fl}} \approx 7.2$ cm^2 and the flow pulse length was $l_{\text{fl}} \approx 4.2$ cm, which yields $V_{\text{fl}} = S_{\text{fl}} l_{\text{fl}} \approx 30$ cm^3. Data on the SF_6 molecular flow parameters are summarized in Table 1.

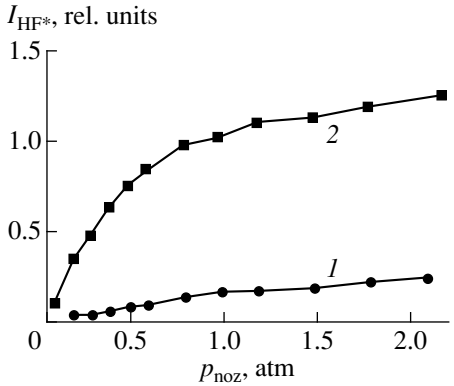

Fig. 2. The plots of HF* luminescence intensity versus gas pressure at the nozzle entrance for SF_6 molecules laser-excited (*1*) in the unperturbed flow and (*2*) in the flow incident onto a solid surface (see the text for comments).

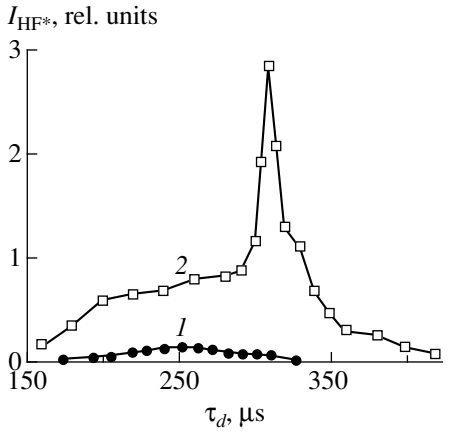

Fig. 3. The plots of HF* luminescence intensity versus delay time τ_d between nozzle pulse and laser pulse for SF_6 molecules excited in the presence of CH_4 ($p_{SF_6}/p_{CH_4} = 1:1$) (*1*) in the unperturbed flow and (*2*) in the flow incident onto a solid surface. Total gas pressure at the nozzle entrance, $p_{noz} = 2.4$ atm.

4.2. SF_6 Dissociation Yield in the Flow Incident onto the Solid Surface

Recently [19], it was demonstrated that an HF* luminescence pulse measured at a distance of $\Delta x \geq 30$ mm from the excitation zone has the form of two peaks separated in time, this time delay increasing with the Δx value. The first peak was attributed to the dissociation of molecules in the excitation zone and the second peak, to that in the region of interaction with the solid surface (i.e., in the shock wave). For $\Delta x \leq 20$ mm, the peaks are not resolved in time because of a rather narrow IR detector transmission bandwidth (≈16 kHz), but the HF* luminescence intensity strongly depends on whether or not the surface is placed in the flow.

Figure 2 shows typical plots of the HF* luminescence intensity versus gas pressure at the nozzle entrance for SF_6 molecules laser-excited in the unperturbed flow (curve *1*) and in the flow incident onto a solid surface (curve *2*) spaced by $x = 51$ mm from the nozzle ($\Delta x = 2.5$ mm). A delay time between the nozzle open pulse and the laser excitation pulse was $\tau_d = 260$ μs, which corresponds to the irradiation of the most intense part of the flow pulse (see also Fig. 3). The SF_6 molecules were excited at a frequency of 947.74 cm^{-1} (laser line 10*P*(16)), which is in close resonance with the ν_3 vibration of SF_6 ($\nu_3 \approx 948$ cm^{-1} [36]). The excitation energy density was $\Phi_{av} \approx 7.3$ J/cm^2. As seen from Fig. 2, the luminescence intensity in the flow incident onto the solid surface is 5–8 times that in the unperturbed flow.

Table 1. Parameters of the flow of SF_6 molecules at a distance of $x \approx 50$ mm from the nozzle

Pulse duration	≈100 μs
Pulse length	≈4.2 cm
Velocity v_x	420 ± 20 m/s
Number of molecules N_{fl}	5×10^{15}–1.5×10^{17}
Concentration of molecules N_1	1.7×10^{14}–5×10^{15} cm^{-3}
Translational temperature $T_{1,\,tr}$	≤40 K*
Rotational temperature $T_{1,\,rot}$	≤40 K*
Vibrational temperature $T_{1,\,vib}$	≤150 K*

* Data from [30].

The increase in the intensity of the HF* luminescence is explained by an increase in the SF_6 dissociation yield related to the shock wave formation in front of the solid surface. The multiphoton IR excitation of gas in this region produces an ensemble of highly excited molecules with a rather broad distribution over the vibrational states [1, 2]. As a result, a part of the excited molecules exhibit dissociation by the radiation mechanism, while another part of the highly excited molecules dissociate as a result of mutual collisions. In the unperturbed flow, the molecules are subject only to the radiation-induced dissociation: the collisional dissociation of highly excited molecules (the contribution of which to the total yield is usually quite large—see, e.g., [1, 2]) is absent because of the deficit of collisions. In the presence of a solid surface, the excited molecules fall within the shock wave formed in front of this surface, where the gas density and temperature are markedly greater as compared to those in the incident flow. Therefore, conditions in this region favor the collisions of highly excited molecules leading to an increase in the dissociation yield.

In the experiments reported in this article, the distance Δx from the solid surface to the excitation zone was shorter than 1 cm. For this reason, the luminescence signal represented a single peak with the intensity proportional to the total SF_6 dissociation yield.

4.3. Time-of-Flight Spectra of SF_6 in the Flow Incident onto the Solid Surface

Figure 3 shows typical plots of the HF* luminescence intensity as a function of the delay time τ_d between nozzle open pulse and laser pulse for SF_6 molecules excited in the presence of CH_4 ($p_{SF_6}/p_{CH_4} = 1:1$) in the unperturbed flow (curve *1*) and in the flow incident onto a solid surface (curve *2*). The SF_6 molecules were excited by the laser line 10*P*(16) (at a frequency of 947.74 cm^{-1}). The nozzle to target surface distance was $x = 51$ mm and $\Delta x = 2.5$ mm. As seen from this figure, the maximum intensity of luminescence from SF_6 excited in the flow interacting with the solid surface is more than ten times that in the case of molecules excited in the unperturbed flow. A sharp shock wave front appears in the zone of excitation ($\Delta x = 2.5$ mm from the surface) at $\tau_d = 310$ μs; the molecular flow velocity in this experiment was $v_x = 500 \pm 20$ m/s.

As the distance Δx from the excitation zone to the solid surface decreases (increases), the HF* luminescence intensity in the shock wave grows (drops), while the delay time τ_d corresponding to the maximum intensity of luminescence in the excitation zone decreases (increases). Figure 4 shows a plot of the HF* luminescence intensity versus the distance Δx for SF_6 molecules laser-excited in the shock wave. This curve characterizes the width and slope of the shock wave front. As seen, the profile width is approximately 3 mm for a an SF_6 pressure at the nozzle entrance $p_{noz} = 1.25$ atm and a nozzle to target surface distance $x = 51$ mm. Figure 5 presents typical plots of the HF* luminescence intensity versus delay time τ_d between nozzle and laser pulses. The curves were measured at various SF_6 pressures at the nozzle entrance for the nozzle to target surface distance $x = 51$ mm and $\Delta x = 2.5$ mm. As seen, the luminescence intensity in the shock wave sharply increases, while the delay time τ_d corresponding to the maximum luminescence intensity decreases. The increase in the luminescence intensity is caused by a growth in the gas density in the shock wave, while the decrease in τ_d is related to an increase both in the flow velocity and in the shock wave front slope (also caused by the growth in the gas density). When the gas pressure at the nozzle entrance is low ($p_{noz} \le 0.2$ atm), in which case the concentration N_1 of molecules in the flow does not exceed 3×10^{14} cm^{-3}, the shock wave weakly affects the luminescence signal. On the contrary, for comparatively high gas pressures ($p_{noz} \ge 1$ atm), the intensity of luminescence in the shock wave is markedly (20–30 times) greater than that in the unperturbed flow.

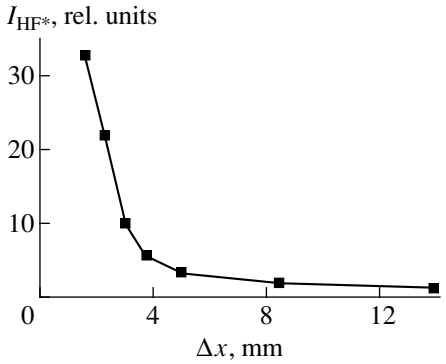

Fig. 4. The plots of HF* luminescence intensity versus x distance for SF_6 molecules laser-excited in the shock wave. The SF_6 pressure at the nozzle entrance, $p_{noz} = 1.25$ atm; nozzle to target surface distance, $x = 51$ mm.

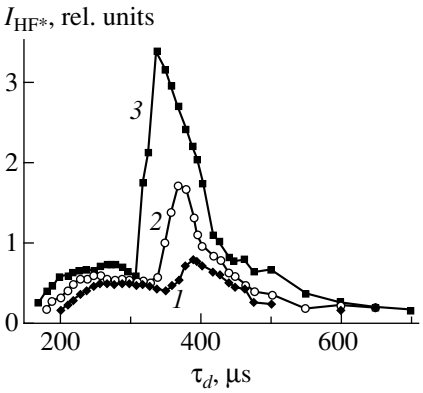

Fig. 5. The plots of HF* luminescence intensity versus delay time τ_d for various SF_6 pressures at the nozzle entrance $p_{noz} = 0.5$ (*1*); 1 (*2*); 2.5 atm (*3*). Nozzle to target surface distance, $x = 51$ mm; $\Delta x = 2.5$ mm; excitation frequency, 947.74 cm^{-1} (laser line 10*P*(16)); excitation energy density, $\Phi_{av} \approx 7$ J/cm^2.

4.4. Spectral and Energy Characteristics of the SF_6 Dissociation Process

Figure 6 shows a series of plots of the HF* luminescence intensity versus excitation energy density for SF_6 molecules excited in the presence of CH_4 ($p_{SF_6}/p_{CH_4} = 1:1$) in the unperturbed flow (curve *1*), in the flow incident onto a solid surface (curve *2*), and in the shock wave (curve *3*). These curves were measured at the laser pulse delay times corresponding to maxima in the TOF spectra of molecules (see Fig. 3): $\tau_d = 240$ μs for curves *1* and *2*, $\tau_d = 310$ μs for curve *3*. As seen from Fig. 6, the intensity of HF* luminescence from SF_6 excited in the flow interacting with the solid surface is approximately four times, and that in the shock wave–more than 30 times greater than the luminescence intensity in the case of molecules excited in the unperturbed flow. The difference is even more pronounced at

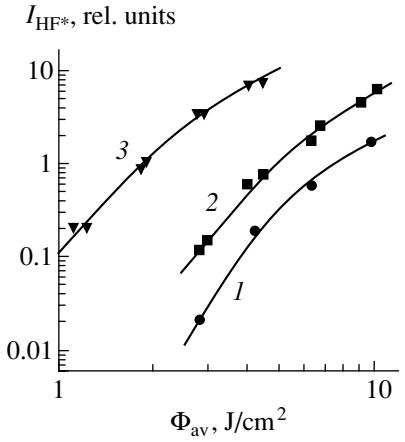

Fig. 6. The plots of HF* luminescence intensity versus excitation energy density for SF_6 molecules excited in the presence of CH_4 (p_{SF_6}/p_{CH_4} = 1 : 1) (*1*) in the unperturbed flow, (*2*) in the flow incident onto a solid surface, and (*3*) in the shock wave. Total gas pressure at the nozzle entrance, p_{noz} = 2.4 atm; nozzle to target surface distance (for curves *2* and *3*), x = 51 mm; Δx = 2.5 mm; excitation frequency, 945.98 cm^{-1} (laser line 10*P*(18)).

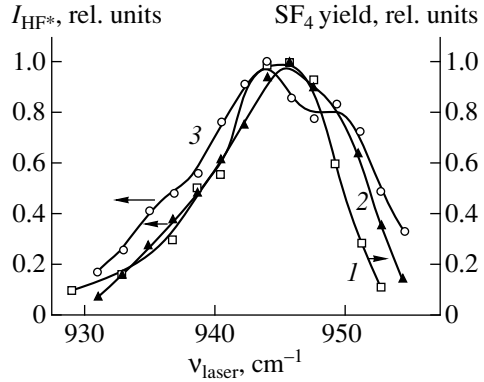

Fig. 7. The plots of HF* luminescence intensity versus excitation radiation frequency for SF_6 molecules excited (*2*) in the flow incident onto a solid surface and (*3*) in the shock wave. Total SF_6 pressure at the nozzle entrance, p_{noz} = 1.25 atm; nozzle to target surface distance, x = 51 mm; Δx = 2.5 mm; excitation energy density, $\Phi_{av} \approx 10$ J/cm^2. For comparison, curve *1* shows the frequency dependence of the SF_4 product yield [14].

$\Phi_{av} \leq 3$ J/cm^2, which indicates that a contribution due to the collisional dissociation in the shock wave to the total dissociation yield at small laser energy densities is very large. Note also a greater slope of the HF* versus Φ_{av} curve for the unperturbed flow, where the dissociation proceeds predominantly via the radiation mechanism.

All the results presented above show that the excitation of SF_6 molecules in the shock wave and in the flow incident onto the surface leads to a greater yield of the dissociation products as compared to the case of excitation in the unperturbed flow. Therefore, it would be of interest to study the selectivity of the dissociation process in the SF_6 flow interacting with the solid surface. The results of these investigations are presented below.

First, we have studied the dependence of the HF* luminescence intensity on the frequency of the exciting laser radiation (i.e., the spectral dependence of the SF_6 dissociation yield). Figure 7 shows the plots of the HF* luminescence intensity versus excitation radiation frequency for SF_6 molecules excited in the flow incident onto a solid surface (curve *2*) and in the shock wave (curve *3*). Also shown for comparison (curve *1*) is an analogous curve of the SF_4 product yield in a molecular flow of SF_6 obtained previously [14] under identical experimental conditions, which can probably be considered as the frequency dependence of the SF_6 dissociation yield in the unperturbed flow. The curves are normalized to the maximum intensity. The ratio of the maximum intensities of curves *2* and *3* was $I_2 : I_3 = 1 : 3.9$. The widths at half height are approximately 11, 12.5, and 16.5 cm^{-1} for curves *1–3*, respectively. Note a greater intensity in the region of wings at 937 and 953 cm^{-1} (especially pronounced in the high-frequency wing) for curve *3* in comparison with curve *1*. This is related to a higher rotational temperature of SF_6 molecules in the shock wave as compared to the unperturbed flow, and to the presence of the collisional dissociation component in the former case.

It is seen that, although the widths of spectra *2* and *3* in Fig. 7 are greater than the width of curve *1*, the relative intensities of the low-frequency wings (near the absorption band corresponding to the ν_3 vibration in $^{34}SF_6$ at $\nu_3 \approx 930.5$ cm^{-1} [37]) in all curves are rather close. This fact indicates that the selectivity of dissociation in these cases must not significantly differ as well. However, a comparison of the spectra presented in Fig. 7 suggests that the selectivity of dissociation in the shock wave must be somewhat lower than that in the unperturbed and incident. This is confirmed by data presented in Section 4.5.

Let us compare the results obtained for the dissociation of SF_6 molecules in the shock wave to the data available on the dissociation of these molecules in molecular beams. This comparison would be of interest for elucidating the role of the rotational and vibrational temperatures of molecules as factors determining the selectivity of dissociation. Continuous molecular beams are usually characterized by rather low rotational (≤50 K) and comparatively high vibrational (≥250 K) temperatures of SF_6 molecules. However, the shock wave formed in our experiments is distinguished by the vibrational temperature of SF_6 molecules being lower than their rotational temperature. The phenomenon of SF_6 dissociation in a molecular beam was most thoroughly studied in [11], where the frequency dependences of the SF_6 dissociation yield were obtained for

various gas temperatures at the nozzle entrance (and, hence, at various vibrational temperatures of SF_6 molecules in the beam). A comparison of curve *3* in Fig. 7 to the frequency dependence of the SF_6 dissociation yield reported in [11] (for the gas temperature $T \approx 300$ K and $\Phi_{av} \approx 7$ J/cm^2) shows that the low-frequency wing of our curve *3* (in the region of the absorption band of $^{34}SF_6$) is less intense than the corresponding wing in the spectrum of a molecular beam. Thus, the selectivity of SF_6 dissociation in the shock wave is probably somewhat greater as compared to that in the molecular beam.

4.5. Determination of the SF_4 Product Yield and Its Enrichment with ^{34}S Isotope

We have performed a series of experiments on the direct determination of the SF_4 (final product) yield and the selectivity of SF_6 dissociation for the molecules excited in an unperturbed flow and in the same flow interacting with a solid surface. The measurements were conducted using a method described elsewhere [14, 33]. The SF_4 yield in the unperturbed flow was measured at $\tau_d = 260$ μs, while the measurements in the flow incident onto the solid surface were performed for $\tau_d = 260$ and 370 μs; these delay times correspond to maximum intensities in the TOF spectra of molecules [20]. The nozzle to surface distance was $x = 51$ mm, $\Delta x = 2.5$ mm, and the SF_6 pressure at the nozzle entrance was $p_{noz} = 1.25$ atm. It was found that the yield of SF_4 excited at $\tau_d = 260$ μs in the flow incident onto the solid surface was 2.5 times, and that in the shock wave (at $\tau_d = 370$ μs) it was approximately 12 times that in the unperturbed flow.

The dissociation process selectivity was studied by measuring the coefficient of enrichment of the SF_4 product with the ^{34}S isotope. The measurements were performed in the flow incident onto the sold surface, in the shock wave, and in the unperturbed flow. The SF_6 molecules were excited at a frequency of 929 cm^{-1} (CO_2 laser line 10*P*(36)), which is in resonance with the ν_3 vibration of the $^{34}SF_6$ molecule [37]. The results of these experiments are summarized in Table 2, together with data on the SF_4 yield. For the SF_6 molecules excited in the unperturbed flow, the isotope enrichment coefficient at $\Phi_{av} \approx 10$ J/cm^2 was $K_{34}^{prod} = 17 \pm 4$, while the analogous value for molecules excited in the shock wave was $K_{34}^{prod} = 14 \pm 3$.

Table 2. Data on the yield of SF_4 and its enrichment with ^{34}S isotope for laser-induced SF_6 dissociation in an unperturbed flow and in the same flow interacting with a solid surface

p_{noz}, atm	CO_2 laser line	Φ_{av}, J/cm^2	Unperturbed flow	Incident flow	Shock wave
			SF_4 yield, rel. units		
1.25	10*P*(16)	12	1.0 ± 0.2	2.5 ± 0.5	12 ± 3
			K_{34}		
1.25	10*P*(36)	10	17 ± 5	15 ± 3	14 ± 3

4.6. Discussion of Results. Estimation of the SF_6 Density and Temperature in the Shock Wave

It should be noted that strongly inhomogeneous, nonstationary, and nonequilibrium conditions in the shock wave, as well as the numerous processes occurring in this region of the flow, markedly complicate the interpretation of the experimental data. For this reason, we will only qualitatively discuss the results of our experiments and roughly estimate the density and effective temperature of the gas in the shock wave.

An increase in the yield of dissociation products for molecules excited in the shock wave is related to several factors: (i) an increase in the gas density; (ii) an increase in the rate of chemical reactions; (iii) an increase in the dissociation yield caused by the more effective excitation and the additional collisional dissociation of molecules excited at the IR laser pulse energy below the dissociation threshold (the latter mechanism is inoperative in the unperturbed flow because of the deficit of collisions). A comparatively high selectivity of dissociation in the shock wave is related to a comparatively low vibrational temperature of molecules in this region.

The limiting increase in density in the direct shock wave for a gas with constant heat capacity is given by the relationship [23– 25]

$$\frac{\rho_2}{\rho_1} = \frac{\gamma + 1}{\gamma - 1},$$

where ρ_1 and ρ_2 are the gas densities in the incident flow and in the shock wave, respectively; $\gamma = c_p/c_v$ is the ratio of specific heat capacities. For SF_6 at $T \approx 300$ K, $\gamma \approx 1.1$ [38, 39] and the formula yields $\rho_2/\rho_1 \approx 21$. However, this estimate is likely to be incorrect because not all of the molecular degrees of freedom contribute to the heat capacity of molecules under the experimental conditions studied. Let us estimate the limiting increase in the gas density and the average concentration of molecules in the shock wave using the molecular flow parameters obtained in Section 4.1. For a rough estimate, the ratio ρ_2/ρ_1 can be taken equal to the ratio of the flow pulse length ($l_{fl} \approx 4.2$ cm in the excitation zone at a distance of $x = 51$ mm from the nozzle) to the shock wave front width (≈3 mm, Fig. 4): $\rho_2/\rho_1 \approx 14$. For an SF_6 pressure at the nozzle entrance $p_{noz} = 1.25$ atm, the total number and concentration of molecules in the flow were $N_{fl} \approx 4.2 \times 10^{16}$ and $N_1 \approx 1.4 \times 10^{15}$ cm^{-3}, respectively. Therefore, the average concentration of molecules in the shock wave is $N_2 \approx 2 \times 10^{16}$ cm^{-3}. Note that the limiting gas density is not reached at a distance of $\Delta x = 2.5$ mm from the surface (as seen in Fig. 4, the HF* luminescence intensity at shorter distances is

higher). For this reason, a increase in the SF_4 yield in the shock wave is probably related not only to the increase in the gas density, but to some other factors considered above as well.

The gas heating due to retardation in the shock wave was evaluated [20, 21] using the formula [24, 25]

$$\Delta T = v_0^2/2c_p,$$

where v_0 is the flow velocity and c_p is the heat capacity. Using the SF_6 flow parameters $v_0 \approx 420$ m/s; $c_p \approx 665$ J/(kg K) [38, 39]), we obtain $\Delta T \approx 130$ K. However, it must be noted that the heat capacity of SF_6 under our experimental conditions is lower than the value for $T \approx 300$ K and, hence, the additional gas heating can be markedly greater. Taking into account the law of energy conservation for SF_6 molecules in the incident flow and in the shock wave (assuming that the vibrational degrees of freedom are not activated), we may conclude that the temperature of translational and rotational degrees of freedom increased by $\Delta T \approx 530$ K. Thus, $T_{2,\,tr} \approx T_{2,\,rot} \approx (T_{1,\,tr} + \Delta T) \approx 570$ K, while the vibrational temperature in the shock wave is $T_{2,\,vib} \approx T_{1,\,vib} \leq 150$ K.

Therefore, once the vibrational temperature of molecules is a dominating factor, the SF_6 dissociation selectivity in the shock wave must not differ significantly from that in the unperturbed flow, which is really observed in experiment. Note also that a decrease in the selectivity of dissociation in the shock wave related to an increase in the gas temperature can sometimes be compensated for by a growth in the selectivity due to the increasing concentration of irradiated molecules [40].

5. CONCLUSION

Now we will summarize the main results of our investigation and formulate the conclusions.

We have studied the isotope-selective multiphoton IR dissociation of SF_6 molecules in a pulsed gasdynamic flow incident onto a solid surface. It was found that there is a considerable (severalfold) increase in the product yield in this system as compared to the case of unperturbed flow, without significant changes in the selectivity of the dissociation process.

It was established that an increase in the efficiency of the multiphoton IR dissociation of SF_6 is related to the formation of a shock wave at the target surface, which provides for the conditions favoring collisions between the highly excited SF_6 molecules (this factor leading to an increase in the dissociation yield).

We have also studied the selective multiphoton IR dissociation of SF_6 molecules excited under nonequilibrium conditions in the shock wave. It was found that there is a considerable (more than tenfold) increase in the product yield as compared to the case of excitation in the unperturbed flow at an insignificant (25–30%) decrease in selectivity in the former case.

A method proposed for the shock wave formation in a gas flow provides for the possibility of studying the selective photochemical processes under nonequilibrium conditions ($T_{tr} \geq T_{rot} \geq T_{vib}$, $T_{vib} \leq 150$ K) that are inverse to those occurring in the gasodynamically cooled jets and flows.

The experimental results obtained for SF_6 show that the dissociation under conditions when $T_{vib} \leq 150$ K and $T_{rot} \leq 570$ K is characterized by a rather high selectivity. This fact indicates that, from the standpoint of selectivity, the vibrational temperature is a more important factor than the rotational temperature.

The method proposed for increasing the efficiency of the multiphoton dissociation is useful in the case of the selective dissociation of large polyatomic molecules possessing long lifetimes (exceeding 100–200 μs) with respect to the monomolecular decay (even for the vibrational excitation levels markedly exceeding the dissociation threshold). Examples are offered by $(CF_3)_3CX$ molecules where X is a halogen of hydrogen [41, 42]. Possessing long lifetimes with respect to the decay, the overexcited molecules may reach the chamber walls and relax there, not forming the dissociation products. By creating a shock wave, it is possible to provide conditions for the collisional dissociation of excited molecules and, hence, to increase the product yield.

ACKNOWLEDGMENTS

The authors are grateful to V.N. Lokhman and S.A. Mochalov for their help in the manuscript preparation. The work was partly supported by the Russian Foundation for Basic Research, project no. 00-03-33003a). The authors are grateful to the referee for his valuable critical remarks.

REFERENCES

1. V. N. Bagratashvili, V. S. Letokhov, A. A. Makarov, and E. A. Ryabov, *Multiple Photon Infrared Laser Photophysics and Photochemistry* (Academic, New York, 1985).
2. *Multiple Photon Excitation and Dissociation of Polyatomic Molecules*, Ed. by C. D. Cantrell (Springer-Verlag, Berlin, 1986).
3. *Laser Induced Chemical Processes*, Ed. by Y. Steinfeld (Plenum, New York, 1981).
4. J. L. Lyman, in *Laser Spectroscopy and Its Applications* (Marcel Dekker, New York, 1987).
5. R. V. Ambartsumyan, Yu. A. Gorokhov, V. S. Letokhov, *et al.*, Zh. Éksp. Teor. Fiz. **71**, 440 (1976) [Sov. Phys. JETP **44**, 231 (1976)].
6. M. C. Gower and K. W. Billman, Opt. Commun. **20**, 123 (1977).
7. U. Del Bello, V. Churakov, W. Fuss, *et al.*, Appl. Phys. B: Photophys. Laser Chem. **B 42**, 147 (1987).

8. R. V. Ambartsumyan, Yu. A. Gorokhov, V. S. Letokhov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **23**, 217 (1976) [JETP Lett. **23**, 194 (1976)].
9. V. Yu. Baranov, E. P. Velikhov, Yu. R. Kolomiĭskiĭ, *et al.*, Kvantovaya Élektron. (Moscow) **6**, 1062 (1979).
10. F. Brunner and D. Proch, J. Chem. Phys. **68**, 4936 (1978).
11. P. A. Schulz, A. S. Sudbo, E. R. Grant, *et al.*, J. Chem. Phys. **72**, 4985 (1980).
12. E. Borsella, R. Fantoni, L. Yu-Shen, and M. Nardelli, Nuovo Cimento D **4**, 548 (1984).
13. S. S. Alimpiev, G. S. Baronov, S. M. Karavaev, *et al.*, Kvantovaya Élektron. (Moscow) **10**, 376 (1983).
14. G. N. Makarov and A. N. Petin, Khim. Vys. Énerg. **34**, 448 (2000).
15. G. N. Makarov, E. Ronander, S. P. van Heerden, *et al.*, Appl. Phys. B: Lasers Opt. **B 65**, 583 (1997).
16. G. N. Makarov, V. N. Lokhman, D. E. Malinovskiĭ, and D. D. Ogurok, Kvantovaya Élektron. (Moscow) **25**, 545 (1998).
17. G. N. Makarov, D. E. Malinovsky, and D. D. Ogurok, Laser Chem. **17**, 205 (1998).
18. G. N. Makarov, V. N. Lokhman, D. E. Malinovskiĭ, and D. D. Ogurok, Khim. Fiz. **18**, 71 (1999).
19. G. N. Makarov and A. N. Petin, Kvantovaya Élektron. (Moscow) **30**, 738 (2000).
20. G. N. Makarov and A. N. Petin, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 583 (2000) [JETP Lett. **71**, 399 (2000)].
21. G. N. Makarov and A. N. Petin, Chem. Phys. Lett. **323**, 345 (2000).
22. J. B. Anderson, in *Gas Dynamics, Molecular Beams, and Low- Density Gas Dynamics* (Marcel Dekker, New York, 1974), Vol. 4, p. 1.
23. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics* (Nauka, Moscow, 1986; Pergamon, New York, 1987).
24. Ya. B. Zel'dovich and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* (Nauka, Moscow, 1966; Academic, New York, 1966, 1967).
25. G. N. Abramovich, *Applied Gas Dynamics* (Nauka, Moscow, 1991), Part 1.
26. E. V. Stupochenko, S. A. Losev, and A. I. Osipov, *Relaxation Processes in Shock Waves* (Nauka, Moscow, 1965).
27. J. I. Steinfeld, I. Burak, D. G. Sutton, and A. V. Nowak, J. Chem. Phys. **52**, 5421 (1970).
28. W. R. Gentry and C. F. Giese, Rev. Sci. Instrum. **49**, 595 (1978).
29. V. M. Apatin, L. M. Dorozhkin, G. N. Makarov, and G. M. Pleshkov, Appl. Phys. B: Photophys. Laser Chem. **B 29**, 273 (1982).
30. V. M. Apatin and G. N. Makarov, Zh. Éksp. Teor. Fiz. **84**, 15 (1983) [Sov. Phys. JETP **57**, 8 (1983)].
31. C. A. Quick, Jr. and C. Wittig, Chem. Phys. Lett. **48**, 420 (1977).
32. S. S. Alimpiev, Izv. Akad. Nauk SSSR, Ser. Fiz. **45**, 1070 (1981).
33. G. N. Makarov, D. E. Malinovskiĭ, and D. D. Ogurok, Zh. Tekh. Fiz. **69** (1), 35 (1999) [Tech. Phys. **44**, 31 (1999)].
34. I. W. Levin and C. V. Berney, J. Chem. Phys. **44**, 2557 (1966).
35. K. O. Christe, E. C. Curtis, C. J. Schack, *et al.*, Spectrochim. Acta A **32**, 1141 (1976).
36. R. S. McDowell, B. J. Krohn, H. Flicker, and C. Vásquez, Spectrochim. Acta A **42**, 351 (1986).
37. G. Baldacchini, S. Marchetti, and V. Montelatici, J. Mol. Spectrosc. **91**, 80 (1982).
38. *CRC Handbook of Chemistry and Physics*, Ed. by David R. Lide (CRC Press, Boca Raton, 1993–1994).
39. *Tables of Physical Quantities: Handbook*, Ed. by I. K. Kikoin (Atomizdat, Moscow, 1976).
40. G. N. Makarov, Pis'ma Zh. Tekh. Fiz. **24** (23), 35 (1998) [Tech. Phys. Lett. **24**, 921 (1998)].
41. V. M. Apatin and G. N. Makarov, Pis'ma Zh. Éksp. Teor. Fiz. **38**, 120 (1983) [JETP Lett. **38**, 141 (1983)].
42. V. N. Bagratashvili, S. I. Ionov, V. S. Letokhov, *et al.*, Zh. Éksp. Teor. Fiz. **93**, 1188 (1987) [Sov. Phys. JETP **66**, 670 (1987)].

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 10–19.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 16–26.

ATOMS, SPECTRA, RADIATION

Wave Packets in Smoothly Inhomogeneous Dispersive Media

G. V. Permitin and A. I. Smirnov*

Institute of Applied Physics, Russian Academy of Sciences, ul. Ul'yanova 46, Nizhniĭ Novgorod, 603600 Russia
**e-mail: smirnov@appl.sci-nnov.ru*

Received May 17, 2000

Abstract—In the quasi-optical aberration-free approximation, a parabolic equation for the envelope of an electromagnetic wave packet propagating along a geometric optical ray in a smoothly inhomogeneous isotropic medium with time dispersion is obtained. The corresponding Green's function is found whose parameters are determined by integrating a system of ordinary differential equations. The effects of the combined influence of the refraction, diffraction, and dispersion on the evolution of the packet are analyzed; in particular, the effects that cause precession of the envelope waves about the binormal to the propagation trajectory are considered.

1. INTRODUCTION

Under certain conditions, the propagation of wave packets in smoothly inhomogeneous media assumes an approximate quasi-optical description by means of an abridged parabolic equation for the "envelope" waves (complex amplitudes). On the one hand, the reference gauge of the envelope (both transverse to the direction of the wave propagation, $\Lambda_\perp$, and longitudinal, $\Lambda_\parallel$) must be much greater than the length of the carrier wave λ. These conditions are equivalent to constraints on the width of the angular and frequency spectrum of the process: $\Delta k_\perp/k \ll 1$, $\Delta\omega/\omega \ll 1$. On the other hand, the cross size of the packet, $\Lambda_\perp$, must be small as compared to the size of the medium inhomogeneity L_ε. Upper constraints on the longitudinal size of the packet (or, in other words, on the pulse duration T_p) are immaterial for quasi-optics; moreover, the continuous passage to the steady-state or time-dependent wave beams is possible as $T_p \longrightarrow \infty$. Steady-state beams ($\Delta\omega = 0$) in inhomogeneous media are thoroughly studied in [1], of which the present paper is a generalization. In addition, for the quasi-optical approach to be applicable, it is required that the dispersion curve of the medium be reasonably close to a parabola within the frequency band $\Delta\omega$.

A theorem proved by Ehrenfest in wave mechanics asserts that the trajectory of the center of a wave packet is quasiclassical. By analogy, we assume that the packet of electromagnetic waves propagates with the group speed v_g along a certain curvilinear geometric optics ray $\mathbf{r} = \mathbf{r}_c(s)$ (in what follows, it is called the reference ray). Then, the wave field can be represented as

$$\mathbf{E}(t, \mathbf{r}) = \mathbf{A}(t, \mathbf{r})\exp[-i\omega t + i\phi(\mathbf{r})], \qquad (1.1)$$

where ω is the central frequency of the packet, ϕ is the phase of the "carrier wave," and $A(t, \mathbf{r})$ is the complex envelope. In homogeneous media, the eigenmode of the system is used as the carrier wave; as a rule, it is a plane monochromatic wave. In inhomogeneous media, the carrier wave is chosen such that it is not the eigenwave, but has plane phase fronts (ϕ= const) perpendicular to the reference ray. $\phi(\mathbf{r}_c)$ is calculated along the ray in the geometric optics approximation.

Due to the small size of the packet as compared to L_ε, the tendency towards transverse "diffusion"[1] caused by diffraction and towards longitudinal diffusion caused by dispersion are retained, this being characteristic of homogeneous media [2–4]. The diffusion of the packet is accompanied by both phase and amplitude modulation of the envelope. However, the dispersion in inhomogeneous media is manifested in one more way, namely, in the transverse divergence of the packet due to the frequency dependence of the ray curvature. The refractive (or Newtonian) dispersion[2] leads, in the small-angle approximation, to a relationship between the transverse and longitudinal field structure [6].

Diffraction and dispersion effects in inhomogeneous media compete with refraction effects that manifest themselves not only in the curving of the trajectory of the packet propagation, but in its transverse focusing (or defocusing) as well.[3] Many studies are devoted to the quasi-optics of focusing systems. In particular,

[1] The analogy between diffraction and diffusion phenomena dates to the works of Th. Young (1800) and received widespread use in quasi-optics (M.A. Leonovich and V.A. Fok, 1944–1946) [5]. It is clear that the wave analogue differs from the ordinary diffusion (of particles or heat) in that it is reversible, i.e., invariant under the change $t \longrightarrow -t$ and under the reversal of the propagation direction.

[2] As a matter of fact, the term *dispersion* was introduced in physics by Newton exactly to describe the expansion of a solar beam into the color spectrum as a result of the refraction in a piecewise homogeneous medium (a prism).

[3] By the way, curving the propagation trajectory results in additional focusing [7].

these studies concern optical lines[4] and lens-like media (see, e.g., [9–11]) of the type of profiled optical fibers. It is worth noting that the linear oscillator [12] is a quantum-mechanical analogue of focusing systems.

The quasi-optical equation for the packet envelope in an arbitrary smoothly inhomogeneous medium is almost identical to the equation for the lens-like medium. The difference is in the appearance of an additional operator (an analogue of the rotation operator in the Schrödinger equation for the electron in a magnetic field [12]) that leads to an unexpected interference phenomenon—the wave precession of the packet about the binormal to the propagation trajectory.

2. EQUATION FOR THE COMPLEX AMPLITUDE OF QUASI-HARMONIC WAVE FIELDS

In a stationary isotropic dielectric (without the magnetic susceptibility) medium with time dispersion, the Maxwell equations can be reduced to the form [13]

$$\mathrm{curl}\,\mathrm{curl}\,\mathbf{E} + \frac{1}{c^2}\frac{\partial^2 \mathbf{D}}{\partial t^2} = 0, \tag{2.1}$$

where the electric field strength $\mathbf{E}$ and the induction vector $\mathbf{D}$ are related as

$$\mathbf{D}(t, \mathbf{r}) = \int_0^\infty \hat{\varepsilon}(\tau, \mathbf{r})\mathbf{E}(t-\tau, \mathbf{r})d\tau. \tag{2.2}$$

For quasi-harmonic processes, we have

$$\begin{aligned} \mathbf{E} &= \tilde{\mathbf{E}}(t, \mathbf{r})\exp(-i\omega t), \\ \mathbf{D} &= \tilde{\mathbf{D}}(t, \mathbf{r})\exp(-i\omega t), \end{aligned} \tag{2.3}$$

where the complex amplitudes $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{D}}$ are slow functions as compared to $1/\omega$ and the constitutive Eq. (2.2) can be written in the differential form (see, e.g., [4])

$$\tilde{\mathbf{D}} = \varepsilon(\mathbf{r}, \omega)\tilde{\mathbf{E}} + i\frac{\partial \varepsilon}{\partial \omega}\frac{\partial \tilde{\mathbf{E}}}{\partial t} - \frac{1}{2}\frac{\partial^2 \varepsilon}{\partial \omega^2}\frac{\partial^2 \tilde{\mathbf{E}}}{\partial t^2} + \ldots. \tag{2.4}$$

Series (2.4) provides an expansion in the small parameter $\mu \sim \Delta\omega/\delta\omega$, which equals to the ratio of the width of the process spectrum $\Delta\omega$ to the reference dispersion gauge of the permittivity $\varepsilon(\omega, \mathbf{r})$.

In the second approximation of dispersion theory [4], Eq. (2.1) is transformed into the following equation for the complex amplitude of the electric field (up to the terms of order $\sim\mu^2$):

$$\begin{aligned} \mathrm{curl}\,\mathrm{curl}\,\mathbf{E} - \frac{1}{c^2}\Big(\omega^2\varepsilon\tilde{\mathbf{E}} + i\frac{\partial \omega^2\varepsilon}{\partial \omega}\frac{\partial \tilde{\mathbf{E}}}{\partial t} \\ - \frac{1}{2}\frac{\partial^2 \omega^2\varepsilon}{\partial \omega^2}\frac{\partial^2 \tilde{\mathbf{E}}}{\partial t^2}\Big) = 0. \end{aligned} \tag{2.5}$$

For a wave packet with a narrow angular spectrum, Eq. (2.5) can be simplified and reduced to a parabolic equation in a curvilinear frame of reference attached to the central (both in the spatial and spectral sense) geometric optics ray.

3. THE FRAME OF REFERENCE ATTACHED TO THE REFERENCE RAY

The canonical (Hamiltonian) equations for a ray in an inhomogeneous isotropic medium are written as [14]

$$\frac{d\mathbf{r}}{d\varsigma} = \mathbf{p}, \quad \frac{d\mathbf{p}}{d\varsigma} = \frac{1}{2}\nabla\varepsilon(\omega, \mathbf{r}), \tag{3.1}$$

where $\mathbf{r}$ is the radius vector of a point on the ray, $\mathbf{p}$ is the current wave vector $\mathbf{k} = \mathbf{s}k_0\sqrt{\varepsilon(\mathbf{r})} = k_0\mathbf{p}$ normalized with respect to $k_0 = \omega/c$ ($\mathbf{s}$ is the unit vector tangent to the ray), and the variable ς is related to the length of a ray arc $\mathbf{s}$, as $d\varsigma = ds/\sqrt{\varepsilon}$.

Let $\mathbf{r} = \mathbf{r}_c(\varsigma)$, $\mathbf{p} = \mathbf{p}_c(\varsigma)$ be a solution of Eq. (3.1). We take it as the reference trajectory for the curvilinear frame of reference. (In what follows, all quantities and their derivatives calculated for the points on the reference ray are marked by a subscript c.)

The geometric characteristics of the reference ray, the Darboux trihedral ($\mathbf{s}$, $\mathbf{n}$, $\mathbf{m}$), the curvature K, and the torsion T, are unambiguously determined at every point by the propagation direction s, the local values of permittivity, and its directional derivatives [1, 14]. The principal normal $\mathbf{n}$ to the ray lies in the plane $\{\mathbf{s}, \nabla\varepsilon\}$, and the binormal is $\mathbf{m} = \mathbf{s} \times \mathbf{n}$. The curvature of the ray and its torsion are expressed by the formulas

$$K = \left[\frac{\partial \varepsilon}{\partial n}/2\varepsilon\right]_{\mathbf{r}=\mathbf{r}_c}, \quad T = \left[\frac{\partial^2 \varepsilon}{\partial s \partial m}/\frac{\partial \varepsilon}{\partial n}\right]_{\mathbf{r}=\mathbf{r}_c}. \tag{3.2}$$

The normal and binormal are functions of the length along the ray, ($\mathbf{n}(s)$, $\mathbf{m}(s)$); however, this dependence is unrelated to formulas (3.2), which involve the directional derivatives at a fixed point.

The reference frame attached to ($\mathbf{s}$, $\mathbf{n}$, $\mathbf{m}$) is affine (for $T \neq 0$); this is seen from the Frenet–Serret formulas $d\mathbf{n}/ds = -K\mathbf{s} + T\mathbf{m}$, $d\mathbf{m}/ds = -T\mathbf{n}$. However, one can define an orthogonal frame of reference (s, ξ_1, ξ_2) whose basis ($\mathbf{s}$, $\mathbf{g}_1$, $\mathbf{g}_2$) rotates relative to ($\mathbf{s}$, $\mathbf{n}$, $\mathbf{m}$)

[4] The small-angle (or paraxial) approximation of the diffraction theory of optical systems was developed well before the term *quasi-optics* became conventional in physics. A rather detailed presentation of this theory can be found in [8].

according to the translation of the vector along the curve (in the Levi-Civita sense):

$$\begin{aligned} \mathbf{g}_1 &= \mathbf{n}\cos\theta + \mathbf{m}\sin\theta, \\ \mathbf{g}_2 &= \mathbf{m}\cos\theta - \mathbf{n}\sin\theta, \\ \frac{d\theta}{ds} &= -T. \end{aligned} \tag{3.3}$$

The Lamé coefficients of this frame of reference are

$$h_1 = h_2 = 1, \quad h_s = h = 1 - K(\xi_1\cos\theta - \xi_2\sin\theta).$$

It is well known [14, 15] that in the geometric optics approximation the polarization vector of the electric field rotates with respect to **n** in accordance with (3.3); this is the so-called Rytov law. Below we will show that the Rytov law remains valid in the quasi-optical approximation. This means that the projections of the polarization vector onto the unit vectors $\mathbf{g}_1$ and $\mathbf{g}_2$ are constant and can be factored outside the differentiation operators.

4. QUASI-OPTICAL EQUATION FOR THE ENVELOPE OF A WAVE PACKET (ABERRATION-FREE APPROXIMATION)

We assume that a wave packet is propagating along the reference ray and let the reference gauge of the packet satisfy the conditions formulated above; i.e., $\Lambda_\perp$, $\Lambda_\parallel \gg \lambda(\mathbf{r})$, $\Lambda_\perp \ll L_\varepsilon$, $\Delta\omega \ll \delta\omega, \omega$. It is seen that the problem involves several small dimensionless parameters that can significantly differ from one another in magnitude. However, since the aberration-free quasi-optical approximation is quadratic in each of these parameters, we denote all of them by the same symbol μ:

$$\mu \sim \frac{\lambda}{\Lambda_\perp}, \ \frac{\lambda}{\Lambda_\parallel}, \ \frac{\Lambda_\perp}{L_\varepsilon}, \ \frac{\Delta\omega}{\delta\omega}, \ \frac{\Delta\omega}{\omega}.$$

We represent the complex amplitude of the electric field as

$$\tilde{\mathbf{E}}(t, \mathbf{r}) = E_\perp(t, \mathbf{r})\mathbf{e}_\perp + E_\parallel(t, \mathbf{r})\mathbf{s}, \tag{4.1}$$

where $\mathbf{e}_\perp = a\mathbf{g}_1 + b\mathbf{g}_2$ is the polarization vector and a and b are complex (in the general case) numbers such that $|a|^2 + |b|^2 = 1$. In the case when the wave packet is inhomogeneously polarized with respect to the transverse coordinates, it can be represented as a superposition of two wave packets with homogeneous mutually perpendicular polarizations (linear, circular, or elliptic).

Substituting (4.1) into (2.5) and writing the differential operators componentwise in the curvilinear frame of reference (s, ξ_1, ξ_2) introduced above, we obtain (accurate to terms that are cubic in μ) the following equation for the transverse component of the electric field $E_\perp$:

$$\begin{aligned} &\frac{1}{h}\frac{\partial}{\partial s}\left(\frac{1}{h}\frac{\partial E_\perp}{\partial s}\right) + \frac{1}{h}\frac{\partial}{\partial \xi_m}\left(h\frac{\partial E_\perp}{\partial \xi_m}\right) - \frac{1}{c^2} \\ &\times\left(\omega^2\varepsilon E_\perp + i\frac{\partial \omega^2\varepsilon}{\partial\omega}\frac{\partial E_\perp}{\partial t} - \frac{1}{2}\frac{\partial^2\omega^2\varepsilon}{\partial\omega^2}\frac{\partial^2 E_\perp}{\partial t^2}\right) = 0. \end{aligned} \tag{4.2}$$

Here the summation is performed over the repeating index $m = 1, 2$.

In the same approximation, vector $\mathbf{e}_\perp$ is constant (accurate to μ^3) in the orthogonal frame of reference (s, ξ_1, ξ_2); i.e., it rotates with respect to the natural trihedral obeying the Rytov law.

The longitudinal component of the field $E_\parallel(t, \mathbf{r})$ can be determined from the condition $\mathrm{div}\mathbf{D} = 0$:

$$ik\varepsilon E_\parallel \approx -\mathrm{div}(\varepsilon E_\perp \mathbf{e}_\perp).$$

Neglecting the dispersion at the angles $\beta_s \gg \mu$ to the propagation trajectory and separating the rapidly oscillating factor in field $E_\perp$, one can pass from (4.2) to a parabolic equation for the slow and smooth envelope of the packet $W(t, \mathbf{r})$:

$$E_\perp = (\varepsilon_c(s))^{-1/4} W(t, \mathbf{r}) \exp\left[ik_0 \int_0^s \sqrt{\varepsilon_c(s')}\,ds'\right]. \tag{4.3}$$

In the first-order approximation in μ, the wave packet is carried along the ray with the group speed v_g without changing its shape:

$$\left(\frac{\partial}{\partial s} + \frac{1}{v_g}\frac{\partial}{\partial t}\right)W = O(\mu^2),$$

$$v_g(s) = \frac{\partial\omega}{\partial k} = c\left[\frac{\partial\sqrt{\omega^2\varepsilon_c(\omega)}}{\partial\omega}\right]^{-1}.$$

Let us write down the terms of the second order of smallness in μ. It is convenient to change the variables (s, t) to the variables (τ, ς) that take this drift into account. We have

$$\begin{aligned} &d\varsigma = ds/\sqrt{\varepsilon_c}, \quad d\tau = dt - ds/v_g, \\ &\frac{\partial}{\partial\varsigma} = \sqrt{\varepsilon_c}\left(\frac{\partial}{\partial s} + \frac{1}{v_g}\frac{\partial}{\partial t}\right), \quad \frac{\partial}{\partial\tau} = \frac{\partial}{\partial t}. \end{aligned} \tag{4.4}$$

Such a choice of variables corresponds to the following observation method. In the plane s = const, receivers are placed that yield the scan of the signal in time. The local time reference τ is shifted by the magnitude of the delay time

$$t_c(s) = \int_0^s \frac{ds'}{v_g(s')}.$$

Substituting (4.3) into (4.2) and retaining only the zero- and first-order terms, we obtain the abridged parabolic equation

$$2ik_0\left(\frac{\partial W}{\partial \varsigma} + k_0\delta_p(\varsigma)\xi_p\frac{\partial W}{\partial \tau}\right) + \frac{\partial^2 W}{\partial \xi_1^2} + \frac{\partial^2 W}{\partial \xi_2^2} + \gamma(\varsigma)\frac{\partial^2 W}{\partial \tau^2} - k_0^2\alpha_{pq}(\varsigma)\xi_p\xi_q W = 0, \tag{4.5}$$

where

$$\gamma(\varsigma) = \frac{1}{v_g^2}\left[1 - \frac{v_g^2}{2c^2}\frac{\partial^2\omega^2\varepsilon}{\partial\omega^2}\right]_{r=r_c} = \frac{k}{v_g^2}\frac{\partial v_g}{\partial\omega},$$

$$\delta_p(\varsigma) = \frac{1}{2}\left[\frac{\partial^2\varepsilon}{\partial\xi_p\partial\omega} - \frac{1}{\varepsilon}\frac{\partial\varepsilon}{\partial\xi_p}\frac{\partial\varepsilon}{\partial\omega}\right]_{r=r_c} = \varepsilon\frac{\partial K}{\partial\omega}(\mathbf{n}\cdot\mathbf{g}_p),$$

$$\alpha_{pq}(\varsigma) = \left[\frac{3}{4\varepsilon}\frac{\partial\varepsilon}{\partial\xi_p}\frac{\partial\varepsilon}{\partial\xi_q} - \frac{1}{2}\frac{\partial^2\varepsilon}{\partial\xi_p\partial\xi_q}\right]_{r=r_c}, \quad p, q = 1, 2.$$

The meaning of the approximation in which Eq. (4.5) is derived can be explained by invoking the analogy between an inhomogeneous medium and an optical line consisting of discrete phase correctors—linear (prisms), quadratic (ideal lenses), and higher-order correctors—that are responsible for aberrations (cubic, spherical, etc.).[5] The linear correctors can be eliminated by changing to the curvilinear frame of reference attached to the ray. However, a trace of "prisms" remains in (4.5) in the form of the operator $\delta_m\xi_m\partial_\tau$ that accounts for the refractive dispersion in the second-order approximation with respect to μ. Although approximation (4.5) accounts for one form of aberration (astigmatism), it is, nevertheless, called aberration-free in quasi-optics.

The quadratic form $U_{\text{eff}} = \alpha_{pq}\xi_p\xi_q$, which determines the optical properties of the lenses, can be reduced to the canonical form for every section $\varsigma = \text{const}$ by the rotation by the angle $\varphi(\varsigma)$ in the plane (ξ_1, ξ_2):

$$U_{\text{eff}} = \alpha_1(\varsigma)\bar{\xi}_1^2 + \alpha_2(\varsigma)\bar{\xi}_2^2.$$

The directions of the axes $\bar{\xi}_1^2$ and $\bar{\xi}_2^2$ are called the principal directions of the astigmatic lens. The behavior of the wave beam or packet essentially depends on the speed of rotation of the principal axes as they are moving along the curve $T^* = d\varphi/d\varsigma$.[6]

[5] A series of ideal lenses is described by the same Eq. (4.5) with $\alpha_{pq}(z) = \sum_i \alpha_{pq}(z)\Delta z_i\delta(z - z_i)$.

[6] We designate the degree of twist of the optical curve (or its "shear") by the same symbol T as the torsion of the propagation trajectory (we only mark it by an asterisk). The torsion of the reference ray always results in a shear of the equivalent lens system; however, the shear can appear ($T^* \neq 0$) for $T = 0$ as well.

If the degree of twist of the line is large ($|T^*|^2 \gg |\alpha_1|, |\alpha_2|$), we can average over the period of the lens rotation and pass to an effective axially symmetric optical line. This assertion follows from the equivalence of two identical orthogonal thin cylindrical axially symmetric lenses pressed to one another:

$$\exp\left(\frac{ikx^2}{2F}\right)\exp\left(\frac{iky^2}{2F}\right) = \exp\left(\frac{ikr^2}{2F}\right).$$

When the opposite inequality $|T^*|^2 \ll |\alpha_1|, |\alpha_2|$ holds, the shear of the optical line can be accounted for in the adiabatic approximation.

The intermediate case $\alpha_1 < |T^*|^2 < \alpha_2$ is most interesting. Under these conditions, the wave packet becomes parametrically unstable (this situation is similar to the parametric resonance of a pendulum). The size of the packet increases exponentially in the direction that rotates with the displacement along ς; moreover, defocusing occurs in the locally focusing system: $\alpha_1, \alpha_2 > 0$. By contrast, the torsion enhances the focusing in the orthogonal direction and the size of the packet oscillates. This effect was thoroughly studied in paper [1] devoted to wave beams.

5. GREEN'S FUNCTION OF THE QUASI-OPTICAL EQUATION

Equation (4.5) possesses a remarkable property: its Green's function can be exactly determined in the geometric optics approximation [1]. The system of ray equations corresponding to (4.5) is linear, which makes the task of "aiming" the ray originating from an arbitrary point source ($\varsigma' = 0$, ξ', τ') at an arbitrary observation point (ς ξ, τ) and the subsequent calculation of the phase and amplitude of Green's function $G(\varsigma', \xi', \tau'; \varsigma, \xi, \tau)$ trivial. However, since the coefficients in (4.5) are asymmetric, an attempt to present this procedure in a compact form fails. For this reason, we take advantage of the fact that the coefficients in (4.5) do not depend explicitly on τ. We use the Fourier method; i.e., we seek the solution in the form of an expansion in partial monochromatic beams (at the frequencies $\omega + \Omega$). Representing the envelope of the packet in the form

$$W(\varsigma, \xi_1, \xi_2, \tau) = \frac{1}{2\pi} \times \int_{-\infty}^{\infty}\tilde{W}(\varsigma, \xi_1, \xi_2, \Omega)\exp(-i\Omega\tau)d\Omega, \tag{5.1}$$

we obtain

$$2ik_0\frac{\partial\tilde{W}}{\partial\varsigma} + \frac{\partial^2\tilde{W}}{\partial\xi_1^2} + \frac{\partial^2\tilde{W}}{\partial\xi_2^2} - k_0^2\left(\alpha_{mn}\xi_m\xi_n - 2\delta_m\xi_m\Omega + c^2\gamma\frac{\Omega^2}{\omega^2}\right)\tilde{W} = 0. \tag{5.2}$$

The expression in parentheses can be reduced to a quadratic form by the change of variables

$$\eta_m = \xi_m + q_m(\varsigma)\Omega, \quad V(\varsigma, \eta_1, \eta_2, \Omega) = \tilde{W}\exp(ik_0\Phi),$$

$$\Phi = (\dot{q}_1\xi_1 + \dot{q}_2\xi_2)\Omega + \frac{1}{2}\Omega^2\int_0^\varsigma \beta(\varsigma)d\varsigma, \qquad (5.3)$$

$$\beta = \frac{\gamma}{k_0^2} - \alpha_{mn}q_m q_n + \dot{q}_1^2 + \dot{q}_2^2,$$

where the dot denotes the derivative with respect to ς. The parameters q_m that characterize the lateral displacement of the partial beams with the frequency distinct from ω are the constraint solution to the system

$$\ddot{q}_m + \alpha_{mn}(\varsigma)q_n = -\delta_m(\varsigma) \qquad (5.4)$$

with the zero initial conditions $q_m(0) = \dot{q}_m(0) = 0$.

As a result of the change of variables (5.3), Eq. (5.2) is reduced to the form

$$2ik_0\frac{\partial V}{\partial \varsigma} + \frac{\partial^2 V}{\partial \eta_1^2} + \frac{\partial^2 V}{\partial \eta_2^2} - k_0^2\alpha_{mn}(\varsigma)\eta_m\eta_n V = 0. \quad (5.5)$$

In the initial section $\varsigma = 0$, the function $V(\varsigma, \eta, \Omega)$ coincides with $\tilde{W}_0 = \tilde{W}(\varsigma = 0, \xi, \Omega)$, since $\eta_m|_{\varsigma=0} = \xi_m$ and $\Phi|_{\varsigma=0} = 0$.

The partial beams on the frequencies $\omega + \Omega$ are described by the same Eqs. (5.5) with the coefficients independent of the frequency shift Ω. However, this does not mean that all beams are identical. Their fields depend on frequency through the variables $\eta_{1,2}$ and (parametrically) through the initial conditions $\tilde{W}_0(\xi, \Omega)$.

Equation (5.5) was thoroughly studied in [1], where its solution in the form of an expansion in Green's functions, in local plane waves, and in Hermitian functions was derived. Assuming that the functions $V(\varsigma, \eta, \Omega)$ are determined in one way or another, the complex amplitude of the packet can be written as

$$W(\varsigma, \xi, \tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} V(\varsigma, \xi + q\Omega, \Omega) \times \exp[-ik_0\Phi(\varsigma, \xi, \Omega) - i\Omega\tau]d\Omega, \qquad (5.6)$$

where ξ and q denote the two-dimensional vectors (ξ_1, ξ_2) and (q_1, q_2).

The representation of a field as a superposition of steady-state wave beams is very descriptive and makes it possible to predict some physical phenomena. For example, it is rather evident that the packet is decomposed in the transverse direction due to refractive dispersion (the dependence of the curvature of rays on frequency). Moreover, the longitudinal (time) structure of the packet's envelope and its transverse structure turn out to be interrelated; namely, the diffraction influences the dispersive longitudinal extension (smearing) [6]. It must be noted that the exact wave description shows that the diffraction and dispersion phenomena are interrelated even for homogeneous media; this is the so-called Wolf effect [16], which is beyond the scope of the applicability of the aberration-free quasi-optical approximation.

Green's function (5.5) obtained in [1] is written as

$$G_V(\varsigma, \eta', \eta) = \frac{k_0}{2\pi i\sqrt{D_S}} \times \exp\left\{\frac{ik_0}{2}S_{mk}^{-1}(P_{mn}\eta_k'\eta_n' - 2\eta_m\eta_k' + \dot{S}_{nk}\eta_n\eta_m)\right\}, \qquad (5.7)$$

where (ς, η) is the source point, $(\varsigma = 0, \eta')$ is the observation point, and $D_S = \det\|\hat{S}\|$. The matrices $\hat{P}$ and $\hat{S}$ define an arbitrary ray $\eta_m = P_{mn}\eta_n' + S_{mn}\dot{\eta}_n'$ (certainly, in the small-angle approximation) that originates from the plane of the initial aperture ($\varsigma = 0$). These matrices are the solutions of the identical differential equations

$$\ddot{P}_{mn} + \alpha_{mk}P_{kn} = 0, \quad \ddot{S}_{mn} = \alpha_{mk}S_{kn} = 0, \qquad (5.8)$$

with different initial conditions

$$P_{mn}(0) = \delta_{mn}, \quad \dot{P}_{mn}(0) = 0,$$

$$S_{mn}(0) = 0, \quad \dot{S}_{mn}(0) = \delta_{mn}.$$

Green's function of the quasi-optical Eq. (4.5) for the packet envelope (i.e., the solution with the initial conditions $W_0 = \delta(\xi - \xi')\delta(\tau - \tau')$) is obtained by substituting the function $V(\varsigma, \eta, \Omega) = G_V\exp(i\Omega\tau')$ into (5.6), where the exponential factor is the Fourier transform of the unit impulse. Integrating with respect to Ω, we obtain

$$G_W = \frac{1}{(2\pi)^{3/2}}\frac{k_0}{\sqrt{iD_S B}} \times \exp\left\{\frac{ik_0}{2}S_{mk}^{-1}(P_{mn}\xi_k'\xi_n' - 2\xi_m\xi_k' + \dot{S}_{nk}\xi_m\xi_n) - \frac{i}{2B}\left[\tau - \tau' + k_0\left(\dot{q}_i\xi_i + S_{mk}^{-1}q_m\xi_k' - \frac{1}{2}S_{mk}^{-1}\dot{S}_{nk}(\xi_m q_n + \xi_n q_m)\right)\right]^2\right\}, \qquad (5.9)$$

where

$$B = k_0\left(S_{mk}^{-1}\dot{S}_{nk}q_m q_n - \int_0^\varsigma \beta(\varsigma)d\varsigma\right).$$

If we recall that $W(\varsigma, \xi, \tau)$ is the slow and smooth envelope of the packet, it may seem that the point source function G_W has no physical meaning. However, this is generally not the case. The function G_W provides a rather good approximation of Green's function of the exact wave equation in a certain neighborhood of the reference ray. However, the main property of G_W is the possibility of representing the solution of (4.5) in the form of the integral expansion

$$W(\varsigma, \xi, \tau) = \int G_W(\xi', \tau'; \varsigma, \xi, \tau) \times W_0(\xi', \tau') d\xi_1' d\xi_2' d\tau', \tag{5.10}$$

where $W_0(\xi, \tau) = W(\varsigma = 0, \xi, \tau)$.

Thus, the initial value problem for the partial differential Eq. (4.5) is reduced to solving the system of ordinary differential Eqs. (3.1), (5.4), (5.8) (which are independent of $W_0(\xi, \tau)$) with the subsequent calculation of integral (5.10) for all sections of the propagation trajectory $\varsigma = \varsigma^*$ for "arbitrary".[7] This procedure is convenient for numerical computer calculations of wave fields; however, it is inconvenient for analytical analysis aimed at revealing the characteristic features of the propagation of packets in inhomogeneous media. In part, these features, which are common for packets and beams, were discussed at the end of Section 4. Below, we discuss one more effect that is specific to short packets ($\Lambda_\parallel \ll L_\varepsilon$).

6. WAVE PRECESSION OF A PACKET IN MEDIA WITH A NEGATIVE DISPERSION OF THE GROUP VELOCITY

In this section, we consider a particular case of the propagation of a short wave packet along a plane trajectory in a medium with a negative dispersion of the group velocity ($\partial v_g/\partial\lambda < 0$, $\gamma > 0$). In this case, $T = 0$, $\mathbf{g}_1 = \mathbf{n}$, $\mathbf{g}_2 = \mathbf{m} = \text{const}$, $\alpha_{12} = 0$, $\delta_2 = 0$, and Eq. (4.5) allows for the separation of variables

$$W(\varsigma, \xi_1, \xi_2) = W_1(\varsigma, \xi_1, \tau) W_2(\varsigma, \xi_2),$$

$$2ik_0\left(\frac{\partial W_1}{\partial \varsigma} + k_0\delta(\varsigma)\xi_1\frac{\partial W_1}{\partial \tau}\right) + \frac{\partial^2 W_1}{\partial \xi_1^2} + \gamma(\varsigma)\frac{\partial^2 W_1}{\partial \tau^2} - k_0^2\alpha_1(\varsigma)\xi_1^2 W_1 = 0, \tag{6.1}$$

$$2ik_0\frac{\partial W_2}{\partial \varsigma} + \frac{\partial^2 W_2}{\partial \xi_2^2} - k_0^2\alpha_2(\varsigma)\xi_2^2 W = 0,$$

$$\delta = \varepsilon\frac{\partial K}{\partial \omega}, \quad \alpha_1 = 3\varepsilon K^2 - \frac{1}{2}\frac{\partial^2 \varepsilon}{\partial n^2}, \tag{6.2}$$

$$\alpha_2 = -\frac{1}{2}\frac{\partial^2 \varepsilon}{\partial m^2}.$$

As is seen, in the case when the variables in (4.5) can be separated, the problem on the three-dimensional wave packet is reduced to solving Eq. (6.2) for a two-dimensional wave beam and Eq. (6.1) for a two-dimensional packet. Formally, (6.1) can be considered as an equation for the three-dimensional wave beam (where ς is the longitudinal and ξ and τ are the transverse coordinates) propagating in a lens-like medium. In this case, the effective distributed lens is cylindrical (collecting in the plane (**n**, **s**) for $\alpha_1 > 0$ and diverging for $\alpha_1 < 0$.[8] With respect to the time coordinate τ, the equivalent lens is neither focusing nor defocusing.

Were it not for the operator $\hat{G} = \xi_1\partial_\tau$ in Eq. (6.1), the variables could be separated further, and the diffraction and dispersion effects could be completely uncoupled as in the case of a homogeneous medium. The operator $\hat{G}$, which is proportional to $\partial K/\partial\omega$, is responsible for the Newtonian dispersion of waves in inhomogeneous media; it relates the longitudinal evolution of the packet with the transverse one, and the type of this relationship essentially depends on the sign of γ. For $\gamma > 0$, the differential operators in (6.1) can be made symmetric by the change of variables

$$\sqrt{\varepsilon}d\varsigma = ds, \quad \xi_1 = u_1, \quad \tau = -\sqrt{\gamma}u_2,$$

$$W = \left(\frac{1}{\gamma}\right)^{1/4}\Psi(s, u_1, u_2) \times \exp\left\{ik\left(\frac{\dot{\gamma}}{4\gamma}u_2^2 + \kappa(s)u_1u_2\right)\right\},$$

$$k = k_0\sqrt{\varepsilon}, \quad \kappa(s) = \frac{k}{2\sqrt{\gamma}}\frac{\partial K}{\partial \omega},$$

where the dot denotes the derivative with respect to the arc length s, and coordinate u_2 is directed along the group velocity vector. This change of variables reduces (6.1) to the form

$$2ik\left(\frac{\partial \Psi}{\partial s} - \kappa(s)e_{mn}u_m\frac{\partial \Psi}{\partial u_n}\right) + \Delta_u\Psi - k^2\beta_{mn}(s)u_mu_n\Psi = 0,$$

[7] The word "arbitrary" is in double quotation marks for the following reason. Formally, for every function $W_0(\xi, \tau)$, (5.10) is a solution of Eq. (4.5). However, only the initial conditions satisfying the constraints formulated at the beginning of the paper have a physical meaning.

[8] α_1 is negative when the packet propagates along the so-called Pedersen ray. This case is not very interesting from the practical point of view, since the signal along this ray attenuates exponentially.

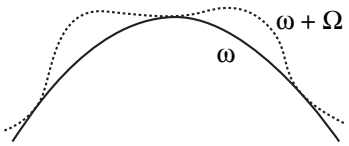

Fig. 1. Oscillation of the central ray of a partial wave beam (dotted line) at the frequency $\omega + \Omega$ relative to the reference trajectory (solid line) described by the equation $\xi = q\Omega$, $\ddot{q} + \alpha q = -\delta$.

$$e_{mn} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \beta_{11} = 3(K^2 - \kappa^2) - \frac{1}{2\varepsilon}\frac{\partial^2 \varepsilon}{\partial n^2}, \quad (6.3)$$

$$\beta_{12} = \beta_{21} = -\frac{3\kappa\dot{\gamma}}{4\gamma} + \frac{1}{2}\dot{\kappa},$$

$$\beta_{22} = \kappa^2 + \frac{\ddot{\gamma}}{2\gamma} - \frac{3\dot{\gamma}^2}{4\gamma^2}.$$

Passage to the new variables in (6.3) corresponds to making a "photograph" of the distributions of the field. It seems that this method of observation is more difficult to implement in experiments than the signal scanning with respect to time; however, it is more convenient for verbal descriptions of changes of the wave packet envelope.

After symmetrizing the operator $\hat{G} = \xi_1 \partial_\tau$, a quadratic phase corrector appears in (6.3), which is proportional to κ^2. It plays the role of a focusing lens in the plane (u_2, s); in other words, it compresses the packet in the longitudinal direction. The "focal power" of the effective lens in the plane (u_1, s) is reduced.[9]

The fact that the longitudinal diffusion of the packet is compensated seems rather strange. Indeed, in a homogeneous medium with $\gamma > 0$, the high-frequency harmonics of the packet spectrum propagate with a higher group speed; due to this fact, the packet smears and becomes frequency-modulated (the current frequency at the leading edge of the impulse is greater and at the trailing edge less than the average one). In a straight wave duct, the packet bleeds even more rapidly, since the dispersion due to the medium is combined with the waveguide dispersion of the same type. A bend of the wave duct results in an additional focusing [7], the duct becomes narrower, and its waveguide dispersion enhances even more. Then, which is the cause of the longitudinal diffusion compensation? The cause is in the transverse refractive dispersion. The trajectories of the steady-state phase of high-frequency harmonics deviate from the reference trajectory and pass a longer way until they touch it. As a result, an additional delay of high-frequency harmonics occurs, which leads to the compensation of the longitudinal diffusion of the packet. Figure 1 illustrates this compensating mechanism by way of a two-dimensional ($\partial/\partial\xi_2 = 0$) packet. It is seen that the refractive (transverse) dispersion does not lead, in focusing media, to a regular deviation of the packet from the reference trajectory.

The operator $\hat{D} \sim e_{mn} u_m \partial/\partial u_n$ in (6.3) is similar to the projection operator of the angular momentum in quantum mechanics (in cylindrical coordinates, $\hat{D} \sim \partial/\partial\varphi$). It is clear that it can be eliminated by changing the frame of reference to the reference frame (x_1, x_2) that rotates with the angular speed $-\kappa(s)$:

$$x_m = R_{mn}(\vartheta) u_n, \quad \hat{R} = \begin{pmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{pmatrix}, \quad (6.4)$$

$$\frac{d\vartheta}{ds} = -\kappa = -\frac{1}{2} k v_g \frac{\partial K}{\partial \omega} \Big/ \sqrt{k \frac{\partial v_g}{\partial \omega}},$$

$$2ik\frac{\partial \Psi}{\partial s} + \Delta\Psi - k^2 a_{mn}(s) x_m x_n \Psi = 0, \quad (6.5)$$

$$a_{mn} = R_{mp}(\vartheta) R_{nq}(\vartheta) \beta_{pq}.$$

Thus, a two-dimensional packet is described by the same equation as the wave function of an electron in a parabolic potential well. However, the well rotates with frequency $-\kappa$, and its depth and eccentricity generally varies. The solution of Eq. (6.5) is thoroughly studied in [1].

If we return to the frame of reference (u_1, u_2) associated with the normal **n**, then the rotation of the envelope structure superimposes on the evolutionary picture of the packet propagation and the variation of its longitudinal and transverse size. Since **n** rotates with the angular speed $\mathbf{\Omega}_n = v_g K \mathbf{m}$ with the motion along the reference ray, then the speed of rotation of the wave packet envelope relative to the medium is

$$\mathbf{\Omega}_p = \left(1 + \frac{\kappa}{K}\right)\mathbf{\Omega}_n. \quad (6.6)$$

In order to separate the wave precession[10] in the pure form eliminating the accompanying effects (oscillation of the transverse size and longitudinal smearing), we consider, by way of example, a packet captured by a plasma circular wave duct.

[9] We can give little attention to the dependence of the tensor β_{mn} on $\dot{\gamma}$ and $\dot{\kappa}$, since it is a result of a nonuniform scale normalization and does not lead to a longitudinal compression of the packet nor to its smearing (extension).

[10] The term *precession* is used for the description of various phenomena related to the superposition of one rotational motion on another. These are the stability of a top, evolution of the orbit of Mercury, precedence (the word-for-word translation of *praecessio*) of equinoxes, etc. The wave precession does not fall out of this list, since in this case we deal with the rotation of the packet envelope structure against the background of the rotation of the propagation direction.

6.1. A Packet in a Circular Duct

Consider the propagation of a packet in axially symmetric isotropic plasma; the approximate dependence of its permittivity on the radius, $\varepsilon(\omega, r) = 1 - \omega_p^2(r)/\omega^2$, is shown in Fig. 2.

A medium of this type is a wave duct for the modes that propagate in the azimuth direction and are located in a neighborhood of the ray $r = \rho$ with the radius defined by the equation $\varepsilon'(\rho) + 2\varepsilon/\rho = 0$. If we take this ray as the reference one for the curvilinear reference frame ($s = \rho\varphi$, $\xi = \rho - r$), then the coefficients in (6.1) and (6.3) are constants

$$\gamma = \frac{1}{v_g^2}\left(1 - \frac{v_g^2}{c^2}\right), \quad v_g = c\sqrt{\varepsilon(\rho)},$$

$$\delta = -\frac{2}{\rho\omega}, \quad \alpha_1 = \varepsilon\left(\frac{3}{\rho^2} + \alpha\right),$$

$$\alpha = -\frac{\varepsilon''(\rho)}{2\varepsilon(\rho)} > 0, \quad \kappa = -\frac{1}{\rho\sqrt{1 - v_g^2/c^2}}, \tag{6.7}$$

$$\beta_{11} = \alpha - \frac{3}{\gamma c^2\rho^2}, \ \beta_{22} = \frac{1}{\gamma v_g^2\rho^2}, \ \beta_{12} = \beta_{21} = 0.$$

It is seen that, due to dispersion, the medium is a "potential well" for the packet not only in the radial, but also in the azimuth direction. If

$$\alpha = \frac{1 + 3v_g^2/c^2}{\rho^2(1 - v_g^2/c^2)},$$

then this "well" is symmetric ($\beta_{mn} = \beta\delta_{mn}$) and, therefore, invariant under rotation (6.4). Taking this fact into account, Eq. (6.5) for the two-dimensional wave packet can be written as

$$2i\frac{\partial Q}{\partial l} + \frac{\partial^2 Q}{\partial y_1^2} + \frac{\partial^2 Q}{\partial y_2^2} - (y_1^2 + y_2^2)Q = 0, \tag{6.8}$$

where the dimensionless coordinates

$$l = \frac{s}{\rho\sqrt{(1 - v_g^2/c^2)}},$$

$$y_{1,2} = \frac{\sqrt{k/\rho}}{(1 - v_g^2/c^2)^{1/4}}x_{1,2}$$

are used.

The properties of Eq. (6.8), which is similar to the Schrödinger equation for the two-dimensional linear oscillator, are thoroughly studied. Its solutions can be represented as a superposition of steady-state modes (described by two-dimensional Hermitian functions),

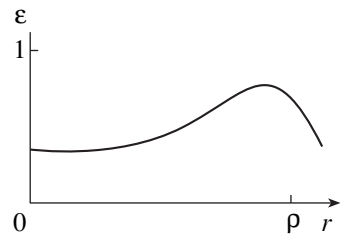

Fig. 2. The radial profile of permittivity in a circular plasma wave duct.

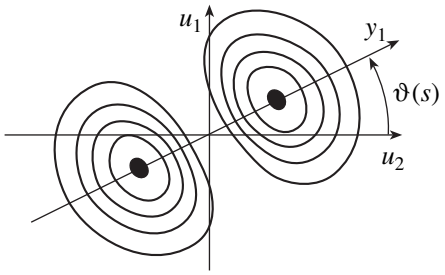

Fig. 3. Wave precession of the mode Q_{10} whose structure in the direction of the axis y_1 is described by the first Hermitian function and in the direction of y_2 by the zero Hermitian function: $Q_{10} \sim y_1\exp(-y_1^2/2 - y_2^2/2)\exp(-2il)$.

which precess in the coordinate representation (u_1, u_2) without changing their shape (Fig. 3).

The angular speed of the wave packet envelope (6.6) with respect to the medium (both for the example above and in the general case of the isotropic inhomogeneous plasma) can be written as

$$\mathbf{\Omega}_p = \left(1 - \frac{1}{\sqrt{1 - v_g^2/c^2}}\right)\mathbf{\Omega}_n, \tag{6.9}$$

where $\mathbf{\Omega}_n$ is the rotation speed of the principal normal $\mathbf{n}$ to the reference ray.

6.2. Analogy with the Thomas Precession

Expression (6.9) has a simple geometric meaning in four-dimensional space-time. It is easy to show that $\mathbf{\Omega}_p$ coincides with the angular speed of the rotation of the vector as it is pseudoparallelly translated along the world line (or along the ray of the space-time geometrical optics). Let a material point (or the center of the packet) move on a circle of the radius a with the angular speed $\mathbf{\Omega}_n$ (the tangential speed is $v = \mathbf{\Omega}_n a$). Consider the imaginary variable $z = ut$, $u = ic$. Then, the world line is described by the equation of a spiral in the space with a formally Euclidean metric:

$$x = a\cos(\Omega_n t), \quad y = a\sin(\Omega_n t), \quad z = ut.$$

The length element and the torsion of this curve are given by equations

$$ds = \sqrt{u^2 + v^2}dt, \quad T = \frac{1}{a}\frac{uv}{u^2 + v^2}.$$

An orthogonal curvilinear frame of reference can be defined in the neighborhood of this curve such that its basis rotates according to the Rytov law (3.3); in the new notation, this law has the form

$$\frac{d\theta}{dt} = -T\sqrt{u^2 + v^2} = -\frac{\Omega_n}{\sqrt{1 - v^2/c^2}}. \qquad (6.10)$$

The description of physical processes in the orthogonal frame of reference associated with the world line is equivalent to changing to the noninertial frame of reference S' attached to the particle (or packet) with the diagonal metric tensor g^{ik}. The condition $g^{0k} = 0$ ($k \neq 0$) testifies that there is no field of Coriolis forces in S'. This fact gave Thomas [17] (see also [18]) a reason to assert that the axis of a gyroscope with three degrees of freedom must be at rest in S' and must precess with the frequency $\mathbf{\Omega}_p$ in the laboratory frame of reference.

The question of the response of actual gyroscopes to geometric properties of the world lines of their suspension points remains open. As for the wave packets, an "adequate response" is characteristic only of those that propagate in inhomogeneous plasma. It is easy to show that in the Minkowski world plasma is the only isotropic medium (for electromagnetic waves). Let us write down the collinearity condition for two four-dimensional vectors, the wave vector $(k_0; \mathbf{k})$ and the group velocity $(\tilde{\gamma}; \tilde{\gamma}\mathbf{v}_g/c)$, where $\tilde{\gamma} = (1 - v_g^2/c^2)^{-1/2}$, as

$$\frac{k}{k_0} = \frac{v_g}{c} \longrightarrow \omega\frac{\partial\omega}{\partial k} = c^2 k \longrightarrow \omega^2 = c^2k^2 + \omega_0^2, \quad (6.11)$$

where ω_0^2 is the integration constant, which must be positive due to the relativistic causality principle ($v_g < c$). The last of the equalities (6.11) expresses the dispersion law of electromagnetic waves in plasma and of fast de Broglie waves.

In the general case, the wave precession is not a relativistic phenomenon: it manifests itself in arbitrarily slow motions. For example, a packet of Langmuir (longitudinal) waves in plasma with the dispersion law of the form $\omega^2 = \omega_{pe}^2 + 3k^2V_{Te}^2$ ($V_{Te} \ll c$) precesses with the frequency Ω_p determined by the same formula (6.9) in which c must be replaced by $\sqrt{3}\,V_{Te}$. In media with a positive dispersion of the group velocity, the wave precession is absent. The rotational motion is replaced by "rotation by an imaginary angle" as in the Lorentz transformation, and the refractional (transverse) dispersion results in an enhancement of the longitudinal diffusion of the packet.

7. CONCLUSION

Equation (4.5) obtained in the aberration-free quasi-optical approximation is very convenient for describing quasi-harmonic fields of wave packets and time-dependent beams in arbitrary (without one or another symmetric property) smoothly inhomogeneous media. It makes it possible to recalculate the wave process from one section of the curvilinear propagation trajectory $s = s_b$ into any other section $s = s_e$ ($s_e \gtrless s_b$) without calculating the fields in intermediate domains. It is sufficient to solve on the interval $s \in (s_b, s_e)$, the system of ordinary differential equations that determine the parameters of Green's function or of eigenfunctions of other expansions (for example, in locally plane waves [1] or in generalized three-dimensional Hermitian functions [19]).

It is nontrivial to determine the applicability conditions of the aberration-free quasi-optical approximation. In this paper, we gave only the necessary local constraints on the reference spatial and time gauge of the processes whose violation can be discovered within the framework of approximation (4.5). However, there exist more insidious, gradually accumulating, violations of the aberration-free quasi-optical approximation. They can be detected only by comparing with the results of a stricter theory. A similarly difficult problem of obtaining sufficient applicability conditions is also characteristic of geometrical optics and its generalizations, such as the Kravtsov–Ludwig reference function method [14] and the Maslov method [20]. It must be noted that quasi-optics has an advantage over other short-wave asymptotics consisting in the following: for quasi-optics, a way is directed not only for detecting integral violations, but also for their elimination.

For optical systems, the powerful diffraction theory of aberrations [8] has been developed. In this theory, the wave field is represented as an integral convolution of an "ideal image" (the aberration-free approximation W) with the so-called pass function of system F. The result given by the theory can greatly differ from W both quantitatively and qualitatively (structurally). Quasi-optics can be adapted to the generalization of the diffraction theory of aberrations to arbitrary smoothly inhomogeneous media (the requirement of the smallness of aberrations of discrete phase correctors is replaced by their smoothness). An example of such a generalization to the case of two-dimensional wave beams was made in [21], where it was shown that the parameters of the pass function F (as well as the parameters of Green's function) could be determined by solving an extended system of ordinary differential equations.

Thus, (4.5) is not only the equation that approximately describes wave fields. Its solutions remain useful even beyond the scope of their immediate applicability. It must be noted that such inhomogeneous media as the ionosphere, the magnetosphere, or plasma in controlled thermonuclear reactors differ substantially from optical lines and light guides, whose construction falls in the field of high technology. Aberrations of "lenses" and

"prisms" that are equivalent to such media are often large and highly diversified. Moreover, no complete classification of such aberrations exists. For these reasons, the diffraction theory of aberrations cannot be directly extended to quasi-optics of inhomogeneous media. The development of a quasi-optical analogue of the diffraction theory of aberrations requires additional justification and somewhat different approaches.

Another advantage of quasi-optical equations is that they can be extended to nonlinear (originally inhomogeneous) media. However, in this case, the reference trajectory depends on the intensity and structure of the wave field [22–24].

In conclusion, we note that the applicability of Eq. (4.5) is not restricted to electrodynamics; it remains valid (up to the notations) for waves of arbitrary nature.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project nos. 98-02-17019 and 00-15-96734.

The authors are grateful to V.B. Gil'denburg, V.E. Semenov, and the participants of the seminar of Miller's school for helpful remarks.

REFERENCES

1. G. V. Permitin and A. I. Smirnov, Zh. Éksp. Teor. Fiz. **109**, 736 (1996) [JETP **82**, 395 (1996)].
2. L. A. Ostrovskiĭ, Zh. Éksp. Teor. Fiz. **51**, 1189 (1966) [Sov. Phys. JETP **24**, 797 (1967)].
3. A. G. Litvak and V. I. Talanov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **10**, 539 (1967).
4. S. A. Akhmanov, V. A. Vysloukh, and A. S. Chirkin, *The Optics of Femtosecond Pulses* (Nauka, Moscow, 1988).
5. I. G. Kondrat'ev and G. D. Malyuzhinets, in *Physical Encyclopedia* (Sov. Éntsiklopediya, Moscow, 1988), Vol. 1, p. 664.
6. A. I. Smirnov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **24**, 1503 (1981).
7. V. M. Babich and V. S. Buldyrev, *Asymptotic Methods in Problems of Diffraction of Shot Waves* (Nauka, Moscow, 1972).
8. M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford, 1969; Nauka, Moscow 1970).
9. V. I. Talanov, Doctoral Dissertation in Mathematical Physics (Gork. Gos. Univ., NIRFI, Gorki, 1967).
10. H. Haus, *Waves and Fields in Optoelectronics* (Prentice-Hall, Englewood Cliffs, 1984; Mir, Moscow, 1988).
11. G. P. Agrawal, *Nonlinear Fibre Optics* (Academic, New York, 1989).
12. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1989, 4th ed.; Pergamon, New York, 1977, 3rd ed.).
13. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media* (Nauka, Moscow, 1982; Pergamon, New York, 1984).
14. Yu. A. Kravtsov and Yu.I. Orlov, *Geometrical Optics of Inhomogeneous Media* (Nauka, Moscow, 1980).
15. S. M. Rytov, Dokl. Akad. Nauk SSSR **18**, 263 (1938).
16. E. Wolf, Phys. Rev. Lett. **56**, 1370 (1986); Yu. E. D'yakov, Kvantovaya Élektron. (Moscow) **20**, 1068 (1993).
17. L. H. Thomas, *The Kinematics of an Electron with an Axis*, Philos. Mag., Jan. (1927).
18. M. Möller, *Theory of Relativity* (Clarendon Press, Oxford, 1972); J. D. Jeckson, *Classical Electrodynamics* (Wiley, New York, 1975).
19. G. V. Pereverzev, Phys. Plasmas **5**, 3529 (1998).
20. V. P. Maslov, *Perturbation Theory and Asymptotic Methods* (Nauka, Moscow, 1988).
21. I. G. Kondrat'ev, G. V. Pepmitin, and A. I. Smipnov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **23**, 1195 (1980).
22. A. I. Smirnov and G. M. Fraĭman, Zh. Éksp. Teor. Fiz. **83**, 1287 (1982) [Sov. Phys. JETP **56**, 737 (1982)].
23. E. V. Vanin and A. I. Smirnov, Zh. Éksp. Teor. Fiz. **110**, 1136 (1996) [JETP **83**, 628 (1996)].
24. A. I. Smirnov, Fiz. Plazmy **13**, 811 (1987) [Sov. J. Plasma Phys. **13**, 468 (1987)].

Translated by A. Klimontovich

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 100–107.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 115–124.

SOLIDS
Structure

Separation of Magnetic Phases in $La_{1-x}Ca_xMnO_3$ Manganites with $0.6 \le x \le 0.9$

T. I. Arbuzova,* I. B. Smolyak, S. V. Naumov, A. A. Samokhvalov, and A. V. Korolev

Institute of Metal Physics, Ural Division, Russian Academy of Sciences,
ul. S. Kovalevskoi 18, Yekaterinburg, 620219 Russia
**e-mail: magsemi@ifm.e-burg.su*

Received May 30, 2000

Abstract—The magnetic properties of $La_{1-x}Ca_xMnO_3$ solid solutions were studied in wide temperature and magnetic field ranges, $T = 2$–600 K and $H \le 50$ kOe. Semiconducting compositions with $0.6 < x < 0.9$ exhibited an unusual temperature behavior of magnetization and susceptibility with well-defined Curie and Néel temperatures. The simultaneous manifestation of both ferromagnetic and antiferromagnetic properties is evidence of the coexistence of two collinear magnetic phases. The content of impurity ions was shown to have a strong influence on the T_C and T_N values, whereas the ratio between the ferro- and antiferromagnetic phase volumes heavily depended on thermal treatment conditions. The magnetic properties of the whole series of $La_{1-x}Ca_xMnO_3$ solid solutions can be explained in terms of the phase separation model. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The interest of researchers in perovskite-like manganites based on $LaMnO_3$ stems from the discovery of the giant magnetic reluctance effect in these compounds and their potential use, in particular, in various switching devices, magnetic storage and recording units, etc. The doping of a $LaMnO_3$ antiferromagnetic semiconductor with divalent nonmagnetic ions A = Sr, Ba, and Ca is known to increase the conductivity and ferromagnetic ordering of Mn ions. The highest T_C, conductivity, and reluctance values are characteristic of $La_{1-x}A_xMnO_3$ solid solutions in the composition range $0.2 < x < 0.4$, and most attention has therefore been given precisely to these compositions.

The complex of such manganite properties as the metal–insulator transition, giant reluctance, ferromagnetic order, and the formation of magnetic polarons is explained by either the double exchange mechanism [1] or the model of spatial phase separation [2]. Both mechanisms predict qualitatively similar properties of conducting compositions in close agreement with experiment. According to the double exchange model, compositions with low ion A contents, $x < 0.2$, should, however, have a uniform magnetic state. In compositions with $x > 0.5$, the double exchange mechanism should not play an important role because of the low conductivity. According to the spatial phase separation model, the magnetic state of solid solutions with $0.2 < x < 0.5$ should be uniform and ferromagnetic, and, at $x < 0.2$ and $x > 0.5$, the magnetic state should be nonuniform with a collinear ordering of spins in the ferro- and antiferromagnetic regions. Compositions with high and low x values should largely differ in the volume ratio between the ferro- and antiferromagnetic regions. The $LaMnO_3$–$CaMnO_3$ system, as distinguished from the other systems, forms a continuous series of solid solutions, which allows the magnetic properties of compositions with large x values to be studied. In the preceding paper [3], we studied the $La_{1-x}Ca_xMnO_3$ system in the whole composition range $0 \le x \le 1$ and observed unusual magnetic properties at $x > 0.5$. In this work, we concentrate on solid solutions with $0.6 \le x \le 0.9$. The effects of magnetic fields on the temperature behavior of magnetization and susceptibility and the influence of sample preparation conditions on the magnetic properties of manganites have been studied.

2. SAMPLES AND PROCEDURE FOR MEASUREMENTS

Polycrystalline $La_{1-x}Ca_xMnO_3$ samples with $0 \le x \le 1$ were prepared from La_2O_3 (99.9%), Mn_3O_4, and $CaCO_3$ powders of OSCh (special purity) grade. The samples were synthesized in several stages. First, the initial mixture of powders was annealed in air at 1300°C for 30 h with one intermediate grinding. Next, the samples were reground, and the powders were pressed into pellets, which were annealed at 1300°C for 50 h. The X-ray powder patterns were obtained on a DRON-2.0 diffractometer, CrK_α radiation. The samples were single-phase. All compositions except $LaMnO_3$ and $La_{0.2}Ca_{0.8}MnO_3$ had perovskite structures of cubic symmetry, and the samples with $x = 0$ and 0.8 crystallized in the orthorhombic system. The lattice parameters decreased as x increased [3]. As is known, deviations from stoichiometry in both the cationic and anionic sublattices can strongly affect the physical properties of manganites. The $La_{1-x}Ca_xMnO_3$

samples with $0.6 \le x \le 0.9$, which exhibited unusual magnetic properties, were subjected to additional annealing at 1200–1400°C followed by quenching in air. Quenching samples from high temperatures should decrease the content of oxygen. The lattice parameters of quenched $La_{0.2}Ca_{0.8}MnO_3$ samples are listed in Table 1. Annealing followed by quenching did not change lattice symmetry. Lattice parameters decreased as the temperature of quenching grew, which might be evidence of an approach to stoichiometry compared with the initial samples.

Magnetic measurements in a wide temperature range $T = 2$–600 K were performed on three units, namely, a vibrating-coil magnetometer, a Faraday magnetic balance, and a SQUID magnetometer (Quantum Design Model 1822).

Several samples were magnetically nonuniform and exhibited diffuse magnetic transitions. The determination of the Curie temperature by the Belov–Goryaga method then gave incorrect results. To reveal the trends of Curie and Néel temperature variations with the concentration of Ca ions and depending on thermal treatment conditions, the T_C values were estimated by extrapolating the sharpest decrease in $\sigma(T)$ and $\chi(T)$ in a weak magnetic field ($H < 10$ Oe), and the T_N temperatures were determined from $\sigma(T)$ and $\chi(T)$ maxima at $H = 9$ kOe.

3. EXPERIMENTAL RESULTS

Perovskite-like manganites $LaMnO_3$ and $CaMnO_3$ form a continuous series of solid solutions. The extreme stoichiometric compositions with $x = 0$ and 1 in the $La_{1-x}Ca_xMnO_3$ system show semiconductor-type conductivity and an antiferromagnetic ordering of the magnetic moments of Mn^{3+} ($S = 2$) or Mn^{4+} ($S = 3/2$) ions below T_N. Their magnetic properties are determined by superexchange interactions between Mn ions through the 3*p* oxygen wave functions. The replacement of a part of La^{3+} ions in $LaMnO_3$ by Ca^{2+} ions results in an increase in conductivity and the appearance of ferromagnetism [4]. Our $La_{1-x}Ca_xMnO_3$ samples with $x < 0.5$ had magnetic characteristics typical of ferromagnets [3]. At low temperatures, the magnetization was fairly high, which was evidence of a ferromagnetic ordering of magnetic ions. An increase in temperature caused a sharp decrease in magnetization in the region of T_C. In the same temperature region, the metal–insulator transition occurred, and reluctance reached a maximum. In the paramagnetic region ($T > T_C$), magnetization linearly increased with the magnetic field strength. The Curie and Néel temperatures characterizing the magnetic properties of the $La_{1-x}Ca_xMnO_3$ samples are listed in Table 2.

The temperature and external magnetic field dependences of solid solution magnetization in the concentration range $0.6 \le x \le 0.9$ have a more complex character. The temperature dependences of magnetization

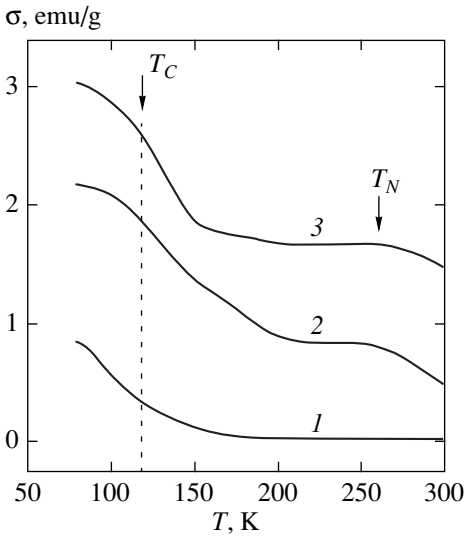

Fig. 1. Temperature dependences of magnetization for $La_{0.4}Ca_{0.6}MnO_3$ in magnetic fields $H = 50$ Oe (scale 1 : 3) (*1*), 5 kOe (*2*), and 10 kOe (*3*).

for $La_{0.4}Ca_{0.6}MnO_3$ measured in various magnetic fields and similar $\sigma(T)$ dependences for $La_{0.2}Ca_{0.8}MnO_3$ are shown in Figs. 1 and 2, respectively. In weak fields ($H < 100$ Oe), the $\sigma(T)$ dependences are typical of fer-

Table 1. Crystal lattice parameters of quenched $La_{0.2}Ca_{0.8}MnO_3$ samples

Quenching temperature, °C	Lattice parameters			Unit cell volume
	a, Å	*b*, Å	*c*, Å	*V*, Å^3
1200	5.334	7.542	5.334	214.984
1350	5.339	7.547	5.330	214.764
1400	5.320	7.523	5.326	213.159

Table 2. Curie and Néel temperatures for $La_{1-x}Ca_xMnO_3$

Composition, *x*	Quenching, °C	T_C, K	T_N, K
0.1	–	170	–
0.3	–	205	–
0.4	1350	172	–
0.6	–	111	260
	1400	115	264
0.7	1400	Diffuse transition	260
0.8	–	93	183
	1200	95	200
	1350	93	215
	1400	98	205
0.9	1400	98	129
1	–	115	–

Note: The T_C values were determined by the kink method in a weak magnetic field, and the T_N values were found from the positions of σ maxima in field $H = 9$ kOe.

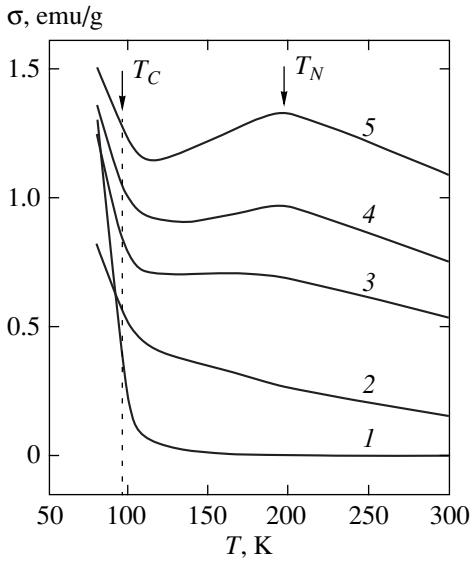

Fig. 2. Temperature dependences of magnetization for $La_{0.2}Ca_{0.8}MnO_3$ in magnetic fields H = 13 Oe (scale, 1 : 10) (*1*); 2 kOe (*2*); 5 kOe (*3*); 8 kOe (*4*); 10 kOe (*5*).

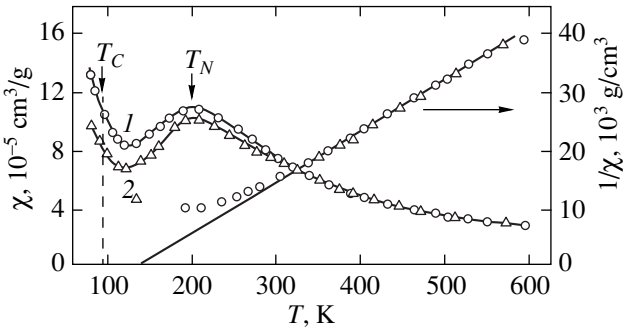

Fig. 3. Temperature dependences of susceptibility and reciprocal susceptibility for $La_{0.2}Ca_{0.8}MnO_3$ in fields H = 4.45 (*1*) and 8.9 kOe (*2*).

romagnetic substances and exhibit a sharp decrease in σ near T_C = 115 and 98 K for x = 0.6 and 0.8, respectively. An increase in field strength results in the appearance of a magnetization maximum at $T > T_C$, which becomes more pronounced as H increases. These trends are most noticeable for the composition with x = 0.8.

The temperature dependences of susceptibility, $\chi(T)$, and reciprocal susceptibility, $1/\chi(T)$, obtained for $La_{0.2}Ca_{0.8}MnO_3$ at high temperatures in fields H = 4.45 and 8.9 kOe are shown in Fig. 3. A susceptibility maximum is observed near T = 200 K, and the susceptibility is smaller in the stronger field, as in collinear antiferromagnetic compounds NiO and MnF_2 [5]. In the region of high temperatures ($T > 320$ K), the sample is in the paramagnetic state, because its susceptibility is independent of the applied magnetic field and obeys the Curie–Weiss law $\chi = N\mu_{eff}^2/3k(T - \Theta)$ with Θ = 142 K and $\mu_{eff} = 3.93\mu_B$, which is close to the calculated value $\mu_{eff} = 4.10\mu_B$.

The presence of a susceptibility maximum in the $\chi(T)$ dependence is characteristic of collinear antiferromagnetic substances near the Néel temperature T_N, of ferrimagnets with two and more magnetic sublattices and $T_N > T_C$ [6], and of the spin glass state near the freezing point T_f. To check whether or not the sample was in the spin glass state, we performed magnetization and susceptibility measurements in the temperature range 100 K < T < 300 K; the sample was cooled in the presence and in the absence of a magnetic field. No hysteresis phenomena typical of spin glasses were observed for $\sigma(T)$ and $\chi(T)$.

Theoretically, $La_{1-x}Ca_xMnO_3$ manganites can occur in the ferrimagnetic state, because the Mn^{3+} and Mn^{4+} ions have different magnetic moments, their concentrations are different in samples of all compositions except when x = 0.5, and they can occupy sites of two or several magnetic sublattices. However, if our samples with $0.6 \le x \le 0.9$ were ferrimagnets with $T_N > T_C$ (that is, if they underwent the transition from the paramagnetic to the antiferromagnetic state near T_N and from the antiferromagnetic to the ferrimagnetic state near T_C during cooling from high temperatures), then the extrapolation of the temperature dependence of reciprocal susceptibility from high temperatures $T > 320$ K would give a negative asymptotic Curie temperature Θ value, and the $1/\chi(T)$ dependence itself would be nonlinear with a concavity toward the temperature axis.

Figure 3 shows that Θ is positive, and the $1/\chi(T)$ dependence is convex toward the T axis in the critical region of temperatures of the transition from the paramagnetic to the magnetically ordered state. The shape of this dependence is characteristic of ferromagnets whose long-range order is destroyed but short-range order is retained. It would, therefore, be groundless to suggest that there are several nonequivalent magnetic sublattices whose magnetizations are balanced at $T_C < T < T_N$, or that one magnetic sublattice becomes paramagnetic above T_C, whereas antiferromagnetic ordering is retained in the other up to T_N. We believe that the maxima of $\sigma(T)$ and $\chi(T)$ are related to the Néel temperature of the antiferromagnetic collinear phase. The large positive Θ value, the sharp increase in magnetization observed when the temperature decreases below 110 K, and the nonlinear field dependences of magnetization are evidence of the important role played by ferromagnetic interactions in the region of low temperatures ($T < T_C$).

It appears that, in the intermediate ($T_C < T < 300$ K) temperature region, antiferromagnetic interactions prevail. As a result, $La_{1-x}Ca_xMnO_3$ manganites with $0.6 \le x \le 0.9$ have the properties of both ferromagnets and antiferromagnetic substances with pronounced Curie T_C and Néel T_N temperatures. They are magnetically

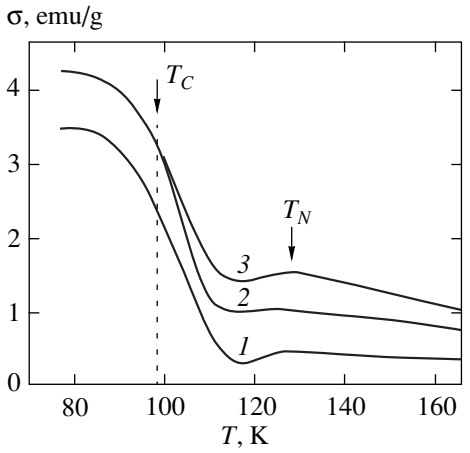

Fig. 4. Temperature dependences of magnetization for $La_{0.1}Ca_{0.9}MnO_3$ in magnetic fields H = 1.7 (*1*), 5 (*2*), and 9 kOe (*3*).

nonuniform because of spatial phase separation into the ferromagnetic and antiferromagnetic components [2], each with T_C and T_N temperatures of its own. The magnetic nonuniformity of these samples also follows from a somewhat diffuse character of the phase transition near T_C compared with a similar transition in $0.2 < x < 0.4$ samples.

In the region of $x > 0.5$, an increase in the content of Ca^{2+} ions shifts T_C and T_N to the lower temperatures. The Néel temperature decreases as x increases much faster than the Curie temperature, and the difference between T_C and T_N therefore also decreases. The temperature dependences of magnetization obtained for $La_{0.1}Ca_{0.9}MnO_3$ in various magnetic fields are shown in Fig. 4. For this sample, T_C = 98 K, and the Néel temperature T_N = 129 K is close to the T_N = 131 and 139 K values for stoichiometric $CaMnO_3$ [7] and $LaMnO_3$ [8].

The magnetic properties of manganites can be influenced not only by impurities but also by intrinsic defects. The compositions with x = 0.6 and 0.8 were subjected to thermal treatment, which decreased the content of oxygen in the samples. The temperature dependences of $La_{0.4}Ca_{0.6}MnO_3$ magnetization in field H = 9 kOe obtained prior to annealing and after annealing at 1400°C followed by quenching are shown in Fig. 5. Similar $\sigma(T)$ dependences for $La_{0.2}Ca_{0.8}MnO_3$ in field H = 10 kOe are plotted in Fig. 6. Thermal treatment has a weak effect on maximum magnetization and, accordingly, on the antiferromagnetic phase, but effectively suppresses the ferromagnetic contribution. The $\sigma(T)$ dependence for the quenched $La_{0.2}Ca_{0.8}MnO_3$ sample has a form characteristic of collinear antiferromagnets. The results of magnetization measurements in a weak field are, however, indicative of the presence of a ferromagnetic phase with a Curie temperature close to that of the initial sample.

Additional information about the presence of ferromagnetic and antiferromagnetic phases can be obtained by magnetic measurements at low temperatures. This prompted us to record the field and temperature dependences of magnetization for the $La_{0.2}Ca_{0.8}MnO_3$ sample quenched from 1350°C with the use of the SQUID magnetometer. The $\sigma(T)$ dependences obtained for this sample in fields H = 100 Oe and 10 kOe are shown in Figs. 7 and 8. In the weak field, magnetization sharply increases with decreasing temperature at T = 100 K, which closely agrees with the data given in Fig. 2 and the theory of paramagnetic–ferromagnetic state phase transitions. At H = 10 kOe, a maximum of magnetiza-

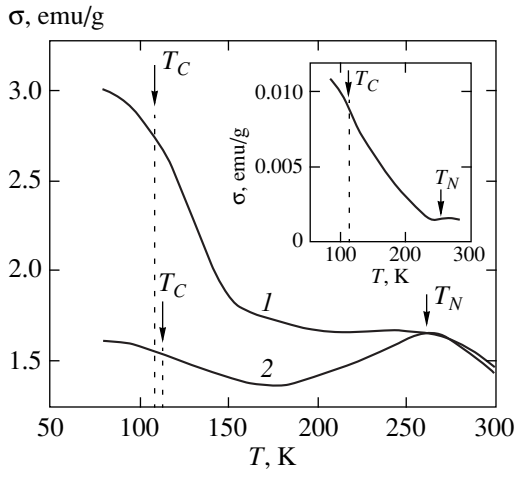

Fig. 5. Temperature dependences of magnetization in field H = 10 kOe for $La_{0.4}Ca_{0.6}MnO_3$ recorded prior to (curve *1*) and after (curve *2*) annealing at 1400°C with quenching; the $\sigma(T)$ dependence in field H = 60 Oe for the quenched sample is shown in the inset.

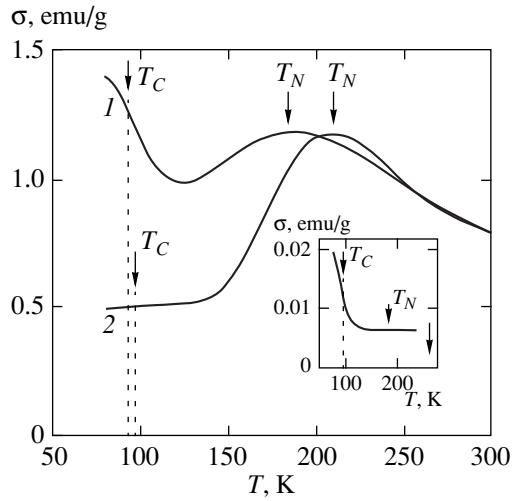

Fig. 6. Temperature dependences of magnetization in field H = 9 kOe for $La_{0.2}Ca_{0.8}MnO_3$ recorded prior to (curve *1*) and after (curve *2*) annealing at 1400°C with quenching; the $\sigma(T)$ dependence in field H = 30 Oe for the quenched sample is shown in the inset.

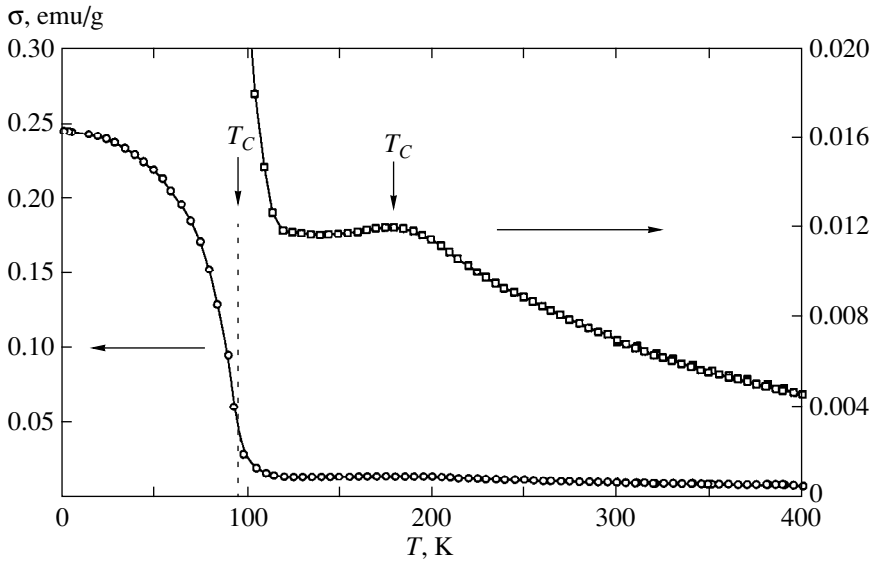

Fig. 7. Temperature dependence of magnetization in field $H = 100$ Oe for $La_{0.2}Ca_{0.8}MnO_3$ sample quenched from 1350°C; the curve at the right is drawn on an enlarged $\sigma(T)$ scale.

tion is observed at about 210 K (see Figs. 2 and 6). Below 100 K, magnetization begins to increase and reaches a value of $\sigma = 0.99$ Gs cm^3/g at $T = 2$ K. Such a magnetization value corresponds to ferromagnetic ordering of about 1% of the total number of manganese ions. Note that magnetization values at low temperatures (2 K < T < 40 K) and near $T_N = 207$ K are fairly close to each other.

The field dependences of magnetization are different for ferromagnets and antiferromagnets. The $\sigma(H)$ field dependences measured up to 50 kOe at low and medium temperatures for the polycrystalline $La_{0.2}Ca_{0.8}MnO_3$ sample quenched from 1350°C are shown in Fig. 9. At $T = 2$ K, the $\sigma(H)$ dependence is nonlinear, and magnetization hysteresis is observed. The nonzero magnetization value at $H = 0$ was obtained because the hysteresis loop at $T = 2$ K was recorded after cooling the sample at $H = 10$ kOe; that is, the sample was not demagnetized. At $T = 150$ K ($T_C < T < T_N$), the ferromagnetic phase in the paramagnetic state, like the antiferromagnetically ordered phase, should exhibit a linear $\sigma(H)$ dependence tending to zero, which agrees with our experiment. It follows that, according to the results of our magnetic measurements, the ferromagnetic properties of $La_{1-x}Ca_xMnO_3$ samples with high Ca^{2+} ion concentrations ($x > 0.5$) manifest themselves at low temperatures $T < 100$ K, whereas the antiferromagnetic properties prevail in the temperature range 100 K < T < 300 K.

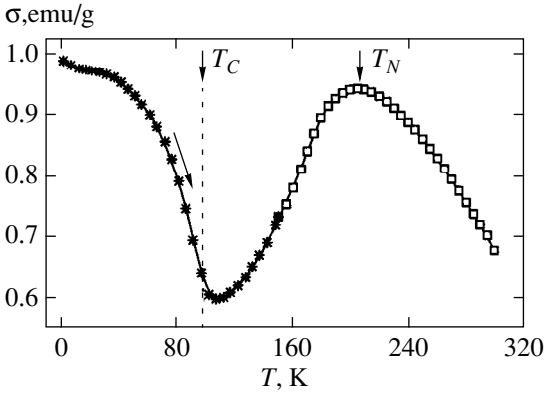

Fig. 8. Temperature dependence of magnetization in field $H = 10$ kOe for $La_{0.2}Ca_{0.8}MnO_3$ sample quenched from 1350°C.

4. DISCUSSION

Manganites $AMnO_3$ of the Rudlesden–Popper series have three-dimensional perovskite structures consisting of interpenetrating MnO_6 octahedra and A_2O_2 blocks with a NaCl structure. Magnetic manganese ions are situated in the centers of octahedra, whose vertices are occupied by oxygen ions. There can be no direct exchange between manganese ions because of their small ionic radius (about 0.7 Å). Magnetic interactions between Mn ions in nonconducting manganites involve 180° superexchange through oxygen ions. The sign and the magnitude of the superexchange interaction depend on lattice parameters, the Mn–O–Mn bond angle, and the overlap of the 3*d*–2*p* wave functions [9, 10]. The degree of covalence has a strong influence on the exchange interaction and *p*-type conductivity. The higher this degree, the stronger the superexchange interactions are.

In stoichiometric $LaMnO_3$, all manganese ions should be in the trivalent state (Mn^{3+} is a Jahn–Teller ion). The replacement of a part of La^{3+} ions by divalent ions and

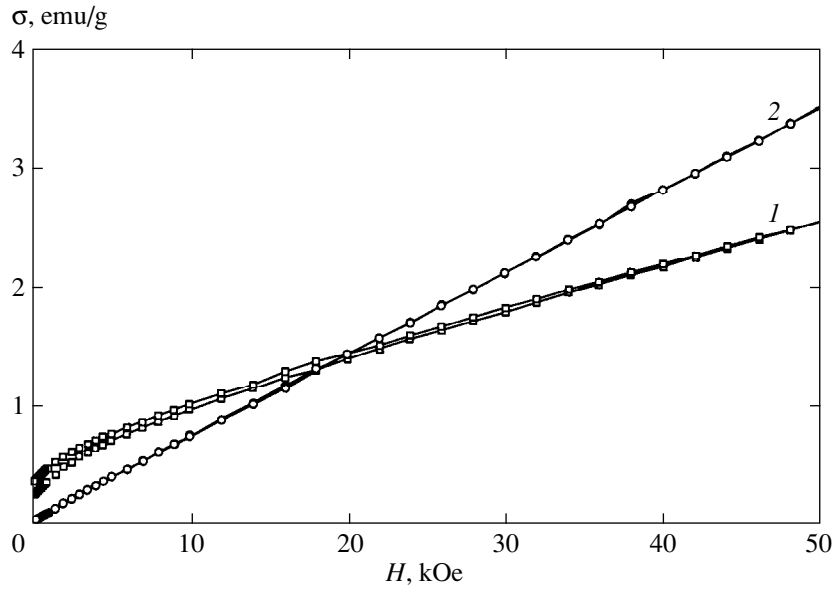

Fig. 9. Field dependences of magnetization at $T = 2$ (*1*) and 150 K (*2*) for $La_{0.2}Ca_{0.8}MnO_3$ sample quenched from 1350°C.

deviations from stoichiometry result in the appearance of Mn^{4+} ions. In perovskite-like compounds, the Mn^{3+}–O^{2-}–Mn^{3+} interactions in a cubic environment and the Mn^{4+}–O^{2-}–Mn^{4+} interactions are antiferromagnetic, whereas the Mn^{3+}–O^{2-}–Mn^{3+} interactions in an octahedral environment and the Mn^{3+}–O^{2-}–Mn^{4+} interactions are ferromagnetic. In the presence of Mn ions in different valence states, the 3*d* electron of a Mn^{3+} ion can be delocalized, and a charge transfer can occur with electron jumps between Mn^{3+} and Mn^{4+} ions. Manganites then become conducting, and the double exchange mechanism results in the ferromagnetic ordering of magnetic moments [11]. If the concentration of mobile charge carriers is high, indirect exchange of the Ruderman–Kittel–Kasuya–Yosida type should also lead to ferromagnetism. Magnetic ordering in manganites is determined by competition between ferromagnetic and antiferromagnetic interactions.

Goodenough [9] used the theory of superexchange through 180° and the experimental data on the magnetic, electric, and crystallographic properties of perovskite to construct a semiempirical phase diagram for the $La_{1-x}Ca_xMnO_3$ system. According to this diagram, compositions with $x < 0.2$ must exhibit noncollinear antiferromagnetism, the compounds with $0.2 < x < 0.4$ should be ferromagnetic, and compositions with $0.5 < x < 0.9$, antiferromagnetic. The phase diagram of the $La_{1-x}Ca_xMnO_3$ system at $T = 0$ was studied in [12] by the method of heavy fermions. When the concentration of Ca increases over the whole concentration range $0 < x < 1$, competition between double ferromagnetic exchange and antiferromagnetic superexchange results in the occurrence of several magnetic phase transitions, namely, antiferromagnet–helicoidal structure–ferromagnet–skew antiferromagnet–antiferromagnet. At elevated temperatures, kinetic energy should be taken into account. At $0.2 < x < 0.4$, the electron-phonon interaction makes the helicoidal magnetic structure less stable than the skew antiferromagnet. In antiferromagnets with nonequivalent magnetic ions, even weak magnetic fields at low temperatures can disturb the collinearity of magnetic moments. Unsaturated magnetic moments in compositions with $x < 0.2$ are often explained precisely by skew antiferromagnetism.

The magnetic properties of manganites and the giant reluctance effect are at present described by two alternative mechanisms, namely, the modified double exchange mechanism [1] and the Nagaev model of spatial separation of phases [2]. According to [2], phase separation occurs because the ferromagnetic ordering of magnetic moments is energetically more favorable for free carriers than antiferromagnetic ordering. At a not too high concentration of charge carriers, they tend to assemble in separate crystal regions and to establish ferromagnetic ordering in these regions. As a result, the crystal becomes separated into conducting ferromagnetic and insulating antiferromagnetic regions. The volume of the ferromagnetic phase increases with the concentration of charge carriers. At the critical concentration corresponding to the percolation threshold, the metal–insulator concentration transition occurs, and the whole crystal becomes ferromagnetic. The double exchange and spatial phase separation mechanisms are qualitatively similar and well explain the giant reluctance effect and ferromagnetism of conducting manganites with $0.2 \le x \le 0.4$. At the same time, in semiconducting systems with $x < 0.2$ and $x > 0.5$, the double exchange model with noncollinear ordering of magnetic moments predicts a uniform magnetic state over the whole crystal, whereas according to the Nagaev

model, the magnetic state should be nonuniform (two-phase) with a collinear arrangement of magnetic moments in each phase.

The experimental data on the magnetic properties of $La_{1-x}Ca_xMnO_3$ manganites can be explained based on the spatial phase separation model. Precisely the separation of phases results in the simultaneous manifestation of both ferromagnetic and antiferromagnetic properties in samples with $0.6 < x < 0.9$. These compositions are semiconducting, because the larger fraction of the crystal is in the insulating antiferromagnetic state. Ferromagnetic conducting regions are formed near defects such as impurity ions and vacancies in the cationic and anionic sublattices. The total magnetization of samples includes the magnetization contributions of ferromagnetic and antiferromagnetic regions. The form of the temperature dependences of magnetization and susceptibility is determined by the volume ratio between the ferro- and antiferromagnetic phases. Experiments with quenching are indicative of a decrease in the volume of ferromagnetic regions. The magnetic properties of the ferro- and antiferromagnetic regions themselves remain unchanged; indeed, annealing virtually does not change the Curie and Néel temperatures. It appears that our samples contained excess oxygen. Quenching from high temperatures decreased the content of oxygen, and the composition of the samples approached stoichiometry. This is substantiated by a decrease in lattice parameters after thermal treatment (see Table 1). Quenching from low temperatures had a weak effect on the ferromagnetic contribution and lattice parameters.

In the magnetically ordered region, the magnetization of ferromagnets is known to be at least two–three orders of magnitude larger than the magnetization of antiferromagnets. At low temperatures $T < T_C$, the total magnetization of $La_{1-x}Ca_xMnO_3$ samples is therefore largely determined by the magnetization of ferromagnetic regions. Near T_C, the magnetization of ferromagnets sharply decreases. The different shapes of the temperature dependences of magnetization in weak (H = 100 Oe) and medium (H = 10 kOe) fields are caused by the small susceptibility of antiferromagnets. On a larger $\sigma(T)$ scale, an antiferromagnetic maximum of σ near T_N in field H = 100 Oe is clearly seen (see Fig. 7). At intermediate temperatures $T_C < T < T_N$, the contributions of the ferromagnetic and antiferromagnetic phases become comparable. At high temperatures $T > T_N$, both phases become paramagnetic. Note that the asymptotic Curie temperature, Θ, determined from the experimental $1/\chi(T)$ dependences is a fictitious value. It does not characterize long-range magnetic order because it represents a superposition of the $\Theta_a < 0$ and $\Theta_f > T_C$ values for the antiferromagnetic and ferromagnetic phases, respectively. The large positive Θ value is only indicative of strong ferromagnetic couplings in solid solutions. The T_C and T_N critical temperatures are not related to structural phase transitions because the compositions with x = 0.6 and 0.8 have qualitatively similar $\sigma(T)$ dependences but differ in crystal lattice symmetry. According to [13], the lattice parameters of $La_{0.35}Ca_{0.65}MnO_3$ do not change near T_N = 160 K. The following observations lend support to the model of a nonuniform magnetic state. In a noncollinear antiferromagnet (weak ferromagnetism or a magnetic spiral), magnetization should decrease in the $T < T_N$ region rather than increase with temperature lowering. The field dependences of magnetization should either obey the law $\sigma = \sigma_s + \chi H$, where χ is the field-independent susceptibility, and σ_s is a small ($\leq 0.1\mu_B$) spontaneous magnetization, or exhibit a sharp change in σ at the critical field. These rules do not hold with $La_{1-x}Ca_xMnO_3$ ($0.6 < x < 0.9$) solid solutions.

In ferromagnets, long-range magnetic order is destroyed near T_C, whereas short-range order is retained. Applying a magnetic field favors the ferromagnetic ordering of spins and shifts T_C to the higher temperatures. In antiferromagnets, magnetic field effects are somewhat different, namely, magnetic fields favor magnetic disordering. The susceptibilities in different fields should coincide at $T = T_N$. The differences between the susceptibilities increase with temperature lowering, because spins tend to be orientated normally to the field. In antiferromagnets, the T_N temperature remains unchanged or even decreases in a strong magnetic field [5]. Such a behavior of T_N was observed for the majority of $La_{1-x}Ca_xMnO_3$ solid solutions with $0.6 < x < 0.9$. The suppression of antiferromagnetic order in solid solutions with $0.50 < x < 0.68$ by magnetic fields was also observed in [14], where an increase in H from 1 to 2 T decreased T_N approximately by 20 K. In several samples, we, however, observed shifts of σ and χ maxima to the higher temperatures in stronger fields (see Figs. 2, 3, 7, and 8). The reasons for an increase in T_N in a magnetic field and for different susceptibility values at T_N are not quite clear. These effects may, for instance, be caused by a change in the nonuniform magnetic state of a substance under field action, which would result in changes in the contributions of the ferro- and antiferromagnetic phases.

The separation of magnetic phases is also possible in solid solutions with $x < 0.2$, but the properties of the antiferromagnetic phase can then be obscured by the ferromagnetic phase. The simultaneous manifestation of ferro- and antiferromagnetic properties requires that the T_C temperature be lower than T_N, and the volume of the ferromagnetic phase be not larger than several percent. Precisely these conditions are met in compositions with $0.6 < x < 0.9$. As mentioned above, the Néel temperature lowers with a decreasing concentration of Ca^{2+} ions and becomes lower than T_C. For instance, the $La_{0.9}Ca_{0.1}MnO_3$ single crystal has T_N = 118 K and T_C = 138 K [15]. The results obtained in [16–18] substantiate the suggesting of a nonuniform magnetic state of manganites and phase separation in them. The neutron diffraction data [16] on $LaMnO_{3+\delta}$ show that, depending on δ, this compound can have antiferromag-

netic or ferromagnetic properties or occur in a mixed state with ferromagnetic clusters in an antiferromagnetic matrix. The volume of ferromagnetic regions depends on the concentration of defects (δ), which is determined by the conditions of quenching. This is in agreement with our data on the thermal treatment of $La_{1-x}Ca_xMnO_3$ samples with $0.6 < x < 0.9$. In [17], the magnetic properties of $LaMnO_{3+\delta}$ are also explained by a mixed two-phase magnetic state. In [18], the presence of ferromagnetic clusters against the background of an antiferromagnetic matrix was observed for the other limiting composition, namely, for quenched $CaMnO_{3-\delta}$ samples.

5. CONCLUSIONS

The unusual temperature behavior of magnetization and susceptibility with well-defined Curie and Néel temperatures observed for $La_{1-x}Ca_xMnO_3$ manganites in the composition range $0.6 < x < 0.9$ is determined by the spatial separation of the magnetic system into collinear antiferromagnetic and ferromagnetic phases. The volume of the ferromagnetic phase depends on the concentrations of both specially introduced admixtures and intrinsic defects. The magnetic properties of the whole $La_{1-x}Ca_xMnO_3$ system can be explained in terms of the unified phase separation model.

ACKNOWLEDGMENTS

This work was financially supported by the INTAS (grant no. 97-30253) and the Russian Federal Program "Surface Atomic Structures" (project no. 2.4.99).

REFERENCES

1. X. Wang and A. F. Freeman, J. Magn. Magn. Mater. **171**, 103 (1997).
2. É. L. Nagaev, Usp. Fiz. Nauk **166**, 833 (1996) [Phys. Usp. **39**, 781 (1996)]; Fiz. Tverd. Tela (St. Petersburg) **40**, 2069 (1998) [Phys. Solid State **40**, 1873 (1998)].
3. T. I. Arbuzova, I. B. Smolyak, S. V. Naumov, *et al.*, Zh. Éksp. Teor. Fiz. **116**, 1664 (1999) [JETP **89**, 899 (1999)].
4. A. P. Ramírez, J. Phys.: Condens. Matter **9**, 8171 (1997).
5. *Antiferromagnetism: Collection of Articles*, Ed. by S. V. Vonsovskiĭ (Inostrannaya Literatura, Moscow, 1956).
6. J. Smit and H. P. J. Wijn, *Ferrites* (Wiley, New York, 1959; Inostrannaya Literatura, Moscow, 1962).
7. E. O. Wollan and W. C. Kochler, Phys. Rev. **100**, 545 (1955).
8. G. J. Matsumoto, J. Phys. Soc. Jpn. **29**, 606 (1970).
9. J. B. Goodenough, *Magnetism and the Chemical Bond* (Interscience, New York, 1963; Metallurgiya, Moscow, 1968).
10. J. B. Goodenough, Phys. Rev. **100**, 564 (1955).
11. C. Zener, Phys. Rev. **82**, 403 (1951).
12. Jie Jiang, Jinming Dang, and D. J. Hing, Phys. Rev. B **55**, 8973 (1997).
13. M. R. Ibarra, J. M. De Teresa, J. Blasco, *et al.*, Phys. Rev. B **56**, 8252 (1997).
14. P. Shiffer, A. P. Ramírez, W. Bao, *et al.*, Phys. Rev. Lett. **75**, 3336 (1995).
15. G. Biottean, M. Hennion, F. Moussa, *et al.*, Physica B (Amsterdam) **259–261**, 826 (1999).
16. A. N. Pirogov, A. E. Teplykh, V. I. Voronin, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **41**, 103 (1999) [Phys. Solid State **41**, 91 (1999)].
17. I. O. Troyanchuk, H. Szymczak, N. V. Kosper, *et al.*, J. Solid State Chem. **130**, 171 (1997).
18. Z. Zeng, M. Greenblatt, and M. Croft, Phys. Rev. B **59**, 8784 (1999).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 108–122.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 125–141.

SOLIDS
Electronic Properties

Vectorial Wave Interaction in Cubic Photorefractive Crystals

B. I. Sturman[a, *], E. V. Podivilov[a], V. P. Kamenov[a], E. Nippolainen[b], and A. A. Kamshilin[b, **]

[a]*International Institute of Nonlinear Studies, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia*
[b]*Department of Physics, Joensuu University, P.O. Box 111, Joensuu, Finland*
**e-mail: sturman@iae.nsk.su*
***e-mail: kamshilin@ns.joensuu.fi*
Received January 11, 2000

Abstract—An analytical theory of the polarization-degenerate interaction of light waves in cubic photorefractive crystals of the 23 and $\bar{4}3m$ space symmetry groups is proposed. The theory, based on a systematic use of the spatial symmetry properties and the σ-matrix apparatus, provides for a unified description of the polarization and orientation properties of the wave interactions, including the AC and DC methods for enhancement of the nonlinear response and the influence of optical activity, elastooptical effect, and spatial inhomogeneity. The application of the theory to the description of the properties of photorefractive crystals $Bi_{12}TiO_{20}$ and $Bi_{12}SiO_{20}$ shows a good agreement with experimental data. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The interaction of light waves under the conditions of a photorefractive nonlinearity was extensively studied by theoretical and experimental methods [1–3]. The most pronounced photorefractive interactions take place in ferroelectric crystals ($LiNbO_3$, $BaTiO_3$, SBN, etc.) possessing significantly anisotropic properties. Accordingly, the wave surfaces of the two intrinsic optical modes (usually, the ordinary and extraordinary waves) are well separated in the **k**-space. This feature allows the nonlinear interactions in ferroelectrics to be described by equations for the scalar amplitudes, thus speaking of a scalar interaction of the light waves.

The situation dramatically changes in the case of isotropic photorefractive media, the most characteristic examples being offered by cubic crystals of the sillenite family ($Bi_{12}SiO_{20}$, $Bi_{12}TiO_{20}$, $Bi_{12}GeO_{20}$) and some cubic semiconductor crystals (GaAs, CdTe). In these media, where a distance between the wave surfaces is (even in the presence of an electric field) considerably smaller than in ferroelectrics, the introduction of scalar wave amplitude is often senseless. Instead of the scalar interaction of waves, we have to consider their vectorial interaction in which changes in the energy and polarization of the interacting waves are nonseparable.

The problem of describing the vectorial interaction of waves initially appeared as being not of large significance, since the photorefractive nonlinearity of cubic crystals is usually rather small. However, subsequent investigations [2–5] showed that the nonlinearity can be markedly increased for the waves interacting in the presence of an alternating electric field (AC method) or for the waves with a small frequency detuning interacting in the presence of a constant field (DC method). At present, the gain in cubic photorefractive crystals reaches a level of 10^2 cm^{-1}, which is comparable to the values typical of ferroelectrics. The phenomenon of the enhanced photorefractive nonlinearity of cubic crystals is already manifested in a number of strong optical effects such as phase conjugation, optical oscillations, generation of surface light waves, nonlinear scattering etc. [6–9].

An important advantage of the cubic crystals is a fast photorefractive response [2, 3]. The response time under continuous illumination conditions usually does not exceed 10^{-2} s, which is smaller by at least two orders of magnitude as compared to the values in ferroelectrics. The fast photorefractive response, in combination with the possibility of increasing its magnitude, makes the cubic crystals promising media for a number of applications.

Owing to the results of investigations conducted for may years, the principal elements that have to be included into the theory of vectorial interactions in cubic crystals have became clear. Besides the aforementioned increase in the nonlinear response, these elements are as follows:

(i) The so-called elastooptical contributions to the nonlinear variation of the optical permittivity, which must be taken into account together with the traditional electrooptical contributions to photorefraction [10–14].

(ii) The effect of optical activity for crystals of the $\bar{4}3m$ group (sillenites).

(iii) Various optical configurations, differing by the orientation of external electric fields and wave vectors relative to the crystallographic axes.

(iv) The arbitrary selection of the input light beam polarization.

Despite all clarity with respect to separate aspects of the description of vectorial interaction, no consistent theory has been developed. Several existing models, which either ignore the polarization degree of freedom [15, 16] or appeal to thin nonlinear crystals [17–19], can by no means provide for the interpretation of the available experimental material. Nor does the numerical modeling approach [7, 20–22]. The presence of many variable parameters (different in physical meaning) can hardly allow us to consistently consider realistic variants and formulate adequate notions of the properties of these interactions. Moreover, numerical methods are usually insensitive to small parameters involved in a theory.

This study presents an attempt at combining the aforementioned fragments so as to obtain a consistent analytical theory of vectorial interaction and apply the resulting model to the most important AC and DC methods for enhancement of the photorefractive response. The relationships derived below are compact and sufficiently general. This is achieved by using the properties of spatial symmetry, applying the σ-matrix apparatus, and minimizing the number of independent parameters. Equations derived within the framework of this theory are convenient for considering various approximations and particular cases.

The structure of this article is as follows. Section 2 introduces the main relationships for the optical permittivity of cubic crystals, including linear and nonlinear terms, derives equations for the vectorial amplitudes of light waves (describing their propagation in and diffraction from a light-induced space charge grating), and describes the basic properties of these waves. The equations contain a minimum number of parameters, which are compatible with the spatial symmetry properties of cubic crystals. These parameters are specified for two basic optical configurations (equivalent configurations are indicated) and some practically important cases are considered. These data are sufficient for describing all real configurations.

In Section 3, we will derive relationships for a nonlinear photorefractive response to the application of the AC and DC enhancement methods. These relationships express the amplitude of the space charge field through the wave amplitudes, thus closing the system of nonlinear equations for the light waves. We will also discuss the applicability of the material relationships and consider the physical meaning of the parameters involved.

Section 4 is devoted to analysis of the interaction of two light waves in a cubic photorefractive crystal. First, the nonlinear system is simplified to a maximum extent by passing (quantum-mechanical analogy) to the interaction representation. Then the AC amplification of a weak wave in the presence of a strong wave is considered in the undepleted pump approximation (not trivial in the case of vectorial interaction). Finally, the features of spatial amplification related to a strong influence of the linear absorption are considered for the DC method of the photorefractive response amplification.

Section 5 gives several examples of applications of the proposed theory to the study of the angular and polarization properties of spatial amplification in crystals of the sillenite family. We will also present original experimental results for the $Bi_{12}TiO_{20}$ crystals and some published data to demonstrate the agreement between the proposed theory and the available experimental data.

Section 6 presents discussion of the results and formulates the conclusions.

2. BASIC RELATIONSHIPS

Let two light waves with wavevectors $\mathbf{k}_1$ and $\mathbf{k}_2$ propagate in a photorefractive cubic crystal along the z axis in a laboratory frame of reference. The electric field $\mathscr{E}$ of the light waves can be presented in the following form:

$$\mathscr{E} = (\mathbf{A}_1 e^{i\mathbf{k}_1\cdot\mathbf{r} - i\Omega t} + \mathbf{A}_2 e^{i\mathbf{k}_2\cdot\mathbf{r}})e^{-i\omega t} + \text{c.c.}, \quad (1)$$

where Ω is a small ($\lesssim 10^2\ \text{s}^{-1}$) frequency difference. The light intensity is modulated at a spatial frequency $\mathbf{K} = \mathbf{k}_1 - \mathbf{k}_2$ (grating vector). The photogenerated charge carriers, diffusing and drifting (migrating) in the applied field $\mathbf{E}_0$, form a lattice of the space charge field $\mathbf{E}_{sc}(\mathbf{r})$ with a period of $\Lambda = 2\pi/K$. The field vector $\mathbf{E}_{sc}$ is parallel to $\mathbf{K}$, and the field strength is determined by the interference pattern contrast $m = 2(\mathbf{A}_1 \cdot \mathbf{A}_2^*)/|\mathbf{A}_1|^2 + |\mathbf{A}_2|^2$. In a linear (with respect to m) approximation, the space charge field obeys a relationship

$$\mathbf{E}_{sc} = \mathbf{n}E_{\mathbf{K}} e^{i(\mathbf{K}\cdot\mathbf{r} - \Omega t)} + \text{c.c.}, \quad (2)$$

where $\mathbf{n} = \mathbf{K}/K$ is the unit grating vector and $E_{\mathbf{K}} \propto m$ is the lattice amplitude. The $\mathbf{E}_{sc}(\mathbf{r})$ field modulates the optical permittivity tensor $\hat{\epsilon}$ and gives rise to the nonlinear optical effects.

In cubic crystals of the 23 and $\bar{4}3m$ space symmetry groups, the permittivity $\hat{\epsilon}$ for a wave with wavevector $\mathbf{k}$ can be presented in the following form [23, 24]:

$$\begin{aligned}\epsilon_{ij} = {} & n_0^2(1 + i\alpha k^{-1})\delta_{ij} + 2i\rho n_0^2 k^{-2}\delta_{ijl}k_l \\ & + \delta\epsilon_{ij}(\mathbf{E}_0) + \delta\epsilon_{ij}(\mathbf{E}_{sc}),\end{aligned} \quad (3)$$

where n_0 is the index of refraction, α is the absorption coefficient, ρ is the optical activity coefficient, and δ_{ijl} is the antisymmetric isotropic tensor. The terms $\delta\epsilon_{ij}(\mathbf{E}_0)$ and $\delta\epsilon_{ij}(\mathbf{E}_{sc})$ describe contributions to the permittivity induced by the homogeneous field $\mathbf{E}_0$ and the spatially-oscillating field $\mathbf{E}_{sc}$. By introducing the scalar field amplitudes E_0 and E_{sc}, such that $\mathbf{E}_0 = E_0\mathbf{n}^0$ and $\mathbf{E}_{sc} = E_{sc}\mathbf{n}$, these contributions are conveniently written as follows:

$$\begin{aligned}\delta\epsilon_{ij}(\mathbf{E}_0) &= -n_0^4 r_{41} E_0 H_{ij}^{(0)}, \\ \delta\epsilon_{ij}(\mathbf{E}_{sc}) &= -n_0^4 r_{41} E_{sc} H_{ij},\end{aligned} \quad (4)$$

where r_{41} is the independent component of the electrooptical tensor r_{ijl} and $H_{ij}^{(0)} = r_{ijl}n_l^0/r_{41}$ is the real symmetric dimensionless matrix with the elements $H_{ij}^{(0)}$ depending on the direction of the unit grating vector $\mathbf{n}^0$. These elements can be determined using the fact [23, 24] that $r_{ijl} = r_{41}|\delta_{ijl}|$ (in the crystallographic system of coordinates).

The elements of the real dimensionless matrix $\hat{H}$ depend on the orientation of the unit grating vector $\mathbf{n}$. If only the obvious electrooptical contribution to $\delta\hat{\epsilon}(\mathbf{E}_{sc})$ is taken into account, then $\hat{H}(\mathbf{n}) = \hat{H}^{(0)}(\mathbf{n})$. Allowance for the elastooptical effect makes the $\hat{H}(\mathbf{n})$ expression more complicated [11, 12, 14]. In some cases (see below), the contributions to $\hat{H}$ due to the elastooptical and electrooptical effects are comparable.

Below we will restrict the consideration to a usual paraxial approximation, assuming that the angles between $\mathbf{k}_{1,2}$ and the z axis are small. Accordingly, only the x- and y-components of the wave amplitudes $\mathbf{A}_{1,2}$ and vectors $\mathbf{n}^0$ and $\mathbf{n}$ are nonzero. Using the Maxwell equations and relationships (1)–(4), one may readily derive reduced equations for $\mathbf{A}_{1,2}$ describing the diffraction of the two light waves from the field grating:

$$\begin{aligned}(\partial_z + 0.5\alpha - i\hat{G})\mathbf{A}_1 &= iE_{\mathbf{K}}\hat{V}\mathbf{A}_2,\\ (\partial_z + 0.5\alpha - i\hat{G})\mathbf{A}_2 &= iE_{\mathbf{K}}^*\hat{V}\mathbf{A}_1.\end{aligned} \tag{5}$$

Here, $\hat{G}$ is a Hermitian (2×2)-matrix describing the linear optical properties of the medium, while $\hat{V}$ is a real symmetric interaction (2×2)-matrix responsible for the wave coupling. The matrix elements G_{ij} and V_{ij} are determined by the relationships

$$G_{ij} = sE_0H_{ij}^{(0)} + i\rho\delta_{ijz}, \quad V_{ij} = sH_{ij}, \quad s = -\pi n_0^3 r_{41}/\lambda, \tag{6}$$

where ρ and sE_0 are the dimensional coefficients, λ is the light wavelength, and the indices i, j refer to the x and y axes. According to Eqs. (5), the diffraction only leads to a redistribution of the energy between two waves, their total intensity I_0 being proportional to $\exp(-\alpha z)$.

Formulation of the theory and the operations with two-dimensional vectors can be considerably simplified by using the Pauli matrix (σ-matrix) apparatus, which is widely employed in quantum mechanics [25, 26] and optics. According to this approach, three Hermitian matrices

$$\hat{\sigma}_1 = \begin{pmatrix} 0 & 1\\ 1 & 0\end{pmatrix}; \quad \hat{\sigma}_2 = \begin{pmatrix} 0 & -i\\ i & 0\end{pmatrix}; \quad \hat{\sigma}_3 = \begin{pmatrix} 1 & 0\\ 0 & -1\end{pmatrix}, \tag{7}$$

are considered as components of a vector matrix $\hat{\boldsymbol{\sigma}}$. The σ-matrices possess a number of remarkable properties [25, 26], which reduce some cumbersome calculations involving two-dimensional vectors to simple algebraic operations.

The $\hat{G}$ and $\hat{V}$ matrices can be represented as linear combinations of $\hat{\boldsymbol{\sigma}}$ and the unit matrix $\hat{1}$:

$$\hat{G} = \kappa\cdot\hat{\boldsymbol{\sigma}}, \quad \hat{V} = \nu_0\hat{1} + \boldsymbol{\nu}\cdot\hat{\boldsymbol{\sigma}}. \tag{8}$$

Here, the scalar quantity ν_0 and the three-dimensional vectors $\boldsymbol{\kappa}$ and $\boldsymbol{\nu}$ (all being real values) are defined by the relationships

$$\begin{aligned}\nu_0 &= 0.5s(H_{xx}+H_{yy}), \quad \nu_1 = sH_{xy}, \quad \nu_2 = 0,\\ \nu_3 &= 0.5s(H_{xx}-H_{yy}), \quad \kappa_1 = sE_0H_{xy}^{(0)},\\ \kappa_2 &= -\rho, \quad \kappa_3 = 0.5sE_0(H_{xx}^{(0)} - H_{yy}^{(0)}).\end{aligned} \tag{9}$$

Accordingly, set (5) acquires a canonical form

$$\begin{aligned}(\partial_z + 0.5\alpha - i\boldsymbol{\kappa}\cdot\hat{\boldsymbol{\sigma}})\mathbf{A}_1 &= iE_{\mathbf{K}}(\nu_0\hat{1} + \boldsymbol{\nu}\cdot\hat{\boldsymbol{\sigma}})\mathbf{A}_2,\\ (\partial_z + 0.5\alpha - i\boldsymbol{\kappa}\cdot\hat{\boldsymbol{\sigma}})\mathbf{A}_2 &= iE_{\mathbf{K}}^*(\nu_0\hat{1} + \boldsymbol{\nu}\cdot\hat{\boldsymbol{\sigma}})\mathbf{A}_1.\end{aligned} \tag{10}$$

Vector $\boldsymbol{\kappa}$ characterizes the linear optical properties of the medium: the $\kappa_{1,3}$ components describe the anisotropy induced by the external field, while the κ_2 component represents the optical activity. The ν_0 and $\boldsymbol{\nu}$ values determine the nonlinear coupling of the light waves. The presence of the scalar quantity ν_0 indicates that the interaction matrix $\hat{V}$ has an isotropic part not affecting the polarization properties of the waves. The $\boldsymbol{\nu}\cdot\hat{\boldsymbol{\sigma}}$ product matrix corresponds to the anisotropic interaction component. The description of the nonlinear interaction markedly simplifies in the case when $\nu \equiv |\boldsymbol{\nu}| \ll \nu_0$.

Strictly speaking, matrix $\hat{G}$ (as well as $\hat{V}$) also has an isotropic part. This part leads only to a trivial phase shift, which is the same for both waves (and is omitted in Eq. (10)).

We should like to emphasize the difference between vectors $\mathbf{A}_{1,2}$ (with the x- and y-components) and the vectors (such as $\boldsymbol{\kappa}$ and $\boldsymbol{\nu}$) defined in the three-dimensional configuration space (components of the latter values are denoted by integers 1, 2, and 3).

First, we will use Eqs. (10) to obtain the necessary information concerning the optical eigenmodes in the absence of interaction between waves (that is, for $E_{\mathbf{K}} = 0$). Assuming $\mathbf{A}_{1,2} \propto \exp(-0.5\alpha z + i\delta k z)$, we arrive at the problem of determining the eigenvalues for the correction δk to the wavevector:

$$(\boldsymbol{\kappa}\cdot\hat{\boldsymbol{\sigma}})\mathbf{A}^{\pm} = \delta k\mathbf{A}^{\pm}. \tag{11}$$

The Hermitian operator $\boldsymbol{\kappa}\cdot\hat{\boldsymbol{\sigma}}$ has two real eigenvalues, $\delta k_{\pm} = \pm\kappa$, so that the distance between two wave surfaces is 2κ. This distance is determined by the optical

activity and induced birefringence. The unit polarization vectors $\mathbf{e}^{\pm} = \mathbf{A}^{\pm}/|\mathbf{A}^{\pm}|$ corresponding to eigenvalues $\delta k_{\pm}$ are given by the relationship

$$\mathbf{e}_{\pm} = \frac{1}{\sqrt{2\kappa}}\begin{pmatrix} \sqrt{\kappa \pm \kappa_3} \\ \pm\dfrac{\kappa_1 + i\kappa_2}{\sqrt{\kappa \pm \kappa_3}} \end{pmatrix}. \quad (12)$$

These vectors are in fact orthogonal ($\mathbf{e}_+^* \cdot \mathbf{e}_- = 0$). In the general case, the optical eigenmodes possess an elliptic polarization. The orientation of the principal axes of the polarization ellipse depends on the choice of the optical configuration (see below), while the eccentricity depends on the relationship between $|sE_0|$ and $|\rho|$. For $|E_0| \ll |\rho/s|$, the polarization is close to circular, whereas in the opposite limit, the polarization is almost linear. The data presented in Table 1 show that the $|\rho/s|$ value for $Bi_{12}TiO_{20}$ is ~2.8 kV/cm, which is considerably lower than the field strengths typically used in experiment. In other words, the optical activity of $Bi_{12}TiO_{20}$ is easily suppressed by an external field. On the contrary, in experiments with $Bi_{12}SiO_{20}$, the field strength usually does not exceed the value $|\rho/s| \approx 14$ kV/cm.

If the input amplitude $\mathbf{A}(0)$ is real (that is, the incident wave is linearly polarized), then $\mathbf{A}(-E_0, z) = \mathbf{A}^*(E_0, z)$. This means that a change in the sign of E_0 leads to a reversal of the direction of polarization rotation at each point of the crystal.

Using the σ-matrix representation introduced as described above, we may readily obtain a general solution to the problem of Bragg vector diffraction from the grating with a constant amplitude. This grating can be recorded using noninteracting waves. Substituting the grating amplitude in the form $E_{\mathbf{K}} = |E_{\mathbf{K}}|e^{i\phi_0}$ into Eq. (10), we obtain

$$\mathbf{A}_{1,2}(z) = e^{-\alpha z/2} \times [\hat{T}_+(z)\mathbf{A}_{1,2}(0) + e^{\pm i\phi_0}\hat{T}_-(z)\mathbf{A}_{2,1}(0)], \quad (13)$$

where $\mathbf{A}_{1,2}(0)$ are the entering amplitudes, $\hat{T}_{\pm}(z) = (e^{i\hat{g}_+ z} \pm e^{i\hat{g}_- z})/2$, and $\hat{g}_{\pm} = \boldsymbol{\kappa} \cdot \hat{\boldsymbol{\sigma}} \pm |E_{\mathbf{K}}|(\nu_0\hat{1} + \boldsymbol{\nu} \cdot \hat{\boldsymbol{\sigma}})$. The initial functions of σ-matrices are readily reduced to linear [25, 26].

The general solution to Eq. (13) contains a large amount of information concerning the diffraction efficiency and polarization properties. This solution can be generalized so as to include the case of a partly polarized light. Earlier attempts at solving the problem of vector diffraction were restricted to strongly simplified analytical models [27, 28] and numerical computational procedures [29, 30]. Direct calculations in particular cases lead to very cumbersome expressions [31].

Table 1. Optical characteristics of $Bi_{12}SiO_{20}$ and $Bi_{12}TiO_{20}$ crystals

Parameter	$Bi_{12}SiO_{20}$ (λ = 514 nm)	$Bi_{12}TiO_{20}$ (λ = 633 nm)
n_0	2.6	2.58
ρ	38.6 deg/mm ≈ 6.74 cm^{-1}	6.5 deg/mm ≈ 1.13 cm^{-1}
r_{41}	4.51×10^{-12} m/V	4.74×10^{-12} m/V
$\|\rho/s\|$	14 kV/cm	2.8 kV/cm

In the experiment, the z axis usually coincides with a crystal symmetry axis. If this is the principal axis (e.g., [001]), the $\hat{H}^{(0)}$ and $\hat{H}$ matrices turn zero. This geometry is convenient for nonperturbing grating recording.

Optical configurations important for the wave interaction can be reduced to a basic configuration depicted in Fig. 1a. Here, the z axis coincides with the $[1\bar{1}0]$ direction, while the x and y axes are directed in the [001] and $[\bar{1}\bar{1}0]$ directions, respectively. The angles ζ and ζ_0, characterizing orientation of the $\mathbf{K}$ and $\mathbf{E}_0$ vectors in the xy plane, are measured from the [001]direction. The angle $\psi = \zeta - \zeta_0$ is convenient for the comparison with experiment (see Section 5). For the basic configuration considered, we obtain

$$\hat{H}^{(0)}(\zeta_0) = \begin{pmatrix} 0 & \sin\zeta_0 \\ \sin\zeta_0 & \cos\zeta_0 \end{pmatrix}. \quad (14)$$

Using this representation and formulas (9), components $\kappa_{1,3}$ are expressed as

$$\kappa_1 = sE_0\sin\zeta_0, \quad \kappa_3 = -0.5sE_0\cos\zeta_0. \quad (15)$$

Neglecting the elastooptical contribution, whereby $\hat{H} = \hat{H}^{(0)}(\zeta)$. In this approximation

$$\nu_0 = 0.5s\cos\zeta, \quad \nu_1 = s\sin\zeta, \quad \nu_3 = -0.5s\cos\zeta. \quad (16)$$

In [11, 13], elements of the $\hat{H}$ matrix were measured and calculated for $Bi_{12}TiO_{20}$ and $Bi_{12}SiO_{20}$ crystals with

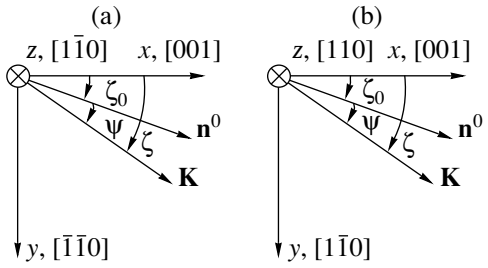

Fig. 1. Two basic optical configurations.

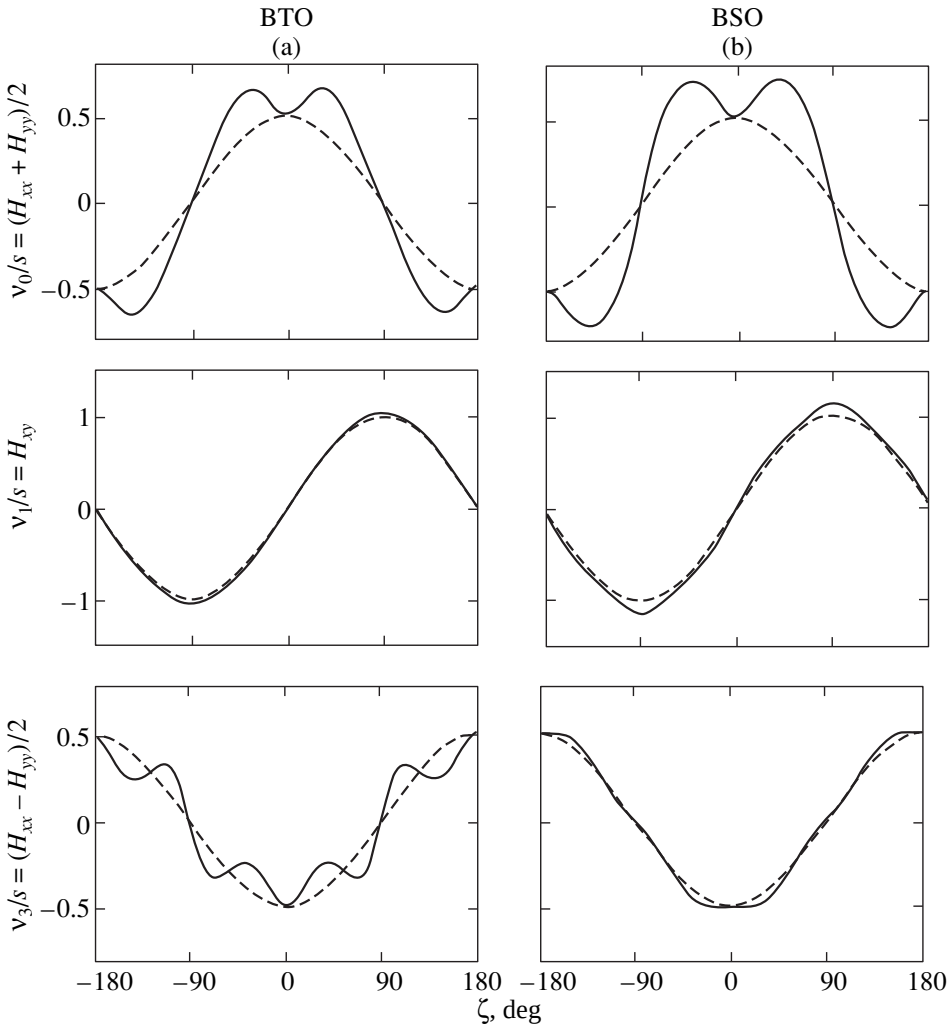

Fig. 2. The plots of $\nu_{0,1,3}(\zeta)$ corresponding to Fig 1 for $Bi_{12}TiO_{20}$ and $Bi_{12}SiO_{20}$ crystals.

allowance for the elastooptical contributions. Figure 2 shows a summary of the results obtained for $\nu_{0,1,3}(\zeta)$. As seen, neglect of the elastooptical contributions is often justified.

A configuration depicted in Fig. 1b is generally similar to that considered above (Fig. 1a). However, the two configurations are not equivalent and the transition between them must be accompanied by a change in the sign of the electrooptical coefficient r_{41}.

There are twelve symmetry transformations not affecting the electrooptical properties of crystals, which include identical transformation, 180° rotations about principal axes, ±120° rotations about principal diagonals, and their combinations [25]. Table 2a gives a set of 12 equivalent configurations, the first of which corresponds to Fig. 1a; by the same token, Table 2b gives another set of 12 equivalent configurations, the first of which is depicted in Fig. 1b.

The symmetry properties lead to the following general relationships

$$\hat{H}^{(0)}(\zeta_0) = -\hat{H}^{(0)}(\zeta_0 + \pi),$$
$$\hat{H}(\zeta) = -\hat{H}(\zeta + \pi), \quad H_{xx}(\zeta) = H_{xx}(-\zeta), \qquad (17)$$
$$H_{yy}(\zeta) = H_{yy}(-\zeta), \quad H_{xy}(\zeta) = -H_{xy}(-\zeta).$$

In other words, matrix $\hat{H}^{(0)}$ (or $\hat{H}$) changes sign upon the inversion of $\mathbf{E}_0$ (or $\mathbf{K}$). Upon reflection from the *xz* plane, the nondiagonal matrix elements change sign, while the diagonal elements remain unchanged.

There are three optical configurations [2, 11, 29] corresponding to various values of angle ζ_0 (Fig. 1a) that are of special importance for experiment:

(i) Longitudinal configuration, $\mathbf{E}_0 \parallel [001]$ ($\sin\zeta_0 = 0$);

(ii) Transverse configuration, $\mathbf{E}_0 \perp [001]$ ($\cos\zeta_0 = 0$);

(iii) Diagonal configuration, $\mathbf{E}_0 \parallel [\bar{1}\bar{1}1]$ ($\tan\zeta_0 = \sqrt{2}$).

For the longitudinal and transverse configurations, the elastooptical contributions to $\nu_{0,3}$ are absent for $\zeta = \zeta_0$ ($\mathbf{K} \parallel \mathbf{E}_0$), while the contribution to ν_1 is insignificant (see Fig. 2). The diagonal configuration corresponds to a maximum value of parameter $\nu_0(\zeta_0)$, which characterizes the isotropic interaction; here, the elastooptical contribution to ν_0 is significant.

3. NONLINEAR PHOTOREFRACTIVE RESPONSE

In order to obtain a closed system of equations for the light waves in a cubic crystal, it is necessary to supplement Eqs. (10) with relationships describing the grating amplitude $E_\mathbf{K}$ as function of the interference pattern contrast $m = 2(\mathbf{A}_1 \cdot \mathbf{A}_2^*)/(|\mathbf{A}_1|^2 + |\mathbf{A}_2|^2)$. These relationships are usually derived based on a simple semiconductor model describing the photogeneration of charge carriers from deep centers followed by the migration and recombination of carriers [1–3]. This model is justified for sillenites, where the photogenerated carriers are electrons, and works satisfactorily in many semiconductors. In a linear (with respect to m) approximation, this model leads to the following equation for $E_\mathbf{K}$:

$$\frac{\partial E_\mathbf{K}}{\partial t} + \left(\omega_0 \frac{E_q + E_D - i\tilde{E}_0}{E_M + E_D - i\tilde{E}_0} - i\Omega\right) E_\mathbf{K} = -\frac{m\omega_0}{2}\frac{E_q(\tilde{E}_0 + iE_D)}{E_M + E_D - i\tilde{E}_0}. \tag{18}$$

Here, $\tilde{E}_0 = E_0\cos\psi = E_0(\mathbf{n}_0 \cdot \mathbf{n})$ is the drawing field and $E_{D,M,q}$ are the characteristic fields defined by the formulas

$$E_D = \frac{KT}{e}, \quad E_M = \frac{1}{K\mu\tau}, \quad E_q = \frac{4\pi e N_t}{\varepsilon K}, \tag{19}$$

where e is the elementary charge, T is the absolute temperature, ε is the static dielectric permittivity, N_t is the effective concentration of traps, and $\mu\tau$ is the mobility-lifetime product for photoelectrons. The fields E_D, E_M, and E_q characterize the diffusion and drift of electrons and the saturation of traps, respectively. Finally, $\omega_0 = \alpha I_0/\hbar\omega N_t$ is the frequency characterizing the rate of charge carrier photoexcitation, which is proportional to the number of light quanta $\alpha I_0/\hbar\omega$ absorbed per unit time in a unit volume.

An important feature of the "fast-response" photorefractive crystals, including sillenites and most semiconductors, is the presence of a low-frequency branch of weakly-damped space charge waves (SCWs) [32, 33]. It is this feature that provides for the possibility of enhancement of the photorefractive response. A sufficient condition for the SCWs to exist is provided by inequalities

Table 2. Triplets of basis vectors x, y, z equivalent to (a) [001], [$\bar{1}\,\bar{1}$ 0], [1$\bar{1}$ 0] and (b) [001], [1$\bar{1}$ 0], [110]

a			*b*		
x	*y*	*z*	*x*	*y*	*z*
[001]	[$\bar{1}\,\bar{1}$ 0]	[1$\bar{1}$ 0]	[001]	[1$\bar{1}$ 0]	[110]
[001]	[110]	[$\bar{1}$ 10]	[001]	[$\bar{1}$ 10]	[$\bar{1}\,\bar{1}$ 0]
[00$\bar{1}$]	[$\bar{1}$ 10]	[110]	[00$\bar{1}$]	[110]	[1$\bar{1}$ 0]
[00$\bar{1}$]	[1$\bar{1}$ 0]	[$\bar{1}\,\bar{1}$ 0]	[00$\bar{1}$]	[$\bar{1}\,\bar{1}$ 0]	[$\bar{1}$ 10]
[100]	[0$\bar{1}\,\bar{1}$]	[01$\bar{1}$]	[100]	[01$\bar{1}$]	[011]
[100]	[011]	[0$\bar{1}$ 1]	[100]	[0$\bar{1}$ 1]	[0$\bar{1}\,\bar{1}$]
[$\bar{1}$ 00]	[0$\bar{1}$ 1]	[011]	[$\bar{1}$ 00]	[011]	[01$\bar{1}$]
[$\bar{1}$ 00]	[01$\bar{1}$]	[0$\bar{1}\,\bar{1}$]	[$\bar{1}$ 00]	[0$\bar{1}\,\bar{1}$]	[0$\bar{1}$ 1]
[010]	[$\bar{1}$ 0$\bar{1}$]	[$\bar{1}$ 01]	[010]	[$\bar{1}$ 01]	[101]
[010]	[101]	[10$\bar{1}$]	[010]	[10$\bar{1}$]	[$\bar{1}$ 0$\bar{1}$]
[0$\bar{1}$ 0]	[10$\bar{1}$]	[101]	[0$\bar{1}$ 0]	[101]	[$\bar{1}$ 01]
[0$\bar{1}$ 0]	[$\bar{1}$ 01]	[$\bar{1}$ 0$\bar{1}$]	[0$\bar{1}$ 0]	[$\bar{1}$ 0$\bar{1}$]	[10$\bar{1}$]

$E_q \gg E_0 \gg E_{D,M}$, while the necessary limitation of the material parameters is given by the inequality $\pi e N_t \mu\tau\varepsilon^{-1} \gg 1$ (see [33]). Once these conditions are satisfied, the first term in the parentheses of Eq. (18) can be replaced by the combination $i\omega_\mathbf{K} + \gamma_\mathbf{K}$, where

$$\omega_\mathbf{K} = \omega_0\frac{E_q}{\tilde{E}_0}, \qquad \gamma_\mathbf{K} = \omega_0\left(1 + \frac{E_q E_M + E_q E_D}{E_0^2\cos^2\psi}\right) \tag{20}$$

are the frequency and damping factor of a space charge wave with wavevector $\mathbf{K}$. Note that the parameter $Q = |\omega_\mathbf{K}|/\gamma_\mathbf{K}$ is the quality factor for SCW. Within the framework of our theory, this parameter (depending on the product $|\mathbf{K} \cdot \mathbf{E}_0|$ and independent of I_0) is markedly greater than unity. The large values of Q are necessary to explain the parametric excitation of SCWs in sillenite crystals [33, 34].

Table 3 presents typical values of the material parameters for $Bi_{12}TiO_{20}$ and $Bi_{12}SiO_{20}$ crystals and gives estimates of the factor $\pi e N_t \mu\tau\varepsilon^{-1}$ and the characteristic fields. As is seen, the above approximations are well justified in a broad range of the grating period Λ, where the most important phenomena related to the nonlinear photorefractive response enhancement take place.

The further analysis of the features of the photorefractive response is expediently performed for the DC and AC cases separately.

Table 3. Material parameters of $Bi_{12}TiO_{20}$ and $Bi_{12}SiO_{20}$ crystals and the characteristic field strengths for $\Lambda = 20$ μm

Parameter	$Bi_{12}TiO_{20}$	$Bi_{12}SiO_{20}$
N_t, cm^{-3}	2×10^{16}	10^{16}
$\mu\tau$, cm^2/V	2×10^{-7}	5×10^{-7}
ε	47	56
E_q, kV/cm	250	105
E_M, kV/cm	1.5	0.6
E_D, kV/cm	0.08	0.08
$\pi e N_t \mu\tau/\varepsilon$	38.5	40.4

3.1. DC Enhancement Method

Assuming that field E_0 is time-independent, we obtain from Eq. (18) a relationship for the grating amplitude

$$E_{\mathbf{K}} \approx \frac{\omega_{\mathbf{K}}|\tilde{E}_0|}{\Omega - \omega_{\mathbf{K}} + i\gamma_{\mathbf{K}}} \frac{\mathbf{A}_1 \cdot \mathbf{A}_2^*}{|\mathbf{A}_1|^2 + |\mathbf{A}_2|^2}. \quad (21)$$

The dependence of the grating amplitude on the frequency difference Ω exhibits a resonance character. In the region of resonance, $|\Omega - \omega_{\mathbf{K}}| \lesssim \gamma_{\mathbf{K}}$, we obtain a relationship

$$E_{\mathbf{K}} \approx -iQ|\tilde{E}_0|(\mathbf{a}_1 \cdot \mathbf{a}_2^*). \quad (22)$$

For convenience, we have introduced dimensionless amplitudes of the light waves $\mathbf{a}_{1,2} = \mathbf{A}_{1,2}/\sqrt{|\mathbf{A}_1|^2 + |\mathbf{A}_2|^2}$. Note that linear absorption does not lead to the decay of these amplitudes: $|\mathbf{a}_1|^2 + |\mathbf{a}_2|^2 = 1$.

The large value of Q in Eq. (22) is the essence of the DC enhancement method. The presence of an imaginary unit in this relationship indicates that the response has a gradient character, whereby the grating $E_{sc}(\mathbf{r})$ is shifted by a quarter period relative to the light interference pattern.

It should be recalled that $\omega_{\mathbf{K}}, \gamma_{\mathbf{K}} \propto I_0$, while intensity I_0 decreases in proportion to $e^{-\alpha z}$ as a result of the linear absorption. Therefore, the resonance excitation condition $|\Omega - \omega_{\mathbf{K}}(z)| \lesssim \gamma_{\mathbf{K}}$ is satisfied in a layer of thickness $\delta z \approx 2/\alpha Q$. As the frequency difference Ω decreases from $\omega_{\mathbf{K}}(0)$ to $\omega_{\mathbf{K}}(l) = \omega_{\mathbf{K}}(0)\exp(-\alpha l)$, the resonance layer shifts from the input ($z = 0$) to output ($z = l$) faces of the crystal. If the crystal thickness l is markedly greater than δz, the wave interaction in most parts of the crystal becomes ineffective. For $Q \approx 6$ and $\alpha \approx 1$ cm^{-1}, we obtain an estimate $\delta z \approx 0.3$ cm.

In the general case, relationship (21) can be conveniently written in the following form

$$E_{\mathbf{K}} \approx -iQR|\tilde{E}_0|(\mathbf{a}_1 \cdot \mathbf{a}_2^*),$$
$$R = [1 + iQ(1 - \delta\exp(\alpha z))]^{-1}, \quad (23)$$

by introducing a resonance factor $R(z)$ and a dimensionless detuning $\delta = \Omega/\omega_{\mathbf{K}}(0)$.

Finally, note that the approximation linear with respect to contrast is valid for $|m| \lesssim Q^{-1} \ll 1$. For the greater $|m|$ values, we must take into account the higher spatial harmonics of the field ($E_{2\mathbf{K}}$, $E_{3\mathbf{K}}$, …). Thus, the DC amplification of the photorefractive response is achieved at the expense of certain restrictions on the crystal length l and the light pattern contrast m.

3.2. AC Enhancement Method

In this case, $\Omega = 0$ and field E_0 is a rapidly oscillating function of time. The oscillation period T_0 is assumed to be shorter than $2\pi/\omega_{\mathbf{K}} \approx 10^{-2}$–$10^{-1}$ s but longer than the charge carrier lifetime $\tau \lesssim 10^{-7}$ s. These conditions ensure that the high-frequency component of $E_{\mathbf{K}}(t)$ is small compared to the constant component. We will restrict the consideration to the case of periodic alternating-sign $E_0(t)$ with a constant amplitude $|E_0|$, which corresponds to the most effective AC enhancement [3, 35].

Upon averaging over the high-frequency oscillations in Eq. (18) and taking into account inequalities $E_q \gg |E_0| \gg E_{D,M}$, we obtain

$$E_{\mathbf{K}} \approx -i|\tilde{E}_0|Q\langle \mathbf{a}_1 \cdot \mathbf{a}_2^* \rangle, \quad (24)$$

where the averaging $\langle \ldots \rangle$ reduces to taking a half sum of two values calculated for $|E_0|$ and $-|E_0|$.

The response enhancement factor Q has the same form as that in the DC method. However, the AC amplification (in contrast to the DC case) is weakly sensitive with respect to the I_0 decrease as a result of the light absorption. Moreover, the validity of Eq. (24) is restricted to even smaller contrast variations: $|m| \lesssim Q^{-2} \ll 1$. Otherwise, it is necessary to take into account the parametric generation of SCWs [33, 36].

Using Eq. (10), we can make another step in simplifying relationship (24). Assuming (which is usually justified in experiment) that the entrance light beams 1 and 2 are linearly polarized, we may conclude that a nonlinear evolution of the amplitudes $\mathbf{a}_{1,2}(z)$ is compatible with the linear property $\mathbf{a}_{1,2}(E_0) = \mathbf{a}_{1,2}^*(-E_0)$ (see Section 2). This implies that $\langle \mathbf{a}_1 \cdot \mathbf{a}_2^* \rangle = \mathbf{a}_1 \cdot \mathbf{a}_2^*$ and relationship (24) for the AC response reduces to formula (22).

4. INTERACTION OF TWO LIGHT WAVES

Substituting relationship (23) for the nonlinear response into Eq. (10), we obtain a closed system of equations for $\mathbf{a}_{1,2} = \mathbf{A}_{1,2}/\sqrt{|\mathbf{A}_1|^2 + |\mathbf{A}_2|^2}$ in the DC case. By the same token, an analogous closed system of equations for the AC case is obtained upon combining Eqs. (24) and (10). An advantage of the σ-matrix representation used here is the possibility to simplify the nonlinear equa-

tions by excluding terms describing linear wave propagation.

Let us pass from amplitudes $\mathbf{a}_{1,2}$ to the new dimensionless amplitudes $\mathbf{b}_{1,2}$ defined as

$$\mathbf{a}_{1,2} = e^{i(\boldsymbol{\kappa}\cdot\hat{\boldsymbol{\sigma}})z}\mathbf{b}_{1,2}, \tag{25}$$

which is equivalent to the interaction representation used in quantum mechanics [25]. In applying to the DC case, the system of equations describing the interaction of two light waves takes the form

$$\begin{aligned}\frac{d\mathbf{b}_1}{dz} &= RQ|\tilde{E}_0|(\mathbf{b}_2^*\cdot\mathbf{b}_1)(\nu_0\hat{1}+\mathbf{q}\cdot\hat{\boldsymbol{\sigma}})\mathbf{b}_2,\\ \frac{d\mathbf{b}_2}{dz} &= -R^*Q|\tilde{E}_0|(\mathbf{b}_1^*\cdot\mathbf{b}_2)(\nu_0\hat{1}+\mathbf{q}\cdot\hat{\boldsymbol{\sigma}})\mathbf{b}_1.\end{aligned} \tag{26}$$

Here, $R = R(z)$ is the factor defined in (23) and $\mathbf{q} = (q_1, q_2, q_3)$ is a real vector expressed as

$$\begin{aligned}\mathbf{q} = \frac{\boldsymbol{\kappa}(\boldsymbol{\nu}\cdot\boldsymbol{\kappa})}{\kappa^2} + \left[\boldsymbol{\nu} - \frac{\boldsymbol{\kappa}(\boldsymbol{\nu}\cdot\boldsymbol{\kappa})}{\kappa^2}\right]\cos(2\kappa z)\\ + \frac{(\boldsymbol{\kappa}\times\boldsymbol{\nu})}{\kappa}\sin(2\kappa z).\end{aligned} \tag{27}$$

The isotropic part of the wave interaction, characterized by parameter ν_0, does not change upon going to the interaction representation. The anisotropic part, which is proportional to $\mathbf{q}$, has acquired an oscillating part related to the interference of the eigenmodes.

Expressions for the AC case exhibit an analogous structure and are obtained from (26) by substituting $R(\mathbf{b}_2^*\cdot\mathbf{b}_1) \longrightarrow \langle\mathbf{b}_2^*\cdot\mathbf{b}_1\rangle$. In other words, equations for AC include averaging over period T_0 instead of the resonance factor $R(z)$.

4.1. Undepleted Pump Approximation

The condition $|m| \ll 1$ determining the applicability of Eqs. (23) and (24) for describing the nonlinear photorefractive response can be fulfilled either for a large difference in intensities of the interacting waves, or at the expense of their polarizations being almost strictly orthogonal. The former case, which is more important, corresponds to the spatial amplification of a weak light beam in the presence of a strong beam. This case, known as the undepleted pump approximation will be considered below.

4.1.1. AC amplification. Assuming for certainty that $|\mathbf{a}_1|/|\mathbf{a}_2| \equiv |\mathbf{b}_1|/|\mathbf{b}_2| \ll 1$, we may neglect the variation of amplitude $\mathbf{b}_2$ in Eq. (26) for $\mathbf{b}_1$ and take $\mathbf{b}_2 = \mathbf{b}_2(0) \equiv \mathbf{a}_2(0) \approx \mathbf{e}_p$, where $\mathbf{e}_p$ is the unit polarization vector of the pump wave (wave 2) at the input crystal face. Then the weak wave (wave 1) amplitude is described by the equation

$$\frac{d\mathbf{b}_1}{dz} = Q|\tilde{E}_0|\langle\mathbf{e}_p^*\cdot\mathbf{b}_1\rangle(\nu_0\hat{1}+\mathbf{q}\cdot\hat{\boldsymbol{\sigma}})\mathbf{e}_p. \tag{28}$$

Multiplying this equation (scalar product) by $\mathbf{e}_p^*$ and averaging with respect to time, we obtain

$$\frac{d\langle\mathbf{e}_p^*\cdot\mathbf{b}_1\rangle}{dz} = Q|\tilde{E}_0|(\nu_0 + \langle\mathbf{q}\cdot\boldsymbol{\xi}_p\rangle)\langle\mathbf{e}_p^*\cdot\mathbf{b}_1\rangle, \tag{29}$$

where $\xi_{pj} = (\mathbf{e}_p^*\,\hat{\sigma}_j\,\mathbf{e}_p)$ are components of the Stokes vector $\boldsymbol{\xi}_p = (\xi_{p1}, \xi_{p2}, \xi_{p3})$ of the incident pump wave [37]. For a completely polarized light wave, this vector has a unit length: the $\xi_{p1,3}$ components describe the degree of linear polarization and the ξ_{p2} component, the degree of circular polarization. On averaging the vector $\mathbf{q}$ value, we should take into account that $\kappa_{1,3}(E_0) = -\kappa_{1,3}(-E_0)$ and that the component $\kappa_2 = -\rho$ is independent of E_0 (see Eqs. (9) and (15)). For this reason, $\langle q_{1,3}(z)\rangle = q_{1,3}(z)$, $\langle q_2\rangle = 0$, and $\langle\mathbf{q}\cdot\boldsymbol{\xi}_p\rangle = q_1\xi_{p1} + q_3\xi_{p3}$. These relationships and Eq. (29) lead to the conclusion that, for a linearly polarized pumping wave ($\xi_{p2} = 0$), the $\mathbf{e}_p^*\cdot\mathbf{b}_1$ component does not exhibit jumps upon switching the field. This case is most important for experiments.

Integrating Eq. (29), we obtain

$$\begin{aligned}\langle\mathbf{e}_p^*\cdot\mathbf{b}_1(z)\rangle &= \mathbf{e}_p^*\cdot\mathbf{b}_1(0)\\ &\times\exp[\Gamma z + C_1\sin(2\kappa z) + C_2\sin^2(\kappa z)],\end{aligned} \tag{30}$$

where increment Γ (the main characteristic of spatial amplification) and coefficients $C_{1,2}$ are given by the formulas

$$\begin{aligned}\Gamma &= Q|\tilde{E}_0|\left[\nu_0 + \frac{(\boldsymbol{\nu}\cdot\boldsymbol{\kappa})}{\kappa^2}(\kappa_1\xi_{p1} + \kappa_3\xi_{p3})\right],\\ C_1 &= \frac{Q|\tilde{E}_0|}{2\kappa^3}[\kappa^2(\boldsymbol{\nu}\cdot\boldsymbol{\xi}_p) - (\boldsymbol{\nu}\cdot\boldsymbol{\kappa})(\kappa_1\xi_{p1} + \kappa_3\xi_{p3})],\\ C_2 &= \frac{Q|\tilde{E}_0|\rho}{\kappa^2}[\boldsymbol{\nu}\times\boldsymbol{\xi}_p]_2.\end{aligned} \tag{31}$$

For $\Gamma l \gg 1$, the amplification is very large. Note some important properties of the increment:

(i) Formula (31) contains the isotropic ($\propto \nu_0$) and anisotropic parts. The isotropic part is independent of the pump wave polarization, while both the magnitude and sign of the anisotropic contribution are polarization-dependent. For a circular polarization ($\xi_{p1,3} = 0$, $\xi_{p2} = \pm 1$), the anisotropic contribution is zero.

(ii) Substitution $\mathbf{K} \longrightarrow -\mathbf{K}$ leads to the reversal of the increment sign: $\Gamma(\zeta) = -\Gamma(\zeta \pm \pi)$. This behavior follows from Eqs. (9), (14), and (17).

(iii) The increment is independent of the signs of E_0 and $\mathbf{n}^0$, which follows from the definition of vector $\boldsymbol{\kappa}$.

Using relationship (31), we can maximize the increment with respect to the pump wave of polarization. The maximum of Γ corresponds to a linear polarization such that $\xi_{p1,3} = \kappa_{1,3}\,\mathrm{sgn}(\boldsymbol{\nu}\cdot\boldsymbol{\kappa})/\sqrt{\kappa_1^2 + \kappa_3^2}$. This condition can be readily rewritten in terms of the inclination

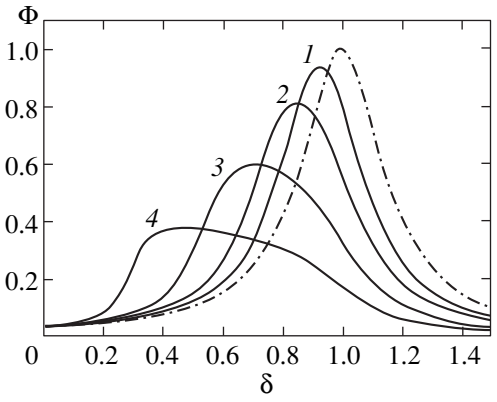

Fig. 3. The curves of $\Phi(\delta)$ for $\alpha = 0.75$ cm^{-1}, $Q = 6$, and various crystal thicknesses $l = 0.2$ (*1*); 0.4 (*2*); 0.8 (*3*); 1.6 (*4*); the dash–dot curve corresponds to the ideal case of $\alpha l = 0$.

angle of the polarization plane. The maximum increment for the optimum polarization is

$$\Gamma = Q|\tilde{E}_0|\left(\nu_0 + \frac{|\boldsymbol{\nu}\cdot\boldsymbol{\kappa}|\sqrt{\kappa_1^2+\kappa_3^2}}{\kappa_1^2+\kappa_3^2+\rho^2}\right). \qquad (32)$$

As is seen from this expression, the optical activity negatively affects the amplification. The right-hand side of Eq. (32) can be further optimized with respect to $|\mathbf{K}|$ (that is, with respect to the angle $\theta = \lambda|\mathbf{K}|/2\pi$ between the light beams) and with respect to angles ζ and ζ_0 determining the orientation of $\mathbf{K}$ and $\mathbf{E}_0$. Examples of such optimization, which can be important in particular cases, are given in Section 5. Here we should like to note that the absolute maximum of the function $\Gamma(\zeta, \zeta_0)$ is attained at $\psi \equiv \zeta - \zeta_0 = 0$, that is, at $\mathbf{K} \parallel \mathbf{E}_0$. For a fixed value of ζ_0, the maximum is observed in the general case at $\zeta \neq \zeta_0$.

Finally, we should like to mention the characteristics of spatial amplification which are not related to the properties of increment Γ. Additional contributions to the exponents in Eq. (30) are mostly important when $|C_{1,2}| \gg 1$, whereas in many cases (see also Section 5) $|C_{1,2}| \lesssim 1$. The $\mathbf{b}_1$ vector component orthogonal to $\mathbf{e}_p$ (necessary for the complete description of the polarization properties of the amplified wave 1) can be also calculated using Eq. (28). The main feature of the spatial variation of this component is the exponential growth of its magnitude with increment Γ. The initial polarization of the weak beam enters only into the preexponential factor in Eq. (31). The optimum is attained for $\mathbf{A}_1(0) \parallel \mathbf{A}_2(0) \propto \mathbf{e}_p$.

4.1.2. DC amplification. Assuming $|\mathbf{b}_1| \ll |\mathbf{b}_2|$, we obtain the following expression for the weak wave amplitude from Eq. (26):

$$\frac{d\mathbf{b}_1}{dz} = RQ|\tilde{E}_0|(\mathbf{e}_p^*\cdot\mathbf{b}_1)(\nu_0\hat{1} + \mathbf{q}\cdot\hat{\boldsymbol{\sigma}})\mathbf{e}_p, \qquad (33)$$

where $R = R(z)$ is the resonance factor defined by Eq. (23). This quantity determines special features of the DC enhancement.

In order to illustrate distinctive features of the DC enhancement, let us omit the spatially-oscillating part $\mathbf{q}$ in Eq. (27) by taking $\mathbf{q} = \mathbf{q}_0 = \text{const}$. This simplified model provides for a rough quantitative description of the wave interaction (see also Section 5). Within the framework of this approximation, $\ln(|\mathbf{a}_1(l)|/|\mathbf{a}_1(0)|) \approx \Gamma l\Phi$, where $\Gamma = Q|\tilde{E}_0|(\nu_0 + \mathbf{q}_0\cdot\boldsymbol{\xi}_p)$ is the spatial enhancement increment (essentially the same as considered above for a linear polarization of the pumping wave) and

$$\Phi = \frac{1}{l}\int_0^l [1 + Q^{-2}(1-\delta e^{\alpha z})^2]^{-1} dz \qquad (34)$$

is the factor of loss in the exponential enhancement rate as compared to that in the AC case. For $\alpha = 0$, the peak value of $\Phi(\delta) = 1$ is attained with $\delta = 1$.

Figure 3 (solid curves) shows the curves of $\Phi(\delta)$ for $\alpha = 0.75$ cm^{-1}, $Q = 6$, and several values of the crystal thickness l. The dash–dot curve corresponds to the ideal case of $\alpha = 0$. As seen, a decrease in the peak amplitude, the shift toward small detunings, and deterioration of the resonance are significant even in rather thin crystals. This behavior is more pronounced with increasing α. Taking into account that the absorption is manifested by the saturation of the enhancement with increasing l: the higher the α value, the earlier the saturation onset.

5. APPLICATIONS OF THE THEORY

The most important and simple applications are offered by calculations of the angular and polarization characteristics of spatial amplification for the basic optical configurations in particular crystals and by the comparison of the results of such calculations with experiment. Below we illustrate these applications of the theory to calculations for $Bi_{12}TiO_{20}$ and $Bi_{12}SiO_{20}$ crystals.

5.1. AC Enhancement in $Bi_{12}TiO_{20}$ Crystals

The AC field amplitude in the experiments with these crystals is usually varied from 10 to 50 kV/cm and the laser beams typically possess a linear polarization. Using the data from Table 1 and assuming $|E_0| = 30$ kV/cm, we obtain the estimate $s|E_0| \approx 12$ cm^{-1}. This implies that $(\rho/\kappa)^2 \approx 10^{-2} \ll 1$, that is, the optical activity is strongly suppressed by the external field. This situation favors the enhancement. Assuming $Q = 6$ and using Eq. (31), we obtain the estimate $\Gamma \approx sQ|E_0| \approx 70$ cm^{-1}. Such a high increment implies the possibility of provid-

ing for a giant enhancement of weak waves even in millimeter-thick crystals.

Under the conditions of strong spatial amplification, the presence of two incident light beams is not necessary for studying the interaction of two waves. Consider a single pump wave incident onto a crystal. The very weak scattered seed waves can also be amplified due to their coupling to the pump wave, which leads to the appearance of light-induced scattering [1–3]. Measurements of the angular characteristics of this scattering provide direct information concerning the dependence of the degree of spatial amplification on the direction of weak wave propagation. Let us consider the application of this concept to various particular cases.

5.1.1. Longitudinal configuration, $\mathbf{E}_0 \parallel [001]$. According to the results obtained in Section 2 and the data presented in Fig. 1a, this case is characterized by $\zeta_0 = 0$, $\psi = \zeta$, and $\kappa_1 = 0$. Denoting the angle between $\mathbf{e}_p$ and the horizontal axis [001] $\parallel \mathbf{E}_0$ by φ_p, we may write expressions for the nonzero components of the Stokes vector of the pump wave in the following form: $\xi_{p1} = \sin 2\varphi_p$, $\xi_{p3} = \cos 2\varphi_p$. Using Eqs. (9) and (31), we obtain a formula for the increment:

$$\Gamma = s|E_0 \cos\psi| Q(H_{xx}\cos^2\varphi_p + H_{yy}\sin^2\varphi_p). \quad (35)$$

According to Eq. (20), the Q value is a function of $|E_0|$, $|\mathbf{K}|$, and the angle ψ between $\mathbf{K}$ and the z axis. The optimum polarization is vertical ($\varphi_p = 90°$).

Figure 4 shows the patterns of the increment distribution $\Gamma(\theta_\parallel, \theta_\perp)$ in the region of its positive values for the polarization angles $\varphi_p = \pi/2$ and 0. The components $\theta_\parallel = \theta\cos\psi$ and $\theta_\perp = \theta\sin\psi$ ($\theta = \lambda|\mathbf{K}|/2\pi$ is the polar angle of scattering) characterize the horizontal and vertical angular deviations of the output weak wave. These parameters are convenient for comparison with experiment. The behavior of the increment in the region of negative values is determined by the symmetry properties: $\Gamma(\mathbf{K}) = -\Gamma(-\mathbf{K})$.

As seen, the effective enhancement has to be observed in the region of small θ values, which is explained by the behavior of $Q(|E_0|, K)$. The increase in $|E_0|$ leads to a decrease in the characteristic θ values. For $\varphi_p = 90°$ (Fig. 4a), the elastooptical contributions to Γ are insignificant. A maximum value of the increment ($\Gamma_{max} \approx 75$ cm^{-1}) corresponds to the azimuthal angle $\psi = 0$. For $\varphi_p = 0$ (Fig. 4b), nonzero values of the increment are due to the elastooptical effect; once this effect is absent, $\Gamma(\theta_\parallel, \theta_\perp) = 0$. This angular dependence represents two rather narrow lobes and is characterized by the Γ_{max} values that are markedly smaller as compared to those in the case of $\varphi_p = 90°$. Note that, in Eq. (30), $C_{1,2} = 0$ for $\varphi_p = 0$ and $\pi/2$.

Figure 5 presents the angular distributions of the light-induced scattering experimentally measured for $\varphi_p = 0$ and $\pi/2$ in the longitudinal optical configuration. In accordance with the theory, we observe a single-lobe pattern for $\varphi_p = 90°$ and a two-lobe pattern for $\varphi_p = 0$. The orientation of lobes also agrees with the theoretical predictions. As expected, the scattering intensity is markedly lower for the horizontal polarization of the pump wave than for vertical polarization.

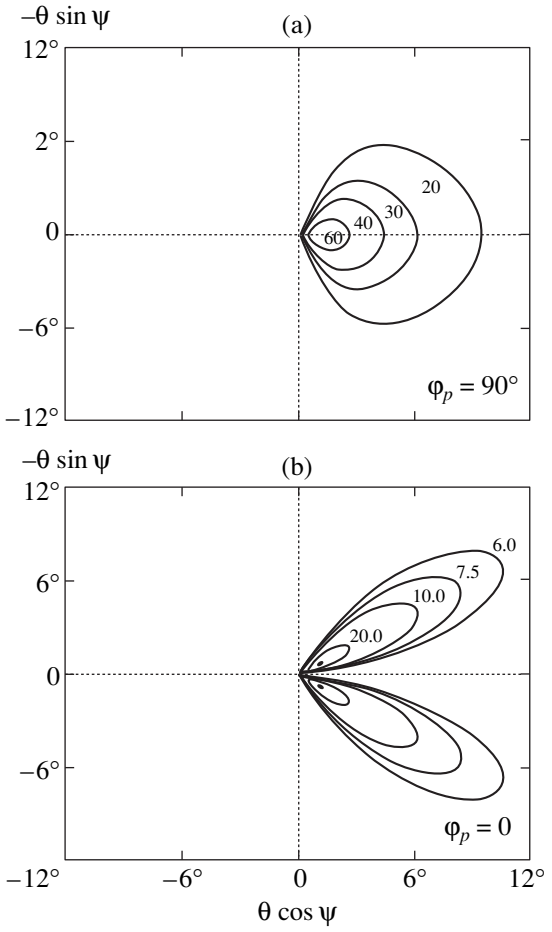

Fig. 4. The patterns of increment isolines $\Gamma(\theta_\parallel, \theta_\perp) = \text{const}$ in the longitudinal configuration for (a) vertical and (b) horizontal polarizations of the pump wave relative to the $\mathbf{E}_0$ direction. The curves are calculated for $|E_0| = 30$ kV/cm (other parameters taken for $Bi_{12}TiO_{20}$ from Tables 1 and 3).

5.1.2. Transverse configuration $\mathbf{E}_0 \perp [001]$. Upon selecting $\zeta_0 = \pi/2$, we have $\zeta = \psi + \pi/2$ and $\kappa_1 = sE_0$, $\kappa_3 = 0$. Denoting the polarization angle (measured from the applied field direction) of the pump wave by φ_p and using Eqs. (9) and (31), we obtain an expression for the increment

$$\Gamma = s|E_0 \cos\psi| Q(\nu_0 - \nu_1 \sin 2\varphi_p). \quad (36)$$

The values of $\nu_{0,1}(\zeta)$ were calculated with and without allowance for the elastooptical effects in Section 2. Figure 6 shows the distributions of $\Gamma(\theta_x, \theta_y)$ for the polarization angles $\varphi_p = 0$, $\pi/4$, and $-\pi/4$ (here, the values of $\varphi_p = 0$ and $\pi/2$ are equivalent). In the case of $\varphi_p = 0$ (or $\pi/2$), the theory predicts two symmetric lobes in the

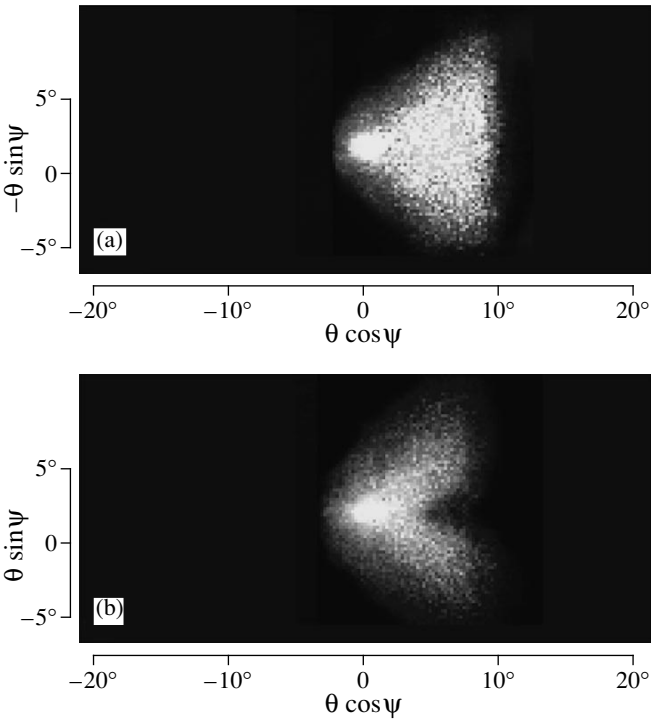

Fig. 5. Experimental angular distributions of the light-induced scattering intensity in $Bi_{12}TiO_{20}$ crystals measured in the longitudinal geometry for $\varphi_p = \pi/2$ (a) and 0 (b). Experimental parameters: $\lambda = 632.8$ nm; $I_0 = 0.8$ W/cm^2; $|E_0| = 20$ kV/cm; $l \approx 2$ cm.

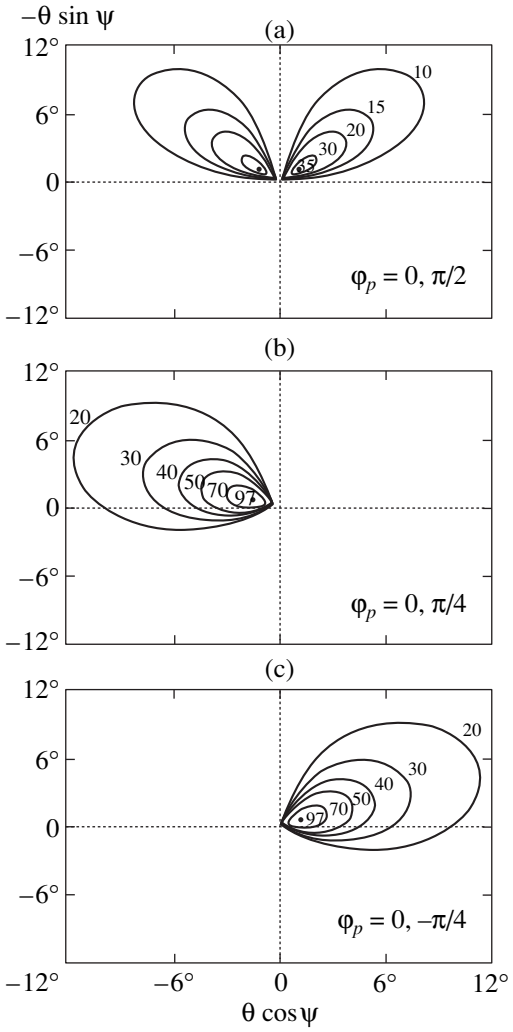

Fig. 6. The patterns of increment isolines for the transverse configuration with $l \approx 2$ cm; $\varphi_p = 0$ (a), $\pi/4$ (b), and $-\pi/4$ (c) calculated for the same parameters as in Fig. 5.

upper quadrants. For $\varphi_p = \pm\pi/4$, the diagram shows single lobes centered at $\psi \approx -90° \mp 64°$. A maximum value of the increment ($\Gamma \approx 100$ cm^{-1}) for the transverse configuration is greater than that for the longitudinal scheme. The elastooptical contributions do not strongly affect the increment variation. Parameter C_1 in this case is zero.

Figure 7 shows the patterns of light-induced scattering experimentally measured in the transverse configuration for $\varphi_p = 0$, $\pi/2$, $\pi/4$ and $-\pi/4$. In accordance with the theory, the patterns for $\varphi_p = 0$ and $\pi/2$ are virtually identical. The angular distributions of the scattering intensity well correlate with the increment variations.

5.1.3. Transverse configuration, $\mathbf{E}_0 \parallel [\bar{1}\bar{1}1]$. According to the formulas derived in Section 2, this configuration is characterized by $\zeta = \psi + \zeta_0$, $\zeta_0 = \arctan\sqrt{2} \approx 54.7°$, $\kappa_1 = \sqrt{2}\,sE_0/\sqrt{3}$, and $\kappa_3 = -sE_0/2\sqrt{3}$. The polarization angle φ_p is measured from the $[\bar{1}\bar{1}1]$ axis parallel to $\mathbf{E}_0$. Using Eq. (31), we obtain the following expression for the increment

$$\Gamma = \frac{1}{3}|E_0\cos\psi|Q[3\nu_0 + (2\sqrt{2}\nu_1 - \nu_3)\cos 2\varphi_p]. \quad (37)$$

Figure 8a shows the isolines of increment for a horizontal polarization of the pump wave calculated for $\varphi_p = 0$. This case corresponds to the maximum possible amplification ($\Gamma_{max} \approx 110$ cm^{-1}), with a single lobe extended along the $[\bar{1}\bar{1}1]$ axis. The elastooptical contributions markedly increase the Γ value, while not significantly affecting the shape of the angular distribution. Figure 8b shows an analogous pattern of isolines for $\Gamma(\theta_\parallel, \theta_\perp)$ in the case of $\varphi_p = \pi/2$. Here we observe the main lobe extended in the direction of $\psi \approx 120°$ and an additional small lobe. The elastooptical contributions slightly increase the maximum value of the increment and shift the main lobe toward the horizontal axis.

Figure 9 shows the corresponding experimental patterns of the light-induced scattering. These patterns also show a good agreement with theoretical predictions. As is clearly seen, the spots of induced scattering are sepa-

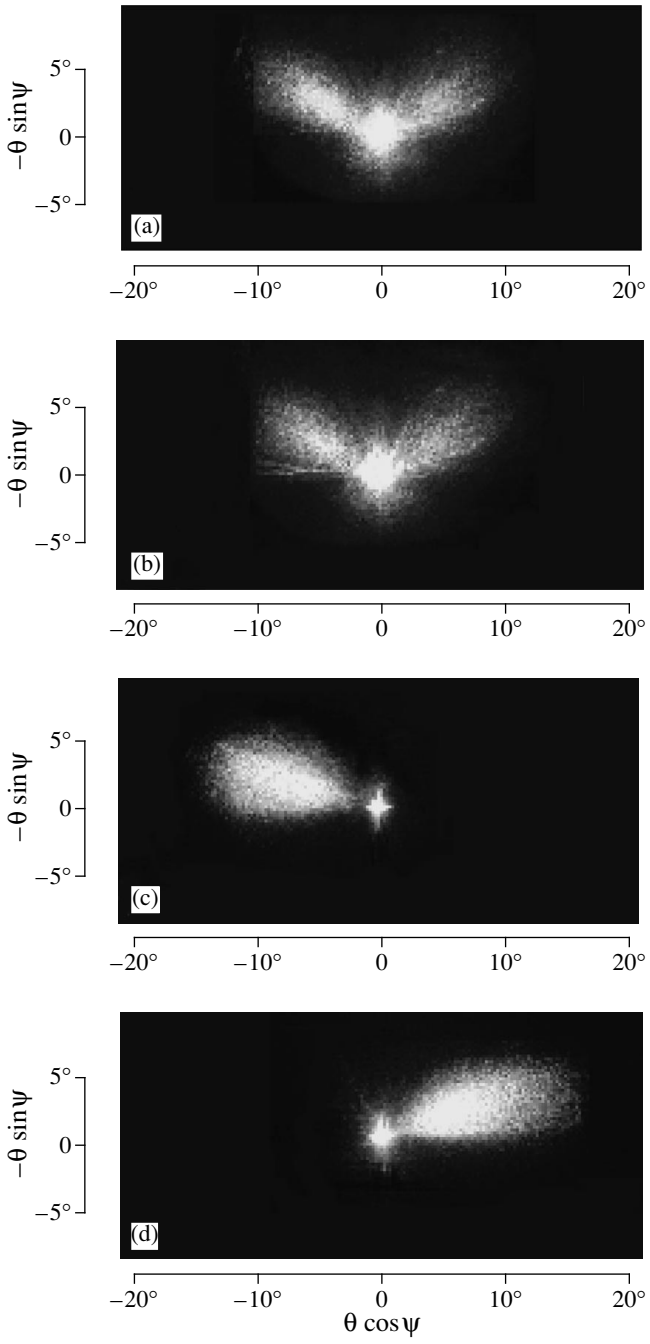

Fig. 7. The experimental patterns of the light-induced scattering measured in the transverse geometry for $|E_0| \approx 20$ kV/cm, $l \approx 2$ cm and $\varphi_p = 0$ (a), $\pi/2$ (b), $\pi/4$ (c), and $-\pi/4$ (d).

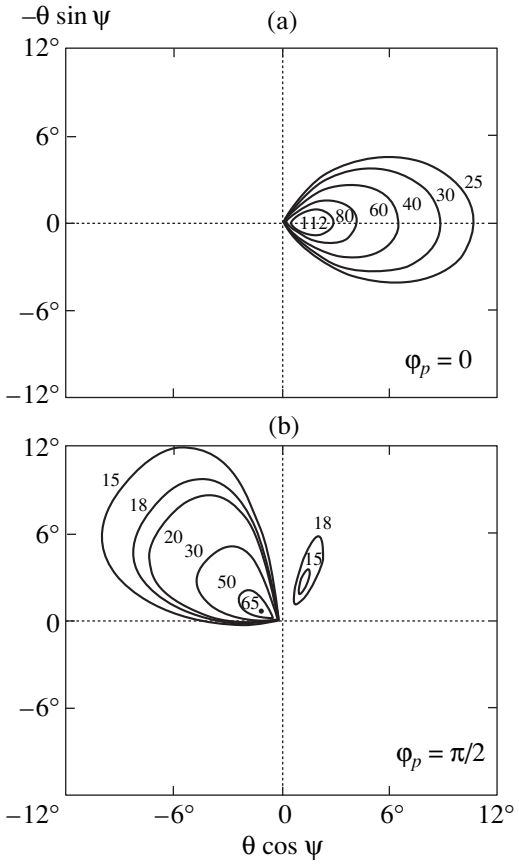

Fig. 8. The patterns of increment isolines for the diagonal configuration with $\varphi_p = 0$ (a) and $\pi/2$ (b) calculated for the same parameters as in Fig. 5.

rated from the origin. This feature is related to the high quality of the crystal sample studied.

5.2. DC Enhancement in $Bi_{12}SiO_{20}$ Crystals

The optical activity observed in the $Bi_{12}SiO_{20}$ crystals is approximately six times that for $Bi_{12}TiO_{20}$, while the applied field strengths usually do not exceed 10 kV/cm.

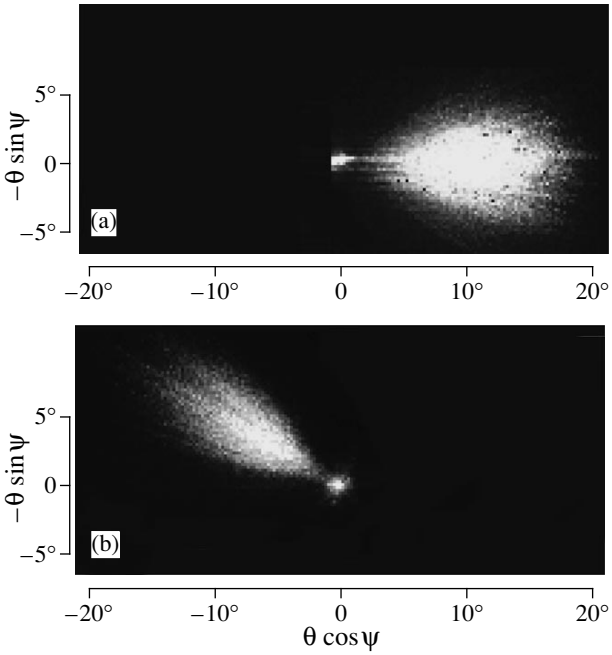

Fig. 9. The patterns of the light-induced scattering measured in the diagonal geometry for $|E_0| \approx 20$ kV/cm, $l \approx 0.8$ cm, and $\varphi_p = 0$ (a) and $\pi/2$ (b).

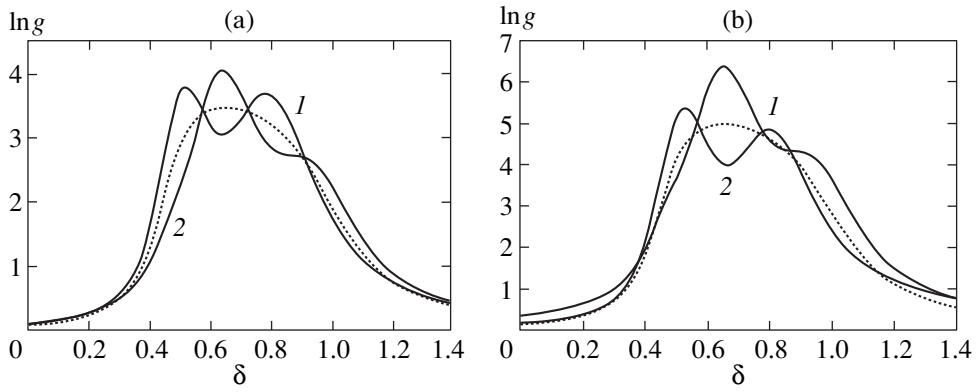

Fig. 10. The plots of $\ln g(\delta)$ calculated for (a) longitudinal and (b) diagonal optical configurations with the parameters taken for $Bi_{12}SiO_{20}$ crystals from Tables 1 and 3 (E_0 = 6 kV/cm; α = 1 cm^{-1}; l = 0.8 cm). The solid curves correspond to (*1*) $\mathbf{a}_{1,2} \parallel \mathbf{E}_0$ and (*2*) $\mathbf{a}_{1,2} \perp \mathbf{E}_0$; the dashed curves correspond to the isotropic (polarization-independent) model.

Under these conditions, the polarization plane rotation tends to average the anisotropic part of the interaction inside the resonance crystal layer. The optical activity and the inhomogeneity related to the light absorption are capable of strongly reducing the level of spatial amplification, which may lead to complicated frequency and angular dependences. The main problems considered below are concerning (a) the efficiency of the simplified model formulated in Section 4 and (b) the optimum conditions for DC amplification under these conditions.

Figure 10 shows the plots of the weak wave amplification factor $g = |\mathbf{a}_1(l)|^2/|\mathbf{a}_1(0)|^2$ calculated as a function of the detuning $\delta = \Omega/\omega_{\mathbf{K}}(0)$ in the longitudinal and diagonal geometries. Here, solid curves *1* and *2* were obtained by numerically integrating Eq. (33) for $\mathbf{a}_{1,2}(0) \parallel \mathbf{E}_0$ and $\mathbf{a}_{1,2}(0) \perp \mathbf{E}_0$, respectively. The dashed curves correspond to an isotropic model in which vector $\mathbf{q}(z)$ is zero (these curves are independent of the polarization). As seen, the solid curves reveal oscillations in the enhancement factor, which are related to the influence of the optical activity. These oscillations were observed in the DC experiments with $Bi_{12}SiO_{20}$ crystals [38]. The details of the oscillatory behavior are strongly polarization-dependent. The greater numerical values of enhancement observed for the diagonal geometry are explained by a greater contribution of the isotropic component ν_0.

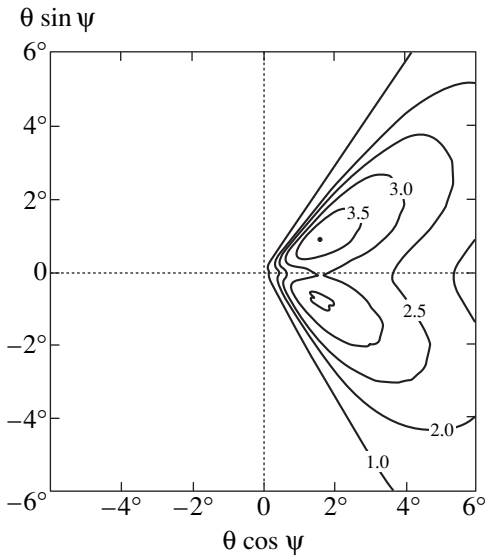

Fig. 11. The pattern of $\ln g(\theta_\parallel, \theta_\perp)$ = const isolines calculated using the isotropic model for the longitudinal geometry with α = 1.5 cm^{-1}, E_0 = 6 kV/cm, l = 4 mm, $\mathbf{a}_{1,2} \parallel \mathbf{E}_0$, and the material parameters from Table 3.

As seen, the simple isotropic model provides for a rough but still reasonable description of the amplification effect. The physical considerations concerning the suppression of the anisotropic contribution to the wave interaction are of considerable heuristic importance. The accuracy of the isotropic model increases with decreasing absorption coefficient α and increasing optical activity ρ.

Now we will consider the influence of the optical activity and spatial inhomogeneity on the angular characteristics of enhancement. Figure 11 shows a distribution of the $\ln g(\theta_\parallel, \theta_\perp)$ value calculated using the isotropic model for the longitudinal geometry with $\mathbf{a}_{1,2}(0) \perp \mathbf{E}_0$, l = 4 mm, and a peak Ω value. As seen, the distribution exhibits two maxima at $\psi \approx \pm 30°$ which correspond to the peaks in the $\nu_0(\zeta)$ curve in Fig. 2 and are related to the elastooptical effect. The peaks in the $g(\theta_\parallel, \theta_\perp)$ distribution possess somewhat different shapes and dimensions. A change in the input wave polarization affects only the details of the split maxima. Beginning with $l \approx 4$ mm, an increase in the crystal thickness leads only to a small increment in $\ln g_{\max}(l)$. Note that the splitting of the angular distribution was also revealed by numerical calculations and observed in experiments with $Bi_{12}SiO_{20}$ crystals [39, 40].

Figure 12 shows a distribution of $\ln g(\theta_{\|}, \theta_{\perp})$ calculated for the diagonal configuration with $l = 8$ mm and $\mathbf{a}_{1,2}(0) \| \mathbf{E}_0$. This pattern differs only slightly from that obtained in the case of $\mathbf{a}_{1,2}(0) \perp \mathbf{E}_0$, which is evidence for the applicability of the isotropic model. The main maximum observed at $\psi \approx -5°$ corresponds to the right peak of $\nu_0(\zeta)$ in Fig. 2, while the secondary maximum ($\psi \approx -130°$) corresponds to the left peak on $\nu_0(\zeta)$. A difference between the amplitudes of maxima in Fig. 12 is related to the different values of $Q(|E_0 \cos\psi|)$. The growth of $\ln g$ as a function of the crystal size l slows down beginning with $l \approx 4$–5 mm.

Finally, we should like to note that the isotropic model is inapplicable to the transverse configuration, where $\nu_0 = 0$ (see Fig. 2). In this case, suppression of the spatial amplification by the inhomogeneity and optical activity is most pronounced.

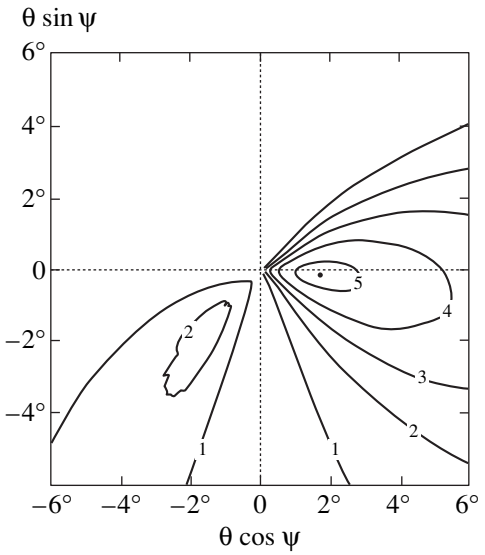

Fig. 12. The pattern of $\ln g(\theta_{\|}, \theta_{\perp}) = \text{const}$ isolines calculated using the isotropic model for the diagonal geometry with $l = 8$ mm and $\mathbf{a}_{1,2} \| \mathbf{E}_0$; the other parameters are the same as in Fig. 11.

6. DISCUSSION AND CONCLUSIONS

First, we will summarize the distinctive features and indicate the advantages of the proposed theory.

The basic relationships of the theory are sufficiently general, covering all the important optical configurations and including factors such as the optical activity, induced birefringence, and the elastooptical effect. The theory also accounts for the AC and DC mechanisms of the photorefractive nonlinear response amplification.

Despite this generality, the main equations are written in a quite compact and informative form. This compromise is achieved owing to the use of the spatial symmetry properties, σ-matrix apparatus, and convenient phenomenological parameters.

The factors of different physical natures are not mixed in the theoretical relationships. The "block" structure allows the data extracted from various experiments to be readily incorporated into the theory. An example is offered by the quality factor Q characterizing the degree of nonlinear photorefractive response enhancement, on the one hand, and determining the observable properties of subharmonics in crystals of the sillenite family, on the other hand [33, 34].

The derived theoretical relationships are readily compatible with various applications and simplified models, including the approximations of undepleted pump, weak and strong optical activity, isotropic model, etc. These relationships help in developing the qualitative notions concerning the general properties of vector interactions.

A systematic consideration of numerous particular cases falls outside the scope of this paper. Sometimes we only indicated the possibility of describing various effects. Nevertheless, we have demonstrated the ability of this theory to predict and explain many distinctive features of the photorefractive amplification in $Bi_{12}TiO_{20}$ and $Bi_{12}SiO_{20}$ crystals.

A practically important conclusion of the proposed theory consists in predicting the possibility of using an external field for suppressing the negative influence of the optical activity on the spatial amplification in $Bi_{12}TiO_{20}$ crystals. The elimination of this negative effect may increase increment Γ and simplify its orientation and polarization properties. This is valid for both AC and DC enhancement methods.

Finally, we will indicate some possibilities for the generalization and application of the proposed theory.

(i) Analysis of the four-wave schemes based on the vector interaction, including optical generation and phase conjugation [41, 42].

(ii) Study of the unusual (critical) optical phenomena arising near the threshold of the parametric generation of space charge waves in crystals of the sillenite family [43], which certainly requires the application of the vectorial interaction theory.

(iii) Description and interpretation of the polarization properties of the spatial enhancement in particular crystals.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, project no. 99-02-17121.

REFERENCES

1. *Photorefractive Materials and Their Applications*, Ed. by P. Günter and J.-P. Huignard (Springer-Verlag, Berlin, 1988, 1989), Top. Appl. Phys., Vols. 61, 62.

2. M. P. Petrov, S. I. Stepanov, and A. V. Khomenko, *Photorefractive Crystals in Coherent Optics* (Nauka, St. Petersburg, 1992).
3. L. Solymar, D. J. Webb, and A. Grunnet-Jepsen, *The Physics and Applications of Photorefractive Materials* (Claredon Press, Oxford, 1996).
4. P. Refregier, L. Solymar, H. Rajbenbach, *et al.*, J. Appl. Phys. **58**, 45 (1985).
5. S. I. Stepanov and M. P. Petrov, Opt. Commun. **53**, 292 (1985).
6. E. Raita, A. A. Kamshilin, V. V. Prokofiev, *et al.*, Appl. Phys. Lett. **70**, 1641 (1997).
7. A. A. Kamshilin, E. Raita, and A. V. Khomenko, J. Opt. Soc. Am. B **13**, 2536 (1996).
8. E. Shamonina, K. H. Ringhofer, B. I. Sturman, *et al.*, Opt. Lett. **23**, 1435 (1998).
9. A. A. Kamshilin, E. Raita, and T. Jaaskelainen, Opt. Rev. **3**, 443 (1996).
10. A. A. Izvanov, A. E. Mandel', D. N. Khat'kov, *et al.*, Avtometriya **2**, 80 (1986).
11. S. I. Stepanov, S. M. Shandarov, and N. D. Khat'kov, Fiz. Tverd. Tela (Leningrad) **29**, 3054 (1987) [Sov. Phys. Solid State **29**, 1754 (1987)].
12. G. Pauliat, P. Mathey, and G. Roosen, J. Opt. Soc. Am. B **8**, 1942 (1991).
13. S. M. Shandarov, A. Emelyanov, O. Kobozev, *et al.*, Proc. SPIE **2801**, 221 (1996).
14. M. Zgonik, K. Nakagava, and P. Günter, J. Opt. Soc. Am. B **12**, 1416 (1995).
15. J. P. Herriau, D. Rojas, J. P. Huignard, *et al.*, Ferroelectrics **75**, 271 (1987).
16. S. F. Lyuksyutov, B. Buchhave, and M. V. Vasnetsov, Phys. Rev. Lett. **79**, 67 (1997).
17. N. V. Kukhtarev, G. E. Dovgalenko, and V. N. Starkov, Appl. Phys. A: Solid Surf. **A 33**, 227 (1984).
18. E. Shamonina, G. Cedilnik, M. Mann, *et al.*, Appl. Phys. B: Lasers Opt. **B 64**, 49 (1997).
19. E. Shamonina, V. P. Kamenov, K. H. Ringhofer, *et al.*, J. Opt. Soc. Am. B **15**, 2552 (1998).
20. V. V. Shepelevich, N. N. Egorov, and V. Shepelevich, J. Opt. Soc. Am. B **11**, 1394 (1994).
21. H. C. Pedersen and P. M. Johansen, J. Opt. Soc. Am. B **12**, 592 (1995).
22. H. Tuovinen, A. A. Kamshilin, and J. Jaaskelainen, J. Opt. Soc. Am. B **14**, 3383 (1997).
23. Yu. I. Sirotin and M. P. Shaskolskaya, *Fundamentals of Crystal Physics* (Nauka, Moscow, 1979; Mir, Moscow, 1982).
24. J. F. Nye, *Physical Properties of Crystals: Their Representation by Tensors and Matrices* (Clarendon Press, Oxford, 1957; Mir, Moscow, 1969).
25. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1996; Pergamon, New York, 1977, 3rd ed.).
26. V. M. Galitskiĭ, B. M. Karnakov, and V. I. Kogan, *Problems in Quantum Mechanics* (Nauka, Moscow, 1981).
27. A. G. Apostolidis, S. Mallik, D. Rouéde, *et al.*, Opt. Commun. **56**, 73 (1985).
28. S. Mallik, D. Rouéde, and A. G. Apostolidis, J. Opt. Soc. Am. B **4**, 1247 (1987).
29. A. Marrakchi, R. V. Johnson, and A. R. Tanguay, J. Opt. Soc. Am. B **3**, 321 (1986).
30. V. V. Shepelevich, S. M. Shandarov, and A. E. Mandel, Ferroelectrics **110**, 235 (1990).
31. J. R. Goff, J. Opt. Soc. Am. B **12**, 99 (1995).
32. R. A. Suris and B. I. Fuks, Fiz. Tekh. Poluprovodn. (Leningrad) **9**, 1717 (1975) [Sov. Phys. Semicond. **9**, 1130 (1975)].
33. B. I. Sturman, M. Mann, J. Otten, *et al.*, J. Opt. Soc. Am. B **10**, 1919 (1993).
34. B. I. Sturman, T. E. McClelland, D. J. Webb, *et al.*, J. Opt. Soc. Am. B **12**, 1621 (1995).
35. C. S. K. Walsh, A. K. Powell, and T. J. Hall, J. Opt. Soc. Am. B **7**, 288 (1990).
36. P. M. Johansen, H. C. Pedersen, E. V. Podivilov, *et al.*, J. Opt. Soc. Am. B **16**, 103 (1999).
37. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Nauka, Moscow, 1973; Pergamon, Oxford, 1975).
38. D. J. Webb and L. Solymar, Opt. Commun. **83**, 287 (1991).
39. H. C. Ellin and L. Solymar, Opt. Commun. **130**, 85 (1996).
40. H. C. Pedersen, P. E. Andersen, and P. M. Johansen, Opt. Lett. **20**, 2475 (1995).
41. M. Cronin-Golomb, B. Fisher, J. O. White, *et al.*, IEEE J. Quantum Electron. **20**, 12 (1984).
42. S. G. Odulov, M. S. Soskin, and A. I. Khizhnyak, *Lasers on Dynamical Lattices* (Nauka, Moscow, 1990).
43. E. V. Podivilov, B. I. Sturman, H. C. Pedersen, *et al.*, Phys. Rev. Lett. **85**, 1867 (2000).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 123–134.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 142–153.

SOLIDS
Electronic Properties

Effective Electrical Characteristics of a Two-Dimensional Three-Component Doubly-Periodic System with Circular Inclusions

B. Ya. Balagurov

Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Moscow, 117997 Russia
e-mail: vkashin@deom.chph.ras.ru

Received May 25, 2000

Abstract—The problem of the electric conductivity of a two-dimensional three-component system containing periodically arranged conducting circular inclusions of two types has been solved. A consistent method for the calculation of the conductivity and other effective electrical characteristics of this model is proposed, which is applicable in the case of arbitrary component concentrations. A complex potential outside the inclusions is expressed in terms of the Weierstrass zeta function and its derivatives. Undetermined coefficients entering into the general expression for the potential are determined from an infinite system of algebraic equations. In the case of a small concentration of inclusions, this system yields a virial expansion for the conductivity. A numerical analysis of this system of equations provides for the principal possibility of investigating various effective characteristics of the model (including the Hall coefficient and thermo emf) within the entire range of the problem parameters. © *2001 MAIK "Nauka/Interperiodica"*.

INTRODUCTION

Investigations of the transfer phenomena in inhomogeneous media mostly concentrate on two-component systems (composites), which are the most simple for analysis. However, solving the "simplest" problem of conductivity (dielectric permittivity, heat transfer, stationary diffusion, etc.) even in this case encounters almost insuperable mathematical difficulties when disordered systems are under consideration. For this reason, data concerning the electric conductivity in such systems were obtained for the most part by model calculations and numerical experiments. Still more complicated are the problems related to description of the thermoelectric, galvanomagnetic, thermogalvanomagnetic, etc., properties of two-component media. Nevertheless, there is certain progress in the study of the electrical properties of these media, which is more significant in the case of two-dimensional systems.

Among the results obtained for two-component media, the following are worth of mentioning. A critical behavior of the conductivity in systems with the metal–dielectric phase transition was described within the framework of a similarity hypothesis [1. 2]. Relationships between the effective conductivities of the initial and the so-called reciprocal system were established in the two-dimensional case [3, 4] (see also [5]). Using the function theory methods, an analytical solution of the conductivity problem was obtained for a series of two-dimensional doubly-periodic models [6–8]. The conductivity [4], galvanomagnetic [9], and thermoelectric [5] characteristics were determined for randomly-inhomogeneous two-dimensional systems with critical compositions. Finally, it was established that, in the two-dimensional case, the description of the thermoelectric, galvanomagnetic, and thermogalvanomagnetic properties is isomorphous to the problem of conductivity [10, 11]. Once an effective conductivity of a two-dimensional system is known, the corresponding isomorphism relationships [10, 11] allow all the main electrical characteristics of this system to be determined and, hence, the whole set of steady-state transfer phenomena in this system to be described. Some results were obtained for the two-component media in the three-dimensional case as well (see, e.g., [1, 2, 10, 12–16]).

A different situation occurs in the study of multicomponent media, the characteristics of which are much more diverse than those of the two-component composites. Although some results from [3–5, 14, 15] and other papers can be applied to multicomponent systems as well, the corresponding theory is yet to be developed. In this field, neither general relationships of the isomorphism type nor the results of consistent calculations for particular models were reported. At the same time, the study of the multicomponent (in particular, three-component) systems is of both theoretical and applied interest. It would be natural to begin the investigation of these systems with periodic objects that are simpler for analysis.

A step in this direction was made in [17], where the conductivity of a two-dimensional three-component system was studied using a generalization of the Ray-

1063-7761/01/9201-0123$21.00

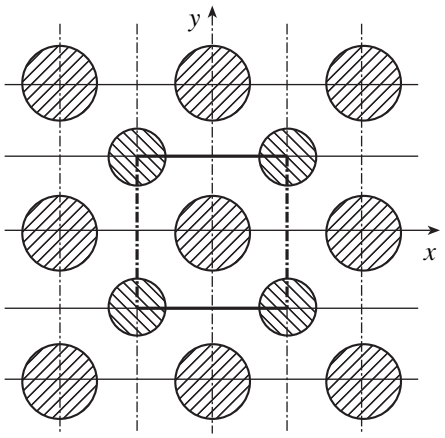

Fig. 1. Schematic diagram of a model two-dimensional three-component lattice.

leigh model [18] representing an isotropic matrix with regularly (chessboard pattern) arranged circular inclusions of two types with different radii and conductivities. However, the method used in [17] was intrinsically approximate and can be applied only to systems with small concentrations of inclusions. Moreover, the exact limits of the validity of the virial expansion for the effective conductivity could not be determined even in this approximation, not to mention the concentrations that cannot be considered as small.

In this study, the problem of the electric conductivity of a two-dimensional three-component model system considered in [17] has been solved using a method developed in [8]. A consistent scheme for calculating the conductivity and other effective electrical characteristics of this model is proposed, which is applicable in the case of arbitrary component concentrations. A complex potential outside the conducting inclusions is expressed (as it was done in [8]) in terms of the Weierstrass zeta function [19, 20] and its derivatives. Undetermined coefficients entering into the general expression for the potential are determined from an infinite system of algebraic equations. In the case of a small concentration of inclusions, this system can be solved by a method of iterations yielding an exact analytical virial expansion for the conductivity and other quantities. In the case of arbitrary concentrations, the system of equations for the coefficients can be solved by numerical methods, which provides for the principal possibility of investigating various effective characteristics of the model within the entire range of the problem parameters.

Within the framework of the conductivity problem, we have also calculated the partial mean-square values of the field strength components ψ_i, which are directly expressed through derivatives of the effective conductivity function σ_e with respect to the variables (component conductivities) [14] (see also [8]). This relationship allows us to study some fine details in the behavior of σ_e (e.g., in vicinity of the metal–insulator phase transition), not revealed by the direct calculation of the effective conductivity. In addition, the knowledge of ψ_i values is necessary for studying the magnetoresistance in weak magnetic fields, the low-frequency dielectric permittivity of metallic conductors, structural fluctuations in the electric field strength and current density, etc.

In addition to the electric conductivity, we have also solved the problem of determining the Hall coefficient in a weak magnetic field **H**. The Hall component σ_{ae} of the effective conductivity tensor $\hat{\sigma}_e$ in a linear (with respect to **H**) approximation is expressed through the coefficients ξ_n and η_n entering into the potential of the conductivity problem at **H** = 0. The same coefficients were used to express the thermo emf of the model with a weak thermoelectric coupling. Note also that a number of exact relationships were established between various quantities, which can be used for checking the correctness of calculations during the numerical analysis of the system of equations for coefficients ξ_n and η_n.

2. ELECTRIC FIELD IN THE MEDIUM

The proposed model represents a two-dimensional matrix possessing an electric conductivity σ_1, containing circular conducting inclusions of two types arranged in a chessboard order (Fig. 1). Inclusions of the first type (radius, R; conductivity, σ_2) form a square lattice with period 2a; inclusions of the second type (radius, ρ; conductivity, σ_3) form a square lattice with the same period shifted by half period in axes x and y. Thus, inclusions of the second type occur at the centers of squares formed by inclusions of the first type and vice versa. Note that this model differs from that considered in [17] only by the coordinate axes being rotated by 45°.

Let us consider a situation when an average electric field vector $\langle \mathbf{E} \rangle$ is directed along the x axis. The field strength $\mathbf{E} = \mathbf{E}(x, y)$ in this system is obviously a periodic function

$$\mathbf{E}(x + 2a, y) = \mathbf{E}(x, y + 2a) = \mathbf{E}(x, y),$$

possessing a certain symmetry (cf. [8]):

$$E_x(-x, y) = E_x(x, -y) = E_x(x, y),$$

$$E_y(-x, y) = E_y(x, -y) = -E_y(x, y).$$

The vertical boundaries of the elementary cell (Fig. 2) and the axis $x = 0$ represent the equipotential lines on which $E_y = 0$, while the horizontal boundaries and the axis $y = 0$ are the current lines on which $E_y = 0$ as well.

The complex potential $\Phi_2(z)$ inside a conducting inclusion with the conductivity σ_2 ($|z| < R$) can be written, with allowance for the field symmetry, in the following form:

$$\Phi_2(z) = \sum_{n=0}^{\infty} A_{2n+1} z^{2n+1}, \quad z = x + iy. \qquad (1)$$

Accordingly, the complex potential $\Phi_2(z)$ inside a conducting inclusion with the conductivity σ_3 ($|z - z_0| < \rho$), centered at the point $z_0 = (1 + i)a$, is

$$\Phi_3(z) = C_0 + \sum_{n=0}^{\infty} C_{2n+1}(z - z_0)^{2n+1}. \quad (2)$$

Similar formulas can be written for the potential of other inclusions centered at the points $z_0 = (1 - i)a$, $z_0 = -(1 + i)a$, and $-(1 - i)a$ belonging to the same elementary cell, but these expressions are not necessary for the consideration.

Outside the conducting inclusions, the complex potential $\Phi_1(z)$ can be presented as a superposition of the potentials of two square lattices [8]

$$\Phi_1(z) = \varphi_0 + \beta z + \sum_{n=0}^{\infty} B_{2n}\zeta^{(2n)}(z) + \sum_{n=0}^{\infty} D_{2n}\zeta^{(2n)}(z - z_0), \quad (3)$$

where $\zeta(z)$ is the Weierstrass zeta function [19, 20], $\zeta^{(2n)}(z)$ are the $2n$-order derivatives of this function, and φ_0 is a constant. In this equation, the linear term z represents an external homogeneous electric field, the terms with $n = 0$ represent the field of induced dipole moments, and the terms with $n \geq 1$ are due to the higher multipoles. For the $\langle \mathbf{E} \rangle$ direction selected, the coefficients φ_0, β, B_{2n}, and D_{2n} in Eq. (3), as well as the quantities A_{2n+1}, C_0, and C_{2n+1} in Eqs. (1) and (2), are real values. Using the known properties of the Weierstrass zeta function [19, 20], one may readily check that the potential described by Eq. (3) satisfies the above conditions at the boundaries and on the symmetry axes of the elementary cell. Note that the real part of the complex potential determines a real electric potential

$$\varphi(\mathbf{r}) = \mathrm{Re}\Phi(z),$$

while the derivative of $\Phi(z)$ is related to the components of the field strength vector:

$$\Phi'(z) = -E_x + iE_y.$$

For the square lattice under consideration, characterized by the half-periods $\omega = a$ and $\omega' = ia$ in the complex plane (the so-called lemniscate case), the Weierstrass zeta function obeys the following relationships [19, 20]:

$$\zeta(z + 2a) = \zeta(z) + \frac{\pi}{2a}, \qquad \zeta(z + 2ia) = \zeta(z) - i\frac{\pi}{2a}, \quad (4)$$

$$g_2 = \frac{1}{a^4}\left[K\left(\frac{1}{\sqrt{2}}\right)\right]^4, \quad g_3 = 0, \quad (5)$$

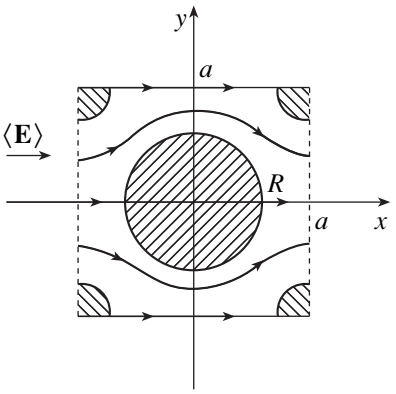

Fig. 2. Schematic diagram of an elementary unit cell of the model lattice. Dashed curves show equipotential lines, solid curves with arrows show the current lines.

where g_2 and g_3 are the invariants of the Weierstrass zeta function and $K(1/\sqrt{2}) = 1.85407\ldots$ is the complete elliptic integral of the first kind $K(k)$ for the modulus $k = 1/\sqrt{2}$.

The electric potentials $\varphi_i(\mathbf{r}) = \mathrm{Re}\Phi_i(z)$ ($i = 1, 2, 3$) must satisfy the usual boundary conditions (see the next section), which can be used to determine the coefficients entering into Eqs. (1)–(3). This procedure is performed with the $\zeta(z)$ function expanded in the vicinity of the point $z = 0$ [20]:

$$\zeta(z) = \frac{1}{z} - \sum_{k=2}^{\infty} \frac{c_k}{2k-1} z^{2k-1}, \quad (6)$$

where

$$c_2 = \frac{g_2}{20}, \quad c_4 = \frac{1}{3}c_2^2, \quad c_6 = \frac{2}{3 \times 13}c_2^3, \quad c_8 = \frac{5}{3 \times 13 \times 17}c_2^4, \quad \ldots. \quad (7)$$

In the above lemniscate case, all the c_k coefficients with odd indexes are zero, while the c_{2k} quantities satisfy a recurrent relationship [20]

$$c_{2k} = \frac{3}{(4k+1)(2k-3)} \sum_{m=1}^{k-1} c_{2m}c_{2k-2m}, \quad k \geq 2, \quad (8)$$

which can be used for determining the sequential coefficients c_{2k} with increasing number k.

Besides expression (6), we will also employ an expansion of the $\zeta(z)$ function in the vicinity of the point $z = z_0$ [20]:

$$\zeta(z) = \zeta(z_0) - \sum_{k=2}^{\infty} \frac{d_k}{2k-1}(z - z_0)^{2k-1}, \quad (9)$$

where

$$\zeta(z_0) = (1-i)\frac{\pi}{4a}, \quad z_0 = (1+i)a, \tag{10}$$

and

$$d_2 = -\frac{g_2}{4}, \quad d_4 = \frac{1}{5}d_2^2,$$
$$d_6 = \frac{2}{75}d_2^3, \quad d_8 = \frac{1}{325}d_2^4, \quad \ldots. \tag{11}$$

Coefficients d_k with odd numbers are zero, while the d_{2k} quantities satisfy a recurrent relationship

$$d_{2k} = \frac{3}{(2k-1)(4k-3)}\sum_{m=1}^{k-1} d_{2m}d_{2k-2m}, \quad k \geq 2, \tag{12}$$

obtained upon substituting expansion (9) into the differential equation [19, 20]

$$2\wp''(z) = 12\wp^2(z) - g_2,$$

where $\wp(z) = -\zeta'(z)$ is the Weierstrass elliptic function. A relationship between the coefficients d_{2k} c_{2k} from Eqs. (6)–(8) is established below (see Section 8, Eq. (99)).

By the same token, we may consider a situation with the average field strength vector $\langle \mathbf{E} \rangle$ directed along the y axis. The corresponding values are denoted by symbols with an upper bar. The complex potentials in the regions with conductivities σ_1, σ_2, and σ_3 are given by the following formulas:

$$\bar{\Phi}_1(z) = \bar{\varphi}_0 - i\left\{\bar{\beta}z - \sum_{n=0}^{\infty}\bar{B}_{2n}\zeta^{(2n)}(z) - \sum_{n=0}^{\infty}\bar{D}_{2n}\zeta^{(2n)}(z-z_0)\right\}, \tag{13}$$

$$\bar{\Phi}_2(z) = -i\sum_{n=0}^{\infty}\bar{A}_{2n+1}z^{2n+1}, \tag{14}$$

$$\bar{\Phi}_3(z) = \bar{C}_0 - i\sum_{n=0}^{\infty}\bar{C}_{2n+1}(z-z_0)^{2n+1}, \tag{15}$$

where the notations are the same as in Eqs. (1)–(3), the quantities $\bar{\varphi}_0$, $\bar{\beta}$, $\bar{B}_{2n}$, $\bar{D}_{2n}$, $\bar{A}_{2n+1}$, $\bar{C}_0$, and $\bar{C}_{2n+1}$ being real.

3. BOUNDARY CONDITIONS

At the boundaries of an inclusion of the first type ($\sigma = \sigma_2$) with $r = R$, the potential must obey the usual conditions

$$\varphi_1 = \varphi_2, \quad \frac{\partial\varphi_1}{\partial r} = h_2\frac{\partial\varphi_2}{\partial r}, \quad h_2 = \frac{\sigma_2}{\sigma_1}. \tag{16}$$

In the vicinity of this inclusion, the electric potential $\varphi_1(\mathbf{r}) = \mathrm{Re}\Phi_1(z)$ can be presented using expansion (6) for the $\zeta(z)$ function and an expansion according to Eq. (9) for the $\zeta(z-z_0)$ function:

$$\zeta(z-z_0) = -\zeta(z_0) - \sum_{k=2}^{\infty}\frac{d_k}{2k-1}z^{2k-1}. \tag{17}$$

The derivatives $\zeta^{(2n)}(z)$ can be expressed using formula [8, (A.2)], while the quantities $\zeta^{(2n)}(z-z_0)$ are determined from Eq. (17) as

$$\zeta^{(2n)}(z-z_0) = -\zeta(z_0)\delta_{n0} - \sum_{m=0}^{\infty}\frac{(2n+2m)!}{(2m+1)!}d_{n+m+1}z^{2m+1}. \tag{18}$$

Substituting these expansions for $\zeta^{(2n)}(z)$ and $\zeta^{(2n)}(z-z_0)$ into Eq. (3), taking $z = r\exp(i\theta)$, and separating the real part, we obtain an expression for $\varphi_1(\mathbf{r})$ in the vicinity of inclusions of the first type. An expression for the potential $\varphi_2(\mathbf{r}) = \mathrm{Re}\Phi_2(z)$ is readily obtained from Eq. (1).

Substituting the expressions for $\varphi_1(\mathbf{r})$ and $\varphi_2(\mathbf{r})$ into the boundary conditions (16) and using a method analogous to that described in [8], we may determine the coefficients A_{2n+1} and B_{2n}:

$$A_{2n+1} = \frac{2}{1-h_2}\frac{(2n)!}{R^{4n+2}}B_{2n}, \quad \varphi_0 = \frac{\pi}{4a}D_0, \tag{19}$$

$$B_{2n} + \frac{1-h_2}{1+h_2}\sum_{m=0}^{\infty}\frac{(2n+2m)!}{(2n)!(2n+1)!} \times R^{4n+2}(B_{2m}c_{n+m+1} + D_{2m}d_{n+m+1}) = \frac{1-h_2}{1+h_2}\beta R^2\delta_{n0}. \tag{20}$$

In the vicinity of an inclusion of the second type ($\sigma = \sigma_3$), the functions $\zeta(z)$ and $\zeta(z-z_0)$ can be expressed using analogous expansions at $z = z_0$ and taking $z - z_0 = r'\exp(i\theta')$. The usual boundary conditions at $r' = \rho$

$$\varphi_1 = \varphi_3, \quad \frac{\partial\varphi_1}{r'} = h_3\frac{\partial\varphi_3}{\partial r'}, \quad h_3 = \frac{\sigma_3}{\sigma_1}, \tag{21}$$

yield

$$\varphi_0 + \beta a + B_0\frac{\pi}{4a} = C_0, \tag{22}$$

$$C_{2n+1} = \frac{2}{1-h_3}\frac{(2n)!}{\rho^{4n+2}}D_{2n}, \tag{23}$$

$$D_{2n} + \frac{1-h_3}{1+h_3}\sum_{m=0}^{\infty}\frac{(2n+2m)!}{(2n)!(2n+1)!} \times \rho^{4n+2}(B_{2m}d_{n+m+1} + D_{2m}c_{n+m+1}) = \frac{1-h_3}{1+h_3}\beta\rho^2\delta_{n0}. \tag{24}$$

Replacing B_{2n} and D_{2n} by the new variables ξ_n and η_n defined as

$$B_{2n} = \beta\frac{R^{2n+2}\delta_2}{\sqrt{(2n)!(2n+1)!}}\xi_n, \quad \delta_2 = \frac{1-h_2}{1+h_2}, \tag{25}$$

$$D_{2n} = \beta\frac{\rho^{2n+2}\delta_3}{\sqrt{(2n)!(2n+1)!}}\eta_n, \quad \delta_3 = \frac{1-h_3}{1+h_3}, \tag{26}$$

we may write Eqs. (20) and (24) in the following form:

$$\xi_n + \sum_{m=0}^{\infty}(M_{nm}\xi_m + P_{nm}\eta_m) = \delta_{n0}, \tag{27}$$

$$\eta_n + \sum_{m=0}^{\infty}(Q_{nm}\xi_m + N_{nm}\eta_m) = \delta_{n0}, \tag{28}$$

where

$$\begin{aligned} M_{nm} &= G_{nm}R^{2(n+m+1)}c_{n+m+1}\delta_2, \\ P_{nm} &= G_{nm}R^{2n}\rho^{2m+2}d_{n+m+1}\delta_3, \\ Q_{nm} &= G_{nm}\rho^{2n}R^{2m+2}d_{n+m+1}\delta_2, \\ N_{nm} &= G_{nm}\rho^{2(n+m+1)}c_{n+m+1}\delta_3, \\ G_{nm} &= \frac{(2n+2m)!}{\sqrt{(2n)!(2n+1)!(2m)!(2m+1)!}}. \end{aligned} \tag{29}$$

Here, matrices $\hat{M}$ and $\hat{N}$ are symmetric ($M_{nm} = M_{mn}$, $N_{nm} = N_{mn}$) and matrices $\hat{P}$ and $\hat{Q}$ are related by the equation

$$P_{nm}R^2\delta_2 = Q_{mn}\rho^2\delta_3. \tag{30}$$

As noted above, coefficients c_k and d_k with odd subscripts are zero. For this reason, only the matrix elements M_{nm}, P_{nm}, Q_{nm}, and N_{nm} with different (even-odd) indexes m and n are nonzero. The coefficients A_{2n+1} and C_{2n+1} are expressed through ξ_n and η_n by the formulas

$$A_{2n+1} = \frac{2\beta}{1+h_2}\frac{1}{R^{2n}}\frac{1}{\sqrt{2n+1}}\xi_n, \tag{31}$$

$$C_{2n+1} = \frac{2\beta}{1+h_3}\frac{1}{\rho^{2n}}\frac{1}{\sqrt{2n+1}}\eta_n. \tag{32}$$

By introducing the column vectors

$$\mathbf{x}_n = \begin{pmatrix}\xi_n\\ \eta_n\end{pmatrix}, \quad \mathbf{1} = \begin{pmatrix}1\\ 1\end{pmatrix}, \tag{33}$$

the set of Eqs. (27) and (28) can be written on a compact "vector" form:

$$\mathbf{x}_n + \sum_{m=0}^{\infty}\hat{S}_{nm}\mathbf{x}_m = \mathbf{1}\cdot\delta_{n0}, \tag{34}$$

where

$$\hat{S}_{nm} = \begin{pmatrix}M_{nm} & P_{nm}\\ Q_{nm} & N_{nm}\end{pmatrix}. \tag{35}$$

Upon solving Eq. (34) by the iteration method with expansion in the powers of matrix elements $\hat{S}_{nm}$, we obtain the following formal solution:

$$\mathbf{x}_n = \left\{\delta_{n0} - \hat{S}_{n0} + \sum_m \hat{S}_{nm}\hat{S}_{m0} - \sum_{lm}\hat{S}_{nl}\hat{S}_{lm}\hat{S}_{m0} + \sum_{klm}\hat{S}_{nk}\hat{S}_{kl}\hat{S}_{lm}\hat{S}_{m0} - \ldots\right\}\cdot\mathbf{1}. \tag{36}$$

For $n = 0$, this yields

$$\mathbf{x}_0 = \left\{1 + \sum_m \hat{S}_{0m}\hat{S}_{m0} + \sum_{klm}\hat{S}_{0k}\hat{S}_{kl}\hat{S}_{lm}\hat{S}_{m0} + \ldots\right\}\cdot\mathbf{1}, \tag{37}$$

where only $\hat{S}_{nm}$ with different (even-odd) indexes m and n are nonzero.

Equations (1)–(3) and (25)–(37), together with an expression for the difference of potentials U_x (see Eq. (43) below) describe an exact solution for the problem of describing electric potential in the medium under consideration.

By the same token, we obtain for the case when the average field $\langle\mathbf{E}\rangle$ is directed along the y axis:

$$\bar{\varphi}_0 = \bar{D}_0\frac{\pi}{4a}, \quad \bar{\varphi}_0 + \bar{\beta}a + \bar{B}_0\frac{\pi}{4a} = \bar{C}_0,$$

$$\bar{A}_{2n+1} = \frac{2\bar{\beta}}{1+h_2}\frac{1}{R^{2n}}\frac{1}{\sqrt{2n+1}}\bar{\xi}_n,$$

$$\bar{C}_{2n+1} = \frac{2\bar{\beta}}{1+h_3}\frac{1}{\rho^{2n}}\frac{1}{\sqrt{2n+1}}\bar{\eta}_n, \tag{38}$$

$$\bar{B}_{2n} = \bar{\beta}\frac{R^{2n+2}\delta_2}{\sqrt{(2n)!(2n+1)!}}\bar{\xi}_n,$$

$$\bar{D}_{2n} = \bar{\beta}\frac{\rho^{2n+2}\delta_3}{\sqrt{(2n)!(2n+1)!}}\bar{\eta}_n.$$

Here $\bar{\xi}_n$ and $\bar{\eta}_n$ satisfy the set of equations

$$\begin{aligned}\bar{\xi}_n - \sum_{m=0}^{\infty}(M_{nm}\bar{\xi}_m + P_{nm}\bar{\eta}_m) &= \delta_{n0},\\ \bar{\eta}_n - \sum_{m=0}^{\infty}(Q_{nm}\bar{\xi}_m + N_{nm}\bar{\eta}_m) &= \delta_{n0},\end{aligned} \tag{39}$$

where M_{nm}, P_{nm}, Q_{nm}, and N_{nm} are the same matrix elements as in Eqs. (27)–(29). Set (39) can also be written in the "vector" form:

$$\bar{\mathbf{x}}_n - \sum_{m=0}^{\infty}\hat{S}_{nm}\bar{\mathbf{x}}_m = \mathbf{1}\cdot\delta_{n0}, \tag{40}$$

where $\hat{S}_{nm}$ are the matrix elements defined by formulas (35).

Using formulas (40) for $\bar{\mathbf{x}}_n$, we may obtain an expansion analogous to (36) upon the substitution $\hat{S}_{nm} \longrightarrow -\hat{S}_{nm}$, which yields

$$\bar{\xi}_n = (-1)^n\xi_n, \quad \bar{\eta}_n = (-1)^n\eta_n, \tag{41}$$

in particular, $\bar{\xi}_0 = \xi_0$ and $\bar{\eta}_0 = \eta_0$. Accordingly, upon comparing Eqs. (38) with (25), (26) and (31), (32) and taking into account (41), we obtain (for $\bar{\beta} = \beta$)

$$\begin{aligned}\bar{B}_{2n} &= (-1)^n B_{2n}, \quad \bar{D}_{2n} = (-1)^n D_{2n},\\ \bar{A}_{2n+1} &= (-1)^n A_{2n+1}, \quad \bar{C}_{2n+1} = (-1)^n C_{2n+1}.\end{aligned} \tag{42}$$

4. EFFECTIVE CONDUCTIVITY

Using the complex potential $\Phi(z)$, we may calculate (using a procedure analogous to that described in [8]) the voltage drop U_x across the elementary cell and the total current I_x through this cell in the direction of the x axis:

$$U_x = -2a\left[\beta + (B_0 + D_0)\frac{\pi}{4a^2}\right], \tag{43}$$

$$I_x = -2\sigma_1 a\left[\beta - (B_0 + D_0)\frac{\pi}{4a^2}\right]. \tag{44}$$

For the effective conductivity defined as $\sigma_{\text{eff}} = I_x/U_x$, formulas (43) and (44) yield

$$\sigma_{\text{eff}} = \sigma_1\frac{\beta - (B_0 + D_0)\pi/4a^2}{\beta + (B_0 + D_0)\pi/4a^2}. \tag{45}$$

Substituting

$$B_0 = \xi_0\beta R^2\delta_2 \text{ and } D_0 = \eta_0\beta\rho^2\delta_3,$$

determined from Eqs. (25) and (26), respectively, we obtain an expression for the dimensionless conductivity defined as $f = \sigma_{\text{eff}}/\sigma_1$:

$$f = \frac{1 - \xi_0 p_2\delta_2 - \eta_0 p_3\delta_3}{1 + \xi_0 p_2\delta_2 + \eta_0 p_3\delta_3}, \tag{46}$$

where $p_2 = \pi R^2/(2a)^2$ and $p_3 = \pi\rho^2/(2a)^2$ are the concentrations of the second and third components (i.e., inclusions of the first and second type), respectively. As can be readily checked, the same expression for f is obtained when the electric field $\langle\mathbf{E}\rangle$ is directed along the y axis. Thus, the effective conductivity in the model under consideration can be calculated as soon as the ξ_0 and η_0 values are determined.

For a small density of both inclusions (i.e., for $R \ll a$ and $\rho \ll a$), coefficients ξ_n and η_n can be determined using Eqs. (36) and (37) in the form of virial series. Restricting these cumbersome expansions to the first terms, we obtain the following expression from Eq. (37):

$$\mathbf{x}_0 = \{1 + \hat{S}_{01}\hat{S}_{10} + \ldots\}\cdot\mathbf{1},$$

which yields

$$\begin{aligned}\xi_0 &= 1 + \frac{1}{3}(R^8 c_2^2\delta_2^2 + R^6\rho^2 c_2 d_2\delta_2\delta_3\\ &\quad + R^2\rho^6 d_2^2\delta_2\delta_3 + \rho^8 c_2 d_2\delta_3^2) + \ldots,\\ \eta_0 &= 1 + \frac{1}{3}(R^8 c_2 d_2\delta_2^2 + R^6\rho^2 c_2 d_2\delta_2\delta_3\\ &\quad + R^2\rho^6 c_2 d_2\delta_2\delta_3 + \rho^8 c_2^2\delta_3^2) + \ldots\end{aligned} \tag{47}$$

By the same token, Eq, (36) gives

$$\mathbf{x}_1 = \{-S_{10} + \ldots\}\cdot\mathbf{1},$$

and

$$\begin{aligned}\xi_1 &= -\frac{1}{\sqrt{3}}(R^4 c_2\delta_2 + R^2\rho^2 d_2\delta_3) + \ldots,\\ \eta_1 &= -\frac{1}{\sqrt{3}}(R^2\rho^2 d_2\delta_2 + \rho^4 c_2\delta_3) + \ldots\end{aligned} \tag{48}$$

In the case when the size of inclusions R and ρ is not small, the ξ_n and η_n values ($n = 0, 1, 2, \ldots$) can be obtained by numerically solving the set of Eqs. (27) and (28).

In the two-dimensional case, the medium satisfies a reciprocation relationship [3, 4] between the effective

conductivity of the initial system (with the local conductivity $\sigma(\mathbf{r})$) and the so-called reciprocal system (with the local conductivity $\lambda^2/\sigma(\mathbf{r})$):

$$\sigma_{\text{eff}}(\{\sigma(\mathbf{r})\})\sigma_{\text{eff}}(\{\lambda^2/\sigma(\mathbf{r})\}) = \lambda^2,$$

where λ is some constant quantity that is a parameter of the corresponding symmetry transformation [5]. In this form, the reciprocation relationship is valid for the isotropic two-dimensional systems with $\sigma(\mathbf{r})$ representing an arbitrary function of the coordinates. For an N-component medium, this relationship is as follows:

$$\sigma_{\text{eff}}(\sigma_1, \sigma_2, \ldots, \sigma_N)\sigma_{\text{eff}}\left(\frac{\lambda^2}{\sigma_1}, \frac{\lambda^2}{\sigma_2}, \ldots, \frac{\lambda^2}{\sigma_N}\right) = \lambda^2. \quad (49)$$

A particular value of parameter λ is insignificant, because this quantity does not enter into the expression for the dimensionless effective conductivity f defined as

$$\sigma_{\text{eff}}(\sigma_1, \sigma_2, \ldots, \sigma_N) = \sigma_1 f(\sigma_2/\sigma_1, \ldots, \sigma_N/\sigma_1),$$

according to which

$$f(h_2, \ldots, h_N) f(1/h_2, \ldots, 1/h_N) = 1, \quad h_i = \sigma_i/\sigma_1. \quad (50)$$

Note that the transition to a reciprocal system ($h_2 \longrightarrow 1/h_2$, $h_3 \longrightarrow 1/h_3$, …) is equivalent to the substitution $\delta_2 \longrightarrow -\delta_2$, $\delta_3 \longrightarrow -\delta_3$, …, where $\delta_i = (1 - h_i)/(1 + h_i)$. For the three-component model under consideration, the matrix $\hat{S}_{nm}$ defined by Eq. (35) changes its sign upon transition, while quantities ξ_0 and η_0 remain unchanged. As a result, f transforms into $1/f$ according to Eq. (46) and the reciprocation relationship (50) for $N = 3$ is automatically satisfied.

5. PARTIAL QUADRATIC CHARACTERISTICS

The effective conductivity σ_{eff} of a composite is directly related to the partial mean-square components of the electric field strength [8, 14]:

$$\psi_i = \langle \mathbf{e}^2 \rangle^{(i)} = \frac{\partial \sigma_e}{\partial \sigma_i}, \quad \mathbf{e}(\mathbf{r}) = \mathbf{E}(\mathbf{r})/|\langle \mathbf{E} \rangle|, \quad (51)$$

where $\langle \ldots \rangle^{(i)}$ denotes the integral over volume (or area in the two-dimensional case) of the ith component, divided by the total sample volume V. Using the well-known identity $\langle \mathbf{j} \cdot \mathbf{E} \rangle = \langle \mathbf{j} \rangle \cdot \langle \mathbf{E} \rangle = \sigma_e(\langle \mathbf{E} \rangle)^2$ [4, 14], we obtain the following relationship for an N-component medium:

$$\sum_{i=1}^{N} h_i \psi_i = f, \quad (52)$$

where the ψ_i and h_i values are determined from Eqs. (51) and (50), respectively. For the three-component system, this yields:

$$\psi_1 = f - h_2\psi_2 - h_3\psi_3, \quad (53)$$

which implies that there is no need to directly calculate ψ_1 by Eq. (51) for the model under consideration.

Using potentials (1) and (2), we obtain

$$\psi_i = \frac{4}{(1 + h_i)^2}\frac{p_i J_i}{\Delta^2}, \quad i = 2, 3, \quad (54)$$

$$J_2 = \sum_{n=0}^{\infty} \xi_n^2, \quad J_3 = \sum_{n=0}^{\infty} \eta_n^2, \quad (55)$$

$$\Delta = 1 + \xi_0 p_2 \delta_2 + \eta_0 p_3 \delta_3. \quad (56)$$

According to expressions (51), the system must obey the following relationships:

$$\psi_2 = \frac{\partial f}{\partial h_2}, \quad \psi_3 = \frac{\partial f}{\partial h_3}. \quad (57)$$

Using formulas (46)–(48) one may readily check that these relationships are really satisfied to within the terms on the order of R^8 and ρ^8. However, the validity of Eqs. (57) for the model under consideration can be proved by direct calculation for arbitrary R and ρ as well.

Indeed, differentiating Eq. (46) with respect to h_2 yields

$$\frac{\partial f}{\partial h_2} = \frac{4}{(1 + h_2)^2} \times \left[\left(\xi_0 + \frac{\partial \xi_0}{\partial \delta_2}\delta_2\right)p_2 + \frac{\partial \eta_0}{\partial \delta_2} p_3 \delta_3\right]\Delta^{-2}. \quad (58)$$

For the first equality in (57) to be valid with ψ_2 from Eq. (54) and $\partial f/\partial h_2$ from (58), it is necessary that

$$\frac{\partial \xi_0}{\partial \delta_2} p_2 \delta_2 + \frac{\partial \eta_0}{\partial \gamma_2} p_3 \delta_3 = p_2(J_2 - \xi_0) \quad (59)$$

with the J_2 value determined according to Eq. (55).

We may rewrite Eq. (36) in the following form:

$$(\mathbf{x}_n)^{\nu} = \sum_{\mu} ((1 + \hat{S})^{-1})^{\nu\mu}_{n0}, \quad (60)$$

where

$$(\mathbf{x}_n)^1 = \xi_n, \quad (\mathbf{x}_n)^2 = \eta_n,$$

$$(\hat{S})^{11}_{nm} = M_{nm}, \quad (\hat{S})^{12}_{nm} = P_{nm},$$

$$(\hat{S})^{21}_{nm} = Q_{nm}, \quad (\hat{S})^{22}_{nm} = N_{nm}.$$

Differentiating Eq. (60) with respect to δ_2 (for $\nu = 1$ and $n = 0$) yields

$$\frac{\partial \xi_0}{\partial \delta_2}\delta_2 = -\sum_{\mu}((1+\hat{S})^{-1}\hat{V}(1+\hat{S})^{-1})^{1\mu}_{00} \tag{61}$$

and an analogous expression for $(\partial\eta_0/\partial\delta_2)\delta_2$. Matrix $\hat{V}$ in Eq. (61) has the following form:

$$\hat{V}_{nm} = \begin{pmatrix} M_{nm} & 0 \\ Q_{nm} & 0 \end{pmatrix}. \tag{62}$$

Using the condition of symmetry of the matrix elements M_{nm} and N_{nm} and Eq. (30), one may readily obtain the following relationships:

$$\begin{aligned} ((1+\hat{S})^{-1})^{11}_{0n} &= ((1+\hat{S})^{-1})^{11}_{n0}, \\ ((1+\hat{S})^{-1})^{12}_{0n} &= \frac{p_2\delta_3}{p_2\delta_2}((1+\hat{S})^{-1})^{21}_{n0}, \\ ((1+\hat{S})^{-1})^{21}_{0n} &= \frac{p_2\delta_2}{p_3\delta_3}((1+\hat{S})^{-1})^{12}_{n0}, \\ ((1+\hat{S})^{-1})^{22}_{0n} &= ((1+\hat{S})^{-1})^{22}_{n0}. \end{aligned} \tag{63}$$

Substituting these formulas, the definition of ξ_n from Eq. (60), and the relationship (30) into Eq. (61) yields

$$\begin{aligned} \frac{\partial \xi_0}{\partial \delta_2}\delta_2 = -\sum_{nm}\{&((1+\hat{S})^{-1})^{11}_{n0}M_{nm} \\ &+ ((1+\hat{S})^{-1})^{21}_{n0}P_{nm}\}\xi_m. \end{aligned} \tag{64}$$

By the same token, we obtain

$$\begin{aligned} \frac{\partial \eta_0}{\partial \delta_2}\delta_2 = -\frac{p_2\delta_2}{p_3\delta_3}\sum_{nm}\{&((1+\hat{S})^{-1})^{12}_{n0}M_{nm} \\ &+ ((1+\hat{S})^{-1})^{22}_{n0}P_{nm}\}\xi_m. \end{aligned} \tag{65}$$

Using these expressions and taking into account the definitions of ξ_n and η_n in Eq. (60), we arrive at

$$\begin{aligned} &\frac{\partial \xi_0}{\partial \delta_2}p_2\delta_2 + \frac{\partial \eta_0}{\partial \delta_2}p_3\delta_3 \\ &= -p_2\sum_{nm}\{\xi_n M_{nm} + \eta_n P_{nm}\}\xi_m. \end{aligned} \tag{66}$$

Multiplying Eq. (27) by ξ_n and summing over n, we obtain

$$J_2 = \xi_0 - \sum_{nm}(\xi_n M_{nm} + \eta_n P_{mn})\xi_m. \tag{67}$$

Substituting Eqs. (66) and (67) into (59) transforms it into an identity. By a similar procedure, one may readily show that the second condition in (57) is valid for the model under consideration as well.

As was pointed out in the introduction, the knowledge of ψ_i values allows us to study the behavior of the effective conductivity σ_{eff} in much detail. In addition, the derivatives $\partial\sigma_{\text{eff}}/\partial\sigma_i = \psi_i$ may appear, for example, in the description of the magnetoresistance in a weak magnetic field, the low-frequency dielectric permittivity of metallic conductors, and structural fluctuations in the electric field strength and current density. For certain relationships between the system parameters, the ψ_i functions appear in the problem of determining thermo emf as well (see Section 7).

6. HALL COEFFICIENT

The effective Hall coefficient R_e in a weak magnetic field **H** is expressed through a non-diagonal (Hall's) component σ_{ae} of the $\hat{\sigma}_{\text{eff}}$ tensor:

$$R_e = \frac{1}{H}\frac{\sigma_{ae}}{\sigma^2_{\text{eff}}}.$$

According to [14], the σ_{ae} values in the linear (with respect to **H**) approximation are given by the general formula

$$\sigma_{ae} = \frac{\langle \sigma_a[\mathbf{E}^{(x)} \times \mathbf{E}^{(y)}]_z\rangle}{\langle E_x^{(x)}\rangle\langle E_y^{(y)}\rangle}, \tag{68}$$

which is valid for an arbitrary inhomogeneous medium (both two- and three-dimensional). In Eq. (68), the angular brackets $\langle\ldots\rangle$ denote averaging over the sample volume V (or area in the two-dimensional case), $\mathbf{E}^{(\nu)} = \mathbf{E}^{(\nu)}(\mathbf{r})$ is the electric field strength in the medium in the absence of the magnetic field (ν indicating the axis ν in which the average field $\langle\mathbf{E}^{(\nu)}\rangle$ is directed), and $[\ldots]_z$ denotes the z-component of the vector product.

For an N-component medium, Eq. (68) acquires the form

$$\sigma_{ae} = \sum_{i=1}^{N}\sigma_{ai}\varphi_i, \tag{69}$$

$$\varphi_i = \frac{\langle E_x^{(x)}E_y^{(y)} - E_y^{(x)}E_x^{(y)}\rangle^{(i)}}{\langle E_x^{(x)}\rangle\langle E_y^{(y)}\rangle}, \tag{70}$$

where $\langle\ldots\rangle^{(i)}$ denotes the same integral as in Eq. (51). As was demonstrated in [14], the $\mathbf{E}^{(\mu)}(\mathbf{r})$ and $\mathbf{E}^{(\nu)}(\mathbf{r})$ values satisfy the identity

$$\langle\mathbf{E}^{(\mu)} \times \mathbf{E}^{(\nu)}\rangle = \langle\mathbf{E}^{(\mu)}\rangle \times \langle\mathbf{E}^{(\nu)}\rangle,$$

which leads to the following "summing rule" for φ_i functions in the N-component medium:

$$\sum_{i=1}^{N}\varphi_i = 1. \tag{71}$$

Upon excluding φ_1 with the aid of Eq. (71), we obtain from Eq. (69) the following relationship for a three-component medium:

$$\sigma_{ae} = \sigma_{a1} + (\sigma_{a2} - \sigma_{a1})\varphi_2 + (\sigma_{a3} - \sigma_{a1})\varphi_3. \quad (72)$$

In the two-dimensional case, the system obeys another relationship for the current density components $\mathbf{j}^{(\mu)}(\mathbf{r})$ and $\mathbf{j}^{(\nu)}(\mathbf{r})$ [14],

$$\langle [\mathbf{j}^{(\mu)} \times \mathbf{j}^{(\nu)}]_z \rangle = [\langle \mathbf{j}^{(\mu)} \rangle \times \langle \mathbf{j}^{(\nu)} \rangle]_z,$$

which leads to a relationship between the φ_i values and the effective conductivity σ_{eff}:

$$\sum_{i=1}^{N} \sigma_i^2 \varphi_i = \sigma_{\text{eff}}^2. \quad (73)$$

Upon excluding φ_1 with the aid of Eq. (71), we obtain from Eq. (73) the following relationship for a three-component medium:

$$1 - (1 - h_2^2)\varphi_2 - (1 - h_3^2)\varphi_3 = f^2. \quad (74)$$

In the two-dimensional case, Eq. (74) allows us to restrict the consideration to a single component φ_i (e.g., φ_2).

Upon determining the field components $\mathbf{E}^{(x)}$ and $\mathbf{E}^{(y)}$ from Eqs. (1), (2), (14), and (15), calculating the integrals entering into the expressions for φ_2 and φ_3, and taking into account relationships (41), we obtain for φ_i the following expressions analogous to Eq. (55) (with J_2 and J_3 replaced by I_2 and I_3):

$$\begin{aligned} I_2 &= \sum_{n=0}^{\infty} \xi_n \bar{\xi}_n = \sum_{n=0}^{\infty} (-1)^n \xi_n^2, \\ I_3 &= \sum_{n=0}^{\infty} \eta_n \bar{\eta}_n = \sum_{n=0}^{\infty} (-1)^n \eta_n^2. \end{aligned} \quad (75)$$

Note that substituting the expressions for φ_2 and φ_3 into Eq. (74) leads to the relationship

$$I_2 p_2 \delta_2 + I_3 p_3 \delta_3 = \xi_0 p_2 \delta_2 + \eta_0 p_3 \delta_3, \quad (76)$$

which can be used to check for the correctness of calculations upon solving Eqs. (27) and (28) by numerical methods.

Now we will demonstrate that the values of I_2 and I_3 from Eq. (75) satisfy relationship (76). Upon multiplying Eq. (27) by $\bar{\xi}_n$ and the first equation in (39) by ξ_n, adding these equations, and summing over all n, we obtain

$$I_2 = \xi_0 + \frac{1}{2}\sum_{nm} [(-1)^m - (-1)^n] P_{nm} \xi_n \eta_m. \quad (77)$$

By the same token,

$$I_3 = \eta_0 - \frac{1}{2}\sum_{nm} [(-1)^m - (-1)^n] Q_{mn} \xi_n \eta_m. \quad (78)$$

Substituting Eqs. (77) and (78) into (76), the latter transforms into an identity by virtue of Eq. (30).

7. THERMO EMF

According to [15], the effective thermo emf coefficient α_{eff} in the case of a weak thermoelectric coupling ($T\sigma\alpha^2/\kappa \ll 1$, where κ is the thermal conductivity) is determined by the formula

$$\alpha_{\text{eff}} = \frac{\langle \alpha \mathbf{j} \mathbf{G} \rangle}{\langle \mathbf{j} \rangle \cdot \langle \mathbf{G} \rangle}, \quad (79)$$

which is valid for an arbitrary isotropic medium. Here, $\mathbf{j} = \sigma(\mathbf{r})\mathbf{E}$ is the electric current density, $\mathbf{G} = -\nabla T$ is the temperature field "strength," and T is the absolute temperature. For an N-component medium, Eq. (79) yields

$$\alpha_{\text{eff}} = \sum_{i=1}^{N} \alpha_i \Psi_i, \quad (80)$$

where

$$\Psi_i = \frac{\langle \mathbf{j} \cdot \mathbf{G} \rangle^{(i)}}{\langle \mathbf{j} \rangle \cdot \langle \mathbf{G} \rangle} = \frac{\sigma_i}{\sigma_e} \frac{\langle \mathbf{E} \cdot \mathbf{G} \rangle^{(i)}}{\langle \mathbf{E} \rangle \cdot \langle \mathbf{G} \rangle} \quad (81)$$

and $\langle \ldots \rangle^{(i)}$ denotes the same integral as in Eqs. (51) and (70).

The $\mathbf{j}$ and $\mathbf{G}$ values satisfy the identity $\langle \mathbf{j} \cdot \mathbf{G} \rangle = \langle \mathbf{j} \rangle \cdot \langle \mathbf{G} \rangle$, which leads to the following "summing rule" for Ψ_i functions in the N-component medium:

$$\sum_{i=1}^{N} \Psi_i = 1. \quad (82)$$

The heat flux density $\mathbf{q} = \kappa(\mathbf{r})\mathbf{G}$ (obeying the equation $\operatorname{div} \mathbf{q} = 0$) and the field strength $\mathbf{E}$ satisfy the identity $\langle \mathbf{q} \cdot \mathbf{E} \rangle = \langle \mathbf{q} \rangle \cdot \langle \mathbf{E} \rangle$, where $\langle \mathbf{q} \rangle = \kappa_{\text{eff}} \langle \mathbf{G} \rangle$ and κ_{eff} is the effective thermal conductivity. This relationship leads to the following sequence of equations:

$$\begin{aligned} \kappa_{\text{eff}} \langle \mathbf{E} \rangle \cdot \langle \mathbf{G} \rangle &= \langle \mathbf{q} \cdot \mathbf{E} \rangle \\ = \sum_{i=1}^{N} \kappa_i \langle \mathbf{E} \cdot \mathbf{G} \rangle^i &= \sum_{i=1}^{N} \frac{\kappa_i}{\sigma_i} \langle \mathbf{j} \cdot \mathbf{G} \rangle^{(i)}, \end{aligned}$$

which is equivalent to an expression

$$\sum_{i=1}^{N} \frac{\kappa_i}{\sigma_i} \Psi_i = \frac{\kappa_{\text{eff}}}{\sigma_{\text{eff}}} \quad (83)$$

relating the Ψ_i values to the effective electric and thermal conductivities of the medium.

Excluding Ψ_1 with the aid of Eq. (82), we obtain for a three-component system from (80)

$$\alpha_{\text{eff}} = \alpha_1 + (\alpha_2 - \alpha_1)\Psi_2 + (\alpha_3 - \alpha_1)\Psi_3. \quad (84)$$

The same procedure with Eq. (83) yields

$$\left(\frac{\kappa_1}{\sigma_1} - \frac{\kappa_2}{\sigma_2}\right)\Psi_2 + \left(\frac{\kappa_1}{\sigma_1} - \frac{\kappa_3}{\sigma_3}\right)\Psi_3 = \frac{\kappa_1}{\sigma_1} - \frac{\kappa_{\text{eff}}}{\sigma_{\text{eff}}}, \quad (85)$$

which allows us to exclude Ψ_3 or Ψ_2 from Eq. (84). Note that, for a two-component medium ($\Psi_3 = 0$), Eqs. (84) and (85) yield an expression for α_{eff} coinciding with that obtained in [10].

In the absence of thermoelectric effects, the problems of describing the thermal and electric conductivities transform one into another upon substituting $\kappa \rightleftharpoons \sigma$. Therefore, the results obtained in Sections 2–4 can be applied to the problem of thermal conductivity by substituting $\sigma_i \longrightarrow \kappa_i$ and $\sigma_{\text{eff}} \longrightarrow \kappa_{\text{eff}}$ (the quantities corresponding to this case will be denoted by double upper bars).

By calculating the bilinear characteristics $\langle \mathbf{E} \cdot \mathbf{G} \rangle^{(i)}$ ($i = 2, 3$) using the potentials (1) and (2) and the corresponding "temperature potentials," we obtain

$$\Psi_i = \frac{\sigma_i}{\sigma_{\text{eff}}} \frac{4}{(1 + h_{\sigma i})(1 + h_{\kappa i})} \frac{p_i \bar{\bar{J}}_i}{\Delta_\sigma \Delta_\kappa}, \quad i = 2, 3, \quad (86)$$

where

$$\bar{\bar{J}}_2 = \sum_{n=0}^{\infty} \xi_n \bar{\bar{\xi}}_n, \quad \bar{\bar{J}}_3 = \sum_{n=0}^{\infty} \eta_n \bar{\bar{\eta}}_n, \quad (87)$$

$h_{\sigma i} = \sigma_i/\sigma_1$, $h_{\kappa i} = \kappa_i/\kappa_1$, and Δ_σ are the same values as in Eq. (56), and Δ_κ is obtained from Δ_σ upon substituting $\sigma_i \longrightarrow \kappa_i$. For a system obeying the Wiedemann–Franz law

$$\frac{\kappa_1}{\sigma_1} = \frac{\kappa_2}{\sigma_2} = \frac{\kappa_3}{\sigma_3},$$

we obtain

$$h_{\kappa i} = h_{\sigma i}, \quad \Delta_\kappa = \Delta_\sigma, \quad \bar{\bar{J}}_i = J_i.$$

In this case, Eqs. (86) and (87) yield

$$\Psi_i = (\sigma_i/\sigma_{\text{eff}})\psi_i,$$

where ψ_i is determined from (54)–(56).

Substituting Ψ_2 and Ψ_3 from Eq. (86), as well as $\sigma_{\text{eff}} = \sigma_1 f$ with f from Eq. (46) and an analogous expression for κ_{eff}, into Eq. (85) leads to the relationship

$$\begin{aligned}(\delta_{\kappa 2} - \delta_{\sigma 2}) p_2 \bar{\bar{J}}_2 + (\delta_{\kappa 3} - \delta_{\sigma 3}) p_3 \bar{\bar{J}}_3 \\ = p_2(\bar{\bar{\xi}}_0 \delta_{\kappa 2} - \xi_0 \delta_{\sigma 2}) + p_3(\bar{\bar{\eta}}_0 \delta_{\kappa 3} - \eta_0 \delta_{\sigma 3}),\end{aligned} \quad (88)$$

where

$$\delta_{\sigma i} = \frac{1 - h_{\sigma i}}{1 + h_{\sigma i}}, \quad \delta_{\kappa i} = \frac{1 - h_{\kappa i}}{1 + h_{\kappa i}}.$$

Upon subtracting from Eq. (27) multiplied by $\bar{\bar{\xi}}_0 \delta_{\kappa 2}$ the corresponding equation for $\bar{\bar{\xi}}_n$ multiplied by $\xi_n \delta_{\sigma 2}$ and summing the difference over all n, we obtain

$$\begin{aligned}(\delta_{\kappa 2} - \delta_{\sigma 2})\bar{\bar{J}}_2 = -\sum_{nm} (P_{nm} \delta_{\kappa 2} \bar{\bar{\xi}}_n \eta_m \\ - \bar{\bar{P}}_{nm} \delta_{\sigma 2} \xi_n \bar{\bar{\eta}}_m) + \bar{\bar{\xi}}_0 \delta_{\kappa 2} - \xi_0 \delta_{\sigma 2}.\end{aligned}$$

By the same token, we obtain

$$\begin{aligned}(\delta_{\kappa 3} - \delta_{\sigma 3})\bar{\bar{J}}_3 = \sum_{nm} (\bar{\bar{Q}}_{mn} \delta_{\sigma 3} \bar{\bar{\xi}}_n \eta_m \\ - Q_{mn} \delta_{\kappa 3} \xi_n \bar{\bar{\eta}}_m) + \bar{\bar{\eta}}_0 \delta_{\kappa 3} - \eta_0 \delta_{\sigma 3}.\end{aligned}$$

Substituting expressions for $\bar{\bar{J}}_2$ and $\bar{\bar{J}}_3$ into Eq. (88) transforms this relationship into an identity. Thus, relationship (85) is satisfied by a solution to the conductivity problem determined in this study. In analyzing the system of Eqs. (27) and (28) by numerical methods, relationship (85) can be used for checking the correctness of the results of calculations.

8. LIMITING CASES

In concluding, we will consider four limiting cases in which some local characteristics possess particular values or satisfy additional relationships.

1. In the limit $\rho \longrightarrow 0$, the model becomes two-component, whereby $P_{nm} = Q_{nm} = N_{nm} = 0$. As a result, $\eta_n = \delta_{n0}$ and $D_{2n} = 0$, so that Eqs. (1), (3), (27), (46) and some others transform into the corresponding formulas derived in [8]. A difference in the values of M_{nm} is related to different definitions of the ξ_n coefficients.

2. For $\delta_3 \longrightarrow 0$, the model also becomes two-component: inclusions of the second type are formally retained ($\rho \neq 0$), but their properties are the same as those of the first component ($\sigma_3 = \sigma_1$). As a result, $P_{nm} = N_{nm} = 0$, Eq. (27) converts into the corresponding expression derived in [8], and Eq. (28) yields

$$\eta_n = \delta_{n0} - \sum_{m=0}^{\infty} Q_{nm} \xi_m.$$

Substituting this expression into (32) (for $h_3 = 1$) and taking into account the explicit form of Q_{nm} from Eq. (29) and definition (25), we obtain

$$C_{2n+1} = \beta \delta_{n0} - \sum_{m=0}^{\infty} B_{2m} \frac{(2n + 2m)!}{(2n + 1)!} d_{n+m+1}. \quad (89)$$

According to Eq. (26), $D_{2n} = 0$ for $\delta_3 = 0$. As can be readily checked, the expansion of the complex potential (3) (for $D_{2n} = 0$) in the vicinity of the point $z = z_0$ with allowance for Eq. (9) coincides with Eq. (2) upon

substituting C_{2n+1} values from Eq. (89). The effective conductivity according to Eqs. (45) and (46) coincides (as in the limiting case 1) with the corresponding formula derived in [8] for $\alpha = 1/\xi_0$.

3. In a more complicated case of $\rho \longrightarrow R$ and $\sigma_3 \longrightarrow \sigma_2$, the tree-component system under consideration transforms into a two-component system as well, differing from the system studied in [8] only by the period ($2\tilde{a} = \sqrt{2}a$ instead of $2a$, which corresponds to the lattice depicted in Fig. 1 with the coordinate axes rotated by 45°). We will demonstrate that, with allowance for this rotation, the solution obtained in our study coincides (in the limit as $\rho \longrightarrow R$ and $\delta_3 \longrightarrow \delta_2$) with that derived in [8].

For $\rho = R$ and $\delta_3 = \delta_2$, we have $N_{nm} = M_{nm}$ and $Q_{nm} = P_{nm}$. In this case, upon subtracting (28) from (27), we obtain a homogeneous system of equations with respect to the "variable" $\xi_n - \eta_n$. Possessing a nonzero determinant, this system has only a zero solution that implies $\eta_n = \xi_n$ (and $D_{2n} = B_{2n}$), where ξ_n satisfies the equation

$$\xi_n + \sum_{m=0}^{\infty}(M_{nm} + P_{nm})\xi_m = \delta_{n0}. \tag{90}$$

Let the field $\langle \mathbf{E} \rangle$ be directed along the quadrant diagonal. The corresponding complex potential $\tilde{\Phi}_1(z)$ is described by a superposition of Eqs. (3) and (13) for $\bar{\beta} = \beta$:

$$\tilde{\Phi}_1(z) = \frac{1}{\sqrt{2}}[\Phi_1(z) - \bar{\Phi}_1(z)] = \beta z e^{i\pi/4} + \sum_{n=0}^{\infty}\frac{1 - i(-1)^n}{\sqrt{2}}B_{2n}\{\zeta^{(2n)}(z) + \zeta^{(2n)}(z - z_0)\}. \tag{91}$$

This equation is derived taking into account the relationships $D_{2n} = B_{2n}$ and $\bar{B}_{2n} = (-1)^n B_{2n}$ [see Eq. (42)]; the constants φ_0 and $\bar{\varphi}_0$ (insignificant for subsequent consideration) are omitted. Using the summation rule and homogeneity relationship for the Weierstrass zeta function [19, 20], as well as the obvious equalities

$$\wp(z_0) = 0, \quad \wp'(z_0) = 0, \quad \zeta(iz) = -i\zeta(z),$$
$$\wp(iz) = -\wp(z), \quad \wp'(iz) = i\wp'(z)$$

(valid in the lemniscate case under consideration), one can readily check that

$$\zeta(z) + \zeta(z - z_0) = -\zeta(z_0) + e^{i\pi/4}\tilde{\zeta}(\tilde{z}), \quad \tilde{z} = ze^{i\pi/4}. \tag{92}$$

Here, $\zeta(z)$ and $\zeta(z - z_0)$ are the Weierstrass zeta functions with halfperiods a and ia, while $\tilde{\zeta}(\tilde{z})$ is the function with the halfperiods $\tilde{a} = a/\sqrt{2}$ and $i\tilde{a} = ia/\sqrt{2}$. Substituting variable $\tilde{z}$ into Eq. (91) and taking into account Eq. (92), we obtain

$$\tilde{\Phi}_1(\tilde{z}) = \beta\tilde{z} + e^{i\pi/4}\sum_{n=0}^{\infty} i^n \frac{1 - i(-1)^n}{\sqrt{2}}B_{2n}\tilde{\zeta}^{(2n)}(\tilde{z}) \tag{93}$$

(an insignificant constant is omitted).

Note that an expansion of the left and right parts of Eq. (92) in powers of z and $\tilde{z}$ using Eqs. (6) and (17) leads to a relationship

$$c_{2n} + d_{2n} = (-1)^n \tilde{c}_{2n}, \tag{94}$$

where $\tilde{c}_{2n}$ differs from c_{2n} by the substitution $a \longrightarrow \tilde{a} = a/\sqrt{2}$. The validity of Eq. (94) for $n = 1, 2, 3, \ldots$ can be directly verified using the expressions for c_{2n} and d_{2n} from Eqs. (7) and (11). Taking into account relationship (94), Eq. (90) can be presented in the following form:

$$\xi_n + \sum_{m=0}^{\infty}(-1)^{(n+m+1)/2}\tilde{M}_{nm}\xi_m = \delta_{n0}, \tag{95}$$

where the matrix elements $\tilde{M}_{nm}$ differs from the M_{nm} given by Eq. (29) only by substituting $c_{n+m+1} \longrightarrow \tilde{c}_{n+m+1}$ (i.e., $\tilde{M}_{nm}$ refers to the square lattice with the halfperiod $\tilde{a} = a/\sqrt{2}$).

As can readily be checked, by introducing the variables $\tilde{\xi}_n$ defined by the relationships $\xi_{4k} = \tilde{\xi}_{4k}$, $\xi_{4k+1} = -\tilde{\xi}_{4k+1}$, $\xi_{4k+2} = -\tilde{\xi}_{4k+2}$, $\xi_{4k+3} = \tilde{\xi}_{4k+3}$ ($k = 0, 1, 2, \ldots$), the set (95) can be converted into the following form:

$$\tilde{\xi}_n + \sum_{m=0}^{\infty}\tilde{M}_{nm}\tilde{\xi}_m = \delta_{n0}. \tag{96}$$

Using this relationship, the final expression for potential (93) can be rewritten as

$$\tilde{\Phi}_1(\tilde{z}) = \beta\tilde{z} + \sum_{n=0}^{\infty}\tilde{B}_{2n}\tilde{\zeta}^{(2n)}(\tilde{z}). \tag{97}$$

By the same token, the potential of an inclusion with $\rho = R$ and $h_3 = h_2$ centered at $z = 0$ converts into

$$\tilde{\Phi}_2(\tilde{z}) = \frac{1}{\sqrt{2}}[\Phi_2(z) - \bar{\Phi}_2(z)] = \sum_{n=0}^{\infty}\tilde{A}_{2n+1}\tilde{z}^{2n+1}. \tag{98}$$

The same form (with the corresponding shift in $\tilde{z}$) is obtained with $\rho = R$ and $h_3 = h_2$ for the potential $\tilde{\Phi}_3(\tilde{z})$. The coefficients $\tilde{B}_{2n}$ and $\tilde{A}_{2n+1}$ entering into Eqs. (97) and (98) are related to $\tilde{\xi}_n$ from Eq. (96) by relationships identical in form to the expressions relating B_{2n} and A_{2n+1} to ξ_n [see Eqs. (25 and 31)].

Thus, the complex potential in the coordinate system $(\tilde{x}, \tilde{y})$ rotated by 45° relative to the coordinate system (x, y) acquires (for $\rho = R$ and $\sigma_3 = \sigma_2$) the same form as that for the two-component model studied in [8], but the lattice halfperiod changes to $\tilde{a} = a/\sqrt{2}$.

Note that relationship (94) allows coefficient d_{2n} to be expressed through c_{2n}. Equations (5)–(8) indicate that $c_{2n} \propto (c_2)^n \propto 1/a^{4n}$. Therefore, $\tilde{c}_{2n}/c_{2n} = (a/\tilde{a})^{4n} = 4^n$ and, hence, Eq. (94) yields as a relationship

$$d_{2n} = [(-4)^n - 1]c_{2n}, \tag{99}$$

which allows us to calculate d_{2n} using the known values of c_{2n}.

4. In the limiting case with $\rho = R$ and $\delta_3 = -\delta_2$ (i.e., $h_3 = 1/h_2$ or $\sigma_2\sigma_3 = \sigma_1^2$), the system under consideration is characterized by $N_{nm} = -M_{nm}$ and $P_{nm} = -Q_{nm}$. Therefore,

$$\hat{S}_{nm} = \begin{pmatrix} M_{nm} & -Q_{nm} \\ Q_{nm} & -M_{nm} \end{pmatrix}.$$

The same structure is inherent in all the odd powers of $\hat{S}$. Therefore, Eq. (36) shows that $\eta_{2k+1} = -\xi_{2k+1}$. All the even powers of $\hat{S}$ also possess identical structures (different from that for the odd powers) such that $\eta_{2k} = \xi_{2k}$. Therefore, the case of $\rho = R$ and $\delta_3 = -\delta_2$ implies that $\eta_n = (-1)^n\xi_n$. In particular, $\eta_0 = \xi_0$ and Eq. (46) yields $f = 1$ and $\sigma_{\text{eff}} = \sigma_1$. The same conclusion was derived in [17] based on an approximate expression for the effective conductivity (dielectric permittivity).

Finally, it should be noted that, in the general case, the equality $f = 1$ takes place provided that the relationship

$$\xi_0 R^2\delta_2 + \eta_0\rho^2\delta_3 = 0$$

is valid. This relationship indicates that it is possible to vary some effective characteristics (Hall coefficient, thermo emf, etc.) of this system without affecting the electric conductivity.

REFERENCES

1. A. L. Efros and B. I. Shklovskii, Phys. Status Solidi B **76**, 475 (1976).
2. J. P. Straley, J. Phys. C **9**, 783 (1976).
3. J. B. Keller, J. Math. Phys. **5**, 548 (1964).
4. A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 110 (1970) [Sov. Phys. JETP **32**, 63 (1971)].
5. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **81**, 665 (1981) [Sov. Phys. JETP **54**, 355 (1981)].
6. Yu. P. Emets, *Electrical Properties of Composites with Regular Structure* (Naukova Dumka, Kiev, 1986).
7. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **79**, 1561 (1980) [Sov. Phys. JETP **52**, 787 (1980)].
8. B. Ya. Balagurov and V. A. Kashin, Zh. Éksp. Teor. Fiz. **117**, 978 (2000) [JETP **90**, 850 (2000)].
9. A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 641 (1970) [Sov. Phys. JETP **32**, 348 (1971)].
10. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **85**, 568 (1983) [Sov. Phys. JETP **58**, 331 (1983)].
11. B. Ya. Balagurov, Fiz. Tverd. Tela (Leningrad) **28**, 2068 (1986) [Sov. Phys. Solid State **28**, 1156 (1986)].
12. B. I. Shklovskiĭ, Zh. Éksp. Teor. Fiz. **72**, 288 (1977) [Sov. Phys. JETP **45**, 152 (1977)].
13. D. J. Bergman and D. Stround, Phys. Rev. B **32**, 6097 (1985).
14. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **93**, 1888 (1987) [Sov. Phys. JETP **66**, 1079 (1987)].
15. B. Ya. Balagurov, Fiz. Tekh. Poluprovodn. (Leningrad) **21**, 1978 (1987) [Sov. Phys. Semicond. **21**, 1198 (1987)].
16. B. Ya. Balagurov, Fiz. Tverd. Tela (Leningrad) **30**, 3501 (1988) [Sov. Phys. Solid State **30**, 2013 (1988)].
17. Yu. P. Emets, Zh. Éksp. Teor. Fiz. **114**, 1121 (1998) [JETP **87**, 612 (1998)].
18. Lord Rayleigh, Philos. Mag. **34**, 481 (1892).
19. *Heigher Transcendental Functions (Bateman Manuscript Project)*, Ed. by A. Erdelyi (McGraw-Hill, New York, 1955; Nauka, Moscow, 1967), Vol. 3.
20. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun (Dover, New York, 1971; Nauka, Moscow, 1979).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 135–145.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 154–165.

SOLIDS
Electronic Properties

Two-Dimensional Electron Gas in Semimagnetic Semiconductor HgMnTe with Inverted Bands¶

V. B. Bogevolnov[a], I. M. Ivankiv[a], A. M. Yafyasov[a], and V. F. Radantsev[b, *]

[a] *Institute of Physics, St. Petersburg State University, St. Petersburg, 198904 Russia*
[b] *Institute of Physics and Applied Mathematics, Ural State University, Yekaterinburg, 620083 Russia*
**e-mail: victor.radantsev@usu.ru*

Received June 7, 2000

Abstract—The two-dimensional electron gas in the surface layers of HgMnTe with inverted bands is studied for the first time experimentally and theoretically. It is shown that the structure of the investigated capacitance magnetooscillations in HgMnTe MOS structures is entirely similar to that observed in the non-magnetic gapless semiconductor HgCdTe and the sole effect of the exchange interaction is the temperature shift of beat nodes. The information about the exchange parameters is obtained only from modeling the oscillations, because no pronounced changes in the position of oscillations are observed and the separate spin components are not resolved. For the description of the spectrum in the magnetic field, we propose a theory that takes the exchange and spin–orbit interactions into account for materials with direct and inverted bands. A comparison between experiment and theory for different temperatures and exchange interaction parameters is reported. The modeling shows that the spin–orbit splitting by far exceeds the contribution of the exchange interaction. The calculated amplitudes of "partial" oscillations for different spin branches of the spectrum are essentially different in accordance with the difference in the intensities of the corresponding lines in the Fourier spectra of the experimental oscillations. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The peculiar features of, and the interest in, the two-dimensional (2D) electron gas in narrow-gap diluted magnetic (semimagnetic) semiconductors (DMSs) are due to two factors. One stems from the *s, p–d* exchange interaction between band electrons and localized magnetic moments [1]. This interaction changes spin splitting of the band states, which can be varied by external factors, e.g., the magnetic field and temperature. The other factor is due to the peculiarities inherent to small-gap Kane semiconductors, leading to the relativistic-type effects of non-parabolicity, kinetic confinement (motional binding [2]), spin–orbit (SO), splitting, and resonant interband mixing by the surface electric field [3–5]. An important property of 2D electronic systems involving DMS is that both the exchange and SO interaction lead to a rearrangement of the spin structure of Landau levels (LLs).

Although historically the first studies of the 2D electron gas in DMSs were performed for metal–insulator–semiconductor (MIS) structures based on HgMnTe [6], the experimental results are mostly available for the grain boundaries in HgMnTe and HgCdMnTe with a positive Kane gap $E_g > 100$ meV at the typical Mn content $x = 0.02$ (for higher x, the exchange interaction exhibits itself poorly, which was attributed to the antiferromagnetic interaction between Mn^{2+} ions) [1, 7–9]. This is due to low electron mobility in the previously investigated MIS structures. At the same time, the inversion layers in MIS structures are of particular experimental interest because of the possibility of controlling the depth of the surface quantum well by gate voltage and because of a relative ease and accuracy of the surface potential description (for bicrystals, additional poorly verified assumptions have to be used to describe the self-consistent potential near the grain boundaries [10]). An important point is that these results can be compared with the data for MIS structures based on narrow-gap HgCdTe [4, 11], which is a non-magnetic analogue of narrow-gap DMSs.

As to the theoretical description, the subband calculations were carried out only for DMSs with direct but not inverted bands and without taking the spinorlike effects into account [1, 6]. However, the SO splitting in asymmetric quantum wells at a zero magnetic field (which itself is currently of great interest [4, 5, 12, 14–19]) leads to a rearrangement of the subband magnetic levels. In narrow-gap semiconductors, the magnetic spectrum perturbation is so drastic that the SO interaction cannot be neglected in the theoretical treatment. It must be stressed that as we see in what follows, the SO splitting by far exceeds the exchange interaction contribution, and therefore, it cannot be considered as a correction to the exchange interaction. It is also clear that a treatment based on the semiclassical quantization in the magnetic field of the subband spectrum (calculated in a zero magnetic field) is unsuitable for the description of exchange interaction effects. A more rigorous

¶This article was submitted by the authors in English.

theoretical consideration of the LL structure is required.

In this paper, the peculiarities of 2D electron gas due to the exchange and SO interaction are studied in inversion layers on $Hg_{1-x}Mn_xTe$ with a small Mn content. In Section 2, we describe the parameters of the samples and the experimental methods used. The experimental data related to the capacitance oscillations in perpendicular magnetic fields versus gate voltage and magnetic field and their temperature evolution are presented in Section 3. The experiments in tilted magnetic fields are discussed in Section 4. In Section 5, we present the theoretical model based on developing the concept that we proposed previously for non-magnetic Kane semiconductors. In Section 6, the results of the computer modeling of capacitance oscillations are presented. The results of comparing the experimental data and theoretical calculations for different temperatures and exchange interaction parameters are discussed.

2. SAMPLES AND EXPERIMENTAL METHODS

MIS structures were fabricated from *p*-$Hg_{1-x}Mn_xTe$ single crystals. We investigated samples with different values of Mn content (x = 0.024, 0.040, 0.060, and 0.1). The Kane gap E_g and the Kane effective mass m_b (and therefore, x) were determined independently by the tunnel spectroscopy method for a comparison of band parameters in the bulk with those in the vicinity of the surface. The discrepancy is within the accuracy of the analysis ($\Delta x \sim 0.002$–0.003). Because the tunnel contacts and the studied MIS capacitors were produced using the same technology (see below) and differ only by the insulator thickness, this agreement testifies that the surface layers are chemically close to the bulk. The similarity of the results for the structures with different insulators (fabricated using different methods) supports this conclusion. The fact that the cyclotron masses measured in 2D subbands for small surface concentrations extrapolate to the bulk value m_b is direct evidence of the absence of a noticeable decomposition in the 2D layer during the structure fabrication process.

At $x < 0.08$, HgMnTe has inverted bands (i.e., becomes semimetal) and traditional galvanometric methods cannot be used because of the shunting of surface conductance by the bulk. We employed the magnetocapacitance spectroscopy method, which is applicable to semiconductors with any sign of the Kane gap. The differential capacitance C of the capacitors and its derivative dC/dV_g with respect to the gate voltage V_g were measured in the dark, typically at 1 MHz and with a test signal amplitude of 5 mV.

Several methods have been used for forming an insulating film in MIS structures: the anodic oxide formation, the silicon oxide and Al_2O_3 deposition, and the Langmuir–Blodgett film technique. The capacitance magnetooscillations due to the magnetic quantization of the 2D electron gas were observed in all the above structures. The general shapes of the oscillations at the same carrier surface density and the same Mn content are similar. In what follows, we present the results for the structures with an ~80 nm thick anodic oxide film. There are several reasons for this choice: (i) the amplitudes of oscillations in these structures are the highest owing to a large insulator capacitance value (this is caused by a large dielectric constant value of the anodic oxide), (ii) the highest surface carrier densities are achieved at low gate voltages V_g = 10–15 V, and (iii) the dielectric constant of the oxide is close to that of a semiconductor, and the contribution of image forces to the surface potential can therefore be neglected in the calculations.

We restrict our consideration to the results for HgMnTe with $x \approx 0.04$ ($E_g = (-100 \pm 5)$ meV). The amplitudes of the oscillations for other samples are much less even at 4.2 K and rapidly decrease with increasing temperature. (For $x \approx 0.024$, this is caused by a small cyclotron energy due to a large Kane gap; for $x \approx 0.06$ and $x \approx 0.1$, this is the result of a large doping level of available materials.) As a result, we could not obtain reasonably accurate information about the oscillation temperature evolution where the specificity of the DMS is manifested. As to the measurements at T = 4.2 K, the subband parameters extracted from oscillations for these samples are similar to those for HgCdTe with the same band parameters and agree well with the theory.

On the other hand, the samples with x = 0.04 are best suited to the purpose of this first study aimed at investigating the peculiarities of the 2D electron gas in DMSs with inverted bands, where (i) the SO and exchange interaction effects are expected to be more clearly pronounced and (ii) the results can be compared with those for well studied surface layers on gapless HgCdTe with $E_g \sim -(50$–$100)$ meV [4, 20]. For a small gap ($|E_g| <$ 100 meV), the parameters of 2D subbands depend only weakly on E_g (except the case of small subband occupancies) [4, 20]. By contrast, the subband parameters are more sensitive to the doping level. For this reason, we present the results for two samples with $N_A - N_D = 1.2 \times 10^{16}$ cm^{-3} (sample S1) and $N_A - N_D = 1.5 \times 10^{17}$ cm^{-3} (sample S2 with the gate area $S = 7.7 \times 10^{-4}$ cm^2 and the insulator capacitance C_{ox} = 155.1 pF).

3. MAGNETOCAPACITANCE IN PERPENDICULAR MAGNETIC FIELDS

Figure 1 shows the capacitance–voltage characteristics at T = 4.2 K in the magnetic field B = 4.5 T perpendicular to the 2D layer for sample S2. The $C(V_g)$ characteristics are typical for the low-frequency behavior. This means that 2D electrons in the inversion layers contribute predominantly to the measured capacitance under the inversion band bending. The low-frequency conditions with respect to the minority carriers are satisfied in the entire investigated frequency range 30 kHz–5 MHz.

The wide hysteresis loop and the dependence of the $C(V_g)$ characteristics on the voltage sweep rate are observed. At a fixed V_g, the capacitance changes in time because of a flat-band voltage shift ΔV_{fb}. The time constant is of the order of several minutes and is almost independent of the temperature. This behavior points to the charge tunnel exchange between the semiconductor and the slow traps in the insulator. The sample prehistory effects and instability are manifested in all the investigated HgMnTe-based MIS structures. This is in contrast with the HgCdTe- and HgTe-based structures with the same insulators that we investigated simultaneously.

The voltage dependence of the charge density $eN_s(V_g)$ induced in the inversion layer is sublinear. This is demonstrated by the non-equidistant of quantum oscillations of the capacitance $C(V_g)$ (see Fig. 1). The tunneling of electrons from the 2D layer into the oxide causes a saturation of the $N_s(V_g)$ dependence at $V_g - V_{fb} \approx$ (10–15) V. As a result, the N_s range accessible for investigations is limited by the value (3–4) $\times 10^{12}$ cm^{-2} (in HgCdTe, the N_s values up to 10^{13} cm^{-2} can be obtained). Although the hysteresis effects hamper the measurements, the discussed physical results are not affected by the band bending instability. This instability is caused by the transient processes but not by degradation. Although the $C(V_g)$ characteristics are history dependent, they are completely repeatable if the voltage range, rate, and direction of the sweep are the same. To assure the stability of the band bending during the measurement of $C(B)$ oscillations, the sample was held at a given voltage for 5–15 min. The identity of $C(B)$ plots registered at increasing and decreasing the magnetic field (i.e., at different times) was examined for each $C(B)$ curve. When the temperature (or angle) dependences of the $C(B)$ oscillations were measured, the long term stability was checked by the repetitive measurement of the initial $C(B)$ plot (for a given measurement cycle).

The $C(B)$ oscillations (and consequently, the subband occupancy and the surface potential) measured at the same capacitance magnitude in zero magnetic field $C(0)$ are identical, irrespective of the voltage (the value of the latter for any given $C(0)$ is determined by the flat-band voltage, which is history- and time-dependent). When the dc gate voltage (or the flat-band voltage at the same V_g) is changed, the filling of interface states is also changed but does not respond to the ac ripple, i.e., the interface states do not contribute to the capacitance. This occurs for all frequencies and temperatures and testifies that the high-frequency conditions with respect to interface states are satisfied. Thus, there is a "one-to-one correspondence" between $C(0)$, the band bending, and the surface density of 2D electrons $N_s = \sum N_i$ (where i is the 2D subband number).

The subband parameters are presented below as functions of N_s. Contrary to the dependence on V_g, their dependence on N_s is not affected by the hysteresis effects or any specific parameters of MOS capacitors; the same N_s dependence is common to a given HgMnTe sample. It may be noted that there are some positive aspects of hysteresis. It is possible to investigate the 2D electron gas in the same surface quantum well on the same sample but with a different interface charge. This is important, in particular, in the investigation of scattering mechanisms.

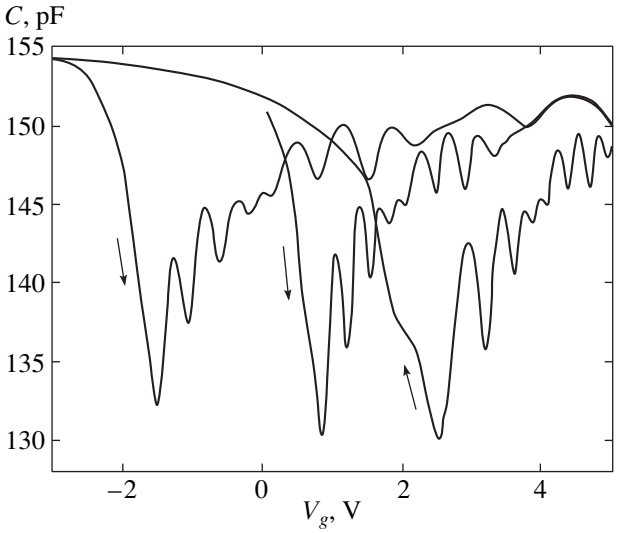

Fig. 1. Capacitance-voltage dependences in the magnetic field $B = 4.5$ T perpendicular to the 2D plane for sample S2 at different gate voltage sweeps. The arrows indicate the sweep direction. The sweep rate is 2 V/mn.

Typical $C(B)$ oscillations are presented in Fig. 2 together with their $1/B$ Fourier transforms. The individual spin components have not been observed in the oscillations at any N_s even for the lowest LLs. On the other hand, the oscillation beats and the Fourier spectra distinctly demonstrate the presence of two frequencies connected with the SO splitting of each 2D subband. The surface densities in the spin-split subbands N_i^+ and N_i^- determined from Fourier transforms are plotted in Fig. 3. The carrier distribution among 2D subbands is different for the two samples. The concentrations N_s corresponding to the "starts" of the excited subbands increase as the doping level increases and agree well with the theoretical calculations in which the bulk values of $N_A - N_D$ are used. This fact also testifies that the disruption of stoichiometry in surface layers that could be caused by the migration of atoms is insignificant. A discrepancy with the theory is detectable only in the relative differences of occupancies $\Delta N_i/N_i = (N_i^- - N_i^+)/(N_i^- + N_i^+)$ in the small N_s range. Similar disagreement also occurs for inversion layers on HgCdTe. Possible reasons for this behavior are discussed in [4]. The intensities of Fourier lines for the high-energy

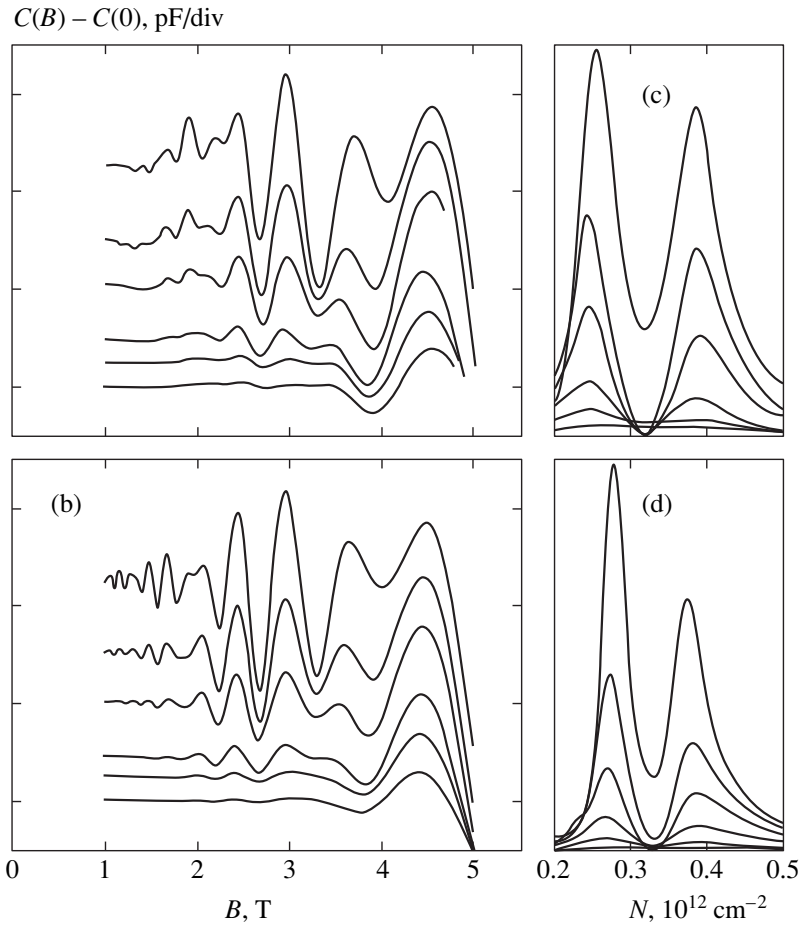

Fig. 2. (a) Experimental and (b) simulated temperature evolutions of $C(B)$ oscillations and (c, d) their Fourier spectra for sample S2. Temperature (from top to bottom): 4.2, 10, 15, 22, 29 K, and 35 K. The values $T_D = 11$ K for $i = 0$, $T_D = 9.5$ K for $i = 1$, and $T_N = 10$ K are used in the calculation.

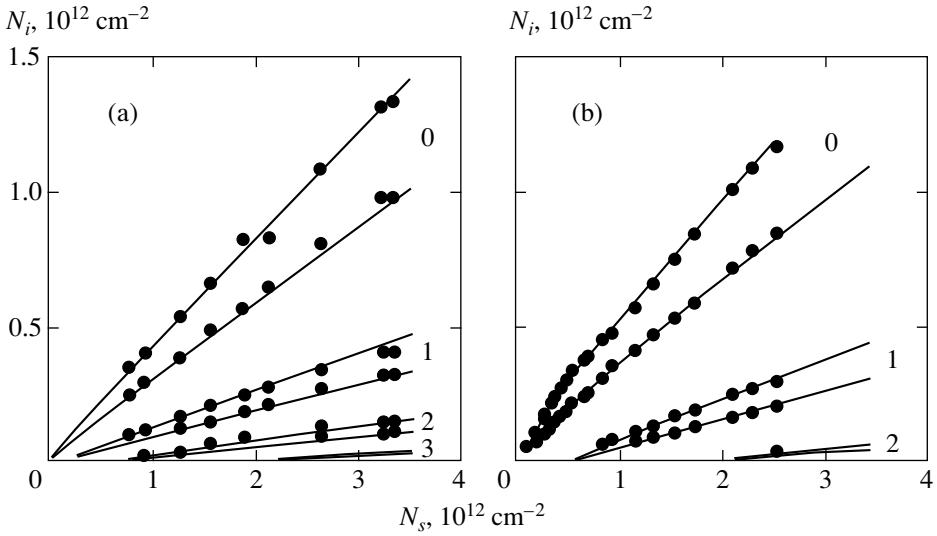

Fig. 3. The calculated (lines) and measured (points) distributions of 2D electrons among the spin-split subbands for (a) S1 and (b) S2 samples. The theoretical dependences are calculated as in [4]. The numbers at lines are the subband numbers.

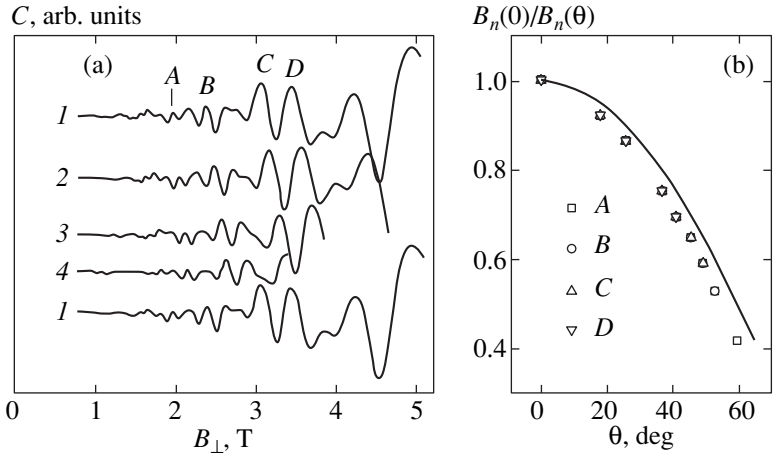

Fig. 4. (a) Capacitance oscillations plotted versus the normal component $B_\perp$ of the applied magnetic field for sample S2 at $N_s = 1.62 \times 10^{12}$ cm^{-2}. The angle θ between B and the normal to the 2D layer: (*1*) 0; (*2*) 26°; (*3*) 42°; (*4*) 50°. To demonstrate the reproducibility of the results, we plotted two $C(B_\perp)$ oscillations for θ = 0. The upper and lower plots are measured before and after the angle dependence measurements respectively. (b) The angle dependences $B_n(0)/B_n(\theta)$ for oscillation maxima marked on the upper $C(B)$ plot. The line is the cosine function.

branch I_i^+ and the low-energy branch I_i^- are different and the ratio I_i^-/I_i^+ decreases with increasing N_s.

The structure of oscillations and the subband parameters extracted from oscillations are identical to those in HgCdTe. No features due to the exchange interaction are manifested. Because the exchange effects are determined by the magnetization and can be varied by the temperature, the investigation of the temperature evolution of oscillations is of primary interest. The results for sample S2 are shown in Fig. 2. As can be seen, no pronounced changes in the position oscillations are observed. The shift of beat nodes to higher gate voltages and to lower magnetic fields (to larger LL numbers) with increasing temperature (and hence, with decreasing magnetization) is the sole temperature effect, besides the usual decrease of oscillation amplitudes. This shift must be attributed to the features inherent to semimagnetic semiconductors because neither the positions of the oscillations nor those of the beat nodes change with the temperature in HgCdTe-based structures.

4. MEASUREMENTS IN TILTED MAGNETIC FIELDS

Although there is no doubt that we are dealing with a 2D system (the existence of the magnetooscillation effect in the capacitance and the observation of magnetooscillations versus gate voltage already testify to it), experiments in a tilted magnetic field were also performed. Some results for sample S2 are presented in Fig. 4. The magnetic field positions of the oscillation extrema and the fundamental fields in the Fourier spectra (to a smaller extent) vary only roughly as the cosine of the angle θ between **B** and the normal to the 2D layer. Clearly distinguishable deviations from this behavior are observed, Namely, the experimental angle dependences are stronger.

There are several reasons for this deviation from the classical cosine dependence, because a number of physical factors are ignored in the simplified model [21]. First, in the strictest sense, this behavior, even for the parabolic dispersion, is valid only for an ideal 2D system. The condition to be satisfied for the cosine dependence is $\langle r\rangle/\langle z\rangle \gg 1$, where $\langle r\rangle$ and $\langle z\rangle$ are the respective mean sizes of the wave function in the 2D plane and in the confinement direction. For narrow-gap semiconductors, the width of the surface quantum well is relatively large and such a strong requirement may not be fulfilled (we also note that $\langle z\rangle$ is energy dependent in this case). In a strong magnetic field and at a small surface concentration, the cyclotron radius and the 2D layer width can be comparably sized (especially, for excited subbands) and the diamagnetic shift must weaken the angle dependence. This is contrary to the experimental behavior. Second, the cosine relationship is obtained for spinless particles, which is not the case in a real system. Third, the SO interaction is neglected in this simple consideration. Undoubtedly, spinlike effects can affect the spectrum in a tilted magnetic field and modify the angle dependence.

Finally, the exchange interaction can also give an additional contribution to the deviation from the simple angle dependence. This assumption has experimental support. For comparison, we investigated the HgCdTe-based samples in a tilted magnetic field. Under the same conditions, they also manifest a deviation from the cosine behavior. However, the deviation is weaker than for gapless HgMnTe and has the opposite sign. At the same time. the samples based on HgCdTe with

$E_g > 0$ show a deviation of the same sign as in HgMnTe, but smaller in magnitude. Contrary to HgCdTe samples, changes in the structure of oscillations are observed in HgMnTe inversion layers. Namely, the beat nodes in oscillations $C(B_\perp)$ (with $B_\perp = B\cos\theta$) are shifted to the lower LL numbers with the increase of θ (i.e., with the increase of the total magnetic field B), see Fig. 4. We note that the direction of the shift occurring with the increase of B (with $B_\perp$ kept fixed) is similar to the one observed with decreasing temperature.

These experimental observations testify that the behavior in tilted magnetic fields is markedly affected by both the SO interaction (which essentially depends on the E_g sign [4]) and the exchange interaction. For narrow-gap semiconductors, the theoretical analysis requires a consideration of spin from the outset. Strong SO and exchange interactions and the resonant effects lead to a serious complication of the theoretical description even for the perpendicular orientation. The calculations in tilted magnetic fields are troublesome even for the simplest parabolic Hamiltonian with a k-linear Rashba term. At present, we cannot make a reasonable theoretical analysis of the effects in tilted fields and we restrict our analysis to the case of perpendicular orientation.

As shown in Section 3, the analysis based on the Fourier transform of oscillations for different temperatures cannot yield any information about the exchange interaction. On the other hand, these data cannot be obtained from spin splitting either because, as noted above, the separate spin components are not observed in the oscillations at any temperatures. Thus, we must settle the question by the capacitance magnetooscillations modeling.

5. THEORETICAL ANALYSIS

To simulate the capacitance magnetooscillations, the density of states (DOSs) must be calculated as a function of B (or V_g). This requires knowing the subband spectrum in the magnetic field and the broadening of the Landau levels. Under the homogeneous magnetic field **B**(0, 0, B) parallel to the confinement direction (the surface potential $V = V(z)$), the motion in the 2D plane can be quantized using the mean field approximation for the exchange interaction. In the framework of a six-band Kane model, the subband LL energy $E_n(B)$ is determined by the matrix equation

$$\begin{bmatrix} -E_- + \alpha & \frac{E_B\sqrt{3(n-1)}}{2} & \frac{E_B\sqrt{n}}{2} & 0 & 0 & s_b\hbar\hat{k}_z \\ \frac{E_B\sqrt{3(n-1)}}{2} & -E_+ + 3\beta & 0 & 0 & 0 & 0 \\ \frac{E_B\sqrt{n}}{2} & 0 & -E_+ - \beta & s_b\hbar\hat{k}_z & 0 & 0 \\ 0 & 0 & s_b\hbar\hat{k}_z & -E_- - \alpha & \frac{E_B\sqrt{3(n+1)}}{2} & -\frac{E_B\sqrt{n}}{2} \\ 0 & 0 & 0 & \frac{E_B\sqrt{3(n+1)}}{2} & -E_+ - 3\beta & 0 \\ s_b\hbar\hat{k}_z & 0 & 0 & -\frac{E_B\sqrt{n}}{2} & 0 & -E_+ + \beta \end{bmatrix} \begin{pmatrix} f_1^{n-1}(z) \\ f_3^{n-2}(z) \\ f_5^{n}(z) \\ f_2^{n}(z) \\ f_4^{n-1}(z) \\ f_6^{n-1}(z) \end{pmatrix} = 0, \quad (1)$$

where $E_\pm(z) = E_n - V(z) \pm E_g/2$, $s_b = \sqrt{|E_g|/2m_b}$ is the Kane velocity, n is the LL number, and $E_B = \sqrt{2m_b s_b^2 \hbar\omega_b} = \sqrt{2}\, s_b\hbar/\lambda$ (with $\hbar\omega_b = \hbar eB/m_b c$ being the cyclotron energy and $\lambda = \sqrt{c\hbar/eB}$ the magnetic length). We denote $\alpha = xN\alpha'\langle S_z\rangle/2$ and $\beta = xN\beta'\langle S_z\rangle/6$, where x is the MnTe mole fraction, N is the number of unit cells per unit volume, and α' and β' are the respective exchange integrals for the Γ_6 and Γ_8 bands. The thermodynamic average $\langle S_z\rangle$ of the z-component of a localized spin S (with 5 = 5/2 for Mn^{2+} ions) defines the magnetic field and temperature dependence of the exchange effects and can be described via the normalized Brillouin function $B_S(x)$,

$$\langle S_z\rangle = -S(1-x)^{18} B_S\left(\frac{2\mu_B B}{k_B(T+T_N)}\right), \quad (2)$$

where T_N is the effective temperature arising from the antiferromagnetic interaction between Mn^{2+} ions [22, 23].

To solve Eq. (1), we use the concept proposed by Zel'dovich and Migdal [24, 25] for the related problem of describing the vacuum condensate of Dirac electrons

near supercritical nuclei; we have used this model for the Kane Hamiltonian [4, 26]. In this scenario, the problem is reduced to a Schrödinger-type equation with the effective potential in which the terms responsible for non-parabolicity, spin–orbit splitting, and the "resonant" shift are easily singled out. The qualitative similarity and the quantitative difference between the cases where $E_g > 0$ and $E_g < 0$ are also clearly seen. As in [4], we use the semiclassical approximation both for the calculation of the surface potential $V(z)$ and for the quantization of the resulting equations. The validity of this approach in narrow-gap semiconductors was argued and demonstrated by comparison with numerical self-consistent calculations in many papers (see [4] and references therein). The simplicity of the method is of considerable advantage for the purposes of oscillations modeling.

Arguing as in [26], we obtain from (1) the semiclassical expression for the "spin-split" z-components of the wave vector

$$k_z^{\pm} = \frac{\sqrt{2m_b s_b^2}}{s_b\hbar}\{K \mp [K^2 - (E_{\text{eff}} - U^+)(E_{\text{eff}} - U^-) + U_{so}^+ U_{so}^-]^{1/2}\}^{1/2}, \quad (3)$$

where

$$K = E_{\text{eff}} - (U^+ + U^-)/2 - m_b s_b^2 C_g^2 (R_n^+ - 1)(R_n^- - 1)$$

and the effective energy is $E_{\text{eff}} = (E^2 - m_b^2 s_b^4)/2m_b s_b^2$. In the effective potential $U^{\pm} = U_0 + U_B^{\pm} + U_{\text{exc}}^{\pm} + U_R^{\pm}$, we single out the "Klein–Gordon" term

$$U_0 = (V^2 - 2EV)/2m_b s_b^2,$$

two spin-like terms, namely the "magnetic potential"

$$U_B^{\pm} = E_B^2[g^2 n R^{\pm} + 3(n \pm 1)(E_+ \pm \beta)/(E_+ \pm 3\beta)]/2m_b s_b^2$$

and the "exchange potential"

$$U_{ex}^{\pm} = [\alpha\beta \pm (\alpha E_+ + \beta E_-)]/2m_b s_b^2,$$

and the "resonant" term describing a "spin-interband" interaction arising from the mixing of the Γ_6 and Γ_8 bands by the electric field,

$$U_R^{\pm} = \frac{s_b^2\hbar^2}{2m_b s_b^2}\left[\frac{3}{4}\left(\frac{1 + L_n^{\pm}}{H_n^{\pm}}\right)^2 + \frac{L_n^{\pm}}{H_n^{\pm}(E_+ \pm 3\beta)}\right]\left(\frac{dV}{dz}\right)^2 + \frac{s_b^2\hbar^2}{4m_b s_b^2}\frac{(1 + L_n^{\pm})}{H_n^{\pm}}\frac{d^2V}{dz^2}.$$

The spin–orbit term involved in Eq. (3) is

$$U_{so}^{\pm} = C_g s_b \hbar \sqrt{R_n^{\pm}}\left[\frac{1 + L_n^{\pm}}{H_n^{\pm}} + \frac{1 + L_n^{\mp}}{2H_n^{\mp}}(R_n^{\pm} - 1)\right]\frac{dV}{dz}.$$

In the above expressions, we used the notations

$$R_n^{\pm} = H_n^{\pm}/H_n^{\mp}, \quad C_g^{\pm} = C_g\sqrt{R_n^{\mp}}(R_n^{\pm} - 1),$$

and

$$C_g = gE_B\sqrt{n}/4m_b s_b^2.$$

We must put

$$L_n^{\pm} = 3E_B^2(n \pm 1)/4(E_+ \pm 3\beta)^2,$$

$$H_n^{\pm} = E_- \pm \alpha - L_n^{\pm}(E_+ \pm 3\beta),$$

and $g = -1$ for surface electrons in Kane semiconductors with $E_g < 0$, and $L_n^{\pm} = 0$, $H_n^{\pm} = E_+ \pm \beta$, and g = +1 in semiconductors with $E_g > 0$. Together with the Bohr-Sommerfeld quantization rule, Eq. (3) defines the magnetic levels $E_n^{\pm}(i, B)$ in the surface quantum well $V(z)$. It must be stressed that the exchange interaction causes not only the appearance of an exchange term in the effective potential, but also a modification of the terms describing the "resonant" and SO interaction.

The calculations show that the SO splitting by far exceeds the contribution of the exchange interaction. Furthermore, the SO interaction also suppresses the splitting due to the exchange interaction. As an example, the SO splitting near the Fermi level corresponding to the first beat node in Fig. 2 is 17.2 meV (at the subband Fermi energy E_{F0} = 78 meV). If we take the exchange interaction into account, the splitting increases by only 4.2 meV even at T = 4.2 K. At the same time, the exchange splitting calculated without taking the SO interaction into account is 5.6 meV. This is why the exchange effects manifest themselves only as a small change in the structure of oscillations near the beat nodes, where the oscillations from different spin branches quench each other.

The SO interaction leads to such a drastic reconstruction of the 2D spectrum in magnetic fields that the description of the spin splitting by the non-relativistic g-value loses its physical meaning. This is also true for narrow-gap DMS with $E_g > 0$. In view of this effect, the results of the analysis of 2D systems in asymmetric quantum wells in these materials are to be revised, because they ignore the SO interaction.

6. RESULTS OF MODELING AND DISCUSSION

In calculating the differential capacitance of the space charge region, the density of states in the magnetic field was described neglecting the mixing between LLs and assuming a Gaussian shape of each level, as we did in [26]. The surface potential and the

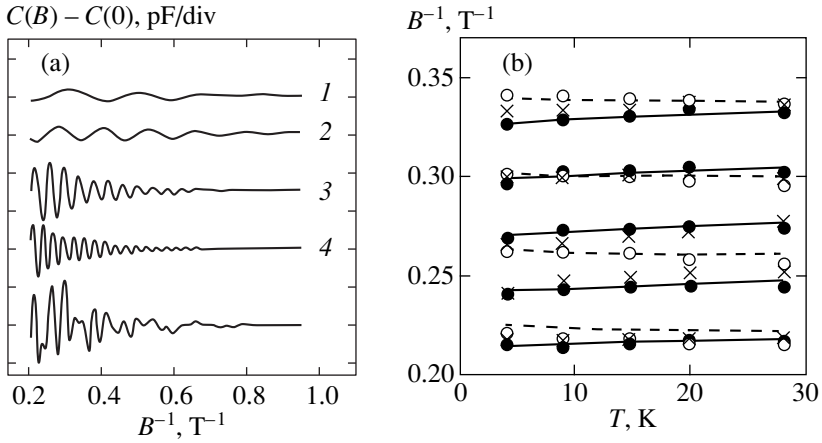

Fig. 5. (a) "Partial" oscillations for different spin sub-subbands $i^{\pm}$ (curve *1*, 1^+; *2*, 1^-; *3*,0^+; *4*, 0^-) extracted from the experimental $C(B)$ trace (the lower plot). (b) Magnetic field positions of (i) the maxima of experimental oscillations (crosses), (ii) "partial" oscillations maxima extracted from the experimental $C(B)$ traces (circles) and (iii) LLs calculated at the Fermi energy (lines) for the ground subband $i = 0$ as functions of temperature. Solid lines and solid circles correspond to the low-energy spin branch (0^-), the dashed line and open circles to the high-energy branch (0^+).

subband Fermi energies are assumed to be constant when the magnetic field is changing. The alternative model is based on the assumption that the surface density is fixed. However, both models give indistinguishable results for a sufficiently large LL broadening (this is manifested by the cosine form of experimental oscillations) [27]. The temperature dependences of band parameters and the bulk Fermi energy are accounted for in the calculations.

Although we performed the calculations for a different net of exchange parameters (literature data vary markedly, see [1, 28–30] and references therein), the results discussed in this section correspond to $N\beta' = 1.5$ eV and $N\alpha' = -0.4$ eV, unless otherwise specified. These values are close to those obtained in [28, 29] by the tunnel spectroscopy method for narrow-gap and gapless HgMnTe with small $|E_g|$. We suppose that these data (with similar values for gapless HgMnTe obtained in many works, see references in [1, 28, 29]) are more suitable for the purposes of this work because the typical electron energies are of the order of or even considerably larger than $|E_g|$ in the studied surface quantum wells. In tunnel experiments, the LL energy positions of "*p*-electrons" as functions of the magnetic field are measured at energies up to 150 meV.

Once the exchange parameters are chosen, two parameters can be obtained when the modeling fits the experimental data: the effective temperature T_N, which describes the temperature shift of beat nodes, and the Dingle temperature, which determines the oscillation amplitudes (and which we use as a characteristic of the scattering).

In calculations, we assume that T_D is the same for both spin–orbit branches. This assumption is supported experimentally. When three or more beat nodes are observed in the oscillations, the "partial" oscillations related to different spin branches can be extracted from the experimental $C(B)$ traces using Fourier filtration and the inverse Fourier transform (see Fig. 5a). The T_D values determined from the fitting of "partial" oscillations turn out to be close for both branches within the accuracy of the analysis.

At the same time, the amplitudes corresponding to these branches can differ considerably (up to several times). This difference is not surprising. Although the DOS is higher at $B = 0$ in the low-energy branch, the corresponding amplitudes can be smaller (even if the relaxation times are equal), because the lower cyclotron energy in this branch leads to a smaller amplitude factor. The calculated amplitude ratio A_c^-/A_c^+ decreases rapidly as N_i increases. This behavior correlates well with the decreasing ratio of the Fourier line intensities observed experimentally. The difference of the amplitudes for different spin components of oscillations mentioned in [11, 14] are therefore expected to be different for 2D systems with a strong SO interaction without invoking the spin-dependent scattering.

Although the general shapes of the simulated and measured oscillations $C(B)$ are well matched, the exact magnetic field positions of the peaks and beat nodes are somewhat different. This is because a number of physical factors are ignored or cannot be exactly taken into account in the theory (the contribution of remote bands, the interface contribution to the SO interaction (see below), the deviation of the real surface potential and the Landau level shape from the calculated ones, etc.). The adjustable phase correction was introduced for convenience for the comparison of the temperature evolution in the measured and calculated oscillations. Its magnitude was chosen to fit the high-field node posi-

tion of the beat pattern at $T = 4.2$ K. None of the physically meaningful results that we discuss are affected by the choice of this factor.

The oscillations calculated with this correction and their Fourier transforms are plotted in Fig. 2. The agreement is quite good with respect to the structure of oscillations and the amplitudes. However, a distinguishable difference in the "number" of oscillations between beat nodes for the measured and calculated plots is observed. These results, as well as similar data on the $dC/dV_g(V_g)$ oscillations, testify to a small (but distinguishable) underestimation of the SO splitting by theory. (We note that in the high-N_s range, the analysis based on the Fourier spectra does not give a clearly detectable discrepancy between experiment and theory.) This inconsistency with theory can be caused by the interface contribution to the SO splitting [16], which cannot be treated in the framework of the effective mass method.

According to the experiment, the individual spin components are not exhibited in simulated $C(B)$ or $dC/dV_g(V_g)$ oscillations even for the lowest LLs at any reasonable broadening parameters, magnetic fields of experimental interest, and temperatures. A decrease of magnetization with increasing temperature results in a slight energy shift of the calculated spin sublevels. However, the position of the resulting oscillations on the magnetic field is almost unchanged (see Fig. 5b), except for the oscillations near the beat nodes (as occurs experimentally). The calculated rate of the temperature shift of beat nodes with the temperature depends on N_i and the node number. At the same time, the value of T_N extracted from the fit of the temperature evolution of the oscillations is almost the same for different nodes and different N_i. The simulation results are not critically dependent on the exact value of T_N chosen. However, the "best fitting" value $T_N = (10 \pm 1.5)$ K must be a reasonably good estimate.

Unfortunately, as far as we know, the low-temperature data on the T_N value for bulk HgMnTe with $x = 0.04$ are absent. Most of the literature data are obtained either for high temperatures or for samples with the Mn content $x \le 0.025$. However, the value $T_N = 10$ K does not contradict other published data. If the sample-independence of the spin-spin interaction is postulated, T_N is nearly proportional to $x(1-x)$[18] [22]. Using the low-temperature data in [22] for a sample with $x = 0.01$ ($T_N = 2.9$ K at $T = 2$ K), we can estimate the value of T_N for samples with $x = 0.04$ as $T_N \approx 8$ K. This is somewhat smaller than the measured value, but T_N can also be temperature dependent [33]. For example, for the same sample with x = 0.01, T_N is equal to 7 K in the high-temperature range [22]. It must be noted that the above estimates are based on assumptions (including the phenomenological expression itself, Eq. (2)) that can be violated for $x > 0.02$ and for low temperatures.

We now turn to the dependence of the observed exchange effects on the value of the exchange parameters. As mentioned above, the exchange interaction is very weakly manifested in the studied system, showing itself as only a slight temperature shift of beat nodes. Because the oscillation amplitudes are small in the neighborhood of nodes even at $T = 4.2$ K and because they decrease drastically with the temperature, the narrow range of $T < 10$–15 K is accessible to the quantitative analysis. Thus, the results are not critically sensitive to the choice of $N\beta'$ and $N\alpha'$. Because the shift rate depends on the product of the exchange parameters $N\beta'$ and $N\alpha'$ and the magnetization $\langle S_z \rangle$, the variations in $N\beta'$ and $N\alpha'$ can be cancelled by the variation in T_N, which is used as an adjustable parameter.

Only the shift of beat nodes to low LL numbers is observed at low temperatures with decreasing $N\beta'$ (the shift is slightly sensitive to the variations of $N\alpha'$ in the –(0.25–0.5) eV range). As a result, the rate of the temperature shift of nodes decreases and becomes less than the one calculated at $N\beta' = 1.5$ eV. However, at $N\beta' > 0.75$ eV, this decrease can be cancelled by a decrease in T_N. For $N\beta' = 1.0$ eV and $N\alpha' = -0.4$ eV, the shifts coincide with those found for $N\beta' = 1.5$ eV and $N\alpha' = -0.4$ eV if the value $T_N = 4$ K is chosen. In both cases, the oscillations are practically the same at all B (including the ranges near the beat nodes) and T. However, the value $T_N = 4$ K seems to be too small for $x = 0.04$.

At the same time, the experimental results cannot be described at $N\beta' < 0.7$ eV. The measured shift rate is nearly twice as large as that calculated at $N\beta' = 0.6$ eV and $N\alpha' = -0.4$ eV (the values given in [30]) even if $T_N = 0$ is chosen. Although the exchange effects in the studied systems with a strong interband mixing are suppressed by the SO splitting, this discrepancy is beyond the limits of experimental error. It is easy to verify that the experimental data (the energy position of LLs and its temperature shift) presented in [28, 29] for bulk HgMnTe with small $|E_g|$ also cannot be explained at $N\beta' < 1.0$–1.2 eV even for $T_N = 0$. As already noted, the value of $N\beta'$ reported in works on gapless HgMnTe falls typically within 0.9–1.6 eV.

The terms in Eq. (3) containing parameter β play the dominant role at the conditions corresponding to a typical experimental situation. On the other hand, the results are only slightly sensitive to reasonable variations of α even in the inversion layers on HgMnTe with $E_g > 0$. It must be stressed that at the energies $E \sim |E_g|$ or higher, the terms involving β must also give the leading contribution in the bulk of DMSs with $E_g > 0$. As a rule, however, the electrons with energy near the band bottom are tested in the investigation of bulk properties. At the same time, in the surface quantum wells on narrow-gap semiconductors, the typical electron energies are of the order of or even considerably larger than $|E_g|$. In this work, the band bending ranges up to 450 meV (this value corresponds to $N_s \approx 4 \times 10^{12}$ cm^{-2}). The above

analyses of the experimental data revealed that the bulk values of the exchange coupling constants obtained at low-energy experiments are workable even at these high energies.

As a related issue, we note that the decrease of $|N\alpha'|$ in a wide-gap CdMnTe-CdMgMnTe quantum well with the increase of energy is reported in a recent paper [31]. The effect is attributed to the admixing of Γ_8 band states to the Γ_6 band at finite **k**-vectors, which leads to switching on a kinetic exchange of the Γ_6 band electrons with the *d* electrons of Mn ions. We note that in narrow-gap semiconductors, the interband mixing described by Kane's Hamiltonian (1) results in a strong (and energy dependent) contribution of the $N\beta'$ containing terms to the spectrum of the Γ_6 band. This is true without taking the energy dependence of the $N\alpha'$ parameter into account. As for the Γ_8 band electrons, the value of the exchange parameter $N\beta'$ is from the outset governed mainly by the kinetic exchange (at any **k**-vector). In this case, an increase of the **k**-vector cannot play a critical role. The absence of an essential change in the value of $N\beta'$ is noted in [31, 32].

The Dingle temperatures T_D determined from the fitting are close to those in HgCdTe-based structures. In the high-N_s range, the T_D values are dictated by the surface roughness scattering. The best agreement between the experimental and calculated values of T_D is achieved at the correlation length $\Lambda \approx (110–120)$ Å and at the average interface displacement $\Delta \approx (20–25)$ Å. Using the T_D values, we can estimate the electron mobility as 0.8×10^4 cm^2/V s in the $i = 0$ subband and 1.5×10^4 cm^2/V s in the $i = 1$ subband for sample S1 at $N_s \sim 10^{12}$ cm^{-2}, which is close to the value 1×10^4 cm^2/V s measured for grain boundaries in *p*-HgMnTe [7].

As in HgCdTe [26], somewhat larger values of T_D are detected at small surface densities $N_s < 5 \times 10^{11}$ cm^{-2}. According to [26], theoretical estimates show that no increase in T_D with the decrease of N_s within this range can be caused by the Coulomb scattering from chargers in the oxide. This conclusion has direct experimental evidence in the present work. It can be seen in Fig. 1 that the charges localized in the oxide differ by a factor of several times for different sweep cycles. If the Coulomb scattering were important, the amplitudes of oscillations corresponding to different cycles (different V_{fb}) but with the same N_s (the same LL number at a fixed magnetic field) would be different. However, the oscillation amplitudes are the same. A possible cause for the increase of the LL broadening at small N_i is the intersubband scattering [34].

This work was supported by the Europen Commission Esprit project no. 28890 NTCONGS EC, by Award no. REC-005 of U. S. Civilian Research and Development Foundation (CRDF) and by the Grant from the Ministry of Education of the Russian Federation.

REFERENCES

1. J. K. Furdina and J. Kossut, *Semiconductors and Semimetals*, Ed. by R. K. Willardson and A. C. Beer (Academic, New York, 1988), Vol. 25; J. K. Furdina, J. Vac. Sci. Technol. A **4**, 2002 (1986).
2. J. Slinkman, A. Zhang, and R. E. Doezema, Phys. Rev. B **39**, 1251 (1989).
3. Y. Takada, K. Arai, and Y. Uemura, in *Proceedings of the Fourth International Conference on the Physics of Narrow-Gap Semiconductors, Linz, Austria, 1981* (Springer-Verlag, Berlin, 1982); Lect. Notes Phys. **125**, 101 (1982).
4. V. F. Radantsev, T. I. Deryabina, G. I. Kulaev, *et al.*, Phys. Rev. B **53**, 15756 (1996).
5. A. V. Germanenko, G. M. Min'kov, V. A. Larionova, *et al.*, Phys. Rev. B **54**, 1841 (1996).
6. G. Grabecki, T. Dietl, J. Kossut, *et al.*, Surf. Sci. **142**, 558 (1984).
7. G. Grabecki, T. Dietl, P. Sobkowicz, *et al.*, Appl. Phys. Lett. **45**, 1214 (1984).
8. T. Suski, P. Wisniewski, L. Dmowski, *et al.*, J. Appl. Phys. **65**, 1203 (1989).
9. G. Grabecki, A. Wittlin, T. Dietl, *et al.*, Semicond. Sci. Technol. **8**, S95 (1993).
10. P. Sobkowicz, G. Grabecki, T. Suski, *et al.*, Acta Phys. Pol. A **75**, 39 (1989).
11. V. F. Radantsev, T. I. Deryabina, L. P. Zverev, *et al.*, Zh. Éksp. Teor. Fiz. **88**, 2088 (1985) [Sov. Phys. JETP **61**, 1234 (1985)].
12. Yu. A. Bychkov and E. I. Rashba, J. Phys. C **17**, 6039 (1984).
13. V. F. Radantsev, Zh. Éksp. Teor. Fiz. **96**, 1793 (1989) [Sov. Phys. JETP **69**, 1012 (1989)].
14. J. Luo, H. Munekata, F. F. Fang, *et al.*, Phys. Rev. B **41**, 7685 (1990).
15. B. Das, S. Datta, and R. Reifenberger, Phys. Rev. B **41**, 8278 (1990).
16. G. Engels, J. Lange, Th. Schapers, *et al.*, Phys. Rev. B **55**, R1958 (1997).
17. J. Nitta, T. Akazaki, H. Takayanagi, *et al.*, Phys. Rev. Lett. **78**, 1335 (1997).
18. J. P. Heida, B. J. van Wees, J. J. Kuipers, *et al.*, Phys. Rev. B **57**, 11911 (1998).
19. L. Wissinger, U. Rössler, R. Winkler, B. Jusserand, and D. Richards, Phys. Rev. B **58**, 15375 (1998).
20. V. F. Radantsev, Semicond. Sci. Technol. **8**, 394 (1993).
21. F. Stern and W. E. Howard, Phys. Rev. **163**, 816 (1967).
22. G. Bastard and C. J. Lewiner, J. Phys. C **13**, 1469 (1980).
23. D. Heiman, Y. Shapira, S. Foner, *et al.*, Phys. Rev. B **29**, 5634 (1984).
24. Ya. B. Zel'dovich and V. S. Popov, Usp. Fiz. Nauk **105**, 403 (1971) [Sov. Phys. Usp. **14**, 673 (1972)].
25. A. B. Migdal, V. S. Popov, and D. N. Voskresenskiĭ, Zh. Éksp. Teor. Fiz. **72**, 834 (1977) [Sov. Phys. JETP **45**, 436 (1977)]; A. B. Migdal, *Fermions and Bosons in Strong Fields* (Nauka, Moscow, 1978).

26. V. F. Radantsev, Zh. Éksp. Teor. Fiz. **115**, 1002 (1999) [JETP **88**, 552 (1999)].
27. N. J. Bassom and R. J. Nicholas, Semicond. Sci. Technol. **7**, 810 (1992).
28. L. P. Zverev, V. V. Kruzhaev, G. M. Min'kov, *et al.*, Zh. Éksp. Teor. Fiz. **86**, 1073 (1984) [Sov. Phys. JETP **59**, 626 (1984)].
29. L. P. Zverev, V. V. Kruzhaev, G. M. Min'kov, *et al.*, Fiz. Tverd. Tela (Leningrad) **26**, 2943 (1984) [Sov. Phys. Solid State **26**, 1778 (1984)].
30. J. Furdyna, J. Appl. Phys. **64**, R29 (1988).
31. I. A. Merkulov, D. R. Yakovlev, A. Keller, *et al.*, Phys. Rev. Lett. **83**, 1431 (1999).
32. A. K. Bhattacharjee, Phys. Rev. B **58**, 15660 (1998).
33. W. Dobrowolski, von M. Ortenberg, and A. M. Sandauer, in *Proceedings of the Fourth International Conference on the Physics of Narrow-Gap Semiconductors, Linz, Austria, 1981* (Springer-Verlag, Berlin, 1982), Lect. Notes Phys. **152**, 302 (1982).
34. V. F. Radantsev, T. I. Deryabina, L. P. Zverev, *et al.*, Zh. Éksp. Teor. Fiz. **91**, 1016 (1986) [Sov. Phys. JETP **64**, 598 (1986)].

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 146–152.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 166–174.

SOLIDS
Electronic Properties

Metamagnetic Transitions in Frustrated Spin-Ladders

D. V. Dmitriev[a], V. Ya. Krivnov[b, *], and A. A. Ovchinnikov[b]

[a]*Emanuel' Institute of Biochemical Physics, Russian Academy of Sciences, Moscow, Russia*
[b]*Max-Planck Institute of Physics of Complex Systems, Dresden, Germany*
**e-mail: krivnov@deom.chph.ras.ru)*

Received July 25, 2000

Abstract—Variational calculations of the magnetization curve at zero temperature are reported for two models of frustrated ladder spin systems with ferro- and antiferromagnetic interactions. The ground state of the models is either ferro- or antiferromagnetic depending on model parameters. The character of the transition from the ferro- to the antiferromagnetic state differs from that of the corresponding transition in the *XXZ* model and is characterized by the appearance of bound multimagnon states. The existence of these states is shown to result in magnetization jumps at certain external field values. The region of the phase diagram where such jumps occur was determined, and the corresponding critical field values were found. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In recent years, spin-ladders have been the focus of considerable attention from both experimental and theoretical physicists. These systems, which comprise bound spin chains and are intermediate between one- and two-dimensional systems, possess certain unique properties [1]. For instance, in some doped ladder systems, superconductivity under compression was observed [2]. The synthesis of new ladder systems is expected to lead to creating high-T_c superconductors based on them.

Of special interest are the properties of ladder systems in magnetic fields. It has, for instance, been found that, under certain conditions, a plateau can appear in the magnetization curves of spin-ladders [3, 4]. Recent discussions have been concerned with another possible anomalous phenomenon, namely, the occurrence of magnetization jumps at certain critical magnetic fields [5]. This phenomenon is sometimes called the metamagnetic transition.

Magnetization is difficult to study even for comparatively simple spin models, because such studies require knowledge of the dependence of energy on the total spin. One of few models that allow this dependence to be determined exactly is the one-dimensional *XXZ* model [6]. For this model, the $m(H)$ magnetization curve (m is the mean spin value per center) is smooth; that is, $m(H)$ continuously varies from zero to the saturation value. Including nearest-neighbor interactions can, however, substantially change the situation and lead to the appearance of magnetization jumps [5].

The magnetic properties of antiferromagnetic spin-ladders have been studied in detail. More complex frustrated system models including both antiferromagnetic and ferromagnetic interactions have received much less attention. Such interactions are present in real ladder systems. Note that the metamagnetic transition at low temperatures was observed in $CsCuCl_3$ [7]. The topology of the structure of $CsCuCl_3$ is close to ladder, and the compound is characterized by both ferro- and antiferromagnetic interactions.

In this work, we consider two spin-ladder models comprising two spin $s = 1/2$ chains with competing ferro- and antiferromagnetic interactions. The phase diagram of these models includes the ferromagnetic (F) and antiferromagnetic (AF) phases. One of the models, which corresponds to a ferromagnetic ladder system with antiferromagnetic diagonals, is gapless in the AF phase, whereas the other, which is a zigzag ladder system, contains a gap in the excitation spectrum.

Generally, the $m(H)$ magnetization curve at zero temperature is characterized by the presence of two critical fields, H_{c1} and H_{c2}. At $0 < H \le H_{c1}$, $m(H)$ equals zero, and at $H \ge H_{c1}$, $m(H) = 1/2$. It is natural to expect that H_{c1} should be zero for the gapless model and should equal Δ for the model with a gap (Δ is the singlet–triplet gap). So far as the upper critical field (H_{c2}) is concerned, it usually equals one-magnon excitation energy ϵ_1, that is, the single-spin-flip energy in the F state. The $m(H)$ dependence exhibits such a behavior when the $\varepsilon(m)$ energy of the ground state at a fixed magnetization is a monotonically increasing and convex function. Precisely such is the $\varepsilon(m)$ dependence for the *XXZ* model if the anisotropy parameter satisfies the inequality $-1 < \gamma < 1$ [6]. However, if $\partial^2\varepsilon/\partial m^2 \le 0$ at certain m values, then the $m(H)$ magnetization curve has jumps, and, generally, the H_{c1} and H_{c2} critical fields are not determined by the Δ and ϵ_1 values. Such an anomalous $m(H)$ behavior is characteristic of the models under consideration at certain parameter values. As will be shown, this is caused by the presence of bound multimagnon states and ensuing phase separation.

The $\varepsilon(m)$ dependence cannot be calculated exactly for ladder systems. For this reason, various approximations are used, such as the mean-field method [8], per-

turbation theory [9], bosonization of spin models [10], and numerical diagonalization of finite systems. In this work, we apply the variational approach based on the use of variational functions of the matrix-product (MP) type. This function has been extensively employed in calculations of one-dimensional and quasi-one-dimensional spin and electronic systems [11, 12] and gives results comparable in accuracy with those of the density matrix renormalization group (DMRG) method [13]. The advantage of the MP approximation is the possibility to apply it to systems containing more than 10^6 spins and, therefore, to virtually pass to the thermodynamic limit. Note also that an MP-type function is the exact wave function of the ground state for certain special models of ladder systems [14–16]. Apart from variational calculations, we used the results of the exact diagonalization of finite systems. In a certain limiting case, zigzag ladder systems can be reduced to the exactly solvable *XXZ* model.

The material is arranged as follows. The results of calculations of magnetization curves for two ladder system models and the details of the approximations that we use are described in Sections 2 and 3. A brief discussion of the results is given in the Conclusions.

2. LADDER MODEL WITH TWO TYPES OF INTERACTIONS

Consider a ladder system with ferromagnetic nearest-neighbor and antiferromagnetic diagonal interactions (Fig. 1). The Hamiltonian of this model in a homogeneous magnetic field has the form

$$\mathcal{H} = \mathcal{H}_0 - H\sum_{i=1}^{2}\sum_{n=1}^{N} S_{in}^z, \tag{1}$$

where

$$\mathcal{H}_0 = J_F\left[\sum_{i=1}^{2}\sum_{n=1}^{N}\left(\mathbf{S}_{in}\cdot\mathbf{S}_{in+1}-\frac{1}{4}\right) + \sum_{n=1}^{N}\left(\mathbf{S}_{1n}\mathbf{S}_{2n}-\frac{1}{4}\right)\right] + J_{AF}\sum_{n=1}^{N}\left(\mathbf{S}_{1n}\cdot\mathbf{S}_{2n+1}+\mathbf{S}_{2n}\cdot\mathbf{S}_{1n+1}-\frac{1}{2}\right), \tag{2}$$

$\mathbf{S}_{1(2)n}$ are the operators of spin $s = 1/2$ that refer to the lower (upper) chain, N is the number of transverse bonds (ladder rungs). Further, exchange integrals are assigned the values $J_F = -1$ and $J_F = J > 0$.

At low J values ($J \ll 1$), the ground state of Hamiltonian $\mathcal{H}_0$ is ferromagnetic. The exact J_c value at which the transition from the ground ferromagnetic to the singlet state occurs (the F–AF transition at zero temperature) is not known. Let J_c be defined as the J value at which the ground F state becomes unstable with respect to the creation of one magnon. As the energy of the one-magnon state $\epsilon_1(J) = E_{\min}(S_{\max} - 1)$ is given by

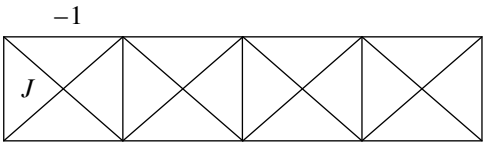

Fig. 1. Ferromagnetic ladder with antiferromagnetic diagonals.

$$\epsilon_1 = (-2J + 1), \quad J \geq 1/2,$$
$$\epsilon \approx N^{-2}, \quad J < 1/2,$$

and as $E_F = E(S_{\max}) = 0$, we have $J_c = 1/2$.

The same J_c value follows from the classical approximation. However, with respect to the creation of two magnons, the ground ferromagnetic state is unstable already at $J = 0.442$. What is more, calculations of finite systems show that $J_c(n+1) < J_c(n)$, where $J_c(n)$ is the J value at which the F state becomes unstable with respect to the creation of n magnons. In this respect, the model under consideration is identical to the two-dimensional ferromagnetic model with antiferromagnetic diagonal interactions [17]. On the other hand, the character of the transition in the present model differs from that of the F–AF transition in the one-dimensional *XXZ* model, in which anisotropy parameter γ tends to –1 and all n-magnon energies simultaneously vanish ($E_F = 0$) at the transition point.

As follows from the aforesaid, the true J_c value in model (2) corresponds to the instability of the F state with respect to the flip of a half of all spins; that is, $J_c = J_c(N)$. At $J = J_c$, the direct transition from the ferromagnetic to the singlet state occurs.

The observation that $\epsilon_n < 0$ at $J > J_c(n)$ and all $\epsilon_i \propto N^{-2}$ $(i < n)$ at $J_c(n) < J < J_c(n-1)$ means that n magnons form a bound state at least at J values within this interval [in reality, the state may also be bound at $J > J_c(n-1)$, and the binding energy reduces to zero at $J = J_c(n)$. It follows that, when $J \longrightarrow J_c$, the bound state of N magnons only "survives." As the total number of spins in this state equals $2N$, the state is singlet ($m = 0$). It is also clear that, when $J \longrightarrow J_c$, the ground state for a fixed m value is two-phase and comprises the ferromagnetic and singlet (bound) states. As the energy of the bound state of a macroscopic number of magnons is proportional to their number, we come to the conclusion that, when $J \longrightarrow J_c$, the $\varepsilon(m) = E_0(m)/N$ energy per cross-piece ($m = S/2N$) is a linear function of m; that is,

$$\varepsilon(m) = \varepsilon(0)(1 - 2m). \tag{3}$$

When magnetic field is switched on or J tends to J_c, magnetization experiences a jump from $m = 0$ to the saturation value, $m = 1/2$, at $H = H_c = |\varepsilon(0)|$. (Note that the spectrum of model (2) is gapless, and $H_{c1} = 0$. We therefore use the denotation H_c instead of H_{c2} in this section.)

The situation becomes more complex at $J > J_c(n)$, because the limiting number of magnons in the bound state, n_c, can be smaller than N, and the magnetization of this state is given by $m_c = 1/2 - n_c/2N$. The $\varepsilon(m)$ dependence is then linear at $m_c < m < 1/2$. In this region, the ground state is two-phase and comprises the ferromagnetic phase and the phase with magnetization m_c. At $H = H_c$, the magnetization changes in a jump from $m = m_c$ to the saturation value $m = 1/2$, and the critical field value is given by $H_c = (2J - 1) + |\epsilon_b|$, where ϵ_b is the binding energy per magnon.

It should, however, be noted that bound magnon states exist not at all J values. Calculations of finite systems (up to $N = 18$) show that increasing J causes a gradual decrease in the limiting number of magnons in the bound state to $n = 2$. The exact solution of the two-magnon problem gives the value $J = J_2 = 2.054$, at which the binding energy vanishes. At $J > J_2$, bound states do not exist, the $\varepsilon(m)$ function is convex, and no magnetization jump occurs. The critical field then equals $H_c = |\epsilon_1(J)| = 2J - 1$.

2.1. MP-Type Variational Function

The considerations given above are, of course, heuristic in character, and their substantiation requires performing calculations of the $m(H)$ dependence. For this purpose, we used an MP-type variational function, which, for the model under consideration, had the form

$$\Psi = \mathrm{Tr}\prod_{n=1}^{N}(A + B_{-1}S_{1n}^{+} + B_1 S_{2n}^{+} + CS_{1n}^{+}S_{2n}^{+})|0\rangle, \quad (4)$$

where index n refers to the nth rung of the ladder, $|0\rangle = |\downarrow\downarrow\ldots\downarrow\rangle$, and A, B_λ, and C are $l \times l$ matrices.

The energy per rung equals

$$\varepsilon = \frac{E}{N} = -\frac{1}{2}\sum_{\lambda=-1}^{1}\frac{\mathrm{Tr}(G_\lambda Q_\lambda T^{N-2})}{\mathrm{Tr}T^N} + \frac{J}{2}\sum_{\lambda=-1}^{1}\frac{\mathrm{Tr}(G_\lambda Q_{-\lambda}T^{N-2})}{\mathrm{Tr}T^N} - \frac{1}{4}\sum_{\lambda=-1}^{1}\frac{\mathrm{Tr}(Z_\lambda Z_\lambda T^{N-2})}{\mathrm{Tr}T^N} + \frac{J}{4}\sum_{\lambda=-1}^{1}\frac{\mathrm{Tr}(Z_\lambda Z_{-\lambda}T^{N-2})}{\mathrm{Tr}T^N} - \frac{1}{2}\sum_{\lambda=-1}^{1}\frac{\mathrm{Tr}(B_\lambda \otimes B_{-\lambda}T^{N-1})}{\mathrm{Tr}T^N} - \frac{1}{4}\frac{\mathrm{Tr}\left[\left(A\otimes A + C\otimes C - \sum_{\lambda=-1}^{1}B_\lambda\otimes B_{-\lambda}\right)T^{N-1}\right]}{\mathrm{Tr}T^N} + \frac{3}{4} - \frac{J}{2}, \quad (5)$$

where

$$G_\lambda = A\otimes B_\lambda + B_{-\lambda}\otimes C,$$

$$Q_\lambda = B_\lambda \otimes A + C\otimes B_{-\lambda},$$

$$Z_{1(-1)} = C\otimes C - A\otimes A \pm \sum_{\lambda=-1}^{1}\lambda B_\lambda\otimes B_{-\lambda},$$

$$T = A\otimes A + C\otimes C + \sum_{\lambda=-1}^{1}B_\lambda\otimes B_\lambda,$$

and symbol $\otimes$ denotes the direct product of the matrices.

The magnetization per center at $N \gg 1$ is given by

$$m = \frac{1}{2N}\sqrt{\langle S^z\rangle^2 + \langle S^+\rangle\langle S^-\rangle}, \quad (6)$$

where

$$\langle S^z\rangle = \sum_{\lambda=-1}^{1}\frac{\mathrm{Tr}(Z_\lambda T^{N-1})}{\mathrm{Tr}T^N},$$

$$\langle S^+\rangle = \sum_{\lambda=-1}^{1}\frac{\mathrm{Tr}(G_\lambda T^{N-1})}{\mathrm{Tr}T^N},$$

$$\langle S^-\rangle = \sum_{\lambda=-1}^{1}\frac{\mathrm{Tr}(Q_\lambda T^{N-1})}{\mathrm{Tr}T^N}.$$

Equations (5) and (6) determine the $\varepsilon(m)$ dependence. The matrix elements of matrices A, B_λ, and C are variational parameters. Formally, there are $4l^2$ variational parameters with respect to which variation should be performed. In reality, one of the matrices under the Tr sign can be selected as diagonal, and one of the elements can be fixed, which leaves $3l^2 + l - 1$ variational parameters. Naturally, the accuracy of calculations with

function (4) depends on the size of the matrices. Note that analytic calculations are only possible at $l = 1$, when the matrices are numbers. For $l > 1$, a numerical minimization procedure should be used.

2.2. Magnetization Curve

The critical J_c value is found from the condition $\varepsilon(m) = 0$. If $l = 1$, $J_c = (\sqrt{7} - 1)/4 \approx 0.411$, $l = 2$ gives $J_c = 0.409$, and with matrices for $l = 4$ and 6, we obtain $J_c = 0.403$ and 0.4029, respectively. Generally, the difference in the results for $l = 4$ and 6 is very insignificant (less than 0.1%). We therefore used 4×4 matrices in our calculations.

It was found that the ground singlet state energy at $J \gtrsim J_c$ was a linear function of $J - J_c$. This was evidence that the F–AF transition in the model under consideration was a first-order transition.

The $m(H)$ magnetization curve was found by minimizing the $e - 2mH$ energy at a fixed magnetic field value. By way of example, the $m(H)$ dependence for $J = 0.5$ is shown in Fig. 2. Note that, when H tends to zero, the $m(H)$ function is linear, because $\varepsilon(m) \sim \varepsilon(0) \propto m^2$ at $m \ll 1$ for the model with a gapless spectrum [18]. The dependences of critical field H_c and critical magnetization m_c on J are shown in Fig. 3. According to this figure, $H_c \propto J - J_c$, at least at small $J - J_c$ values. Such a dependence follows from (3), because $H_c \approx |\varepsilon(0)| \propto J - J_c$ if $J \longrightarrow J_c$.

According to our calculations, critical field H_c is larger than the $|\epsilon_1|$ value at $1/2 < J < 2$. As mentioned above, the upper critical field coincides with this value in models in which the $\varepsilon(m)$ function is everywhere convex. In the model under consideration, the condition of convexity of the $\varepsilon(m)$ function is violated because of the presence of bound multimagnon states, and $H_c = (2J - 1) + |\epsilon_b|$. The ϵ_b energy cannot be found exactly, and we estimated it from the ϵ_{2b} value equal to half the binding energy of the two-magnon state. Clearly, the $|\epsilon_b(J)|$ energy is larger than $|\epsilon_{2b}(J)|$. It follows that the $H_c(J)$ estimate obtained in such a way is the lower bound estimate of the critical field. Figure 3 shows that this estimate is very close to the critical field value found in variational calculations. It follows that the deviation of ϵ_b from ϵ_{2b} is exceedingly small, at least at $J > 1/2$.

As follows from Fig. 3, the magnetization jump disappears when $J \longrightarrow 2$; that is, at J values close to J_2 at which bound states do not exist.

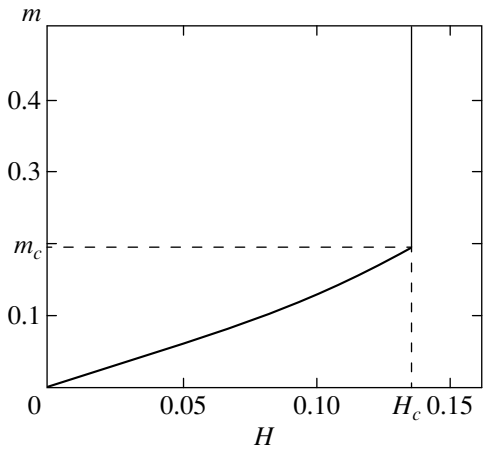

Fig. 2. Magnetization curve of model (2) at $J = 0.5$.

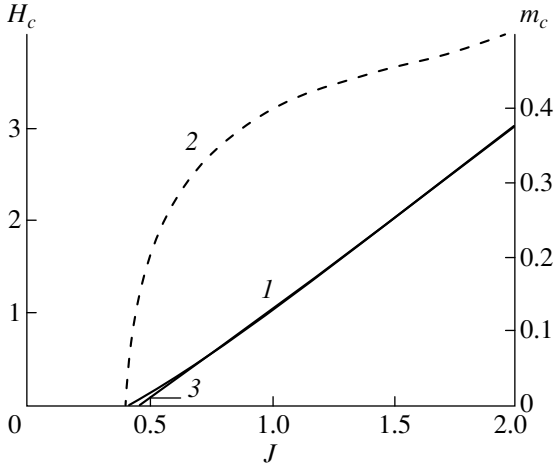

Fig. 3. Critical field H_c (curve *1*) and critical magnetization m_c (curve *2*) as functions of J in model (2). Curve *3* corresponds to $H_c(J)$ estimated from ϵ_{2b}.

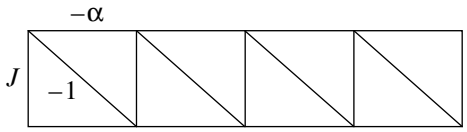

Fig. 4. Zigzag model.

3. THE ZIGZAG MODEL

Consider a ladder model of another type with ferromagnetic interaction $-\alpha$ along chains and antiferromagnetic interaction J in crosspieces. In addition, there is ferromagnetic interaction (assumed to equal -1) along one of ladder diagonals (Fig. 4). The Hamiltonian of this model, which is also called the zigzag ladder system model, or zigzag model, has the form

$$\mathcal{H} = \mathcal{H}_0 - H\sum_{i=1}^{2}\sum_{n=1}^{N} S_{in}^z,$$

$$\mathcal{H}_0 = -\alpha\sum_{i=1}^{2}\sum_{n=1}^{N}\left(\mathbf{S}_{in}\cdot\mathbf{S}_{in+1} - \frac{1}{4}\right) \qquad (7)$$

$$-\sum_{n=1}^{N}\left(\mathbf{S}_{2n}\cdot\mathbf{S}_{1n+1} - \frac{1}{4}\right) + J\sum_{n=1}^{N}\left(\mathbf{S}_{1n}\cdot\mathbf{S}_{2n} - \frac{1}{4}\right),$$

$$\alpha, J > 0.$$

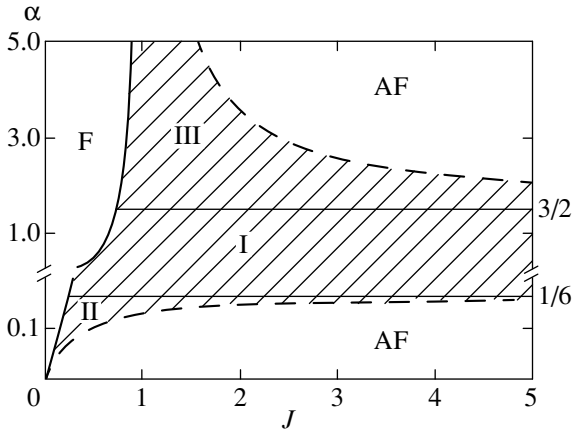

Fig. 5. Phase diagram of the zigzag model. Solid curve is the line of F–AF transitions, dashed lines are the boundaries of the existence of bound two-magnon states. The region of magnetization jumps is hatched.

This model is equivalent to the one-dimensional spin model with alternating ferro- and antiferromagnetic nearest-spin interactions and ferromagnetic next-nearest-spin interactions.

As distinguished from the model considered above, the exact critical $J_c(\alpha)$ value for the F–AF transition in model (7) is known [16, 19], and $J_c(\alpha) = 2\alpha/(2\alpha + 1)$. The corresponding phase diagram (at zero temperature) is shown in Fig. 5. The AF phase has a gap in the excitation spectrum [19]. There is one more feature distinguishing model (7) from that considered above, namely, all n-magnon energies ϵ_n simultaneously vanish at $J = J_c(\alpha)$ [19]. In particular, when $J \longrightarrow J_c(\alpha)$ at a fixed α value, we have

$$\epsilon_1(J) = -\frac{(J - J_c)^2}{4(1 - J_c)J_c^2},$$

and the ground state energy equals $\varepsilon(0) = \varepsilon_N = -(J - J_c)/2J_c$.

This model, however, differs from the one-dimensional XXZ model with $\gamma \gtrsim -1$ by the formation of n-magnon bound states; the binding energies reduce to zero at $J = J_c(\alpha)$. As mentioned in Section 2, the presence of bound states results in the appearance of linear $\varepsilon(m)$ dependence regions and magnetization jumps (metamagnetic transitions). Generally, bound magnon states exist in the AF phase not at all J and α values. The results of the exact diagonalization of finite systems show that, if two-magnon states are not bound, bound states of larger numbers of magnons also do not exist. The boundary of the phase diagram region in which magnetization jumps should be expected is therefore determined from the condition of the vanishing of the two-magnon state binding energy. The exact solution to this problem gives the boundary shown in Fig. 5. To the right of this boundary (that is, in the AF phase outside regions I, II, and III), bound states do not exist.

As for the model considered in Section 2, the $\varepsilon(m)$ and $m(H)$ dependences were found with the use of an MP-type variational function with a 4×4 matrix. The corresponding calculations showed that the $m(H)$ function in the region where bound states do not exist had the form typical of an antiferromagnet with a gap in the spectrum (Fig. 6a); that is, it was a monotonically increasing function characterized by the presence of two critical fields, upper H_{c2} and lower H_{c1}. In addition, $H_{c1} = \Delta$, where Δ is the singlet–triplet gap, and $H_{c2} = |\epsilon_1|$. The region of the existence of bound states is characterized by $m(H)$ dependences of two different types. In regions II and III, the $m(H)$ dependence has the form schematically shown in Fig. 6b, and the H_{c2} critical field is larger than $|\epsilon_1|$. At the boundary between these regions, where the binding energy of two-magnon states vanishes, we have $m_c = 1/2$ and $H_{c2} = |\epsilon_1|$, where

$$\epsilon_1 = \begin{cases} 1 - J, & \alpha > 1/2 \\ 2\alpha - J, & \alpha < 1/2. \end{cases}$$

By way of example, the dependences of H_{c1}, H_{c2}, and m_c on J at a fixed $\alpha = 2$ value are shown in Fig. 7.

Of special interest in the phase diagram is region I with $1/6 < \alpha < 3/2$. In this region, $m(H)$ is a step function (Fig. 6c), and $h_{c1} = h_{c2} = |\varepsilon(0)|$. The $m(H)$ dependence exhibits such a behavior because $\Delta > |\varepsilon(0)|$ in region I, the critical field is smaller than is required for triplet excitation, and, at fields exceeding critical, the transition from the singlet to the ferromagnetic state

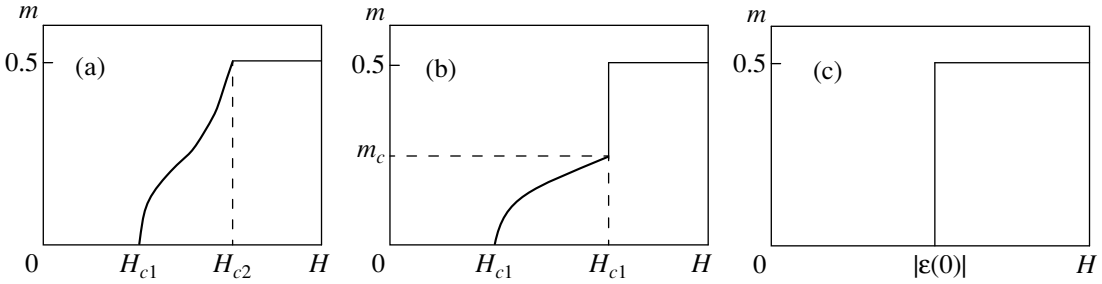

Fig. 6. Behavior of $m(H)$ for the zigzag model (schematically) (a) in the region without bound magnon states, (b) in regions II and III, and (c) in region I.

occurs without the formation of states with intermediate spins. It follows that the metamagnetic transition that occurs in region I is ideal.

These conclusions concerning the behavior of the magnetization curve in the zigzag model follow from variational calculations. At the same time, at $J \gg \max(1, \alpha)$, the zigzag model can be reduced to the exactly solvable effective *XXZ* Hamiltonian. This reduction is performed with the use of the Bose representation of a pair of spins $s = 1/2$ of ladder rungs [20]. According to [20], the S_{1n} and S_{2n} spin operators can be expressed in terms of four Bose operators s_n^+, u_n^+, d_n^+, and t_n^+. The action of these operators on the $|0\rangle$ vacuum state determines four possible spin states of the rung, namely,

$$s_n^2|0\rangle = \frac{1}{\sqrt{2}}(|\downarrow\uparrow\rangle - |\downarrow\uparrow\rangle), \quad u_n^+|0\rangle = |\uparrow\uparrow\rangle,$$
$$d_n^+|0\rangle = |\downarrow\downarrow\rangle, \quad t_n^+|0\rangle = \frac{1}{\sqrt{2}}(|\downarrow\uparrow\rangle + |\downarrow\uparrow\rangle). \tag{8}$$

In the $|\alpha\beta\rangle$ configuration, α and β refer to the lower and upper chains, respectively.

The Bose operators in (8) satisfy the constraint condition

$$s_n^+ s_n + u_n^+ u_n + d_n^+ d_n + t_n^+ t_n = 1. \tag{9}$$

The S_{1n} and S_{2n} operators are written in terms of the introduced Bose operators as follows [20]:

$$S_{1(2)n}^+ = \frac{1}{\sqrt{2}}(\pm s_n^+ d_n \mp u_n^+ s_n + t_n^+ d_n + u_n^+ t_n),$$
$$S_{1(2)n}^- = \frac{1}{\sqrt{2}}(\pm d_n^+ s_n \mp s_n^+ u_n + d_n^+ t_n + t_n^+ u_n), \tag{10}$$
$$S_{1(2)n}^z = \frac{1}{\sqrt{2}}(\pm s_n^+ t_n \pm t_n^+ s_n - d_n^+ d_n + u_n^+ u_n).$$

At $J \gg \max(1, \alpha)$, the singlet–triplet gap and the lower critical field values are proportional J; we can therefore ignore the $t_n^+|0\rangle$ and $d_n^+|0\rangle$ triplet states and only retain the $s_n^+|0\rangle$ singlet state and the $u_n^+|0\rangle$ state with the largest spin projection onto the field direction. These two states correspond to pseudospin $\tilde{s} = 1/2$. Taking into account (9), the S_{1n} and S_{2n} operators can be written through the $\tilde{S}_n$ pseudospin operators as

$$S_{1(2)n}^+ = \mp\frac{1}{\sqrt{2}}\tilde{S}_n^+, \quad S_{1(2)n}^- = \mp\frac{1}{\sqrt{2}}\tilde{S}_n^-,$$
$$S_{1(2)n}^z = \frac{1}{2}\left(\frac{1}{2} + \tilde{S}_n^z\right), \tag{11}$$

where

$$\tilde{S}_n^+ = u_n^+ s_n, \quad \tilde{S}_n^- = s_n^+ u_n,$$
$$\tilde{S}_n^z = \frac{1}{2}(u_n^+ u_n - s_n^+ s_n).$$

Representation (11) corresponds to a perturbation series expansion in $1/J$ to first-order terms [21, 22].

Using (11), let us write the Hamiltonian of the zigzag model at $J \gg \max(1, \alpha)$ in the form

$$\mathcal{H}_{\text{eff}} = J_{xy}\sum_n(\tilde{S}_n^x\tilde{S}_{n+1}^x + \tilde{S}_n^y\tilde{S}_{n+1}^y) + J_z\sum_n\tilde{S}_n^z\tilde{S}_{n+1}^z$$
$$-\tilde{h}\sum_n\tilde{S}_n^z + \frac{3}{8}N\mu - \frac{1}{2}NJ - \frac{1}{2}Nh, \tag{12}$$

where

$$J_{xy} = -\lambda, \quad J_z = -\mu/2, \quad \tilde{h} = h + \mu/2 - J,$$
$$\lambda = \alpha - 1/2, \quad \mu = \alpha + 1/2. \tag{13}$$

The projections of the total spin of the initial model and model (12) are related as

$$S^z = \tilde{S}^z + \frac{N}{2}, \quad S^z = \sum_{i=1}^{2}\sum_n S_{in}^z, \quad \tilde{S}^z = \sum_n\tilde{S}_n^z. \tag{14}$$

It follows that, in the $J \gg \max(1, \alpha)$ limit, the zigzag model reduces to the *XXZ* model with exchange integrals given by (13). As $J_z < 0$, the *zz* interaction in (12) is ferromagnetic.

The properties of the *XXZ* model have been thoroughly studied, and the corresponding exact results can be used to analyze the behavior of the zigzag model in a magnetic field at $J \gg \max(1, \alpha)$.

Clearly, the ground state of the $\tilde{\mathcal{H}}_{\text{eff}}$ Hamiltonian at $H = 0$ corresponds to $\tilde{S}^z = -N/2$, and, for the zigzag model, $S^z = 0$. The ground state energy is given by $\varepsilon(0) = -J + \mu/2$. The gap in the spectrum of $\tilde{\mathcal{H}}_{\text{eff}}$ exci-

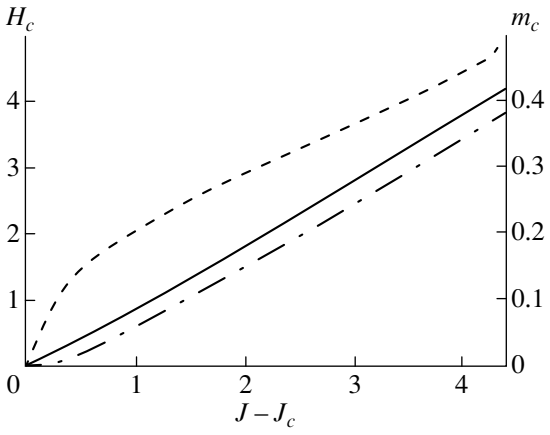

Fig. 7. Critical fields and critical magnetizations for the zigzag model with $\alpha = 2$ ($J_c = 0.8$); solid, dot-and-dash, and dashed lines correspond to H_{c2}, H_{c1}, and m_c, respectively.

tations corresponding to a single-spin flip (the singlet–triplet gap in the initial model) equals $\Delta = J - |\lambda|$.

If

$$|J_z/J_{xy}| > 1 \quad (1/6 < \alpha < 3/2),$$

then the lowest excited states of $\tilde{\mathcal{H}}_{\text{eff}}$ at $H = 0$ are bound states of n flipped spins, and their energies are [23, 24]

$$E_n = \left(-J + \frac{\mu}{2}\right)(N - n).$$

This means that the ground state energy of the zigzag model at a fixed magnetization value is a linear function of m,

$$\varepsilon(m) = -J + \frac{\mu}{2} + 2\left(J - \frac{\mu}{2} - H\right)m. \tag{15}$$

According to (15), a magnetization jump from $m = 0$ to $m = 1/2$ occurs at $H = |\varepsilon(0)|$, and $H_{c1} = H_{c2} = |\varepsilon(0)|$. This substantiates the conclusion of the existence of the metamagnetic transition in region I, which was made based on variational calculations.

At $|J_z/J_{xy}| < 1$ (that is, at $0 < \alpha < 1/6$ and $\alpha > 3/2$), the $\varepsilon(m)$ function is convex [6]. According to [6], $m(H)$ increases from $m = 0$ at $H = H_{c1} = \Delta$ to $m = 1/2$ at $H = H_{c2} = |\epsilon_1|$. The behavior of $m(H)$ at $H \approx H_{c1}$ and $H \approx H_{c2}$ has the root-type singularity

$$\begin{aligned} m &= \frac{1}{2\pi}\sqrt{\frac{2}{|\lambda|}(H - H_{c1})}, \\ \frac{1}{2} - m &= \frac{1}{2\pi}\sqrt{\frac{2}{|\lambda|}(H_{c2} - H)}. \end{aligned} \tag{16}$$

It follows that, outside the $1/6 < \alpha < 3/2$ band, the $m(H)$ dependence has the form typical of an antiferromagnet with a gap in the spectrum. At the same time, the region of the phase diagram where $m(H)$ has jumps (Fig. 5) extends outside this band and covers regions II and III. At $J \gg 1$, these regions, however, have sizes of the order of $1/J$.

4. CONCLUSIONS

We considered two models of frustrated spin-ladders with ferro- and antiferromagnetic interactions. One of them is gapless in the AF phase, whereas the other has a gap in the spectrum of excitations. At certain model parameter values, the magnetization curves of these models have jumps corresponding to metamagnetic transitions. Such jumps are caused by the presence of multimagnon bound states and the appearance of linear $\varepsilon(m)$ dependence regions. These linear regions correspond to the two-phase state of the system. The results of the numerical diagonalization of finite systems lead us to conclude that the boundaries of the phase diagram regions in which magnetization jumps occur are determined by the condition of the disappearance of two-magnon bound states. This conclusion is substantiated by variational calculations of magnetization curves.

It is reasonable to expect that such transitions can occur in a two-dimensional ferromagnet with antiferromagnetic interaction between next neighbor spins.

ACKNOWLEDGMENTS

The authors thank the Max Planck Institute of Complex System Physics (Dresden) for hospitality and financial support. This work was also financially supported by the Russian Foundation for Basic Research (projects nos. 00-03-32981 and 00-15-97334).

REFERENCES

1. E. Dagotto, Rep. Prog. Phys. **62**, 1525 (1999).
2. M. Uehara, T. Nagata, J. Akimitsu, *et al.*, J. Phys. Soc. Jpn. **65**, 2764 (1996).
3. D. C. Cabra, A. Honecker, and P. Pujol, Phys. Rev. Lett. **79**, 5126 (1997).
4. K. Totsuka, Phys. Rev. B **57**, 3454 (1998).
5. C. Gerhardt, K. H. Mutter, and H. Kroger, Phys. Rev. B **57**, 11504 (1998).
6. C. N. Yang and C. P. Yang, Phys. Rev. **150**, 321 (1966); **150**, 327 (1966); **151**, 258 (1966).
7. H. Nojiri, Y. Tokunaga, and M. Motokawa, J. Phys. (Paris) **49**, C8-1459 (1988).
8. S. Copolan, T. M. Rice, and M. Sigrist, Phys. Rev. B **49**, 8901 (1994).
9. Weihong Zheng, V. N. Kotov, and J. Oitmaa, Phys. Rev. B **57**, 11439 (1998).
10. R. Chitra and T. Giamarchi, Phys. Rev. B **55**, 5816 (1997).
11. A. K. Kolezhuk and H.-J. Mikeska, Int. J. Mod. Phys. B **12**, 2325 (1998).
12. V. Ya. Krivnov and A. A. Ovchinnikov, Phys. Lett. A **248**, 453 (1998).
13. S. R. White, Phys. Rev. B **48**, 10345 (1993).
14. A. Klumper, A. Schadschneider, and J. Zittartz, Z. Phys. B **87**, 281 (1992); Europhys. Lett. **24**, 293 (1993).
15. A. K. Kolezhuk and H.-J. Mikeska, Phys. Rev. B **56**, R11380 (1997).
16. D. V. Dmitriev, V. Ya. Krivnov, and A. A. Ovchinnikov, Phys. Rev. B **56**, 5985 (1997).
17. D. V. Dmitriev, V. Ya. Krivnov, and A. A. Ovchinnikov, Phys. Rev. B **55**, 3620 (1997).
18. M. Gaudin, *La fonction d'onde de Bethe* (Masson, Paris, 1983; Mir, Moscow, 1987).
19. D. V. Dmitriev, V. Ya. Krivnov, and A. A. Ovchinnikov, Eur. Phys. J. B **14**, 91 (2000).
20. S. Sachdev and R. N. Bhatt, Phys. Rev. B **41**, 9323 (1990).
21. F. Mila, Eur. Phys. J. B **6**, 201 (1998).
22. H.-T. Wang, H. Q. Lin, and J.-L. Shen, Phys. Rev. B **61**, 4019 (2000).
23. A. A. Ovchinnikov, Pis'ma Zh. Éksp. Teor. Fiz. **5**, 48 (1967) [JETP Lett. **5**, 38 (1967)].
24. I. G. Gochev, Zh. Éksp. Teor. Fiz. **61**, 1674 (1971) [Sov. Phys. JETP **34**, 892 (1972)].

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 153–158.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 175–181.

SOLIDS
Electronic Properties

Spin-Plasmon Oscillations of the Two-Dimensional Electron Gas

L. I. Magarill, A. V. Chaplik, and M. V. Éntin*
Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia
**e-mail: entin@isp.nsc.ru*
Received August 23, 2000

Abstract—Interaction of the spins of 2D electrons with an alternating electric field in the plane of the system is considered. It is assumed that the double spin degeneracy is eliminated by the spin-orbit splitting. It is shown that transitions between different spin states produce a narrow absorption band in the degenerate electron gas. In the frequency domain corresponding to these transitions, those frequencies are combined with two-dimensional plasmons; as a result, the plasmon spectrum is modified, and a new type of oscillations occurs, namely, a spin-plasmon polariton. The dispersion law of these oscillations is derived. The problem of the excitation of spin-plasmon polaritons by an external electromagnetic field is solved. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Recently, the investigation of spin effects in low-dimensional systems has come to the attention of researchers. This is mainly due to the possible application of the spin degree of freedom for creating quantum bits. It is preferable to control the spin subsystem by means of an electric field, since the magnetic field required for this purpose must be rather large, and large fields cannot be varied quickly. Another advantage of the electric control is a high degree of spatial selectivity.

An electric field can affect the spin degree of freedom by means of the spin-orbit interaction. For example, it was shown in [1] that a direct current induced by a static electric field lying in the plane of the system results in the spin polarization of the electrons in the two-dimensional system. In the case of alternating fields, resonance effects occur, and the relatively small spin-orbit interaction can become substantial. It is well known (see [2]) that an alternating electric field can cause transitions between spin sublevels; this phenomenon is called combined resonance. The combined resonance for the one-electron problem in the presence of a magnetic field was considered in [2, 3]. It is clear that similar effects can manifest themselves in the collective oscillations of the electron plasma.

The purpose of this study is to analyze the effect of an alternating electric field on a two-dimensional Fermi gas with account for the spin-orbit interaction. In the absence of the magnetic field, a narrow peak occurs in the absorption spectrum caused by transitions between spin-orbit split subbands in the vicinity of the spin splitting of the Fermi surface. The interaction of such a transition with two-dimensional plasmons induces mixed spin-plasmon waves. The spectrum of these oscillations is found and the possibility of exciting them is considered.

2. DYNAMIC CONDUCTANCE IN THE COLLISION-FREE APPROXIMATION

It is well known that the spin-orbit interaction of 2D electrons on an oriented surface can be described in terms of the Hamiltonian proposed in [3, 4]:

$$\hat{\mathcal{H}}_0 = \frac{\mathbf{p}^2}{2m} + \hat{\mathcal{H}}_{so}, \quad \hat{\mathcal{H}}_{so} = \alpha\boldsymbol{\sigma}\cdot\mathbf{p}\times\mathbf{n} \equiv \boldsymbol{\sigma}\cdot\boldsymbol{\Omega}. \quad (1)$$

Here $\mathbf{p} = (p_x, p_y)$ is the two-dimensional momentum of the electron, axis z is directed perpendicular to the system's plane, α is the spin-orbit interaction constant, and σ_i are the Pauli matrices; here and in what follows, we assume that $\hbar = 1$.

Hamiltonian (1) yields the energy spectrum

$$\varepsilon_\mu(p) = \frac{p^2}{2m} + \mu|\alpha|p, \quad (2)$$

where $\mu = \pm 1$ labels two branches of the spin split spectrum of the two-dimensional electron gas. The splitting of the branches is usually not large and amounts to $\sim 10^{-2}$ of the Fermi energy. The wave functions can be written as

$$\psi_\mu = \frac{\exp(i\mathbf{p}\cdot\mathbf{r})}{\sqrt{2S}}\begin{pmatrix} i\mu e^{-i\varphi_\mathbf{p}}\operatorname{sgn}\alpha \\ 1 \end{pmatrix}, \quad (3)$$

where $\varphi_\mathbf{p}$ is the azimuth angle of vector $\mathbf{p}$ and S is the area of the system.

Resonance transitions between spin-orbit split subbands in the degenerate two-dimensional gas occur in the vicinity of the spin splitting frequency of the Fermi

surface $\omega_0 = 2|\alpha|p_F$ (where p_F is the Fermi momentum). The Kubo formula [5] for the tensor of the dynamic conductivity is written as

$$\sigma_{ij}(\omega) = \frac{\pi G_0}{S}\int_0^\infty dt e^{-\delta t} \times \int_0^{1/T} ds\left\langle \mathrm{Sp}\{ f(\hat{\mathcal{H}})\hat{v}_j(-is)(1-f(\hat{\mathcal{H}}))\hat{v}_i(t)\}\right\rangle. \quad (4)$$

Here T is temperature, $G_0 = e^2/\pi$ is the conductance quantum, $\hat{\mathcal{H}}$ is the complete Hamiltonian of the system, $f(\varepsilon) = 1/(1 + \exp(\varepsilon - \zeta)/T)$ is the Fermi function (ζ is the chemical potential), $\delta \longrightarrow +0$, $\hat{\mathbf{v}}(t)$ is the electron velocity operator in the Heisenberg representation, and ω is the frequency of the electric field; here and in what follows, $\hbar = 1$. Angular brackets denote averaging over impurities.

In the collision-free limit, we obtain from (4) that

$$\sigma_{ij}(\omega) = \frac{G_0}{4}\int d\mathbf{p}\sum_\mu \left\{ \frac{\partial f_{\mu p}}{\partial \zeta}\frac{v^j_{\mu\mu}(\mathbf{p})v^i_{\mu\mu}(\mathbf{p})}{\delta - i\omega} + \frac{f_{\bar{\mu}p} - f_{\mu p}}{\varepsilon_\mu(p) - \varepsilon_{\bar{\mu}}(p)}\frac{v^j_{\mu\bar{\mu}}(\mathbf{p})v^i_{\bar{\mu}\mu}(\mathbf{p})}{\delta - i\omega + \varepsilon_\mu(p) - \varepsilon_{\bar{\mu}}(p)}\right\}. \quad (5)$$

Here $\bar{\mu} = -\mu$ and $f_{\mu p} \equiv f(\varepsilon_\mu(p))$. The infinitely small switching rate of the field δ can be considered as the phenomenological reciprocal relaxation time.

Using the wave functions (3), we can easily find the matrix elements of the velocity operator as

$$\mathbf{v}_{\mu\mu}(\mathbf{p}) = \frac{\mathbf{p}}{pm}(p + m|\alpha|\mu),$$
$$\mathbf{v}_{\mu,\bar{\mu}}(\mathbf{p}) = i|\alpha|\mu\frac{\mathbf{p}\times\mathbf{n}}{p}. \quad (6)$$

In what follows, we assume that the electron gas is degenerate in the sense that the temperature is small as compared to the Fermi energy ε_F. At the same time, the temperature can be comparable with the splitting between spin subbands.

Substituting (2) and (6) in (5), we obtain $\sigma_{ij}(\omega) = \delta_{ij}\sigma(\omega)$

$$\sigma(\omega) = \sigma_D(\omega) + \sigma_s(\omega),$$
$$\sigma_D(\omega) = \frac{ne^2}{m(\delta - i\omega)} = iG_0\frac{\varepsilon_F}{\omega + i\delta}, \quad (7)$$

$$\sigma_s(\omega) = \frac{G_0|\alpha|}{8}\int_0^\infty dp\left\{\left[f\left(\frac{p^2}{2m} - |\alpha|p\right) - f\left(\frac{p^2}{2m} + |\alpha|p\right)\right] \times\left[\frac{1}{\delta - i\omega - 2i|\alpha|p} + \frac{1}{\delta - i\omega + 2i|\alpha|p}\right]\right\}. \quad (8)$$

Here n is the density of the 2D electrons. Magnitude σ_D is the contribution to conductivity due to the diagonal (in μ) matrix elements of the velocity. We included this contribution in the principal order with respect to the parameter $(\alpha/v_F)^2$ (v_F is the Fermi velocity) in which it coincides with the Drude conductivity. Magnitude σ_s is determined by transitions between spin subbands.

We are interested in the domain of frequencies close to ω_0 ($|\omega - \omega_0| \ll \omega_0$). In this domain, formula (8) for σ_s can be simplified as

$$\mathrm{Re}(\sigma_s) = \frac{\pi G_0}{16}\left[\frac{1}{\exp\left(\frac{\eta - 1}{\Theta}\right) + 1} - \frac{1}{\exp\left(\frac{\eta + 1}{\Theta}\right) + 1}\right],$$
$$\mathrm{Im}(\sigma_s) = \frac{iG_0}{16}\mathcal{P}\int_{-\infty}^\infty \frac{dy}{\eta - y} \times\left[\frac{1}{\exp\left(\frac{y - 1}{\Theta}\right) + 1} - \frac{1}{\exp\left(\frac{y + 1}{\Theta}\right) + 1}\right], \quad (9)$$

where $\Theta = T/(\alpha p_F)$ is the dimensionless temperature and $\eta = (\omega - \omega_0)/(2m\alpha^2)$ is the detuning from the center of the spin absorption band. The symbol $\mathcal{P}$ denotes the principal part of the integral.

For low temperatures $\Theta \ll 1$, we obtain from (9) that

$$\sigma_s = \frac{G_0}{16}\left\{ i\ln\left|\frac{\eta + 1}{\eta - 1}\right| + \pi\theta(1 - \eta^2)\right\}. \quad (10)$$

For $T = 0$, $\mathrm{Re}(\sigma(\omega))$ is a step function of frequency, which is not zero for $\omega_0 - 2m\alpha^2 < \omega < \omega_0 + 2m\alpha^2$. This is a result of the energy conservation law under vertical transitions between spin subbands (see Fig. 1). The imaginary part of the conductivity has logarithmic singularities at the endpoints of this interval. The low temperature limit is obtained for the temperature less than that of the splitting of the subbands in the vicinity of the Fermi momentum $2|\alpha|p_F$. At large temperatures, the peak of absorption due to spin transitions is smeared by the magnitude $2Tm\alpha/p_F$.

3. SPECTRUM OF THE PLASMON

We consider the problem of oscillations of the two-dimensional electron gas with allowing for transitions between spin subbands; however, we neglect the displacement currents.

To determine the spectrum of the oscillations, one must solve the system of the continuity equation $\dot{\rho} + \nabla\mathbf{j} = 0$ for the surface charge density ρ and the current $\mathbf{j}$, the Poisson equation $\Delta\phi = -4\pi\rho\delta(z)$ for the potential, and the constitutive equation $\mathbf{j} = -\sigma\nabla\phi$.

After the Fourier transform in the pattern plane, the dispersion equation for the plasmon is written as

$$\frac{2\pi i\sigma(\omega)q}{\omega} = -\kappa. \qquad (11)$$

Here q is the wave vector and κ is the static dielectric constant. If the 2D system is located between two dielectrics with constants κ_1 and κ_2, then $\kappa = (\kappa_1 + \kappa_2)/2$. For the configuration of the type metal–dielectric–semiconductor [6], we have

$$\kappa(q) = (\kappa_1 + \kappa_2 \coth qd)/2, \qquad (12)$$

where d is the distance between the two-dimensional electrons and the gate electrode. Thus, the dependence of the conductivity on frequency implicitly determines the spectrum of the spin-plasmons $\omega(k)$.

In the collision-free limit and the zero temperature, the spectrum in the absence of the gate electrode is determined by the equation

$$\frac{1}{k} = \frac{1}{16}\frac{1}{1+\eta\lambda}\left[\frac{4}{1+\eta\lambda} + \lambda\ln\left(\frac{\eta+1}{\eta-1}\right)\right], \qquad (13)$$

where $k = q\pi G_0/(\kappa m\alpha^2)$ is a dimensionless wave vector and $\lambda = |\alpha|/v_F$ is a parameter. In the absence of the spin-orbit interaction ($\lambda = 0$), (13) yields the well-known root spectrum of the plasmon [7] (see also survey [8] and references therein). Nondamping plasma waves exist for $\eta > 1$ and $\eta < -1$, which corresponds to the frequencies $\omega < 2\alpha p_F - 2m\alpha^2$ and $\omega > 2\alpha p_F + 2m\alpha^2$ (see Fig. 2). Within the frequencies $2\alpha p_F - 2m\alpha^2 < \omega < 2\alpha p_F + 2m\alpha^2$, the damping of the plasmon occurs due to the collision-free transfer of the energy into spin excitation.

For the same limit case, Eq. (13) can be further simplified in the domain of the wave interaction $\eta \ll 1/\lambda$ by introducing the dimensionless variable $\xi = (k-4)/8\lambda$:

$$\ln\left(\frac{\eta-1}{\eta+1}\right) = 8(\xi - \eta). \qquad (14)$$

Equation (14) shows that the domain of characteristic frequencies and momenta in which the plasmon spectrum is modified are determined (in the order of magnitude) by the domain of the intersection of the plasmon dispersion and the domain of spin transitions; they are $|\omega - 2\alpha p_F| \sim 2m\alpha^2$ and $(k-4) \sim 8\lambda$, respectively (see the inset in Fig. 2).

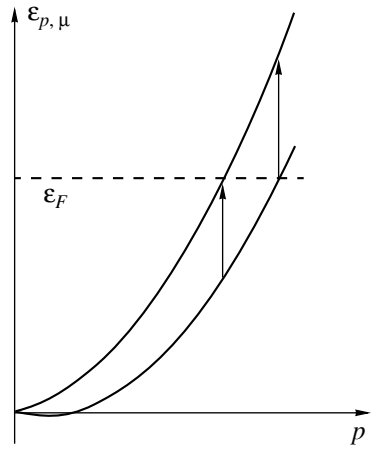

Fig. 1. The scheme of transitions between spin subbands. The arrows correspond to transitions with threshold frequencies of $2(|\alpha p_F| \pm m\alpha^2)$.

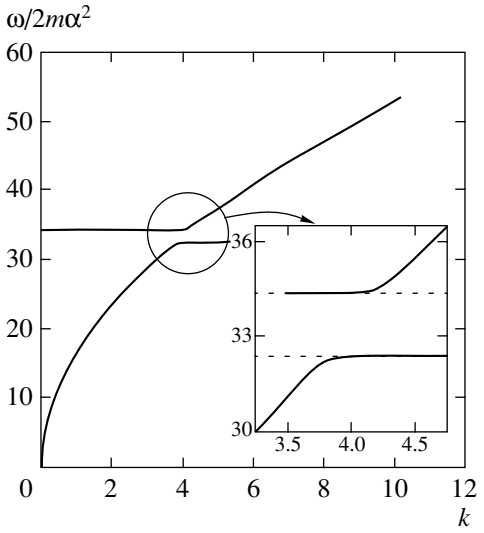

Fig. 2. The spectrum of spin-plasmons in the absence of wave damping for $\lambda = 0.0259$ corresponding to the heterostructure InAs/GaSb with the concentration 7.5×10^{11} cm^{-2} of carriers ($2m\alpha^2 = 0.117$ meV). Outside the domain of the intersection with the spin transitions, the spectrum is described by the function $\omega \sim \sqrt{k}$. The inset shows the domain of the transition interaction in more detail. The dotted lines mark the boundary frequencies for spin transitions. Away from the interaction domain, the branches exponentially approach the dotted lines.

4. QUANTUM KINETIC EQUATION

We have already derived a formula for high-frequency conductivity for the collision-free case. Although magnitude δ can be considered as a phenomenological damping, the following question remains open taking into account the fact that both the translational and spin degrees of freedom are involved: which relaxation time must be used as δ? To answer this question, one must solve the kinetic problem with account of the collisions of electrons; we consider this problem for the case of impurity scattering.

The electron kinetics is described by the single-particle density matrix $\hat{\rho}(\mathbf{p})$, which is diagonal in the momentum $\mathbf{p}$ and nondiagonal in the spin variable. The

quantum kinetic equation linearized with respect to the external electric field is written as

$$\dot{\rho} + i[\hat{\mathcal{H}}_{so}, \hat{\rho}] + \hat{\mathcal{F}} = \mathrm{St}(\hat{\rho}). \tag{15}$$

Here $\mathcal{F}$ is the field term, and $\mathrm{St}(\hat{\rho})$ is the collision term. We write Eq. (15) in the basis of the eigenstates of the Rashba Hamiltonian as

$$\begin{aligned}\dot{\rho}_{\mu'\mu}(\mathbf{p}) + i(\varepsilon_{\mu'}(p) - \varepsilon_{\mu}(p))\rho_{\mu'\mu}(\mathbf{p}) \\ + \mathcal{F}_{\mu'\mu} = \mathrm{St}(\rho)_{\mu'\mu},\end{aligned} \tag{16}$$

where the field and collision terms are written as

$$\mathcal{F}_{\mu'\mu}(\mathbf{p}) = -e\mathbf{E}\mathbf{v}_{\mu'\mu}(\mathbf{p})\frac{f(\varepsilon_{\mu'}(p)) - f(\varepsilon_{\mu}(p))}{\varepsilon_{\mu'}(p) - \varepsilon_{\mu}(p)}, \tag{17}$$

$$\begin{aligned}\mathrm{St}(\rho)_{\mu'\mu} = \pi N_i \sum_{\mathbf{p}', \nu', \nu} \{M_{\mathbf{p}\mu'; \mathbf{p}'\nu'}M_{\mathbf{p}'\nu; \mathbf{p}\mu}\rho_{\nu'\nu}(\mathbf{p}') \\ \times [\delta(\varepsilon_{\nu'}(p') - \varepsilon_{\mu}(p)) + \delta(\varepsilon_{\nu}(p') - \varepsilon_{\mu'}(p))] \\ - M_{\mathbf{p}\mu'; \mathbf{p}'\nu'}M_{\mathbf{p}'\nu'; \mathbf{p}\nu}\rho_{\nu\mu}(\mathbf{p})\delta(\varepsilon_{\mu}(p) - \varepsilon_{\nu'}(p')) \\ - M_{\mathbf{p}\nu'; \mathbf{p}'\nu}M_{\mathbf{p}'\nu; \mathbf{p}\mu}\rho_{\mu'\nu'}(\mathbf{p})\delta(\varepsilon_{\mu'}(p) - \varepsilon_{\nu}(p'))\}.\end{aligned} \tag{18}$$

Here $M_{\mathbf{p}\mu'; \mathbf{p}'\mu}$ is the matrix element corresponding to the interaction with the impurity center and N_i is the number of impurities. For $\mu' = \mu$, the definition of the fraction in (17) must be understood as $f'(\varepsilon_{\mu}(p))$, where $f'(\varepsilon) = \partial f(\varepsilon)/\partial\varepsilon$. Equation (16) with the collision term (18) was derived in [9, 10] for the general form of the energy spectrum of the electron by using the Konstantinov–Perel' diagram technique [9]. For spin-orbit split bands, a similar equation without the collision term was given in [11] (however, with the interband optic generation matrix). The field term linearized in α was presented in [1].

It is convenient to use a fixed basis of states (independent of the direction of p). For the field term F, we have

$$\begin{aligned}\hat{\mathcal{F}} = -\frac{e}{4m\varepsilon}\mathbf{E}\{2\mathbf{p}(2\varepsilon + \hat{\mathcal{H}}_{so})f'(\varepsilon + \hat{\mathcal{H}}_{so}) \\ + i\mathbf{p}\times\mathbf{n}(\boldsymbol{\sigma}\cdot\mathbf{n})(f(\varepsilon + \hat{\mathcal{H}}_{so}) - f(\varepsilon - \hat{\mathcal{H}}_{so}))\},\end{aligned} \tag{19}$$

where $\varepsilon = p^2/2m$. Since parameter λ is assumed to be small, we can confine ourselves to spin-conserving scattering. Then, the collision integral can be written as

$$\mathrm{St}(\hat{\rho}) = 2\pi N_i\sum_{\mathbf{p}'}|V_{\mathbf{p}'-\mathbf{p}}|^2\delta(\varepsilon - \varepsilon')(\hat{\rho}(\mathbf{p}') - \hat{\rho}(\mathbf{p})). \tag{20}$$

Here $V_{\mathbf{p}'-\mathbf{p}}$ is the Fourier component of the impurity potential (we neglected the spin-orbit corrections to the Hamiltonian of the interaction with impurities).

We are interested in the response to the alternating electric field $\mathbf{E}(t) = \mathrm{Re}(\mathbf{E}_{\omega}e^{-i\omega t})$. Using the fact that the identity matrix I and $\boldsymbol{\sigma}$ form a complete basis in the space of the second-order matrices, we seek the solution to Eq. (15) as

$$\begin{aligned}\hat{\rho} = (\hat{\rho}_{\omega}e^{-i\omega t} + \mathrm{H.c.})/2, \\ \hat{\rho}_{\omega} = A_{\omega}(\mathbf{p})\hat{I} + \mathbf{B}_{\omega}(\mathbf{p})\cdot\boldsymbol{\sigma}.\end{aligned} \tag{21}$$

For $A_{\omega}(\mathbf{p})$, we have

$$\begin{aligned}A_{\omega}(\mathbf{p}) \equiv A_1(\varepsilon)\frac{\mathbf{E}_{\omega}\cdot\mathbf{p}}{p} \\ = \tau_{1\omega}\frac{e\mathbf{E}_{\omega}\cdot\mathbf{p}}{2m}\left[f'_{+}(\varepsilon) + \frac{\Omega}{2\varepsilon}f'_{-}(\varepsilon)\right].\end{aligned} \tag{22}$$

Here $\Omega = |\boldsymbol{\Omega}| = |\alpha|p$, $1/\tau_{1\omega} = 1/\tau_1 - i\omega$, and $f_{\pm}(\varepsilon) \equiv f(\varepsilon + \Omega) \pm f(\varepsilon - \Omega)$. Here and in what follows, τ_n is the relaxation time of the nth angular moment of the distribution function (τ_1 is the conventional transport relaxation time for the momentum) determined as

$$\tau_n^{-1} = 2\pi N_i\sum_{p'}|V(\mathbf{p} - \mathbf{p}')|^2\delta(\varepsilon - \varepsilon')(1 - \cos n\phi), \tag{23}$$

where ϕ is the angle between $\mathbf{p}$ and $\mathbf{p}'$.

The quantity $\mathbf{B}_{\omega}(\mathbf{p})$ satisfies the equation

$$\begin{aligned}-i\omega\mathbf{B}_{\omega}(\mathbf{p}) - 2\boldsymbol{\Omega}\times\mathbf{B}_{\omega}(\mathbf{p}) = \mathrm{St}(\mathbf{B}_{\omega}(\mathbf{p})) \\ + \frac{e}{4m\varepsilon}\left\{\boldsymbol{\Omega}(\mathbf{E}_{\omega}\cdot\mathbf{p})\left(f'_{+}(\varepsilon) + \frac{2\varepsilon}{\Omega}f'_{-}(\varepsilon)\right)\right. \\ \left. - \frac{\mathbf{p}}{\Omega}(\mathbf{E}_{\omega}\cdot\boldsymbol{\Omega})f_{-}(\varepsilon)\right\}.\end{aligned} \tag{24}$$

The solution to Eq. (24) is decomposed into the sum of three terms corresponding to the angular harmonics with respect to the momentum:

$$\begin{aligned}\mathbf{B}_{\omega}(\mathbf{p}) = B_0(\varepsilon)\mathbf{E}\times\mathbf{n} + \frac{B_1(\varepsilon)}{p}\mathbf{E}\times\mathbf{p} \\ + B_2(\varepsilon)\left\{\frac{(\mathbf{E}\times\mathbf{n}\cdot\mathbf{p})\mathbf{p}}{p^2} - \frac{\mathbf{E}\times\mathbf{p}}{2p}\right\}.\end{aligned} \tag{25}$$

Separating the angular harmonics out, we obtain a system of algebraic equations for $B_i(\varepsilon)$:

$$\begin{aligned}2\alpha pB_0 + \frac{1}{\tau_{1\omega}}B_1 + \alpha pB_2 = 0, \\ i\omega B_0 + \alpha pB_1 \\ = -\frac{e\alpha}{4\Omega}[\Omega f'_{+}(\varepsilon) + 2\varepsilon f'_{-}(\varepsilon) + f_{-}(\varepsilon)], \\ \frac{1}{\tau_{2\omega}}B_2 - 2\alpha pB_1 = -\frac{e\alpha}{2\Omega}[\Omega f'_{+}(\varepsilon) + 2\varepsilon f'_{-}(\varepsilon) - f_{-}(\varepsilon)].\end{aligned} \tag{26}$$

The dynamic conductivity can be expressed in terms of $A_\omega(\mathbf{p})$ and $B_0(\varepsilon)$ as

$$\sigma(\omega) = -\int_0^\infty \frac{d\varepsilon}{2\pi}(pA_1(\varepsilon) + m\alpha B_0(\varepsilon)). \quad (27)$$

The quantity $B_0(\varepsilon)$ in (27) can be determined from system (26) as

$$B_0(\varepsilon) = \frac{ie\alpha}{4\Omega}(4\Omega^2\tau_{1\omega}\tau_{2\omega} + 1) \times \frac{\Omega f'_+(\varepsilon) + 2\varepsilon f'_-(\varepsilon) + f_-(\varepsilon)}{2\omega\Omega^2\tau_{1\omega}\tau_{2\omega} + \omega + 2i\Omega^2\tau_{1\omega}}. \quad (28)$$

The first term in (27) (in the principal order in λ) coincides with the conventional Drude–Lorentz formula (7) $\sigma_D = ne^2\tau_1(1 - i\omega\tau_1)^{-1}m^{-1}$ for a finite $\delta = 1/\tau_1$.

In order for a plasmon to be weakly damping in the vicinity of the resonance $\omega \sim \omega_0$, the inequality $\Omega\tau_1 \gg 1$ must be satisfied. This makes it possible to simplify the expression for σ_s and write it as

$$\sigma_s(\omega) = -\frac{ie^2m\alpha^2}{8\pi\omega_0}\int_0^\infty d\varepsilon \frac{f_-(\varepsilon)}{\omega - 2\Omega + i\tau_s^{-1}}. \quad (29)$$

Here, $\tau_s^{-1} = (2\tau_1^{-1} + \tau_2^{-1})/4$ is the collision broadening of the spin transitions.

From Eq. (29) in the collision-free limit, we obtain (9) and, for low temperature $\Omega \ll 1$,

$$\sigma_s = \frac{iG_0}{16}\ln\left(\frac{\eta + 1 + i(2m\alpha^2\tau_s)^{-1}}{\eta - 1 + i(2m\alpha^2\tau_s)^{-1}}\right). \quad (30)$$

The order of the magnitude of τ_s is the same as that of τ_1, which determines the mobility. In particular, for the scattering by neutral impurities, we have $\tau_s = (4/3)\tau_1$; for charged unscreened impurities in the plane of the 2D electrons, $\tau_s = \tau_1$; and for the small-angle scattering by charged impurities placed far away from the two-dimensional layer (a thick spacer), $\tau_s = (2/3)\tau_1$.

5. EXCITATION OF PLASMONS

It is well known that the plasmon spectrum can be observed by the absorption of the electromagnetic wave incident onto the two-dimensional system. Since the wavelength of the corresponding frequencies in a vacuum is large, the electric field of the wave is spatially modulated by means of a grating structure [12] (see also survey [8] and references therein). The density of the power absorbed by a two-dimensional system is determined by the Fourier harmonics of the electromagnetic field $\mathbf{E}(\mathbf{q}, z = 0)$ in the plane of the system. For a one-dimensional grating with grooves oriented along axis y with period d, we have $\mathbf{q}_n = (0, 2\pi n/d)$. These harmonics are linearly related to the zero harmonic of the field $\mathbf{E}(\mathbf{q} = 0, z = 0)$ by the coefficients $c_\mathbf{q}$ determined only by the grating:

$$\mathbf{E}(\mathbf{q}, z = 0) = \mathbf{E}(\mathbf{q} = 0, z = 0)c_\mathbf{q}. \quad (31)$$

According to [8], the absorbed power can be written as

$$W = \frac{1}{2}|E_x(\mathbf{q} = 0, z = 0)|^2\mathrm{Re}(\sigma_{\mathrm{eff}}(\omega)). \quad (32)$$

The effective conductivity $\sigma_{\mathrm{eff}}(\omega)$ is determined by the Fourier coefficients of the field and by the partial effective conductivity $\sigma(\mathbf{q}, \omega)$

$$\sigma_{\mathrm{eff}}(\omega) = \sum_\mathbf{q}\sigma(\mathbf{q}, \omega)|c_\mathbf{q}|^2, \quad (33)$$

where

$$\sigma(\mathbf{q}, \omega) = \sigma(\omega)\left(1 + \frac{2\pi iq\sigma(\omega)}{\omega\kappa}\right)^{-1}. \quad (34)$$

Zeros of the denominator in (34) determine the spectrum of the oscillations $\omega(\mathbf{q})$.

Figure 3 illustrates the dependence of the absorption coefficient on frequency for various values of the wave vector. We use the parameters of the heterostructure InAs/GaSb $m = 0.055m_0$ and $\alpha = 9 \times 10^{-10}$ eV cm [13] with the electron mobility $\mu = 10^6$ cm^2/V s (Fig. 3a) and $\mu = 3 \times 10^6$ cm^2/V s (Fig. 3b) and the concentration $n = 7.5 \times 10^{11}$ cm^{-2}. It is seen from the figure that there are two main absorption peaks—a narrow plasma one and a wide spin one; moreover, if the resonances are spaced in terms of frequency, then the spin resonance becomes much weaker (by several orders of magnitude) than the plasma one. The maximum of the absorption in terms of frequency "follows" the spectrum of the spin-plasmon $\omega(q)$ as k varies. For convenience, Fig. 3 presents only the plots for those k for which the plasma resonance lies to the left or in the domain of the spin transitions (if the wave vector determined by the grating structure is greater than $k_c = 4$, then the plasmon peaks lie symmetrically to the right of the domain of spin-flip transitions).

As the domain of the spin transitions is approached, the plasma resonance substantially decreases, while the spin one increases. The transfer of the force of oscillators occurs due to the shielding of the external field by the polarization of the medium caused by the spin transitions. When the resonance frequency falls in the domain of spin transitions, the narrow plasma resonance presses itself against the boundary of this domain in accordance with the spectrum of the spin-plasmon resonance in the collision-free limit (Fig. 2).

For the parameters of the heterostructure InAs/GaSb indicated above, $\hbar\omega_0 = 3.91$ meV ($\omega_0 = 5.94 \times 10^{12}$ s^{-1}) and $q_c = 2.44 \times 10^4$ cm^{-1} (the corresponding period of the grating structure is $a = 2.57$ μm).

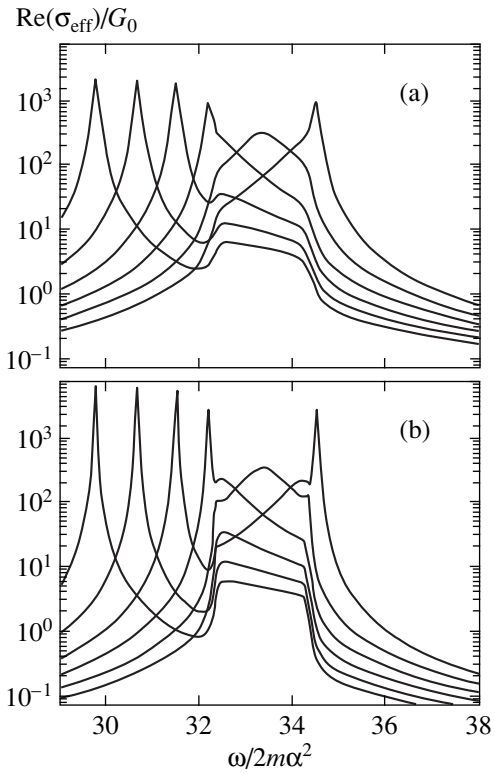

Fig. 3. The dependence of the absorption coefficient on frequency in the domain of the spin-plasmon resonance for the same system as in Fig. 2. The dimensionless wave vector runs through the values k = 3.2, 3.4, 3.6, 3.8, 4, and 4.2. These values correspond to the motion of the principal absorption maximum from left to right; panels (a) and (b) correspond to the mobility $\mu = 10^6$ cm^2/V s and $\mu = 3 \times 10^6$ cm^2/V s, respectively.

We note that the substantial increase of the absorption coefficient on the spin transitions in the vicinity of the resonance with the plasma oscillations makes it easier to observe this phenomenon experimentally.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project nos. 99-02-17127 and 00-02-17658, and the State Program of the Russian Federation "Physics of Solid-State Nanostructures."

REFERENCES

1. L. I. Magarill and M. V. Éntin, Pis'ma Zh. Éksp. Teor. Fiz. **72**, 195 (2000) [JETP Lett. **72**, 134 (2000)].
2. É. I. Rashba, Fiz. Tverd. Tela (Leningrad) **2**, 1224 (1960) [Sov. Phys. Solid State **2**, 1109 (1960)].
3. Yu. A. Bychkov and É. I. Rashba, Pis'ma Zh. Éksp. Teor. Fiz. **39**, 66 (1984) [JETP Lett. **39**, 78 (1984)]; E. I. Rashba and V. I. Sheka, in *Landau Level Spectroscopy*, Ed. by G. Landwehr and E. I. Rashba (Elsevier, New York, 1991), p. 178.
4. F. T. Vas'ko, Pis'ma Zh. Éksp. Teor. Fiz. **30**, 574 (1979) [JETP Lett. **30**, 541 (1979)].
5. R. Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957).
6. A. V. Chaplik, Zh. Éksp. Teor. Fiz. **62**, 746 (1972) [Sov. Phys. JETP **35**, 395 (1972)].
7. R. H. Ritchie, Phys. Rev. **106**, 874 (1957); R. A. Ferrel, Phys. Rev. **111**, 1214 (1958).
8. T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
9. O. V. Konstantinov and V. I. Perel', Zh. Éksp. Teor. Fiz. **39**, 197 (1960) [Sov. Phys. JETP **12**, 142 (1960)].
10. L. É. Gurevich and G. M. Nedlin, Zh. Éksp. Teor. Fiz. **40**, 809 (1961) [Sov. Phys. JETP **13**, 568 (1961)].
11. E. L. Ivchenko, Yu. B. Lyanda-Geller, and G. E. Pikus, Zh. Éksp. Teor. Fiz. **98**, 989 (1990) [Sov. Phys. JETP **71**, 550 (1990)].
12. S. S. Allen, Jr., D. C. Tsui, and R. A. Logan, Phys. Rev. Lett. **38**, 980 (1977).
13. J. Luo, H. Munekata, F. F. Fang, and P. J. Stiles, Phys. Rev. B **38**, 10142 (1988); Phys. Rev. B **41**, 7685 (1990).

Translated by A. Klimontovich

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 159–168.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 182–192.

SOLIDS
Electronic Properties

Low-Field Electrodynamics of Polycrystalline Superconductors $SnMo_6S_8$ and $PbMo_6S_8$

S. L. Ginzburg*, I. D. Luzyanin, I. R. Metskhvarishvili, É. G. Tarovik, and V. P. Khavronin

Institute of Nuclear Physics, Russian Academy of Sciences, Gatchina, Leningradskaya oblast, 188350 Russia
**e-mail: savitska@thd.pnpi.spb.ru*

Received August 22, 2000

Abstract—Linear and nonlinear dynamic response to the penetration of very weak magnetic fields (of the order of 10^{-2} Oe) into polycrystalline superconductors are studied theoretically and experimentally. The results of experiments are found to be in satisfactory agreement with the conclusions of the low-field electrodynamics. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is well known that magnetic fields weaker than the lower critical field do not penetrate in superconductors. However, this statement becomes incorrect if the superconductor is a multiply connected system, e.g., a polycrystal in which crystallites are connected through Josephson junctions. Since Josephson junctions can be easily pierced by very weak magnetic fields, such fields will penetrate into polycrystalline superconductors also. A similar situation is observed in ceramic HTSC in which granules are connected through weak Josephson junctions. In this case, a wide range of irreversible and nonlinear effects is observed experimentally, which can be explained in the frames of low-field electrodynamics [1, 2]. However, to our knowledge, the penetration of very weak magnetic fields into polycrystalline superconductors has not been investigated as yet.

The present work is devoted to an analysis of the penetration of ultraweak magnetic fields into superconducting polycrystalline samples of $SnMo_6S_8$ and $PbMo_6S_8$. This choice was dictated by a very large value of H_{c2}, and accordingly, very small coherence length $\xi \approx 23$ Å for molybdenum chalcogenides. This makes these materials similar to HTSC for which ξ is of the order of a few Ångströms. Hence, we can expect that any (even small) defect will play the role of a Josephson junction for such values of ξ. The results of our experiments prove, first, that very weak magnetic fields of the order of a few millioersteds indeed penetrate into such a superconductor, and second, that the obtained experimental results are in accord with the predictions of low-field electrodynamics.

2. THEORY

It should be recalled that low-field electrodynamics was formulated initially for granular HTSC materials and is based on the following two assumptions.

1. Bean's model of the critical state [3] is applicable to ceramic superconductors. It is well known that the critical state of a superconductor is strongly nonequilibrium and can be described by a spatially inhomogeneous magnetic induction $\mathbf{B}(\mathbf{r})$. The latter is defined by the equation of the critical state which contains, among other things, the dependence of the equilibrium magnetic induction on the magnetic field strength $B_{eq}(H) \approx \mu_{eff}H$ (μ_{eff} is the effective permeability of the ceramic material, taking into account the impermeability of granules to the field). Consequently, the problem boils down to the definition of the nonuniform field $H(r)$ for which the equation of critical state has the form [3–5]

$$\left|\frac{dH}{dx}\right| = 4\pi j_c(H), \qquad (1)$$

where $j_c(H) = \alpha(H)/H$. The average induction for an infinitely large slab of thickness d is given by

$$\bar{B} = \frac{2}{d}\mu_{eff}\int_0^{d/2} H(x)dx.$$

In Eq. (1), $\alpha(H)$ is the pinning force and $j_c(H)$ has the meaning of the critical current density and is a phenomenological function of H.

It should be noted that there exist several models leading to different dependences $j_c(H)$. For instance, our experiments with various YBaCuO ceramic samples [2, 6] revealed that

$$j_c(H) = \frac{j_0H_0}{|h| + H_0}, \quad j_c(H) = \frac{j_0H_0^2}{h^2 + H_0^2}. \qquad (2)$$

Here H_0 is a certain characteristic field, approximately equal to 3 Oe.

The case when $H_0 \longrightarrow \infty$ leads to Bean's model [3] in which j_c is independent of H. The case $H_0 = 0$ and $j_0H_0 = \text{const}$ in the first formula of (2) corresponds to

the Kim–Anderson model [5] in which the pinning force $\alpha(H)$ does not depend on H. Generally speaking, there is no theoretically substantiated choice of the function $j_c(H)$.

2. A Josephson medium can exhibit two qualitatively different patterns of magnetic field penetration, whose realization is determined by the dimensionless parameter β (which will be defined below) characterizing the number of magnetic flux quanta per elementary loop formed by adjacent granules. When $\beta \ll 1$, a granular superconductor behaves as a classical type II superconductor in which the field penetrates in the form of vortices, and anisotropy induced by the field becomes an important factor. It was shown in a previous publication by one of the authors [7] that in the other limiting case, when $\beta \gg 1$, the discreetness of the Josephson medium, which is now described by equations equivalent to those for systems with self-organized critical state [8], becomes significant.

An approach that can be used to distinguish unambiguously between these two limiting cases and completely different physical patterns corresponding to them was considered in [9].

This approach is directly connected with a very important though sparsely studied problem of longitudinal currents in hard type II superconductors. The essence of the problem can be briefly formulated as follows. It is well known [10] that in the presence of uncut vortices and pinning, the critical current $j_{c\|}$, which is longitudinal relative to the magnetic field, is equal to infinity (in fact, it is equal to the depairing current), while the longitudinal electric field $E_\|$ is always equal to zero. However, experiments [11–16] proved that $j_{c\|}$ and $j_{c\perp}$ ($j_{c\perp}$ is the transverse critical current density) are of the same order of magnitude, and $E_\|$ differs from zero. In order to explain these phenomena, a model of flux-line cutting was proposed [17–20] (the macroscopic theory of flux-line cutting was formulated in [21–27]). According to this model, nonparallel external magnetic fields penetrate a superconductor through mutual cutting of the flux lines formed by these fields, followed by their cross restoration. As a result, finite $j_{c\|}$ and $E_\|$ are formed. In such a case, the local current–voltage characteristic (IVC) connecting the electric field **E** and the current density **j** is strongly anisotropic relative to the magnetic induction vector **B**. All that has been said above is also applicable to a Josephson medium for $\beta \ll 1$.

On the other hand, if $\beta \gg 1$, the local IVC does not depend on the angle between the current and the magnetic field and is isotropic (see [9]). Thus, if we can establish experimentally whether or not the IVC is isotropic, we can distinguish between the continual ($\beta \ll 1$) and discrete ($\beta \gg 1$) cases.

2.1. Continual Approximation in the Theory of Josephson Medium

In this subsection, we will show how parameter β appears and derive an expression for the IVC in the flux-line cutting model for $\beta \ll 1$.

If a granule has a phase φ, the current between the ith and jth granules is given by

$$j_{ij} = -4eJ_{ij}\sin(\varphi_i - \varphi_j + 2eA_{ij}),$$
$$A_{ij} = \int_{\mathbf{r}_i}^{\mathbf{r}_j} \mathbf{A}\,d\mathbf{l}, \tag{3}$$

where J_{ij} is the overlap integral and **A** is the vector potential.

We assume that $|\varphi_i - \varphi_j| \ll 1$ and $|eA_{ij}| \ll 1$. Putting

$$\varphi_i - \varphi_j = -\int_{\mathbf{r}_i}^{\mathbf{r}_j} \nabla\varphi\,d\mathbf{l}, \quad J_{ij} = Ja^2, \tag{4}$$

and neglecting fluctuations, we arrive at the Londons equation

$$\mathbf{j}(\mathbf{r}) = \frac{4e}{a}J(\nabla\varphi - 2e\mathbf{A}). \tag{5}$$

The solution of Maxwell's equations

$$\frac{\partial\mathbf{B}}{\partial t} = -\operatorname{curl}\mathbf{E}, \quad 4\pi\mathbf{j} = \operatorname{curl}\mathbf{H} \tag{6}$$

and Londons equation (5) taking into account the fact that the induction **B** for an effective medium is connected with **H** through the simple relation

$$\mathbf{B} = \mu_{\text{eff}}\mathbf{H} \tag{7}$$

leads to the conventional Meissner effect with the effective penetration depth

$$\lambda_{\text{eff}}^2 = \frac{a}{32\pi Je^2\mu_{\text{eff}}} = \frac{\Phi_0}{8\pi^2 a\mu_{\text{eff}}j_c},$$
$$j_c = \frac{4e}{a^2}J, \quad \Phi_0 = \frac{\pi}{e}, \tag{8}$$

where a is the characteristic size of granules and Φ_0 is the magnetic flux quantum.

The obtained expression holds for $\lambda_{\text{eff}} \gg a$. Hence, we can easily obtain the following expression for the dimensionless parameter β:

$$\frac{\lambda_{\text{eff}}^3}{a^2} = \frac{\Phi_0}{8\pi^2\mu_{\text{eff}}j_c a^3} = \frac{1}{\beta},$$
$$\beta = 8\pi^2\mu_{\text{eff}}\frac{j_c a^2}{\Phi_0}. \tag{9}$$

Parameter β is very important since it characterizes the number of flux quanta per elementary loop formed

by adjacent granules. For $\beta \ll 1$ ($\lambda_{\text{eff}} \gg a$), this means that each flux line embraces many granules, and such a medium can be regarded as a conventional type II superconductor. Naturally, the theory of flux-line cutting is applicable in this case. For $\beta \gg 1$, however, $\lambda_{\text{eff}} \ll a$, and the theory presented above obviously becomes inapplicable. (For the sake of simplicity, we will henceforth assume that $\mu_{\text{eff}} = 1$.)

In the electrodynamics of continuous media, a description of electromagnetic field propagation requires, in addition to two Maxwell's equations (6), an explicit expression for $\mathbf{E}(\mathbf{j}, \mathbf{B})$. Following [21–27], we will write the expression for the IVC taking into account the pinning and flux-line cutting model. In this model, the system is strongly anisotropic, and the local IVC splits into two individual equations for longitudinal and transverse (relative to $\mathbf{B}$) components of the current and electric field. We put

$$\begin{gathered}\mathbf{n} = \frac{\mathbf{B}}{B}, \quad E_\parallel = \mathbf{E}\cdot\mathbf{n}, \quad \mathbf{E}_\perp = \mathbf{n}\times[\mathbf{E}\times\mathbf{n}], \\ j_\parallel = \mathbf{j}\cdot\mathbf{n}, \quad \mathbf{j}_\perp = \mathbf{n}\times[\mathbf{j}\times\mathbf{n}], \\ \mathbf{E} = \mathbf{n}E_\parallel + \mathbf{E}_\perp, \quad \mathbf{j} = \mathbf{n}j_\parallel + \mathbf{j}_\perp.\end{gathered} \tag{10}$$

Since we have only one preferred direction ($\mathbf{j}_\perp$) in a plane perpendicular to $\mathbf{B}$, the IVC in the flux-line cutting model has the form

$$\begin{gathered}E_\parallel = 0, \quad |j_\parallel| < j_{c\parallel}, \\ E_\parallel = \rho_\parallel(j_\parallel - j_{c\parallel}\operatorname{sgn} j_\parallel), \quad |j_\parallel| > j_{c\parallel}, \\ \mathbf{E}_\perp = E_\perp\frac{\mathbf{j}_\perp}{j_\perp}, \quad j_\perp = |\mathbf{j}_\perp|, \\ E_\perp = 0, \quad j_\perp < j_{c\perp}, \\ E_\perp = \rho_\perp(j_\perp - j_{c\perp}), \quad j_\perp > j_{c\perp}.\end{gathered} \tag{11}$$

Here, $\rho_{\parallel,\perp}$ are the longitudinal and transverse resistivities in the supercritical ($j_{\parallel,\perp} > j_{c\parallel,\perp}$) state and sgn x is the signum function.

Using (10) and (11), we obtain the final expression for the IVC:

$$\mathbf{E} = \mathbf{n}E_\parallel(j_\parallel) + \frac{\mathbf{j}_\perp}{j_\perp}E_\perp(j_\perp). \tag{12}$$

Maxwell's equations (6) and the IVC (12) form a complete system of equations. It should be noted that $\rho_{\parallel,\perp}$ and $j_{\parallel,\perp}$ are generally functions of the modulus of $\mathbf{B}$.

2.2. Equations of Critical State for the Flux-Line Cutting Model in Planar Geometry

Since we used samples in the form of a thin disk in our experiments, the subsequent analysis will be carried out in the planar geometry.

Let us consider an infinitely large slab in the yz plane, having a thickness d along the x-axis. If the external field $\mathbf{H}$ is parallel to the yz plane, the quantities $\mathbf{B}$, $\mathbf{E}$, and $\mathbf{j}$ also lie in this plane and depend only on x. In this geometry, Maxwell's equations have the form

$$\frac{\partial\mathbf{B}}{\partial t} = -\mathbf{e}_x\times\frac{\partial\mathbf{E}}{\partial x}, \quad 4\pi\mathbf{j} = \mathbf{e}_x\times\frac{\partial\mathbf{B}}{\partial x}. \tag{13}$$

We now introduce the unit vectors $\mathbf{n}(x)$ and $\mathbf{m}(x)$, which are parallel and perpendicular to vector $\mathbf{B}(x)$:

$$\begin{gathered}\mathbf{n}(x) = \mathbf{e}_z\cos\alpha(x) + \mathbf{e}_y\sin\alpha(x), \\ \mathbf{m}(x) = -\mathbf{e}_x\times\mathbf{n}(x) = -\mathbf{e}_z\sin\alpha(x) + \mathbf{e}_y\cos\alpha(x),\end{gathered} \tag{14}$$

where α is the angle between $\mathbf{B}$ and the z axis. We define $E_\parallel$, $E_\perp$, $j_\parallel$, and $j_\perp$ as

$$\mathbf{E} = E_\parallel\mathbf{n} + E_\perp\mathbf{m}, \quad \mathbf{j} = j_\parallel\mathbf{n} + j_\perp\mathbf{m}. \tag{15}$$

In the planar geometry, $E_\perp$ and $j_\perp$ become scalars. It is convenient to write Eqs. (13) in terms of B, α, $E_{\perp,\parallel}$, and $j_{\perp,\parallel}$:

$$\begin{gathered}4\pi j_\parallel = B\frac{\partial\alpha}{\partial x}, \quad 4\pi j_\perp = -\frac{\partial B}{\partial x}, \\ \frac{\partial B}{\partial t} = -E_\parallel\frac{\partial\alpha}{\partial x} - \frac{\partial E_\perp}{\partial x}, \\ B\frac{\partial\alpha}{\partial t} = -E_\perp\frac{\partial\alpha}{\partial x} + \frac{\partial E_\parallel}{\partial x}.\end{gathered} \tag{16}$$

It can be seen from these equations that the longitudinal current determines the change in the direction of **B,** while the transverse current determined the change in its magnitude.

Equations (11) for the IVC in the flux-line cutting model in this case assume the form

$$\begin{gathered}E_\parallel = 0, \quad |j_\parallel| < j_{c\parallel}, \\ E_\parallel = \rho_\parallel(j_\parallel - j_{c\parallel}\operatorname{sgn} j_\parallel), \quad |j_\parallel| > j_{c\parallel}, \\ E_\perp = 0, \quad |j_\perp| < j_\perp, \\ E_\perp = \rho_\perp(j_\perp - j_{c\perp}\operatorname{sgn} j_\perp), \quad |j_\perp| > j_{c\perp}.\end{gathered} \tag{17}$$

It is often more convenient to use the dependence $j(E)$ rather than $E(j)$. In this case, using the fact that $\sigma_{\perp,\parallel} = 1/\rho_{\perp,\parallel}$, we obtain

$$\begin{gathered}j_\perp = j_{c\perp}\operatorname{sgn} E_\perp + \sigma_\perp E_\perp, \\ j_\parallel = j_{c\parallel}\operatorname{sgn} E_\parallel + \sigma_\parallel E_\parallel.\end{gathered} \tag{18}$$

If the external field varies quite slowly, it can be seen from Eqs. (16) that $|E_\parallel|$ and $|E_\perp|$ tend to zero, and we can assume that $j_\parallel$ and $j_\perp$ are close to their critical values. This gives

$$j_\parallel = j_{c\parallel}\operatorname{sgn} E_\parallel, \quad j_\perp = j_{c\perp}\operatorname{sgn} E_\perp, \tag{19}$$

which are equations of critical state for longitudinal and transverse currents defined for $E_{\parallel,\perp} \longrightarrow 0$. If the magnetic field variation ceases, the value of $\mathbf{E}$ vanishes, but the currents do not vanish and are given by

$$j_\parallel = \pm j_{c\parallel}, \quad j_\perp = \pm j_{c\perp}, \tag{20}$$

the plus or minus sign depending on the sign possessed by $E_{\perp,\parallel}$ before vanishing. It is this two-valuedness of the current that is responsible for the hysteresis of the critical state.

Thus, Eqs. (20) are generalizations of the conventional equations for the critical state in the case of pinning and flux-line cutting model, when both longitudinal and transverse critical currents are present.

2.3. Equations of Critical State for $\beta \gg 1$. *Model of Isotropic Local IVC*

The physical meaning of the condition $\beta \gg 1$ is that several magnetic flux quanta are embraced by an elementary loop ($\beta \sim \Phi/\Phi_0$). It was proved earlier [7] that the local IVC in this case does not depend on the angle between **j** and **B**. We considered in [7] the model with a cubic symmetry, and hence the local IVC also possessed the same symmetry. Naturally, ceramics do not possess a cubic symmetry and must have a spherical symmetry. We will assume that the local IVC in this case can be presented in the form

$$\mathbf{E} = E(j)\frac{\mathbf{j}}{j},$$

$$E(j) = \begin{cases} 0, & j < j_c, \\ \rho(j - j_c), & j > j_c. \end{cases} \tag{21}$$

In other words, we have

$$\mathbf{j} = j_c\frac{\mathbf{E}}{E} + \sigma\mathbf{E}, \quad \sigma = \frac{1}{\rho}. \tag{22}$$

A local IVC in the form (21) or (22) will be referred to as an isotropic IVC. In this case, the equation of the critical state for small E has the form

$$\mathbf{j} = j_c\frac{\mathbf{E}}{E}. \tag{23}$$

Note that this is the same model as that proposed by Bean [28], but in a different language.

The expressions for IVC in the isotropic model and in the flux-line cutting model differ radically. First, an isotropic IVC does not depend on the angle between **j** and **B**. Second, in Eq. (23) we have

$$j^2 = j_\parallel^2 + j_\perp^2 = j_c^2, \tag{24}$$

and $j_\parallel^2$ and $j_\perp^2$ are not fixed separately, while for the IVC in the flux-line cutting model, these quantities remain unchanged.

In the case of a slab, we derive from (23)

$$j_\parallel = j_c\frac{E_\parallel}{\sqrt{E_\parallel^2 + E_\perp^2}}, \quad j_\perp = j_c\frac{E_\perp}{\sqrt{E_\parallel^2 + E_\perp^2}}. \tag{25}$$

At first sight, the difference between expressions (24) and (19) is insignificant, but actually this is not so since $j_\parallel$ and $j_\perp$ are now determined not only by their "intrinsic" fields, $E_\parallel$ and $E_\perp$, but also by "alien" fields in contrast to the flux-line cutting model. In simple cases, the solutions in the two models are very close, but in more complex situations these solutions are completely different.

In the following subsections, we will consider the penetration of a linearly polarized varying field of amplitude h_0 in the presence of a constant field **H** for the anisotropic and isotropic models. In all cases, we will assume that, first, $h_0 \ll H$, and second, the constant field is directed strictly along the z axis, while the varying field $h(t) = h_0\cos\omega t$ lies in the zy plane at an angle γ to **H**.Thus, for $x = 0$, we have

$$\begin{aligned} \mathbf{H}(0, t) &= H\mathbf{e}_z + \mathbf{h}(t), \quad h_0/H \ll 1, \\ \mathbf{h} &= h_z\mathbf{e}_z + h_y\mathbf{e}_y, \\ h_z &= h_0\cos\gamma\cos\omega t, \\ h_y &= h_0\sin\gamma\cos\omega t. \end{aligned} \tag{26}$$

2.4. Linearly Polarized Varying Magnetic Field in the Presence of a Strong Constant Field in the Flux-Line Cutting Model

Field $H\mathbf{e}_z$ penetrates over a distance $x_0 = H/4\pi j_{c\perp}$, while field $\mathbf{h}(t)$ penetrates only to a depth of the order of $h_0/(4\pi j_{c1}) = x_1 \ll x_0$, where j_{c1} has the meaning of $j_{c\parallel,\perp}$ in the flux-line cutting model and of j_c in the isotropic model. Obviously, for $x < x_1$, we have

$$\mathbf{H}(x, t) = H\mathbf{e}_z + \mathbf{h}(x, t). \tag{27}$$

Substituting this equation into Maxwell's equations (16) in the planar geometry, we take into account the IVC in the flux-line cutting model and linearize in $\mathbf{h}(x, t)$. Here, we assume that $j_{c\perp}$ is a function of the total field **H**. Linearization in **h** is reduced to the choice of the longitudinal direction along the z-axis and of the transverse direction along the y-axis. Generally speaking, we must use Eqs. (16) and carry out linearization in these equations. It can be proved, however, that the result will be the same. Thus, we arrive at the two pairs of equations

$$\begin{aligned} \frac{\partial E_z}{\partial x} &= \frac{\partial h_y}{\partial t}, \\ \frac{1}{4\pi}\frac{\partial h_y}{\partial x} &= j_{c\parallel}(H)\operatorname{sgn}E_z, \\ \frac{\partial E_y}{\partial x} &= -\frac{\partial h_z}{\partial t}, \\ \frac{1}{4\pi}\frac{\partial h_z}{\partial x} &= -j_{c\perp}(H_z)\operatorname{sgn}E_y, \\ H_z &= H + h_z. \end{aligned} \tag{28}$$

In these equations, we retain only the dependence on h_z in $j_{c\perp}$. Indeed, we can write

$$\begin{aligned} j_{c\perp,\parallel}(H) &= j_{c\perp,\parallel}(H) + \frac{\mathbf{H}\cdot\mathbf{h}}{H}\left(\frac{\partial j_{c\perp,\parallel}}{\partial H}\right) \\ &= j_{c\perp,\parallel}(H) + h_z\left(\frac{\partial j_{c\perp,\parallel}}{\partial H}\right). \end{aligned} \tag{29}$$

Thus, we are left only with h_z in the first order in $\mathbf{h}$, but it can be easily proved that the term containing h_z in the expression for $j_{c\parallel}$ is insignificant and is hence omitted, but it leads to the emergence of even harmonics in $j_{c\perp}$ and is therefore retained in this expression.

It can be seen that system (28) contains two pairs of independent equations in h_y and h_z. This means that the fields h_y and h_z penetrate into the sample independently. The solutions of Eqs. (28) leads to the following expressions for the induction $\mathbf{h}(t) = \langle\mathbf{h}(x, t)\rangle_x$:

$$\begin{aligned} \mathbf{h}(t) &= h_z(t)\mathbf{e}_z + h_y(t)\mathbf{e}_y, \\ h_z(t) &= \sum_k a_{kz}\cos k\omega t + b_{kz}\sin k\omega t, \\ h_y(t) &= \sum_k a_{ky}\cos k\omega t + b_{ky}\sin k\omega t, \\ a_{1z} &= \frac{h_0^2\cos^2\gamma}{4\pi j_{c\perp}(H)d}, \quad a_{1y} = \frac{h_0^2\sin^2\gamma}{4\pi j_{c\parallel}(H)d}, \\ a_{2k+1,z} &= a_{2k+1,y} = 0, \quad k \geq 1, \\ b_{2k+1,z} &= -\frac{h_0^2\cos^2\gamma}{8\pi^2 j_{c\perp}(H)d}\frac{1}{(k^2-1/4)(k+3/2)}, \\ b_{2k+1,y} &= -\frac{h_0^2\sin^2\gamma}{8\pi^2 j_{c\parallel}(H)d}\frac{1}{(k^2-1/4)(k+3/2)}, \\ a_{2z} &= \frac{h_0^3\cos^3\gamma}{32\pi d}\left(\frac{\partial}{\partial H}\frac{1}{j_{c\perp}(H)}\right), \\ a_{2k,z} &= 0, \quad k \geq 2, \\ a_{2k,y} &= 0, \\ b_{2k,z} &= -\frac{h_0^3\cos^3\gamma}{16\pi^2 d}\frac{k}{(k^2-1/4)(k^2-9/4)} \\ &\quad\times\left(\frac{\partial}{\partial H}\frac{1}{j_{c\perp}(H)}\right), \\ b_{2k,y} &= 0. \end{aligned} \tag{30}$$

It can be seen from these equations that the z and y components oscillate independently with their j_c, and odd harmonics are proportional to $h_0^2\cos^2\gamma$ and $h_0^2\sin^2\gamma$, while even harmonics differ from zero only for the z component and are proportional to $h_0^3\cos^3\gamma$.

2.5. Linearly Polarized Varying Field in the Presence of a Strong Constant Field in the Isotropic Model

As in the previous case, we assume that the varying field is directed at an angle γ to the constant field, i.e., we put

$$\begin{aligned} \mathbf{h} &= h(t)\mathbf{n}_0, \quad h(t) = h_0\cos\omega t, \\ \mathbf{n}_0 &= \mathbf{e}_z\cos\gamma + \mathbf{e}_y\sin\gamma. \end{aligned} \tag{31}$$

Since there is no preferred direction in the constant field in the isotropic model, the direction is specified only by vector $\mathbf{n}_0$. This means that

$$\begin{aligned} H(x, t) &= H\mathbf{e}_z + \mathbf{h}(x, t), \quad \mathbf{h}(x, t) = h(x, t)\mathbf{n}_0, \\ \mathbf{E}(x, t) &= E(x, t)\mathbf{m}_0, \quad \mathbf{m}_0 = -\mathbf{e}_x\times\mathbf{n}_0. \end{aligned} \tag{32}$$

In this case, Maxwell's equations (13) in the planar geometry together with (23) lead to the following equations for h and E:

$$\frac{\partial E}{\partial x} = -\frac{\partial h}{\partial t}, \quad 4\pi j_c(B)\operatorname{sgn}E = -\frac{\partial h}{\partial x}. \tag{33}$$

Equations (33) are ordinary equations of a critical state. A difference appears only when we expand $j_c^{-1}(H)$ in h and calculate even harmonics:

$$\begin{aligned} \frac{1}{j_c(H)} &= \frac{1}{j_c(H)} + \mathbf{h}\left(\frac{\partial}{\partial\mathbf{H}}\frac{1}{j_c(H)}\right) = \frac{1}{j_c(H)} + h(\mathbf{n}_0\cdot\mathbf{e}_z) \\ &\quad\times\left(\frac{\partial}{\partial H}\frac{1}{j_c(H)}\right) = \frac{1}{j_c(H)} + h\cos\gamma\left(\frac{\partial}{\partial H}\frac{1}{j_c(H)}\right). \end{aligned} \tag{34}$$

Thus, in this expansion we have an extra $\cos\gamma$ as compared to the case when $\mathbf{h}_0$ and $\mathbf{H}$ are parallel. Ultimately, we obtain

$$\begin{aligned} \mathbf{h}(x, t) &= \sum_k(\mathbf{a}_k\cos k\omega t + \mathbf{b}_k\sin k\omega t), \\ \mathbf{a}_1 &= \frac{h_0^2}{4\pi j_c(H)d}\mathbf{n}_0, \\ a_{2k+1} &= 0, \quad k \geq 1, \\ \mathbf{b}_{2k+1} &= -\frac{h_0^2}{8\pi^2 j_c(H)d}\frac{1}{(k^2-1/4)(k+3/2)}\mathbf{n}_0, \\ \mathbf{a}_{2k} &= \frac{h_0^3\cos\gamma}{32\pi d}\mathbf{n}_0\left(\frac{\partial}{\partial H}\frac{1}{j_c(H)}\right), \end{aligned} \tag{35}$$

$$a_{2k} = 0, \quad k \geq 2,$$

$$\mathbf{b}_{2k} = -\frac{h_0^3 \cos\gamma}{16\pi^2 d} \frac{k}{(k^2 - 1/4)(k^2 - 9/4)} \mathbf{n}_0 \left(\frac{\partial}{\partial H} \frac{1}{j_c(H)} \right).$$

In this case, we obtain the following expressions for the z and y components:

$$a_{1z} = \frac{h_0^2}{4\pi j_c(H) d} \cos\gamma, \quad a_{1y} = \frac{h_0^2}{4\pi j_c(H) d} \sin\gamma,$$

$$a_{2k+1,z} = a_{2k+1,y} = 0, \quad k \geq 1,$$

$$b_{2k+1,z} = -\frac{h_0^2 \cos\gamma}{8\pi^2 j_c(H) d} \frac{1}{(k^2 - 1/4)(k + 3/2)},$$

$$b_{2k+1,y} = -\frac{h_0^2 \sin\gamma}{8\pi^2 j_c(H) d} \frac{1}{(k^2 - 1/4)(k + 3/2)},$$

$$a_{2z} = \frac{h_0^3 \cos^2\gamma}{32\pi d} \left(\frac{\partial}{\partial H} \frac{1}{j_c(H)} \right),$$

$$a_{2y} = \frac{h_0^3 \sin\gamma \cos\gamma}{32\pi d} \left(\frac{\partial}{\partial H} \frac{1}{j_c(H)} \right), \tag{36}$$

$$a_{2k,z} = a_{2k,y} = 0, \quad k \geq 2,$$

$$b_{2k,z} = -\frac{h_0^2 \cos^2\gamma}{16\pi^2 d} \frac{k}{(k^2 - 1/4)(k^2 - 9/4)} \left(\frac{\partial}{\partial H} \frac{1}{j_c(H)} \right),$$

$$b_{2k,y} = -\frac{h_0^2 \sin\gamma \cos\gamma}{16\pi^2 d} \frac{k}{(k^2 - 1/4)(k^2 - 9/4)} \left(\frac{\partial}{\partial H} \frac{1}{j_c(H)} \right).$$

Thus, it can be seen that the isotropic model strongly differs from the flux-line cutting model, first, in the form of the angular dependence of harmonics, and second, in that the even harmonics $a_{2k,y}$, $b_{2k,y}$, which are equal to zero in the flux-line cutting model, now differ from zero.

3. EXPERIMENT

The experimental part of this work was aimed at an analysis of the applicability of the concepts of low-field electrodynamics to polycrystalline superconductors of the molybdenum chalcogenide type and the effect of anisotropy induced by a constant magnetic field on the penetration of a varying magnetic field directed at an angle to the constant field in a superconductor. For this purpose, we studied the dependences of higher harmonics of induction on the amplitudes of the varying and constant magnetic fields and on the angle formed by these fields.

The experiments were made on polycrystalline samples of $SnMo_6S_8$ and $PbMo_6S_8$, having the shape of a disk of diameter 9.4 mm and thickness 3.4 mm. The sample was placed in a measuring coil that also served as a source of the varying magnetic field. The constant field was produced by an external solenoid. The linear and nonlinear responses were measured in the temperature range from 4.2 K to $T > T_c$ at frequencies varying from 10^3 to 10^5 Hz in the field interval $10^{-2} \leq h_0 \leq 1$ Oe and for $H \leq 20$ Oe. We studied the field and temperature dependences of the real and imaginary components of linear susceptibility (χ' and χ'') as well as the moduli of the higher harmonic amplitudes $c_n = (a_n^2 + b_n^2)^{1/2}$ ($n = 2, 3, \ldots$) by using the approach which was applied earlier for studying the low-field electrodynamics of granular HTSC materials (see, for example, [2]).

Let us first consider the results of analysis of the temperature and field dependences of the linear and nonlinear susceptibilities for collinear fields ($\gamma = 0$). In such a configuration, expressions (30) and (36) are obviously identical. Taking into account μ_{eff}, we obtain

$$a_0 = \mu_{\text{eff}} 2H, \quad a_1 = \frac{\mu_{\text{eff}} h_0^2}{4\pi j_c(H) d},$$

$$a_{2k+1} = 0, \quad k \geq 1,$$

$$b_{2k+1} = -\frac{\mu_{\text{eff}} h_0^2}{8\pi^2 j_c(H) d} \frac{1}{(k^2 - 1/4)(k + 3/2)}, \tag{37}$$

$$a_2 = \frac{\mu_{\text{eff}} h_0^3}{32\pi d} \left(\frac{\partial}{\partial H} \frac{1}{j_c(H)} \right), \quad a_{2k} = 0, \quad k \geq 2,$$

$$b_{2k} = \frac{\mu_{\text{eff}} h_0^3}{16\pi^2 d} \left(\frac{\partial}{\partial H} \frac{1}{j_c(H)} \right) \frac{k}{(k^2 - 1/4)(k^2 - 9/4)}.$$

It should be noted that in our experiments we studied in detail the temperature and field behavior of χ' and χ'', which are directly connected with a_1 and b_1, as well as the moduli of amplitudes of the third and fifth harmonics (c_3 and c_5). The values of a_1 and b_1 were determined from the measured values of the real and imaginary susceptibility components. Obviously, $\chi' = -1/4\pi$ in the limit $h_0 \longrightarrow 0$. Consequently, χ' can be presented as $\chi'(h_0) = -1/4\pi + \chi'_h(h_0)$. This gives

$$a_1 = (1 + 4\pi\chi')h_0 = 4\pi h_0 \chi'_h(h_0),$$
$$b_1 = 4\pi h_0 \chi''(h_0). \tag{38}$$

It follows from the temperature dependences of the real component of linear susceptibility (see Fig. 1) that the superconducting transition for the samples under investigation begins at a temperature typical of this class of compounds (see, e.g., [29]). The same figure shows the $\chi''(T)$ dependences having a peak typical of type II superconductors.

The spectrum of higher harmonics demonstrates the clearly manifested nonlinear properties of the polycrystals under investigation. For example, the spectrum obtained for $PbMo_6S_8$ (Fig. 2) indicates a weak

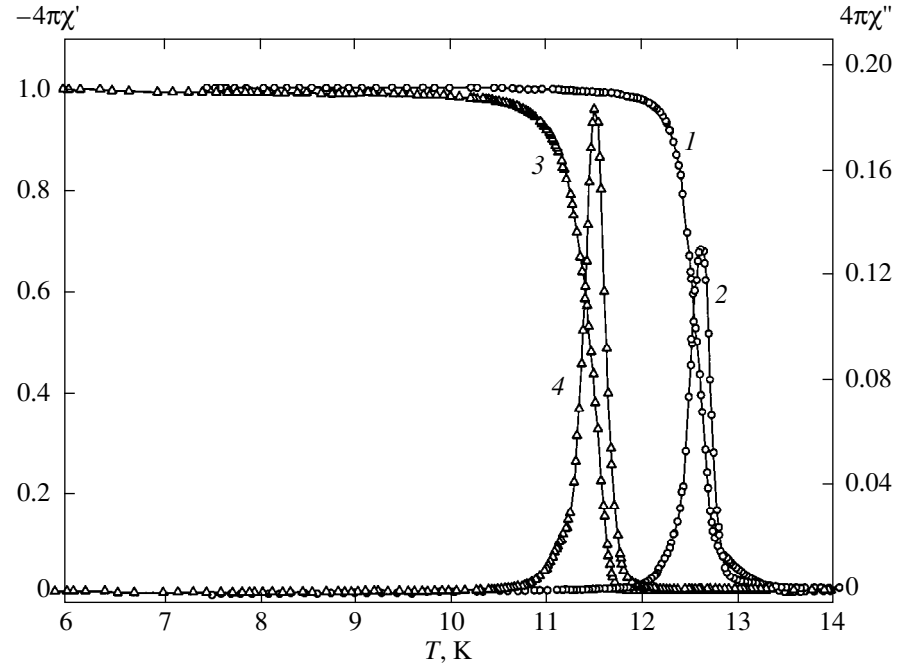

Fig. 1. Temperature dependences of the real (curves *1* and *3*) and imaginary (curves *2* and *4*) components of linear susceptibility for $SnMo_6S_8$ and $PbMo_6S_8$ samples, respectively; $H = 0$, $h_0 = 1$ Oe.

decrease in the amplitude of the harmonics upon an increase in their number. It should be noted that even harmonics were not observed for $SnMo_6S_8$ for $H \neq 0$ either.

Figure 3 shows the temperature dependences of the susceptibility $\chi_3 = c_3/h_0$ for various values of the amplitude of the varying field h_0, which are identical to the corresponding dependences for HTSC ceramics (see, e.g., [2]). A typical feature of these curves is the presence of two peaks. The high-temperature peak can be associated with the penetrating of the field into crystallites, while the low-frequency peak is associated with the penetration only into the Josephson medium formed by weak links between the crystallites (note that no high-temperature peak was observed on the $\chi_3(T)$ dependence for $PbMo_6S_8$). The position of the low-temperature peak is displaced towards lower temperatures upon an increase in h_0, which is associated with a transition from the mode in which the field penetrates to the middle of the sample to a mode in which the field penetrates only to a small depth of the sample. In the latter case, the condition of a weak field ($h_0 \ll H$) can be realized easily, and hence expression (37) can be used to verify the applicability of the conclusions of the critical state theory to the polycrystals under investigation. Henceforth, we will consider the experimental results only for this limiting case.

Figure 4 shows the dependences of a_1, b_1, c_3, and c_2 (for $H \neq 0$) on h_0 for $PbMo_6S_8$. It can be seen that in accordance with the results of the theory, odd harmonics exhibit a quadratic dependence on h_0, while the second harmonic exhibits a cubic dependence. Note that the scale of magnetic induction induced by the varying field (from 0.01 mOe to 1 Oe) varies from 1 μG to 0.1 G. Similar dependences for a_1, b_1, c_3, and c_5 were also obtained for $SnMo_6S_8$.

Equation (37) contains two independent quantities μ_{eff} and j_0, which were not determined independently in our experiments. Consequently, in order to compare the theoretical and experimental results, it is expedient to consider the ratios a_1/b_1, b_1/c_3, etc., which must be independent of μ_{eff} and j_0 in accordance with (37). It follows from the table that these experimentally determined ratios are in good agreement with the theoretical predictions.

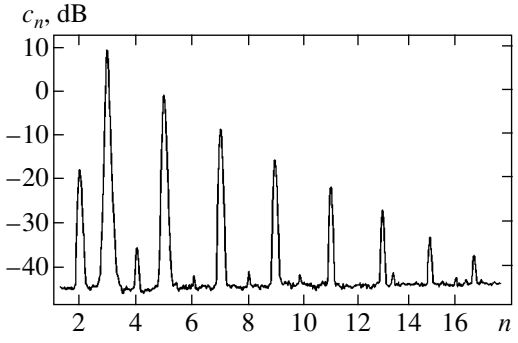

Fig. 2. Spectrum of higher harmonics ($PbMo_6S_8$); $H = 6$ Oe, $h_0 = 0.8$ Oe, and $T = 11$ K.

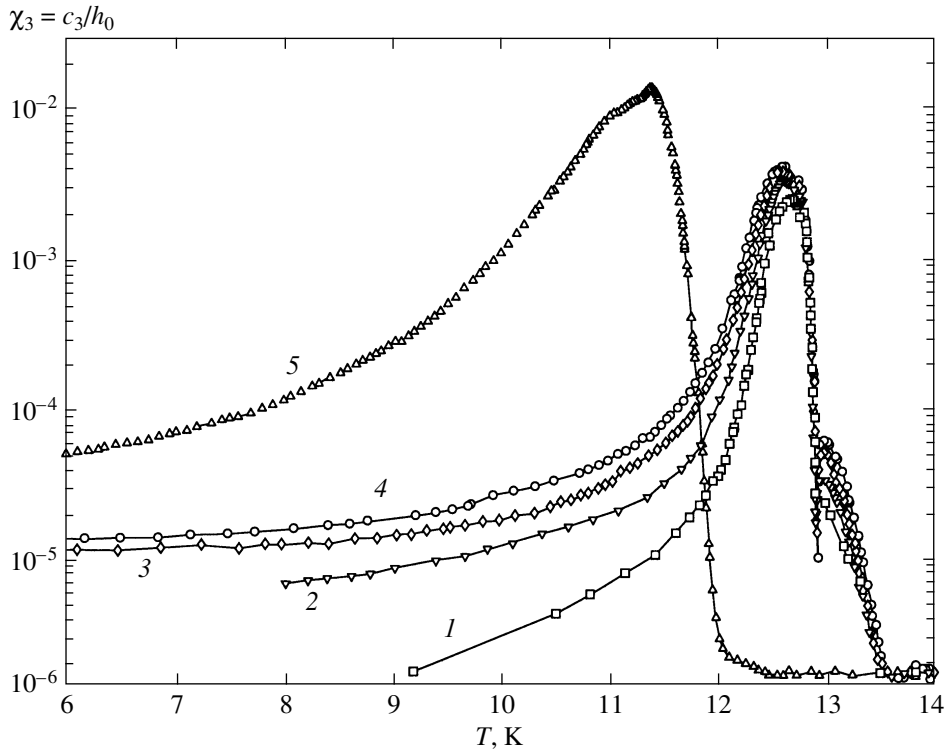

Fig. 3. Temperature dependences of nonlinear susceptibility c_3/h_0, obtained in zero constant magnetic field H for various amplitudes of the varying field h_0: (curve *1*) 0.1, (curve *2*) 0.3, (curve *3*) 0.6, (curve *4*) 0.9 Oe ($SnMo_6S_8$) and (curve *5*) 0.9 Oe ($PbMo_6S_8$).

Analyzing the behavior of even harmonics proportional to $(\partial/\partial H)(1/j_c(H))$ and odd harmonics proportional to $1/j_c(H)$ in a constant field for h_0 = const and using expressions (37), we can determine self-consistently the form of function $j_c(H)$. This is how the explicit form of this function was determined for granular HTSC [see (4)]. However, the situation turned out to be more complicated for the dependence on the constant magnetic field in $PbMo_6S_8$ (Fig. 5). It can be seen from the figure that starting from $H \approx 1$ Oe, c_3 is practically independent of H. Proceeding from formulas (37), we should expect the absence of the second harmonic (as in the case of $SnMo_6S_8$). However, in the experiments with $PbMo_6S_8$ we observed the second harmonic whose magnitude was considerably larger than the value predicted by (37). Using relations (37), we can write the following expression for the ratio c_2/c_3:

$$\frac{c_2}{c_3} \approx 4h_0 j_c(H)\left(\frac{\partial}{\partial H}\frac{1}{j_c(H)}\right), \tag{39}$$

which does not depend on the specific form of function $j_c(H)$, i.e., is independent of the model. If we calculate the derivative in (39) proceeding from the experimental data on the third harmonic [$c_3 \propto 1/j_c(H)$], it turns out that the ratio c_2/c_3 must be two or three orders of magnitude smaller than the observed value equal approximately to 0.1. In this case, our sensitivity (not worse than 0.1 μV) is insufficient for observing the second harmonic (precisely this situation was observed for $SnMo_6S_8$). However, in spite of such a discrepancy, all the remaining (field and angular) dependences are in excellent agreement with the theory.

Ratio of amplitudes of neighboring harmonics for $SnMo_6S_8$

	Ratio		
	a_1/b_1	b_1/c_3	c_3/c_5
Theory	2.3	5	7
Experiment	2.2 ± 0.2	3.5 ± 0.3	9.5 ± 1.2

The next series of experiments was carried out with a linearly polarized varying field forming an angle γ with the constant field. These experiments were aimed at determining the version (the flux-line cutting model or the isotropic model) realized in the polycrystalline superconductors under investigation. It can be seen from formulas (30) and (36) that in the former case, $c_{3z} \propto \cos^2\gamma$, $c_{3y} \propto \sin^2\gamma$, and $c_{2z} \propto \cos^3\gamma$, $c_{2y} = 0$, while in the latter case $c_{3z} \propto \cos\gamma$, $c_{3y} \propto \sin\gamma$, and $c_{2z} \propto \cos^2\gamma$,

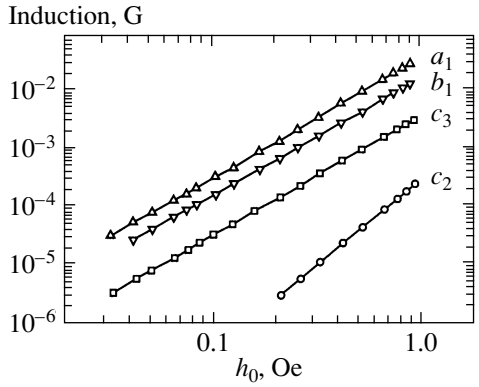

Fig. 4. Dependence of the quantities a_1, b_1, c_3, and c_2 on h_0 for $H = 6$ Oe for $PbMo_6S_8$ at $T = 10.5$ K.

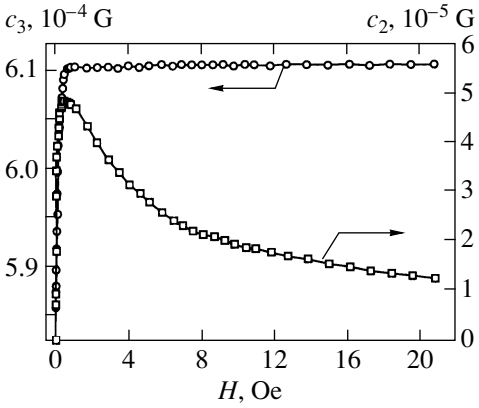

Fig. 5. Dependence of c_3 and c_2 on H for $h_0 = 0.4$ Oe for $PbMo_6S_8$ at $T = 10.5$ K.

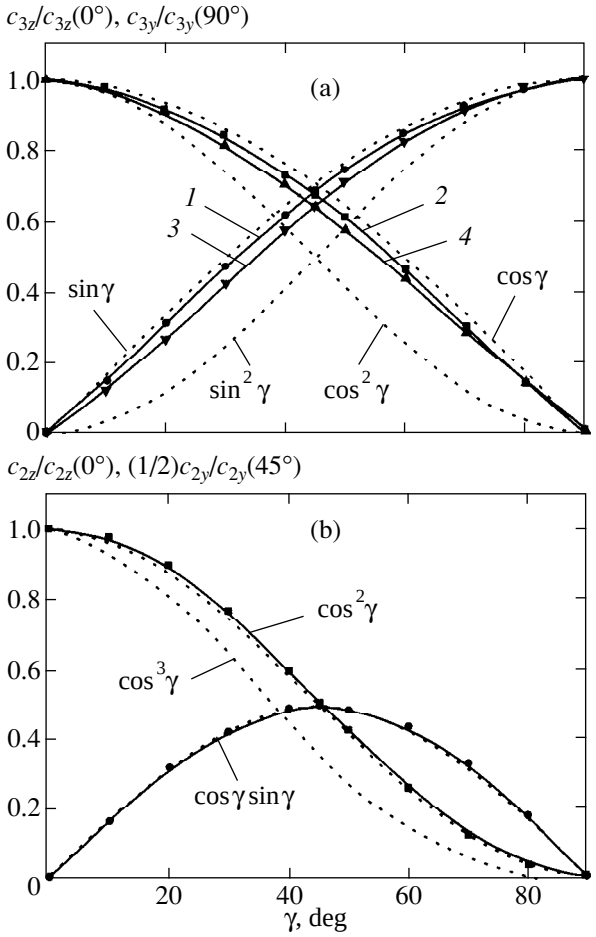

Fig. 6. (a) Angular dependences of c_{3z} and c_{3y} normalized to their values for $\gamma = 0$ and $\gamma = 90°$, respectively. Curves *1* and *2* correspond to $PbMo_6S_8$, $H = 8$ Oe, $h_0 = 0.6$ Oe, and $T = 10.5$ K, while curves *3* and *4* correspond to $SnMo_6S_8$, $H = 20$ Oe, $h_0 = 0.6$ Oe, and $T = 11.5$ K. (b) Angular dependences $c_{2z}/c_{2z}(0)$ and $((1/2)c_{2y}/c_{2y}(45°)$ for $PbMo_6S_8$.

$c_{2y} \propto \sin\gamma\cos\gamma$. Figure 6 shows that the angular dependences for harmonics obtained by us unambiguously correspond to the isotropic model. Besides, the fact that the second harmonic c_{2y} differs from zero for an arbitrary γ (except $\gamma = 0$ and $\gamma = \pi/2$) also speaks in favor of this model.

4. CONCLUSIONS

Thus, the results of our experiments show that a polycrystalline superconductor, as well as an HTSC ceramics material, behaves as a standard Josephson medium in weak fields. Such systems display irreversible and nonlinear effects typical of low-field electrodynamics in Josephson media.

ACKNOWLEDGMENTS

The authors are grateful to A. V. Mitin (Kapitsa Institute of Physical Problems, Russian Academy of Sciences) for fruitful discussions of the results and for providing the samples.

This work was supported financially by the Scientific Council on "Superconductivity" Problem, program "Current Trends in Physics of Condensed Media" (project no. 96021 "Profile"), subprogram "Statistical Physics," of the State Program in Science and Technology "Physics of Quantum and Wave Processes" (project VIII-3), and also by the Russian Foundation for Basic Research (project no. 00-02-16729).

REFERENCES

1. J. R. Clem, Physica C (Amsterdam) **153–155**, 50 (1988).
2. S. L. Ginzburg, V. P. Khavronin, G. Yu. Logvinova, *et al.*, Physica C (Amsterdam) **174**, 109 (1991).
3. C. P. Bean, Rev. Mod. Phys. **36**, 31 (1964).
4. Y. B. Kim, C. F. Hempstead, and A. R. Strnad, Phys. Rev. **131**, 2486 (1963).

5. P. W. Anderson and Y. B. Kim, Rev. Mod. Phys. **36**, 39 (1964).
6. I. D. Luzyanin, S. L. Ginzburg, V. P. Khavronin, and G. Yu. Logvinova, Phys. Lett. A **141**, 85 (1989).
7. S. L. Ginzburg, Zh. Éksp. Teor. Fiz. **106**, 607 (1994) [JETP **79**, 334 (1994)].
8. P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. A **38**, 364 (1988).
9. S. L. Ginzburg, V. P. Khavronin, and I. D. Luzyanin, Supercond. Sci. Technol. **11**, 255 (1998).
10. A. M. Campbell and J. E. Evets, *Critical Currents in Superconductors* (Taylor and Francis, London, 1972), p. 243.
11. R. Boyer and M. A. R. LeBlanc, Solid State Commun. **24**, 261 (1977).
12. G. Fillion, R. Gauthier, and M. A. R. LeBlanc, Phys. Rev. Lett. **43**, 86 (1979).
13. R. Boyer, G. Fillion, and M. A. R. LeBlanc, J. Appl. Phys. **51**, 1692 (1980).
14. J. R. Cave and M. A. R. LeBlanc, J. Appl. Phys. **53**, 1631 (1982).
15. M. A. R. LeBlanc and J. R. Lorrain, J. Appl. Phys. **55**, 4035 (1984).
16. M. A. R. LeBlanc, D. LeBlanc, A. Golebiowski, and G. Fillion, Phys. Rev. Lett. **66**, 3309 (1991).
17. E. H. Brandt, J. R. Clem, and D. G. Walmsley, J. Low Temp. Phys. **37**, 43 (1979).
18. J. R. Clem, J. Low Temp. Phys. **38**, 353 (1980).
19. E. H. Brandt, J. Low Temp. Phys. **39**, 41 (1980).
20. J. R. Clem and S. Yeh, J. Low Temp. Phys. **39**, 173 (1980).
21. J. R. Clem, Phys. Rev. B **26**, 2463 (1982).
22. J. R. Clem and A. Pérez-González, Phys. Rev. B **30**, 5041 (1984).
23. A. Pérez-González and J. R. Clem, Phys. Rev. B **31**, 7048 (1985).
24. A. Pérez-González and J. R. Clem, Phys. Rev. B **32**, 2909 (1985).
25. J. R. Clem and A. Pérez-González, Phys. Rev. B **33**, 1601 (1986).
26. A. Pérez-González and J. R. Clem, Phys. Rev. B **42**, 4100 (1990).
27. A. Pérez-González and J. R. Clem, Phys. Rev. B **43**, 7792 (1991).
28. C. P. Bean, J. Appl. Phys. **41**, 2482 (1970).
29. N. E. Alekseevskiĭ, A. V. Mitin, and E. P. Khlybov, Zh. Éksp. Teor. Fiz. **82**, 927 (1982) [Sov. Phys. JETP **55**, 543 (1983)].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 169–173.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 193–198.

SOLIDS
Electronic Properties

Neutral and Charged Magnetic Excitons in Finite-Width Quantum Rings

A. V. Chaplik

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia
CeNS und Sektion Physik, LMU, München, Deutschland
e-mail: chaplik@isp.nsc.ru

Received August 4, 2000

Abstract—The behavior of electrons, excitons, and charged electron–hole complexes in quantum rings is studied taking into account radial vibrations in a magnetic field. The diamagnetic shift of the exciton luminescence line is found to be positive for a neutral exciton and negative for a trion and all other charged complexes. It is shown that the magnetic-field dependent component of the electron ring energy does not depend on the electron–electron interaction. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The effect of the self-organization of quantum rings in the InAs/GaAs system has been reported recently [1, 2]. Under certain conditions, the evolution of the ensemble of quantum points of InAs located on the GaAs surface as a result of the diffusion of the material to the edges of the points leads to the formation of objects resembling volcano craters. The region formed at the center is free of InAs and has a characteristic diameter of 20 nm. The outer diameter of a "crater" is estimated in [1, 2] as 60 to 120 nm. Thus, an ensemble of rings is formed, their radius being comparable with the effective Bohr radius of an electron in InAs. Such a structure permits the deposition of a field electrode (shutter), owing to which the population of rings by electrons can be controlled. Warburton *et al.* [3] reported a successive population of rings by up to five electrons. In the same publication, the exciton luminescence of charged rings was investigated; i.e., recombination-induced emission by the systems from a neutral exciton X_0 to a quintuply negatively charged complex X^{5-} was observed. The behavior of these systems in a magnetic field is of considerable interest. The present article deals with a theoretical analysis of this problem.

In order to take into account the radial degree of freedom of electrons and holes, we will use the parabolic model of a ring proposed by Chakraborty and Pietilähen [4]. The potential energy of a particle has the form

$$U_i(\rho) = \frac{m_i\omega_i^2(\rho - R_0)^2}{2}, \quad i = e, h, \qquad (1)$$

where the subscripts e and h mark an electron or a hole, and R_0 is the electron (and hole) radius of the ring. Thus, the position of the minimum is assumed to be the same for electrons and holes.[1] In the present work, we will be interested in orbital (diamagnetic) effects, leaving aside the contribution from the spin degree of freedom. The latter depends on the specific form of the system through the g-factors of electrons and holes, and a transition, say, from a quantum point to the ring is not critical for this contribution. On the contrary, the orbital motion is sensitive to a change in the object topology (the Aharonov–Bohm effect), and we will be interested here in such an effect of a magnetic field on the system behavior.

According to estimates of the above parameters of a quantum ring, the radial motion is characterized by a much smaller amplitude as compared to the azimuthal motion, for which the corresponding size is just $2\pi R_0$. In this sense, the ring is narrow, and we will assume that the condition $\hbar\omega_i \gg W_i$ is satisfied, where $W_i = \hbar^2/2m_iR_0^2$ is the rotational motion quantum (an ideal one-dimensional ring corresponds to the limit $W_i/\hbar\omega_i \longrightarrow 0$). In this case, the problem can be solved in the approximation which is well known from molecular theory: first, the rotational energy levels are determined for fixed nuclei (in our case, for fixed radial coordinates ρ_e and ρ_h of an electron and a hole), after which the small vibrations of nuclei are taken into account. In this case, effects of the type of a nonrigid rotator and the interaction between vibration and rotation are observed. Similar corrections to energy levels also appear in the problem on excitons in a finite-width quantum ring.

[1] The absence of a noticeable spatial separation of electrons and holes follows from the fact that the recombination manifested in exciton luminescence is quite effective.

1063-7761/01/9201-0169$21.00

2. ELECTRON RING

We begin our analysis with a system containing N electrons in a ring. This is the final state of the initial system ($Ne + h$) after recombination, and hence the magnetic-field dependence of its energy makes a significant contribution to the diamagnetic shift of the exciton luminescence line. In an ideal one-dimensional ring, the position of electrons is defined by azimuthal angles φ_j ($j = 1, \ldots, N$). It is convenient to introduce the Jacobi variables in accordance with the formulas

$$\varphi_c = \frac{\varphi_1 + \ldots + \varphi_N}{N}, \quad \psi_1 = \frac{\varphi_1 - \varphi_2}{\sqrt{1 \cdot 2}},$$
$$\psi_2 = \frac{\varphi_1 + \varphi_2 - 2\varphi_3}{\sqrt{2 \cdot 3}}, \ldots, \tag{2}$$
$$\psi_{N-1} = \frac{\varphi_1 + \varphi_2 + \ldots - (N-1)\varphi_N}{\sqrt{N(N-1)}}.$$

In new variables, The Hamiltonian of the system taking into account the magnetic field B normal to the ring assumes the form

$$\hat{H}_{1D} = \frac{W_{e0}}{N}\left[-i\frac{\partial}{\partial\varphi_c} + N\tilde{\Phi}(R_0)\right]^2 - W_{e0}\left(\frac{\partial^2}{\partial\psi_1^2} + \ldots + \frac{\partial^2}{\partial\psi_{N-1}^2}\right) + V_{ee}(\psi_1, \ldots, \psi_{N-1}), \tag{3}$$

where $W_{e0} = \hbar^2/2m_eR_0^2$ is the rotational quantum, $\tilde{\Phi} = \pi BR_0^2/\Phi_0$ ($\Phi_0 = hc/e$ is the magnetic flux quantum), and V_{ee} is the electron–electron interaction energy, which is naturally independent of φ_c. Thus, the magnetic-field dependent component of the system energy is defined by the first term in (3):

$$E(B) = \frac{W_e}{N}[J + N\tilde{\Phi}(R_0)]^2, \tag{4}$$

where $J = 0, \pm1, \pm2, \ldots$ is the total angular momentum of the system.

It is important that the value of $E(B)$ is independent of the electron–electron interaction. For this reason, the diamagnetic shift, for instance, for an interacting system with $J = 0$ (the ground state in a weak field) is equal just to the n-fold shift of a single particle:

$$E(B)|_{J=0} = NW_{e0}\tilde{\Phi}^2 = N\frac{(\hbar\Omega_e)^2}{16W_{e0}}, \tag{5}$$

where Ω_e is the cyclotron frequency of an electron.

Let us now take into account the radial degrees of freedom and write the Hamiltonian in the polar coordinates φ_j and ρ_j:

$$\hat{H}_{2D} = \sum_j\left\{\hat{T}(\rho_j) + W_{ej}\left[-i\frac{\partial}{\partial\varphi_j} + \tilde{\Phi}(\rho_j)\right]^2 + U_e(\rho_j)\right\} + V_{ee}, \tag{6}$$

where $\hat{T}(\rho_j)$ is the kinetic energy operator for the radial motion, $W_{ej} = \hbar/2m_e\rho_j^2$, $\tilde{\Phi}(\rho_j) = \pi B\rho_j^2/\Phi_0$, $U_e(\rho_j)$ is defined by formula (1), and V_{ee} is now a function of all φ_j and ρ_j. The second term in (6) containing the magnetic field can be transformed as follows:

$$\sum_j W_{ej}\left[-i\frac{\partial}{\partial\varphi_j} + \tilde{\Phi}(\rho_j)\right]^2 = \sum_j W_{e0}\left[-i\frac{\partial}{\partial\varphi_j} + \tilde{\Phi}(R_0)\right]^2 + \sum_j(W_{e0} - W_{ej})\frac{\partial^2}{\partial\varphi_j^2} + W_{e0}\hat{\Phi}(R_0)\sum_j\delta\hat{\Phi}_j, \tag{7}$$

where $\partial\hat{\Phi}_j = \pi B(\rho_j^2 - R_0^2)/\Phi_0$. We now apply transformation (2) to the right-hand side of Eq. (7). Since φ_c does not appear explicitly in the Hamiltonian (V_{ee} depends only on pairwise difference $\varphi_j - \varphi_k$ in which φ_c does not appear), we can single out the "magnetic" component of the Hamiltonian, i.e.,

$$\Delta\hat{H}(B) = \frac{W_{e0}}{N}\left(-i\frac{\partial}{\partial\varphi_c} + N\tilde{\Phi}(R_0)\right)^2 + W_{e0}\tilde{\Phi}(R_0)\sum_j\delta\tilde{\Phi}_j, \tag{8}$$

which completely determines the diamagnetic shift of energy levels, the effects of the finite ring width appearing only in the second term on the right-hand side of Eq. (8). Introducing the vibrational coordinates $\xi_j = \rho_j - R_0$, we obtain the following expression for perturbation:

$$\Delta H_{\text{pert}}(B) = \pi W_{e0}\tilde{\Phi}(R_0)B\sum_j(2\xi_jR_0 + \xi_j^2). \tag{9}$$

It was mentioned in the Introduction that for $\hbar\omega_e \gg W_{e0}$, the corrections to energy associated with ΔH_{pert} can be calculated in perturbation theory. Strictly speaking, the unperturbed wave function of radial motion must include the interaction V_{ee} between particles. However, the Coulomb interaction strongly perturbs the azimuthal and weakly radial motion of particles in the approximation adopted by us. Indeed, expanding the paired interaction of particles separated by a distance of the order of R_0 from one another into a power series in

their radial coordinates ξ_i, we can easily obtain the estimates

$$\Delta\xi \sim l_0\left(\frac{R_y^*}{\hbar\omega_e}\right)^{1/2}\frac{W_{e0}}{\hbar\omega_e}, \quad (10)$$

where $l_0 = \sqrt{\hbar/m\omega_e}$ is the amplitudes of vibrations in potential (1) and R_y^* is the effective Rydberg energy, for the shift of the equilibrium position, and

$$\Delta\omega^2 \sim \omega_e^2\left(\frac{W_{e0}}{\hbar\omega_e}\right)^{3/2}\left(\frac{R_y^*}{\hbar\omega_e}\right)^{1/2} \quad (11)$$

for the correction to the square of frequency. For the ring parameters indicated at the beginning of the article, $R_y^* \sim \hbar\omega_e$, and hence both corrections are small in the parameter $W_{e0}/\hbar\omega_e$, and the vibrational component of diamagnetic shift can be calculated using the unperturbed vibrational functions. This gives

$$\Delta E_{\text{vibr}}(B) = \left[\frac{(\hbar\Omega_e)^2}{8\hbar\omega_e}\left(\nu_e + \frac{1}{2}\right) - \frac{(\hbar\Omega)^4}{64(\hbar\omega_e)^2 W_{e0}}\right]N, \quad (12)$$

where ν_e is the vibrational quantum number.

In this formula, we have also taken into account the second approximation in the linear term appearing in (9) (the contribution proportional to B^4). It can be seen that the vibrational component of diamagnetic shift is small as compared to the rotational component [formula (5)] in the same parameter $W_{e0}/\hbar\omega_e$.

3. NEUTRAL EXCITON X_0

The rotational component of diamagnetic shift in the energy of a neutral formation X_0 connected with the motion of an exciton as a whole obviously vanishes. For an ideal one-dimensional ring, there exists a contribution to the exciton binding energy ε, which is a periodic function of the magnetic flux and emerges as a result of the tunneling of an electron and a hole towards each other along the ring. In the strong coupling approximation ($a^* \ll 2\pi R_0$, where a^* is the effective Bohr radius) and for zero total angular momentum $J = 0$, the binding energy is given by

$$\varepsilon = \varepsilon_0 - \Delta\cos(2\omega\Phi/\Phi_0), \quad (13)$$

where $\varepsilon_0 < 0$ is the energy level of a one-dimensional exciton and $\Delta > 0$ is the tunneling amplitude (see our previous publication [5]). For weak fields (fluxes), Eq. (13) gives a positive (diamagnetic) energy shift. However, this shift can be considerably smaller than the vibrational diamagnetic shift which does not contain an exponential smallness. Besides, radial vibrations of particles lead to the attenuation of the oscillating energy component due to the destructive interference of trajectories embracing various magnetic fluxes.

In order to take these effects into consideration, we introduce the coordinate φ_c of the exciton "center of gravity" and the angular separation φ between an electron and a hole:

$$\varphi_c = \frac{m_e\varphi_e + m_h\varphi_h}{M}, \quad \varphi = \varphi_e - \varphi_h, \quad M = m_e + m_h, \quad (14)$$

as well as the radial coordinates $\xi = \rho_e - R_0$, $\eta = \rho_h - R_0$.

To avoid cumbersome expressions, we confine our analysis to the case of zero orbital angular momentum of an exciton as a whole (for an optical transition, the intensity of the exciton line is determined by the integral $\int\Psi(\varphi_e = \varphi_\eta)d\varphi_e$ which differs from zero only for $J = 0$). The magnetic-field dependent component of the Hamiltonian has the form

$$\hat{H}(B) = -(W_e + W_h) \times\left[\frac{\partial}{\partial\varphi} + i\frac{W_e\tilde{\Phi}_e + W_h\tilde{\Phi}_h}{W_e + W_h}\right]^2 + \frac{W_e W_h(\tilde{\Phi}_e - \tilde{\Phi}_h)^2}{W_e + W_h}, \quad (15)$$

where $\tilde{\Phi}_{e,h} = \pi B\rho_{e,h}^2/\Phi_0$, $W_e = \hbar^2/2m_e\rho_e^2$, and the expression for W_h has a similar form.

The phase transformation of the wave function

$$\psi = \chi e^{-i\lambda\varphi}, \quad \lambda(\xi,\eta) = \frac{W_e\tilde{\Phi}_e + W_h\tilde{\Phi}_h}{W_e + W_h}, \quad (16)$$

makes it possible to eliminate the second term in the brackets in (15) and to obtain the Schrödinger equation for the relative motion of an electron and a hole in the adiabatic approximation, i.e., for fixed radial coordinates ξ and η. Arguing in the same way as in [5] (the conditions for the wave function periodic in φ_e and φ_h), we obtain the oscillating contribution to energy of the form $\Delta(\xi,\eta)\cos2\pi\lambda(\xi,\eta)$. Averaging this contribution in the vibrational functions of radial motion, we obtain the following expression for attenuation of the Aharonov–Bohm oscillations of the exciton binding energy:

$$\varepsilon = \varepsilon_0 - \bar{\Delta}\exp\left(-\frac{B^2}{B_0^2}\right)\cos2\pi\left(\frac{\Phi}{\Phi_0}\right), \quad (17)$$

where we have used the following expression for the damping decrement of the vibrational ground state:

$$\frac{1}{B_0^2} = \left(\frac{2\pi^2 R_0}{M\Phi_0}\right)^2\hbar\left[\frac{m_h^2}{m_e\omega_e} + \frac{m_e^2}{m_h\omega_h}\right], \quad (18)$$

and $\bar{\Delta}$ is the tunneling amplitude averaged over radial vibrations. Thus, in the case under investigation, oscillations attenuate according to the Gaussian law, and the damping decrement $1/B_0$ increases linearly with the ring radius.

The vibrational diamagnetic shift for an exciton is determined by the last term in Hamiltonian (15). After averaging over vibrations, we obtain

$$\Delta E_{\text{vibr}}(B;\, X_0) \;=\; \frac{\hbar}{2M} \times\left[\frac{m_e\Omega_e^2}{\omega_e}\left(v_e+\frac{1}{2}\right)+\frac{m_h\Omega_h^2}{\omega_h}\left(v_h+\frac{1}{2}\right)\right], \tag{19}$$

where v_e and v_h are the quantum numbers for the vibrational states of electrons and holes, and Ω_e and Ω_h are their cyclotron frequencies.

4. TRION X^- AND MULTIPLY CHARGED EXCITONS

The complex e–e–h (trion) is known to form a bound state of three particles, i.e., is more advantageous from the energy point of view than an individual electron and an exciton. Since a trion has a charge, its motion as a whole along a ring in a magnetic field must lead to a rotational diamagnetic shift of energy levels. Besides, as in the case of an exciton, we can expect an oscillating magnetic-flux dependence of the trion internal energy in a one-dimensional ring. Indeed, introducing the variables

$$\varphi_c \;=\; \frac{m_e(\varphi_1+\varphi_2)+m_h\varphi_h}{M_{\text{tr}}}, \qquad \alpha \;=\; \frac{\varphi_1+\varphi_2}{2}-\varphi_h, \quad \beta \;=\; \varphi_1-\varphi_2, \tag{20}$$

where $M_{\text{tr}} = 2m_e + m_h$ is the trion mass and $\varphi_{1,2}$ are the coordinates of the electrons, we can easily derive the following expression for the total energy of the trion:

$$E_{\text{tr}} \;=\; W_{\text{tr}}(J-\tilde{\Phi})^2+\varepsilon_{\text{int}}. \tag{21}$$

Here, $W_{\text{tr}} = \hbar/2M_{\text{tr}}R_0^2$, ε_{int} is the internal energy, which is the eigenvalue of the equation

$$-\left(\frac{1}{2}W_{e0}+W_{h0}\right)\frac{\partial^2\chi}{\partial\alpha^2}-2W_e\frac{\partial^2\chi}{\partial\beta^2} +\left[V(\beta)-V\left(\alpha+\frac{\beta}{2}\right)-V\left(\alpha-\frac{\beta}{2}\right)\right]\chi \;=\; \varepsilon_{\text{int}}\chi, \tag{22}$$

$$V(\gamma) \;=\; \frac{e^2}{2R_0|\sin(\gamma/2)|}.$$

The wave function ψ of the trion is connected with the solution $\chi(\alpha, \beta)$ of Eq. (22) through the relation

$$\Psi \;=\; \chi\exp i(J\varphi_c+\lambda\alpha), \qquad \lambda\equiv\frac{2(W_{e0}+W_{h0})}{W_{e0}+2W_{h0}}\tilde{\Phi}. \tag{23}$$

The potential energy in Eq. (22) is a periodic function on the plane (α, β). The basis vectors of the corresponding lattice are given by

$$\bar{\mathbf{e}}_1 \;=\; \boldsymbol{\alpha}_0\cdot\pi+\boldsymbol{\beta}_0\cdot 2\pi, \quad \bar{\mathbf{e}}_2 \;=\; \boldsymbol{\alpha}_0\cdot 2\pi+\boldsymbol{\beta}\cdot 0, \tag{24}$$

where $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ are the unit vectors of the Cartesian system of coordinates in the plane (α, β). Equations (24) describe a rectangular body-centered lattice (in the two-dimensional sense). The reciprocal lattice is of the same type with the basis vectors

$$\mathbf{g}_1 \;=\; \boldsymbol{\alpha}_0\cdot 0+\boldsymbol{\beta}_0\cdot 1, \quad \mathbf{g}_2 \;=\; \boldsymbol{\alpha}_0\cdot 1+\boldsymbol{\beta}_0\cdot\left(-\frac{1}{2}\right). \tag{25}$$

The general solution of Eq. (22) is characterized by the quasimomentum vector $\boldsymbol{\alpha}_0 p + \boldsymbol{\beta}_0 q$, and energy ε_{int} is a function of its components p and q. Subjecting the total function φ to the conditions of periodicity in φ_1, φ_2, and φ_h, we can find the allowed values of p and q, and thus determine the energy levels of a trion in a quantum ring. The periodicity conditions have the form

$$\frac{Jm_e}{M_{\text{tr}}}+\frac{\lambda}{2}+\frac{p}{2}+q \;=\; n_1, \quad \frac{Jm_e}{M_{\text{tr}}}+\frac{\lambda}{2}+\frac{p}{2}-q \;=\; n_2, \quad \frac{Jm_h}{M_{\text{tr}}}-\lambda-p \;=\; n_3, \tag{26}$$

were n_1, n_2, and n_3 are integers. It follows hence that J and $2q$ are integers, while $p = Jm_h/M_{\text{tr}} - \lambda +$ integer. The energy ε_{int} is periodic on the reciprocal lattice, and hence it depends periodically on Φ through λ. By way of an example, we write the result in the strong-coupling approximation:

$$\varepsilon_{\text{int}} \;=\; \varepsilon_0-2\Delta(2\pi)\cos 2\pi\left(\frac{2m_e}{M_{\text{tr}}}J+\lambda\right) -4\Delta(\sqrt{5}\pi)\cos\pi\left(\frac{2m_e}{M_{\text{tr}}}J+\lambda\right)-2\Delta(4\pi). \tag{27}$$

The arguments 2π, $\sqrt{5}\,\pi$, and 4π of the tunneling amplitudes indicate the distances from the initial site to their nearest neighbors in the lattice which is defined by Eq. (24), and ε_0 is the binding energy of a trion in a rectilinear quantum wire (i.e., the limit $R_0 \longrightarrow \infty$). All the tunneling amplitudes in Eq. (27) are positive. Thus, the binding energy of the trion oscillates with the magnetic flux (two harmonics are present in the model under investigation). The period of oscillations $\Delta\Phi$ is determined by the value of λ from Eq. (23), i.e., is a function of the ratio of effective masses; for the fundamental harmonic, we have

$$\Delta\Phi \;=\; \Phi_0\frac{2m_e+m_h}{m_e+m_h}. \tag{28}$$

In the opposite limit of weak coupling outside forbidden bands, the Coulomb interaction between particles can be neglected outside the region of forbidden

bands in the spectrum. In the first Brillouin zone, we must put $n_1 = n_2 = n_3 = 0$ in formulas (26), whence we obtain $J = 0$, $q = 0$, and $p = -\lambda$. It follows from (22) that

$$\varepsilon_{\text{int}} = \left(\frac{1}{2}W_{e0} + W_{h0}\right)\lambda^2,$$

while for the total energy we obtain the expression

$$E_{\text{tr}} = W_{\text{tr}}\tilde{\Phi}^2 + \left(\frac{1}{2}W_{e0} + W_{h0}\right)\lambda^2 = (2W_e + W_h)\tilde{\Phi}^2, \quad (29)$$

as should be expected in the case of three free particles. A comparison of this result with E_{tr} from (21) for $J = 0$ shows that the Coulomb interaction reduces considerably the rotational component of diamagnetic shift (to zero for X_0).

The vibrational diamagnetic shift is of the order of $\hbar\Omega_e^2/\omega_e$ as in the case of an exciton, but the main contribution for a trion comes from rotations defined by the first term on the right-hand side of (21). The observable quantity is the shift of the exciton luminescence line, which is equal to the difference between energies (21) and (5) in the principal order in $W_e/\hbar\omega_e$ for $N = 1$:

$$\Delta\nu_{\text{tr}} = (W_{\text{tr}} - W_{e0})\tilde{\Phi}^2, \quad J = 0. \quad (30)$$

Since $M_{\text{tr}} > m_e$, the obtained diamagnetic shift turns out to be negative. Naturally, the total shift contains a contribution from spin splitting; the obtained value of $\Delta\nu_{\text{tr}}$ describes the behavior of the center of gravity of the spin multiplet.

Finally, the same computations (separation of the rotation of the system as a whole and the gradient transformation of the wave function) can be used to obtain the following magnetic-flux dependent contribution to the total energy for a multiply charged complexes (hole + N electrons):

$$E_N(B) = W_M[J + (N-1)\tilde{\Phi}]^2,$$
$$W_M = \frac{\hbar^2}{2MR_0^2}, \quad M = Nm_e + m_h. \quad (31)$$

The corresponding shift of a line upon the transition $(h + Ne) \longrightarrow (N-1)e$ is given by

$$\Delta\nu_N = (N-1)[W_M(N-1) - W_e]\tilde{\Phi}^2$$
$$= -\frac{(N-1)\hbar^2}{2R_0^2}\left(\frac{m_e + m_h}{m_e M}\right)\tilde{\Phi}^2. \quad (32)$$

In formulas (31) and (32), we have omitted small corrections to $E_N(B)$ and $\Delta\nu_N$ associated with radial vibrations and the magnetic- field dependent component of the internal energy of the complex. Numerical estimates for $m_c = 0.07m_0$, $m_h = 0.35m_0$, $\hbar\omega_e = 12$ meV (i.e., close to R_y^*), and $R_0 = 14$ nm give $\Delta\nu(X_0) = 7 \times 10^{-3}B^2$ meV for the frequency shift for a neutral exciton (the ground state in radial vibrations) and $\Delta\nu(X^-) = -5 \times 10^{-2}B^2$ meV for a trion, where B is measured in teslas. The values of ω_e and R_0 are borrowed from [6].

Let us summarize the obtained results. The diamagnetic shift of the energy of a system comprising electrons only in a quantum ring of a small but finite width in a transverse magnetic field does not depend on the electron–electron interaction. Conversely, the Coulomb interaction in the X^{n-} complexes reduces the diamagnetic shift appreciably. In the case of a neutral exciton, we are left only with the vibrational component of the shift and with the exponential small contribution associated with the Aharonov–Bohm effect. The diamagnetic shift of the exciton luminescence line is positive for X_0 (and small in the parameter $W_e/\hbar\omega_e$) and negative for all charged excitons. The period of oscillations of the binding energy of charged complexes in magnetic flux differs from Φ_0 and depends on the ratio of effective masses of electrons and holes.

ACKNOWLEDGMENTS

This work was financed by the Russian Foundation for Basic Research (project no. 99-02-17127). Support of the DFG via SFB and the BMBF via a Max Planck research award is gratefully acknowledged.

REFERENCES

1. J. M. Garsia *et al.*, Appl. Phys. Lett. **71**, 2014 (1997).
2. A. Lorke and R. J. Luyken, Physica B (Amsterdam) **256**, 424 (1998).
3. R. J. Warburton, C. Schäflein, D. Haft, *et al.*, submitted to Nature (2000).
4. T. Chakraborty and L. Pietiläinen, Phys. Rev. B **50**, 8460 (1994).
5. A. V. Chaplik, Pis'ma Zh. Éksp. Teor. Fiz. **62**, 885 (1995) [JETP Lett. **62**, 900 (1995)]; R. Römer and M. Raikh, submitted to Phys. Rev. B (2000).
6. A. Lorke, R. J. Luyken, A. O. Govorov, *et al.*, Phys. Rev. Lett. **84**, 2223 (2000).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 174–178.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 199–204.

MISCELLANEOUS

Physics of the ATPase Molecular Motor¶

S. Pikin[a] and W. Haase[b]

[a]*Institute of Crystallography, Russian Academy of Sciences, Moscow, 117333 Russia*
[b]*Institut für Physikalische Chemie, Technische Universitat Darmstadt, 64287, Darmstadt, Deutschland*

Received June 22, 2000

Abstract—A physical description of the rotary molecular motors is given, the fundamental relationships between the dynamic characteristics of the motors are found, and their basic macroscopic parameters are evaluated. The role of the molecular chirality is explained. The membrane nanostructure and the ionic subsystem as elements of the motors are discussed in much detail. The stochastic features of the thermal noise-assisted processes occurring in the motors are established. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Adenosine triphosphate (ATP) synthase is an exceptional molecular machine acting in the cell respiration in mitochondria and in the photosynthesis in chloroplasts of plants and also in bacteria [1–3]. ATP synthase is called the enzyme "F_0F_1ATPase," in which the F_1 part contains the catalytic center and the F_0 part couples the F_1 part to the membrane. It was shown that a flow of hydrogen ions across the membrane controls the ATP formation and that the binding of ATP synthase to the hydrogen ion transport occurs via the F_0 part. ATP functions as a carrier of energy in all living organisms, it captures the chemical energy released by the combustion of nutrients and transfers it to reactions that require energy. ATP consists of the nucleoside adenosine linked to three phosphate groups. Upon the removal of the outermost phosphate group, adenosine diphosphate (ADP) is formed and the energy released can be used in other reactions. Conversely, with the help of energy, an inorganic phosphate group can be bound to ADP and form ATP.

To understand in detail how the F_0F_1ATPase functions, a detailed chemical and structural knowledge of the enzyme was obtained [2]. The three-dimensional structure of ATP synthase was clarified. The F_1 part consists of sub-units of five types: α, β, γ, δ, and ε. While there are three α and three β units, there is only one unit of each of the γ, δ, and ε types. The ATP synthesis occurs on the β units. The sub-units γ, δ, and ε are not symmetrical, they are in fact chiral [4, 5], a feature of importance for our understanding of how ATP synthase functions.

Much attention has been given to the F_1 part. It was found that the step requiring energy is not the synthesis of ATP from ADP and inorganic phosphate, but the binding of ADP and the phosphate to the enzyme which produces the ATP (where the energy surplus is stored). The γ, δ, and ε sub-units rotate in a cylinder formed of the alternating α and β sub-units. This rotation induces structural changes in β that lead to differences in the binding ability during a cycle; in addition, this rotation is driven by the ion flow through the membrane. The γ sub-unit behaves as an asymmetrical axle. It has unique contacts with the β sub-units and forces their active surfaces to assume different three-dimensional structures. The rotation of the γ sub-unit was demonstrated experimentally [6].

2. GENERALIZED FORCES AND CURRENTS IN THE ISOTHERMAL CHIRAL ROTARY MOTOR

This molecular motor is isothermal, which means that its internal states are locally in equilibrium at a constant temperature *T*. We believe that a certain general physical approach to the description of these motors is possible, similarly to how this was recently done for linear molecular motors [7]. The action of the motor is induced by generalized forces: the electric field **E** acting in the membrane, the mechanical motion velocity **N** relevant to the motor, and the chemical potential difference Δ that measures the free energy change per consumed "fuel" molecule (Δ = 0 at chemical equilibrium, Δ is positive for the ATP hydrolysis process when ATP is in excess and is negative when ADP is in excess). The action of forces **E**, **N**, and Δ leads to the electric current, rotation, and fuel consumption characterized by the generalized "currents": the electric current density **j**, the mechanical force **g** applied to the motor, and the average consumption rate of fuel molecules *u*, i.e., in our case, the average number of ATP molecules bound and hydrolyzed per unit time (or the average number of ADP molecules bound to the enzyme and used for ATP synthesis). The mechanical force **g** includes the viscous friction force relevant to the motor in which the mechanical motion occurs.

¶This article was submitted by the authors in English.

The functions $\mathbf{j}(\mathbf{E}, \mathbf{N}, \Delta)$, $\mathbf{g}(\mathbf{E}, \mathbf{N}, \Delta)$, and $u(\mathbf{E}, \mathbf{N}, \Delta)$ are, in the general case, nonlinear because the motor operates far from the equilibrium ($\Delta > k_B T$). Still it is useful to consider the linear regime ($\Delta < k_B T$), where we can write the relations

$$\mathbf{j} = \sigma\mathbf{E} + \nu\mathbf{n} \times \mathbf{N} + \mathbf{e}\Delta, \tag{1a}$$

$$u = -\mathbf{e} \cdot \mathbf{E} - \mathbf{v} \cdot \mathbf{N} + \lambda\Delta, \tag{1b}$$

$$\mathbf{g} = \nu\mathbf{n} \times \mathbf{E} + \eta\mathbf{N} + \mathbf{v}\Delta. \tag{1c}$$

Relations (1) insure that the dissipation rate $\dot{S}$ is positive,

$$\dot{S} = \mathbf{j} \cdot \mathbf{E} + u\Delta + \mathbf{g} \cdot \mathbf{N} = \sigma E^2 + \lambda\Delta^2 + \eta N^2, \tag{2}$$

where the electric conductivity σ, the viscosity η, and the coefficient λ are positive. The appearance of coefficient ν in Eqs. (1) is directly related to the asymmetry (chirality) of the moving part (the γ sub-unit) of the motor; ν is a pseudo-scalar when the product $\mathbf{n} \times \mathbf{N}$ is a pseudo-vector and $\nu\mathbf{n} \times \mathbf{N}$ is a vector. The component $\mathbf{n}_\perp$ of the director $\mathbf{n} = \mathbf{n}_\parallel + \mathbf{n}_\perp$ describes the inclination of the γ sub-unit axis $\mathbf{n}$ at a small angle Θ from the normal $\mathbf{z}$ to the membrane surface,

$$n_x^2 + n_y^2 = \Theta^2, \quad \mathbf{n}^2 = 1, \quad n_z^2 \approx 1. \tag{3}$$

We have the vector $\mathbf{N} = d\mathbf{n}/dt = \boldsymbol{\Omega} \times \mathbf{n}$, where the pseudo-vector $\boldsymbol{\Omega}$ is the angular velocity of the axle,

$$\begin{gathered} \Omega_z \neq 0, \quad \Omega_x = \Omega_y = 0. \\ N_x = -n_y\Omega_z, \quad N_y = n_x\Omega_z, \quad N_z = 0. \end{gathered} \tag{4}$$

It is seen from Eqs. (1) and (4) that the polar vector $\mathbf{v}$ must be perpendicular to $\mathbf{n}$ and parallel to $\mathbf{N}$ if $\mathbf{vN} \neq 0$. The formalism in Eqs. (1)–(4) based on describing the motor by a director $\mathbf{n}$ is quite similar to the approach developed for liquid crystals [8].

The kinetic coefficient $\mathbf{e}$ is a polar vector. It describes the interaction whereby the fuel consumption induces the electric current and mechanical motion. Without the tilting at a finite angle Θ, this interaction disappears. We, thus, assume $\mathbf{e}$ to be related physically to some polar interactions, for example between dipoles (electric and/or steric) $\mathbf{p}_\beta$ characterizing the deformation of the β-units and the polarization $\mathbf{P}_\gamma$ characterizing the inclined γ-unit. Because of chirality, the two-dimensional vector $\mathbf{P}_y$ is related to $\mathbf{n}$ by

$$\mathbf{P}_\gamma = \mu\xi, \quad \xi_x = n_z n_y, \quad \xi_y = -n_z n_x, \tag{5a}$$

$$\begin{gathered} P_{\gamma x} = \mu n_z n_y, \quad P_{\gamma y} = -\mu n_z n_x, \\ \mathbf{P}_y \times \mathbf{n}_\perp \approx \mu\Theta^2, \end{gathered} \tag{5b}$$

where μ is the "piezomodulus" depending on the enzyme chirality. Equations (5) are similar to the formalism developed for the description of chiral smectic-C liquid crystals [9]. Thus, vector $\mathbf{e}$ can be represented as

$$\mathbf{e} = c\mathbf{P}_\beta \times \boldsymbol{\xi}, \tag{6}$$

which vanishes in the absence of the γ-unit tilting.

The polar vector $\mathbf{v}$, by the same physical reasons, is assumed to be related to director $\mathbf{n}$ and the polarization vectors $\mathbf{P}_\gamma$ and $\mathbf{p}_\beta$. The relation $\mathbf{v} = b\mathbf{P}_\gamma$, for example, is forbidden because, in this case, a part of the dissipation rate is the scalar product

$$\Delta\mathbf{v} \cdot \mathbf{N} = -b\mu n_z \Omega_z \Theta^2 \Delta$$

that depends on the direction of $\mathbf{n}$. However, the physical quantities must not depend on the direction of $\mathbf{n}$ (they depend on even combinations of the $\mathbf{n}$ components). At the same time, the components of $\mathbf{v}$ must be proportional to the components of director $\mathbf{n}$, with $\mathbf{v}$ being parallel to $\mathbf{N}$ in accordance with Eqs. (1). It is possible to write a unique combination (at small tilt angles) for vector $\mathbf{v}$,

$$\begin{aligned} \mathbf{v} &= b\mu[\mathbf{n} \times \mathbf{e}] \\ &= -b\mu\xi(\mathbf{n}_\perp \cdot \mathbf{p}_\beta) = -b\mathbf{P}_\gamma(\mathbf{n}_\perp \cdot \mathbf{p}_\beta). \end{aligned} \tag{7}$$

The products $\mathbf{v} \cdot \mathbf{N}$ and $\mathbf{v} \times \mathbf{n}$ then become

$$\mathbf{v} \cdot \mathbf{N} = b\mu n_z \Omega_z \Theta^2 (\mathbf{n}_\perp \cdot \mathbf{p}_\beta), \tag{8}$$

$$[\mathbf{v} \times \mathbf{N}]_i = b\mu(\mathbf{n}_\perp \cdot \mathbf{p}_\beta)(n_z^2 - \delta_{iz})n_i. \tag{9}$$

These expressions depend only on even combinations of the $\mathbf{n}$ components. The products $\mathbf{n}_\perp \cdot \mathbf{p}_\beta$ and $\mathbf{p}_\beta \times \xi$ must be constant parameters of the system under consideration. This is possible under the assumptions made above: dipoles $\mathbf{p}_\beta$, pseudo-vector ξ, and the director component $\mathbf{n}_\perp$ are in the membrane plane, $\mathbf{n}_\perp$ and $\mathbf{p}_\beta$ are collinear, and $\mathbf{n}_\perp$ and ξ are orthogonal to each other during the interactions of the rotating γ-unit with different β-units. It then follows that $\mathbf{n}_\perp \cdot \mathbf{p}_\beta = \Theta p_\beta$. Thus, these products are constant parameters of the membrane system and are proportional to the tilt angle Θ and the dipole moment p_β.

We note that the tilt angle Θ and the dipole moment p_β describing the inclination of the γ-unit and deformations of β-units respectively, must depend on the energy transferred to these sub-units during the motor work. These quantities vanish in the absence of this energy transfer. The energy transfer can occur due to chemical reactions, ionic currents, and mechanical movements. Constants b and c in relations (6) and (7) must in general describe some energy transfer mechanisms, which are unknown in detail unfortunately. Because these constants involved in $\mathbf{e}$ and $\mathbf{v}$ are related to the same physical process, we can expect them to be related to each other by other parameters of the motor, such as the electric conductivity, chirality, etc.; in other words, the motor is assumed to work as a self-consistent system in which all kinds of motion start and stop simultaneously.

3. NON-PASSIVE REGIMES OF THE MOTOR WORK

The molecular motor stops in the absence of the γ-unit tilting because both the $\mathbf{j} \cdot \mathbf{E}$ and $u\Delta$ terms are positive and $\mathbf{g} \cdot \mathbf{N} = 0$; i.e., there is no energy output from the enzyme: all work performed at the enzyme is dissipated in the thermal bath. When $\Delta = 0$ and $\Theta = 0$, we have the density current $\mathbf{j} = \sigma\mathbf{E}$ if the electric field is present inside the membrane. But without the energy output from the thermal bath, this current decreases with time and vanishes at the equilibrium conditions because of a re-distribution of ions. We assume that $\mathbf{E} = 0$ if the difference Δ is zero.

If $\mathbf{g} \cdot \mathbf{N}$ is negative, mechanical work is performed by the motor. If $u\Delta$ is negative, chemical energy is generated by the motor. If $\mathbf{j} \cdot \mathbf{E}$ is negative, the motor produces electric energy. Thus, the enzyme can work in various non-passive regimes.

1) $u\Delta > 0$, $\mathbf{g} \cdot \mathbf{N} < 0$, $\mathbf{j} \cdot \mathbf{E} > 0$. The motor takes the energy from the heat bath, for instance, it uses the hydrolysis of ATP in excess (i.e., $u > 0$ and $\Delta > 0$) and from the work produced by the electric current to generate mechanical rotation.

2) $u\Delta < 0$, $\mathbf{g} \cdot \mathbf{N} > 0$, $\mathbf{j} \cdot \mathbf{E} > 0$. The system binds ADP and the phosphate to the enzyme and releases ATP already in excess (i.e., $u < 0$ and $\Delta > 0$) from the mechanical and electrical input.

3) $u\Delta > 0$, $\mathbf{g} \cdot \mathbf{N} < 0$, $\mathbf{j} \cdot \mathbf{E} > 0$. The motor uses ADP in excess (i.e., $u < 0$ and $\Delta < 0$) and the electric current to generate mechanical work.

4) $u\Delta < 0$, $\mathbf{g} \cdot \mathbf{N} > 0$, $\mathbf{j} \cdot \mathbf{E} > 0$. The system produces ADP already in excess (i.e., $u > 0$ and $\Delta < 0$) from the mechanical and electric input.

5) $u\Delta > 0$, $\mathbf{g} \cdot \mathbf{N} < 0$, $\mathbf{j} \cdot \mathbf{E} < 0$. The motor (a) uses the hydrolysis of ATP in excess (i.e., $u > 0$ and $\Delta > 0$) or (b) uses ADP in excess for the synthesis of ATP (i.e., $u < 0$ and $\Delta < 0$) to generate mechanical rotation and electric current.

6) $u\Delta < 0$, $\mathbf{g} \cdot \mathbf{N} > 0$, $\mathbf{j} \cdot \mathbf{E} < 0$. The system (a) produces ADP already in excess (i.e., $u > 0$ and $\Delta < 0$) or (b) produces ATP already in excess (i.e., $u < 0$ and $\Delta > 0$) and also the electric current from the mechanical input.

7) $u\Delta > 0$, $\mathbf{g} \cdot \mathbf{N} > 0$, $\mathbf{j} \cdot \mathbf{E} < 0$. The motor (a) uses the hydrolysis of ATP in excess (i.e., $u > 0$ and $\Delta > 0$) or (b) uses ADP in excess (i.e., $u < 0$ and $\Delta < 0$) and also the mechanical input to generate electric current.

8) $u\Delta < 0$, $\mathbf{g} \cdot \mathbf{N} < 0$, $\mathbf{j} \cdot \mathbf{E} > 0$. The motor (a) releases ATP already in excess (i.e., $u < 0$ and $\Delta > 0$) or (b) produces ADP already in excess (i.e., $u > 0$ and $\Delta < 0$) and generates mechanical rotation from the electric input.

Thus, 12 regimes are available in this motor. The rotation of the axle (at $\Theta \neq 0$) with a constant angular velocity $\Omega_z = \phi$ occurs when the viscous-driven torque and the field-driven torque compensate each other (neglecting some elastic forces), i.e., when $\mathbf{g} \times \mathbf{n} = 0$. We then obtain from Eqs. (1) and (5)–(9) that

$$\gamma\Omega_z = \nu E_z - b\mu\Delta(\mathbf{p}_\beta \cdot \mathbf{n}_\perp)n_z. \tag{10}$$

This equation shows that a constant angular velocity arises even in the case where $\mathbf{E} = 0$, but the tilt angle Θ must have a finite value for the effect to occur. Tilting of the γ-unit induced by the energy transfer to this subunit from the heat bath results in the appearance of the electric current $\mathbf{j}$ even at $\mathbf{E} = 0$ (see Eq. (1a)). We note that the motor does not produce mechanical work; i.e., $\mathbf{g} \cdot \mathbf{N} = 0$, when the rotation of the axle occurs with constant velocity (10).

In the general case, the sum

$$\mathbf{j} \cdot \mathbf{E} + \mathbf{g} \cdot \mathbf{N} = \sigma E^2 + \gamma N^2 + \Delta(\mathbf{e} \cdot \mathbf{E} + \mathbf{v} \cdot \mathbf{N}) \tag{11}$$

can change its sign only if the sign of $\Delta(\mathbf{e} \cdot \mathbf{E} + \mathbf{v} \cdot \mathbf{N})$ changes. In the regimes with $u\Delta > 0$ and $\mathbf{j} \cdot \mathbf{E} + \mathbf{g} \cdot \mathbf{N} < 0$, the motor produces mechanical work if $\Delta(\mathbf{e} \cdot \mathbf{E} + \mathbf{v} \cdot \mathbf{N})$ is negative and $\mathbf{g} \cdot \mathbf{N}$ is negative. Thus, we conclude that the sign of the angular velocity can change only if the direction of vector $\mathbf{E}$ and the sign of Δ are reversed to preserve the negative signs of $\mathbf{j} \cdot \mathbf{E} + \mathbf{g} \cdot \mathbf{N}$ and $\mathbf{g} \cdot \mathbf{N}$. This means that to preserve the positive sign of $u\Delta$, the sign of u must change; for example, the hydrolysis of ATP in excess must change to using ADP in excess. At the same time, the direction of $\mathbf{E}$ is reversed. If the motor also generates the electric current, i.e., $\mathbf{j} \cdot \mathbf{E} < 0$, the change of the $\mathbf{E}$ direction results in changing the direction of the electric current $\mathbf{j}$ in the case where energy is taken from the heat bath.

In the acting motor [1], the regime $u\Delta > 0$ seems to determine the binding of ADP in excess and phosphate to the enzyme and the release of ATP (i.e., $u < 0$ and $\Delta < 0$). This results in only one way for the enzyme to react; i.e., the directions of the angular velocity Ω and of current $\mathbf{j}$ are strictly determined. The hydrolysis of ATP in excess and the release of ADP (i.e., $u > 0$ and $\Delta > 0$) must lead to reversing the directions of Ω and $\mathbf{j}$.

We note that at $\mathbf{E} = 0$, the dissipation rate $\dot{S}$ is equal to $\lambda\Delta^2$, with $\mathbf{N} = \mathbf{g} = 0$ and $\mathbf{j} = \mathbf{e}\Delta$; i.e., the motor does not produce work. Thus, to produce mechanical work, a certain electric field inside the biomembrane and the inclination of the γ-unit at a certain angle must occur.

4. ELECTRIC CURRENT WORK AND EFFICIENCY OF THE ROTARY MOTOR

For the motor with a constant rotation velocity ($\mathbf{g} \cdot \mathbf{N} = 0$, with the mechanical work not produced), we see from Eqs. (5)–(11) that $\mathbf{j} \cdot \mathbf{E} < 0$ if

$$\left(\sigma + \frac{\nu^2\Theta^2}{\gamma}\right)E_z^2 - \Delta\left(c + \frac{b\nu\mu\Theta^2}{\gamma}\right) \times (\mathbf{n}_\perp \cdot \mathbf{p}_\beta)(n_z E_z) \leq 0. \tag{12}$$

Thus, the motor generates electric current until $\mathbf{j} \cdot \mathbf{E} < 0$ occurs at a finite value of E_z when condition (12) is satisfied. In this case, the j_z value is

$$j_z = \left(\sigma + \frac{\nu^2\Theta^2}{\gamma}\right)E_z^2 - \Delta\left(c + \frac{b\nu\mu\Theta^2}{\gamma}\right)(\mathbf{n}_\perp \cdot \mathbf{p}_\beta)n_z \quad (13)$$

$$= \sigma E_z - c\Delta p_\beta \Theta n_z + \nu\Theta^2\Omega_z. \quad (14)$$

We conclude from relations (12)–(14) that certain threshold conditions (for the Θ value at a given value of E_z or for the E_z value at a given value of Θ) must occur for the motor to generate the current. In weak fields and at small tilt angles, the threshold condition must have the form

$$\sigma n_z E_z \approx c\Delta p_\beta \Theta. \quad (15)$$

At the threshold condition, we have $j_z = 0$. The motor stops to generate current when E_z overcomes the threshold value determined by Eq. (15) or, in other words, when the tilt angle Θ is less than the threshold value determined by Eq. (15). We also see from relations (12)–(14) that the change of the sign of Δ results in sign changes of E_z, j_z, and Ω_z when the motor generates electric current in another regime with $u\Delta > 0$. Thus, the change of regimes reverses the directions of the electric current and the angular velocity. In the regime with $u < 0$ and $\Delta < 0$, the same threshold conditions take place for the electric field of the opposite direction. Because the experimental data show that the rotation and the current flow start and stop simultaneously, i.e., $j_z = 0$ and $\Omega_z = 0$ at the same time, we conclude from Eqs. (10) and (14) that parameters σ, ν, b, and c are related by

$$b\mu\sigma \approx c\nu. \quad (16)$$

It is useful to note that relation (16) can also be obtained when $E_z = 0$ if we consider Ω_z and j_z in Eqs. (10) and (14) as induced by the effective field

$$E_{z,\,\mathrm{eff}} \approx -\frac{bD\mu p_\beta \Theta}{\nu} n_z \approx -\frac{c\Delta p_\beta \Theta}{\sigma} n_z. \quad (17)$$

This is quite reasonable physically because the current density $j_z = \sigma E_{z,\,\mathrm{eff}}$ and the angular velocity $\Omega_z = \nu E_{z,\,\mathrm{eff}}/\gamma$ must have the same origin: these closely related quantities have the same meaning as in the Lehmann effect in chiral liquid crystals [8]. These relations show again that the flow of ions and the γ-unit rotation are related to each other and exist due to the tilting of the γ-unit. According to the data known today [3–5], both the tilt and the rotation of the γ-unit open the way for the ion motion and the effect of moving ions (the transfer of their energy) on the sequence of β- and α-units; the latter ones are prepared (by a change of the unit shape) for the subsequent processes of the ADP and phosphate binding to the enzyme and of releasing ATP. It is important for the motor that the rotation of the γ-unit involves a cycle: while the γ-unit interacts with one of the β-units (see the products $(\mathbf{n}_\perp \cdot \mathbf{p}_\beta)$ and $(\mathbf{p}_\beta \times \xi)$), two other β-units are deformed by the ion flow to be ready for the next interaction with the γ-unit (with the steric end electric dipoles probably appearing during the process).

To induce the γ-unit tilting in the membrane, the heat bath must transfer a portion of energy to the enzyme, for example, by chemical reactions. The simplest assumption is that this portion of energy, which is proportional to the tilt angle squared, is related to the chemical potential difference Δ via $\Delta = a\Theta^2$, where a is a constant that has the scale of the intermolecular interaction energy. Thus, Θ is determined by Δ and vanishes at $\Delta = 0$.

It is interesting to estimate the efficiency of the considered regime. As for any macroscopic motor, we can define the electric efficiency of the molecular motor η by the ratio of the electric current work performed to the chemical energy consumed,

$$\eta = -\frac{j_z E_z}{u\Delta} \approx \frac{c\Delta p_\beta \Theta n_z E_z - \sigma E_z^2}{\lambda\Delta^2 + c\Delta p_\beta \Theta n_z E_z}. \quad (18)$$

For other regimes, in which the chemical energy is generated from the electric input, the chemical efficiency is the inverse of expression (18). It is seen from Eq. (18) that the efficiency maximum $\eta_{\max}$ occurs at the values of $\Delta(E_z)$ given by

$$\Delta(E_z) \approx \frac{\lambda\sigma + \sqrt{\lambda^2\sigma^2 + \lambda_\sigma (c p_\beta \Theta)^2}}{c\lambda p_\beta \Theta}(n_z E_z). \quad (19)$$

Thus, $\eta_{\max}$ has a constant value along the straight line $\Delta = \mathrm{const} E_z$. It is independent of force E_z, is close to zero at small tilt angles Θ, and is close to 1 at large values of Θ (when $\Theta \gg \lambda\sigma/c^2 p_\beta^2$). This general statement is independent of the assumption about a relation between Δ and Θ. This molecular motor is an isothermal chemical motor working irreversibly far from equilibrium, i.e., it differs principally from the reversible Carnot engines [7].

5. ESTIMATES OF THE MOTOR CHARACTERISTICS AND CONCLUDING REMARKS

We now make rough quantitative estimates of the above parameters using some characteristic values [10, 11] such as the intermolecular interaction energy $\varepsilon \sim 10^2$ kcal/mole $\sim 10^{-19}$ J, the volume per protein molecule $l^3 \sim 10^{-24}$ m^3 (linear dimension $l \sim 10^2$ Å), the dipole moment of proteins $p_\beta \sim 100D \sim 10^{-16}$ esu, the piezoelectric modulus $\mu \sim 10^{-3}$ C m^{-2} (assumed to be of the order of unity in liquid crystals), i.e., $\mu p_\beta \sim 10^{-20}$ J, the rotation viscosity $\gamma \sim 10^{-1}$ Pa s ~ 1 Poise (assumed to be of the order of unity in liquid crystals), and the conductivity $\sigma \sim 10^{-7}$ Ω^{-1} m^{-1} $\sim 10^3$ s^{-1}. From the dimensional considerations, we can conclude roughly

that the coefficient a in the expected relation $\Delta = a\Theta^2$ is of the order ε, i.e., $\Delta \sim \varepsilon\Theta^2$; the coefficient b is of the order $(\varepsilon l^3)^{-1}$, i.e., $ab \sim l^{-3} \sim 10^{24}$ m^{-3}; the coefficient c is of the order $c \sim (\tau\varepsilon l^3)^{-1}$, where τ is a certain relaxation time that probably admits the estimate $\tau \sim \gamma l^3/\varepsilon$, i.e., $ac \sim (\varepsilon/\gamma l^6)$; the parameter ν has the dimension [energy density]$^{1/2}$, i.e., $\nu \sim r(\varepsilon/l^3)^{1/2}$, where r is a factor characterizing a small part of chiral interactions with respect to the total energy of intermolecular interactions.

Thus, we obtain the estimate

$$E_{th} \sim \frac{c\Delta p_\beta \Theta}{\sigma} \sim 10^9 \Theta^3 \text{ V m}^{-1}, \qquad (20)$$

which implies that for a typical value $\Theta \sim 10^{-1}$, we have $E_{th} \sim 10^6$ V m^{-1}, which is close to the membrane parameter [10]. From Eq. (10), we obtain the estimate

$$E_{th} \sim \frac{b\Delta\mu p_\beta \Theta}{\nu} \sim \frac{10^4\Theta^3}{\nu} \text{ J}^{1/2} \text{ m}^{-3/2} \sim \frac{10^5\Theta^3}{\nu} \text{ V m}^{-1}, \qquad (21)$$

where ν is measured in J$^{1/2}$ m$^{-3/2}$. Thus, if the motor stops when the current flow and the angular rotation stop simultaneously, we obtain $\nu \sim 10^{-4}$ J$^{1/2}$ m$^{-3/2}$ from Eqs. (16), (20) and (21). Because $(\varepsilon/l^3)^{1/2} \sim 10^2$ J$^{1/2}$ m$^{-3/2}$, the factor r is of the order $r \sim 10^{-6}$. Because $\nu E_{th} \sim 10$ J m^{-3} $\sim 10^2$ erg/cm^3 and $\gamma \sim 1$ Poise, we obtain the order value of Ω_z as

$$\Omega_z \sim \frac{\nu E_{th}}{\gamma} \sim 10^2 \text{ s}^{-1}, \qquad (22)$$

which is also comparable with the measured values [6]. It was also shown in [6] that, when ATP is absent, there is no rotary motion apart from the Brownian fluctuations (a few turns in either direction). Therefore, the obtained estimates do not look fantastic, they can explain some experimental facts.

No doubt, the problem of molecular motors in biophysical objects is very important and very complicated. It opens many possibilities for the research by various experimental methods and by physical modeling. The present paper proposes a physical model and the description of a rotary (F_1-ATPase) motor type avoiding the purely biological complexity. Our approach is based on the knowledge of the structure and properties of liquid-crystalline substances including the structure of ultrathin films, polarization properties of chiral materials, and the relations between dynamics of structural parameters and ionic currents in films. We have determined:

(i) general symmetry principles for the operation of these motors using the polar and chiral properties of the molecular system under consideration, the order parameter for this molecular system being defined;

(ii) general relationships between ionic currents across the membrane and the velocity characterizing the dynamics of the order parameter;

(iii) relationships between generalized forces inducing the motor, including viscous friction forces between the motor and the surrounding solvent and the free energy change per consumed "fuel" molecule, and generalized currents, including the above-mentioned velocity and the consumption rate of fuel molecules;

(iv) basic macroscopic parameters of the motor and the relationships between them. This may explain the observed action of the motor.

We thank V. Lorman and M. Kleman for stimulating discussions and M. Gorkunov for assistance. Financial support from Russian Foundation for Basic Research (grant 00-02-17801) is acknowledged. S.P. acknowledges the receipt of the research award from the Alexander von Humboldt-Stiftung.

REFERENCES

1. P. D. Boyer, Biochim. Biophys. Acta **1140**, 215 (1993).
2. J. P. Abrahams, A. G. Leslie, R. Lutter, and J. E. Walker, Nature **370**, 621 (1994).
3. P. D. Boyer, Annu. Rev. Biochem. **66**, 717 (1997).
4. W. Junge, D. Sabbert, and S. Engelbrecht, Ber. Bunsenges. Phys. Chem. **100**, 2014 (1996).
5. B. Schulenberg, F. Wellmer, H. Lill, *et al.*, Eur. J. Biochem. **249**, 134 (1997).
6. H. Noji, R. Yasuda, M. Yoshida, and K. Kinosita, Jr., Nature **386**, 299 (1997).
7. F. Julicher, A. Ajdari, and J. Prost, Rev. Mod. Phys. **69**, 1269 (1997).
8. P. G. de Gennes, *The Physics of Liquid Crystals* (Clarendon Press, Oxford, 1974; Mir, Moscow, 1977).
9. S. A. Pikin, *Structural Transformations in Liquid Crystals* (Nauka, Moscow, 1981; Gordon and Breach, New York, 1991).
10. R. Gabler, *Electric Interactions in Molecular Biophysics* (Academic, New York, 1978).
11. R. Pethig, *Dielectric and Electric Properties of Biological Materials* (Wiley, Chichester, 1979).

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 179–193.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 205–220.

MISCELLANEOUS

Big Entropy Fluctuations in the Nonequilibrium Steady State: A Simple Model with the Gauss Heat Bath¶

B. V. Chirikov

Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia
e-mail: chirikov@inp.nsk.su

Received July 6, 2000

Abstract—Large entropy fluctuations in a nonequilibrium steady state of classical mechanics are studied in extensive numerical experiments on a simple two-freedom model with the so-called Gauss time-reversible thermostat. The local fluctuations (on a set of fixed trajectory segments) from the average heat entropy absorbed in the thermostat are found to be non-Gaussian. The fluctuations can be approximately described by a two-Gaussian distribution with a crossover independent of the segment length and the number of trajectories ("particles"). The distribution itself does depend on both, approaching the single standard Gaussian distribution as any of those parameters increases. The global time-dependent fluctuations are qualitatively different in that they have a strict upper bound much less than the average entropy production. Thus, unlike the equilibrium steady state, the recovery of the initial low entropy becomes impossible after a sufficiently long time, even in the largest fluctuations. However, preliminary numerical experiments and the theoretical estimates in the special case of the critical dynamics with superdiffusion suggest the existence of infinitely many Poincaré recurrences to the initial state and beyond. This is a new interesting phenomenon to be further studied together with some other open questions. The relation of this particular example of a nonequilibrium steady state to the long-standing persistent controversy over statistical "irreversibility", or the notorious "time arrow", is also discussed. In conclusion, the unsolved problem of the origin of the causality "principle" is considered. © *2001 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION: EQUILIBRIUM VS. NONEQUILIBRIUM STEADY STATE

Fluctuations are an inseparable part of statistical laws. This has been well known since Boltzmann. What is apparently less known are the peculiar properties of rare big fluctuations (BF) as different from, and in a sense even opposite to, those of small stationary fluctuations. In particular, the former can be perfectly regular on the average, symmetric in time with respect to the fluctuation maximum, and can be described by simple kinetic equations rather than by a sheer probability of irregular "noise". Even though big fluctuations are very rare, they may be important in many various applications (see, e.g., [1] and references therein). In addition, the correct understanding and interpretation of the properties and origin of big fluctuations may help (at last!) to settle a strangely persistent controversy over statistical "irreversibility" and the notorious "time arrow".

In the big fluctuations problem, one must distinguish at least two qualitatively different classes of the fundamental (Hamiltonian, nondissipative) dynamical systems: those with and without the statistical equilibrium, or the equilibrium steady state (ES).

In the former (simpler) case, a big fluctuation consists of the two symmetric parts: the rise of a fluctuation followed by its return, or relaxation, back to ES (see Fig. 1 below). Both parts are described by the same kinetic (e.g., diffusion) equation, the only difference being in the sign of time. This relates the time-symmetric dynamical equations to the time-antisymmetric kinetic (but not statistical!) equations. The principal difference between the two, sometimes overlooked, is that the kinetic equations are widely understood as describing the relaxation only, i.e., the increase of the entropy in a closed system, whereas they actually do so for the rise of the big fluctuation as well, i.e., for the entropy decrease. All this was qualitatively known already to Boltzmann [2]. The first simple example of a symmetric big fluctuations was considered by Schrödinger [3]. A rigorous mathematical theorem for the diffusion (slow) kinetics was proved by Kolmogorov in 1937 in the paper entitled "Zur Umkehrbarkeit der statistischen Naturgesetze" ("Concerning the Reversibility of Statistical Laws in Nature") [4] (see also [5]). Regrettably, the principal Kolmogorov theorem still remains unknown to participants of the heated debate over "irreversibility" (see, e.g., "Round Table on Irreversibility" in [6]) and to the physicists actually studying such big fluctuations [1].

By now, there exists the well developed ergodic theory of dynamical systems (see, e.g., [7]). In particular, it proves that the relaxation (correlation decay, or mixing) proceeds eventually in both directions of time for almost any initial conditions of a chaotic dynamical

¶This article was submitted by the author in English.

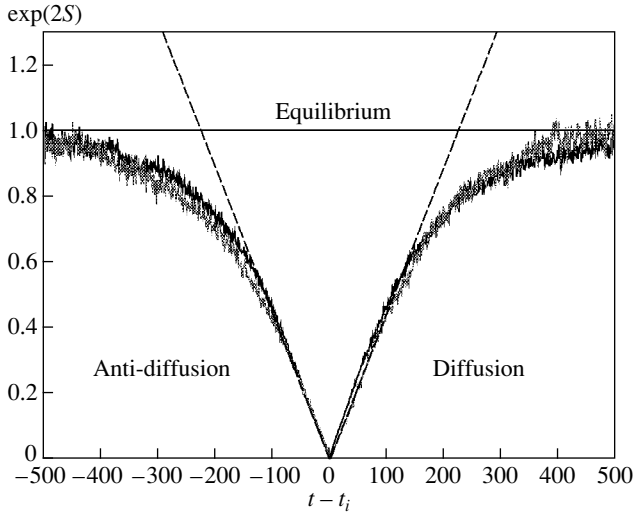

Fig. 1. Boltzmann's diffusive fluctuations in model (1.2) with the parameter $C = 15$: the square of the phase space area occupied by N independent trajectories ("particles") vs. the time (the number of map iterations $t - t_i$) counted from the instant t_i of fluctuation maximum, or of minimal Γ_{fl}, for each of the N_{fl} superimposed big fluctuations separated by the average period $P = \langle (t_i - t_{i-1}) \rangle$. Straight lines show the expected dependence for anti-diffusion and diffusion (see text). Two slightly different curves correspond to $N = 1$ (grey) and $N = 4$ (black) with $\Gamma_{fl} = 0.0001$ and 0.1: $N_{fl} = 3352$ and 2851; $P = 29863$ and 35110, respectively.

system. However, the relaxation must not always be monotonic, which simply means a big fluctuation on the way, depending on the initial conditions. To eliminate this apparently confusing (to many) "freedom," one can take a different approach to the problem: to start at arbitrary initial conditions (most likely corresponding to ES) and see the big fluctuation dynamics and statistics.

At this point, it is essential to recall that the systems with ES allow for very simple models in both the theoretical analysis and numerical experiments (of which the latter are even more important). In this paper, we use one of the most simple and popular models specified by the so-called Arnold cat map (see [8, 9])

$$\begin{aligned} \bar{p} &= p + x \mod 1, \\ \bar{x} &= x + \bar{p} \mod 1, \end{aligned} \tag{1.1}$$

that is a linear canonical map on the unit torus. It has no parameters and is chaotic and even ergodic. The rate of the local exponential instability, the Lyapunov exponent

$$\lambda = \ln(3/2 + \sqrt{5}/2) = 0.96,$$

implies a fast (ballistic) kinetics with the relaxation time $t_r \sim 1/\lambda \approx 1$.

A minor modification of this map,

$$\begin{aligned} \bar{p} &= p + x - \frac{1}{2} \mod C, \\ \bar{x} &= x + \bar{p} \mod 1, \end{aligned} \tag{1.2}$$

where $C \gg 1$ is the circumference of the phase space torus admits a slow (diffusive) relaxation with

$$t_r \sim C^2/4D_p,$$

where $D_p = 1/12$ is the diffusion rate in p. A convenient characteristic of the big fluctuation size is the rms phase space volume (area) $\Gamma(t) = \sigma_p \sigma_x$ for a group of N trajectories. In the ergodic motion at equilibrium, we have

$$\Gamma = \Gamma_0 = C/12.$$

In what follows, we use the dimensionless measure

$$\tilde{\Gamma} = \Gamma/\Gamma_0 \longrightarrow \Gamma$$

and omit the tilde.

The entropy S can be defined by the relation

$$S(t) = \ln\Gamma(t), \tag{1.3}$$

with $S = 0$ at equilibrium. This definition is not identical to the standard one (via the (coarse-grained) distribution function) but it is quite close to the latter if $\Gamma \ll 1$, i.e., for a big fluctuation, which is what we need in the problem under consideration. A great advantage of definition (1.3) is that the computation of S does not require very many trajectories as does the distribution function. In fact, even a single trajectory is sufficient!

A finite number of trajectories used for calculating the phase-space volume Γ is a sort of the coarse-grained distribution, as required in relation (1.3), but with a free bin size that can be arbitrarily small. The detailed study of big fluctuations in this class of ES models will be published elsewhere [10]. Here, we briefly consider the example shown in Fig. 1.

The data were obtained from running 4 and only 1 (!) trajectories for a sufficiently long time in order to collect sufficiently many big fluctuations; they are superimposed in Fig. 1 to clean up the regular big fluctuation from a "podlike trash" of stationary fluctuations. The size of big fluctuation chosen was approximately fixed by the condition $\Gamma(t) \le \Gamma_{fl}$. In spite of the inequality, the mean values $\langle \Gamma(t_i) \rangle = 0.000033$ and 0.069 are close (by the order of magnitude) to the fixed Γ_{fl} values in Fig. 1. We note that for a slow diffusive kinetics, we have

$$\exp(2S) \propto \sigma_p^2 \propto \langle p^2 \rangle$$

and σ_x remains constant.

The probability of big fluctuations can be characterized by the average period between them, for which a very simple estimate

$$P \approx 3\Gamma_{fl}^{-N} \approx 3\exp(-NS_{fl}) \tag{1.4}$$

is in a good agreement with data in Fig. 1 (upon including the empirical factor 3).

In the example presented here, the position of all big fluctuations in the phase space is fixed as $x_{fl} = 1/2$ and $p_{fl} = C/2$. If one lifts this restriction, the probability of big fluctuation increases by the factor $1/\Gamma_{fl}$, or by decreasing N by one ($N \longrightarrow N-1$), due to the arbitrary position of a big fluctuation in phase space. In the former case, a chain of big fluctuations is precisely the well known Poincaré recurrence. It is less known that the latter are a particular and specific case of big fluctuations, and the recurrence of a trajectory in a chaotic system is determined by the kinetics of the system. The recurrence of several ($N > 1$) trajectories can also be interpreted as the recurrence of a single trajectory in N uncoupled freedoms.

As can be seen from Fig. 1, irregular deviations from a regular big fluctuation are rapidly decreasing with the entropy $S \longrightarrow S_{fl}$. It may seem that the motion becomes regular near big fluctuation maximum, hence the term "optimal fluctuational path" [1]. In fact, the motion remains diffusive down to the dynamical scale, that is, $|\Delta p| \sim 1$ independently of parameter C in model (1.2).

Big fluctuations are not only perfectly regular by themselves but also surprisingly stable against any perturbations, both regular and chaotic. Moreover, the perturbations do not need to be small. At first glance, this looks very strange in a chaotic, highly unstable dynamics. The resolution of this apparent paradox is that the dynamical instability of motion affects the big fluctuation time instant t_i only. The big fluctuation shape is determined by the kinetics that can have an arbitrary mechanism, ranging from a purely dynamical one, as in model (1.2), to a completely noisy (stochastic, cf. Fig. 1 above and Fig. 4 in [1]). As a matter of fact, the fundamental Kolmogorov theorem [4] is specifically related to the latter case but remains valid in a much more general situation. The surprising stability of big fluctuations is similar to the full (less known) robustness property of the Anosov (strongly chaotic) systems [11], whose trajectories are only slightly deformed under a small perturbation (for discussion, see [12]). From a different perspective, this stability can be interpreted as a fundamental property of the "macroscopic" description of big fluctuations. In such a simple few-freedom system similar to (1.2), the term "macroscopic" refers to the averaged quantities σ, Γ, S, and similar ones. However, a somewhat confusing result is that the "macroscopic" stability comprises not only the relaxation of big fluctuations but also its rise, because both parts of big fluctuation always appear together. This may lead to another misunderstanding that the fluctuation and relaxation probabilities are the same, which is certainly wrong. The point is that the ratio of both (unequal!) probabilities is determined by the crossover parameter

$$R_{cro}(S_{fl}) = \frac{P}{t_r} \approx \frac{3\exp(-NS_{fl})}{C^2} \gg 1, \qquad (1.5)$$

where the latter expression refers to model (1.2) and the inequality determines the region of a big fluctuation where its waiting time is much longer than that of its immediate relaxation from a nonequilibrium "macroscopic" state (for further discussion, see Section 6 in what follows).

2. A NEW CLASS OF DYNAMICAL MODELS: WHAT ARE THEY FOR?

A relatively simple picture of big fluctuations in systems with the equilibrium steady state is well understood by now, although not yet well known. To Boltzmann, this picture was the basis of his fluctuation hypothesis for our Universe. Again, as is well understood by now, this hypothesis is entirely incompatible with the present structure of the Universe, because it would immediately imply the notorious "heat death" (see, e.g., [13]). For this reason, one may even term such systems the heat death models. Nevertheless, they can be and actually are widely used in the description and study of local statistical processes in thermodynamically closed systems. The latter term means the absence of any heat exchange with the environment. We note, however, that for exponentially unstable motion, the only dynamically closed system is the whole Universe. In particular, this excludes the hypothetical "velocity reversal," which is still popular in debates over "irreversibility" occurring since Loschmidt (for discussion, see, e.g., [12, 14] and Section 6 in what follows).

In any event, dynamical models with ES do not tell us the whole story of either the Universe or even a typical macroscopic process therein. The principal solution of this problem, unknown to Boltzmann, is quite clear by now, namely, the "equilibrium-free" models are wanted. Various classes of such models are intensively studied today. Moreover, the celebrated cosmic microwave background tells us that our Universe was born already in the state of a heat death; fortunately to us, however, it became unstable because of the well-known Jeans gravitational instability [15]. This resulted in developing a rich variety of collective processes, or synergetics, the term recently introduced or, better to say, put in use by Haken [16]. The most important peculiarity of this collective instability is in that the total overall relaxation (to somewhere?) with ever increasing total entropy is accompanied by an also increasing phase space inhomogeneity of the system, particularly in temperature. In other words, the whole system as well as its local parts become more and more nonequilibrium to the extent of the birth of a secondary dynamics that can be, and sometimes is, as perfect as,

for example, the celestial mechanics (for general discussion see, e.g., [17, 18, 12]).

We stress that all these inhomogeneous nonequilibrium structures are not big fluctuations as in ES systems, but are a result of regular collective instability, and therefore, they are immediately formed under a certain condition. In addition, they are typically dissipative structures in Prigogine's terms [19] because of the energy and entropy exchange with the infinite environment. The latter is the most important feature of such processes, and at the same time the main difficulty in studying the dynamics of those models both theoretically and in numerical experiments, which are so much simpler for the ES systems. Usually, the investigations in this field are based upon statistical laws omitting the underlying dynamics from the beginning.

Recently, however, a new class of dynamical models has been developed by Evans, Hoover, Morriss, Nosé, and others [20, 21]. Some researchers still hope that these new models will help to resolve the "paradox of irreversibility." A more serious reason for studying these models is that they allow one to relatively simply include the infinitely dimensional "thermostat," or "heat bath" into a model with a few degrees of freedom. This greatly facilitates both numerical experiments and the theoretical analysis. In particular, a derivation of the Ohm law within this model was presented in [22], thereby solving "one of the outstanding problems of modern physics" [23] (for this peculiar dynamical model only!). The authors of [22] claim that "At present, no general statistical mechanical theory can predict which microscopic dynamics will yield such transport laws…." In my opinion, it would be more correct to inquire which of many relevant models could be treated theoretically, and especially in a rigorous way as was actually done in [22].

The zest of new models is the so-called Gauss thermostat, or heat bath (GHB). In the simplest case, the motion equations of a particle in this bath are [20–22]:

$$\frac{d\mathbf{p}}{dt} = \mathbf{F} - \zeta\mathbf{p}, \quad \zeta = \frac{\mathbf{F}\cdot\mathbf{p}}{p^2}, \tag{2.1}$$

where $\mathbf{F}$ is a given external force and ζ stands for the "friction coefficient." The first peculiarity of this "friction" is in its explicit time reversibility contrary to the "standard friction." The price for reversibility is the strict connection between the two forces, the friction and the external force $\mathbf{F}$. Moreover, and this is most important, the connection is such that

$$|\mathbf{p}|^2 = p_0^2 = \text{const}$$

is the exact motion invariant,

$$\frac{d}{dt}\frac{|\mathbf{p}|^2}{2} = \mathbf{p}\cdot\frac{d\mathbf{p}}{dt} = \mathbf{p}\cdot\mathbf{F} - \mathbf{F}\cdot\mathbf{p}. \tag{2.2}$$

The first of the two identical terms represents the mechanical work of the external regular force $\mathbf{F}$, the spring of the external energy, and the second one describes the sink of energy into GHB. Thus, asymptotically as $t \longrightarrow \infty$, the model describes a steady state only. This is the main restriction of such models. The particle itself does only immediately transfer the energy without any change of its own because of the above constraint

$$|\mathbf{p}|^2 = \text{const}.$$

For a single degree of freedom, the latter would lead to the trivial solution $p = \text{const}$. Therefore, at least two degrees of freedom are required to allow for a variation of vector $\mathbf{p}$ in spite of the constraint. For many interacting particles, the constraint

$$\sum|\mathbf{p}_i|^2 = \text{const}$$

is less stringent, hence the reference to the Gauss "Principle of Least Constraint" [24] for deriving the reversible friction in Eq. (2.1). In the present paper, the simplest case of N noniteracting particles with two degrees of freedom is considered only as in [22].

The next important point is a special form of the energy in GHB, which is the heat. In true heat bath it is given by the chaotic motion of infinitely many particles. This is not the case in GHB, and one needs an additional force in Eq. (2.1) to make the particle motion chaotic, at the same time maintaining the constraint. Whether such an external to GHB chaos is equivalent to the chaos inside the true heat bath, at least statistically, remains an open question, but it seems plausible from the physical viewpoint [22] (see also [25]). If so, the model describes the direct conversion of mechanical work into heat Q, and hence the permanent entropy production. The calculation of the latter is not a trivial question (for discussion, see [20–22]). In our opinion, the simplest way is to use the thermodynamic relation

$$\frac{dS}{dt} = \frac{1}{T}\frac{dQ}{dt}, \quad \frac{dQ}{dt} = \mathbf{p}\cdot\mathbf{F}, \tag{2.3}$$

where $T = p_0^2$ is the effective temperature [22]. Because the input energy is of zero entropy (the formal temperature $T_{in} = \infty$), relation (2.3) determines the entropy production in the whole system (particles + GHB). We note that in Eq. (2.3), as well as throughout this paper, the entropy S is understood to be determined in the standard way via a coarse-grained distribution function.

On the other hand, the usual interpretation of GHB models is quite different [20–22]. Namely, the entropy production in Eq. (2.3) is expressed via the Lyapunov exponents λ_i of the particle motion,

$$\frac{dS}{dt} \equiv \frac{dS_{GHB}}{dt} \equiv -\frac{dS_p}{dt} = -\sum_i \lambda_i, \tag{2.4}$$

where S_{GHB} and S_p are the respective entropy of GHB and of the ensemble of particles. An unpleasant feature

of this relation is in that the latter equality holds for the Gibbs entropy only, which is conserved in the Hamiltonian system modeled by the GHB. As a result, the entropy of the total system (particle + GHB) remains constant (the second equality in Eq. (2.4)), which literally means no entropy production at all! Even though this interpretation can be formally justified, it seems to us to be physically misleading. In our opinion, the application of Lyapunov exponents would be better restricted to characterization of the phase-space fractal microstructure of the particle motion (which is really interesting), retaining the universal coarse-grained definition of the entropy (cf. ES models in Section 1).

As mentioned above, the GHB models describe the nonequilibrium steady states only. Moreover, any collective processes of interacting particles are also excluded, among them those responsible for the very existence of regular nonequilibrium processes, in particular, of field **F** in model (2.1). In a more complicated Nosé–Hoover version of GHB models, these severe restrictions can be partly, but not completely, lifted. Whether this is sufficient for the inclusion of collective processes remains, to my knowledge, an open question.

In any event, even the simplest GHB model like (2.1) represents a qualitatively different type of statistical behavior compared to that in the ES models. The origin of this principal difference is twofold: (i) the external "inexhaustible" spring of energy, if only introduced "by hand", and (ii) a heat sink of infinite capacity that excludes any equilibrium.

In conclusion of this section, we precisely formulate the model considered in the main part of the paper. Choosing the model for numerical experiments, I follow my favored the "golden rule": construct the model as simple as possible but not simpler. In the problem under consideration, the models already studied are mainly based on the well-known and well-studied "Lorentz gas" that is a particle (or many particles) moving through a set of fixed scatterers. A new element is a constant field accelerating the particles. Actually, the Lorentz model becomes the famous Galton Board [26], the very first model of chaotic motion, which was invented by Galton for another purpose, and which has not been studied in detail until recently [20–22]. Our model is still simpler, and is specified by the two maps: (i) the 2D Arnold cat map (1.1) to chaotize particles, and (ii) the 1D map version of Eq. (2.1),

$$\overline{p_1} = p_1 + F - 4Fp_1^2, \tag{2.5}$$

where $p_1 = p - p_0$ and the parameter in Eq. (2.1) is $p_0 = 1/2$. For $|F| < 1/4$, the momentum p remains within the unit interval ($0 \le p \le 1$) as in map (1.1). The principal relation (2.3) for the entropy reduces also to the additional 1D map,

$$\bar{S} = S + (p_1 + F)^2 - p_1^2 = S + 2p_1F + F^2, \tag{2.6}$$

where the entropy unit is changed by the factor 2 for simplicity. Because S is the entropy produced in GHB, the latter map implicitly includes also the motion in the second degree of freedom for each of the noninteracting particles because of the Gauss constraint that guarantees the immediate transfer of energy to GHB.

In numerical experiments considered below, an arbitrary number N of noninteracting particles (trajectories) with random initial conditions was used. In this case, the Gauss constraint remains unchanged, and all the trajectories are run simultaneously.

3. NONMONOTONIC ENTROPY PRODUCTION: LOCAL FLUCTUATIONS

The statistical properties of entropy growth in the model chosen are determined by the first two moments of the p_1 distribution function. In the limit as $t \longrightarrow \infty$ and/or $N \longrightarrow \infty$, they are given by (per iteration and per trajectory)

$$\langle p_1 \rangle = 0, \quad \langle p_1^2 \rangle = \frac{1}{12}, \tag{3.1}$$

where averaging is done over both the motion time t (now the number of the iterations of the map) and N noninteracting particles (particle trajectories). In combination with Eq. (2.6), the first moment in Eq. (3.1) implies the linear growth of the average entropy (per trajectory),

$$\langle S(t) \rangle = tF^2. \tag{3.2}$$

In this section, the statistics of local fluctuations is considered. A similar problem was studied in [27] for a more realistic model with many interacting particles. In the present model, the local fluctuation is defined as follows. The total motion time t_f is subdivided into many segments of equal duration t_1. On each segment $i = 1, \ldots, t_f/t_1$, the total change of the entropy S_i for all N trajectories is calculated using Eq. (2.6) and represented as the dimensionless random variable

$$S_\sigma = \frac{S_i - \langle S_i \rangle}{\sigma} = \frac{S_i - \tau}{\sigma}, \tag{3.3}$$

where

$$\langle S_i \rangle = Nt_1F^2 = \tau$$

(see Eq. (3.2)), and the rms fluctuation σ is given by a simple relation (see Eqs. (2.6) and (3.1))

$$\sigma^2 = \frac{\tau}{3}. \tag{3.4}$$

This relation neglects all the correlations, which implies the standard Gaussian distribution

$$G(S_\sigma) = \frac{\exp(-S_\sigma^2/2)}{\sqrt{2\pi}}. \tag{3.5}$$

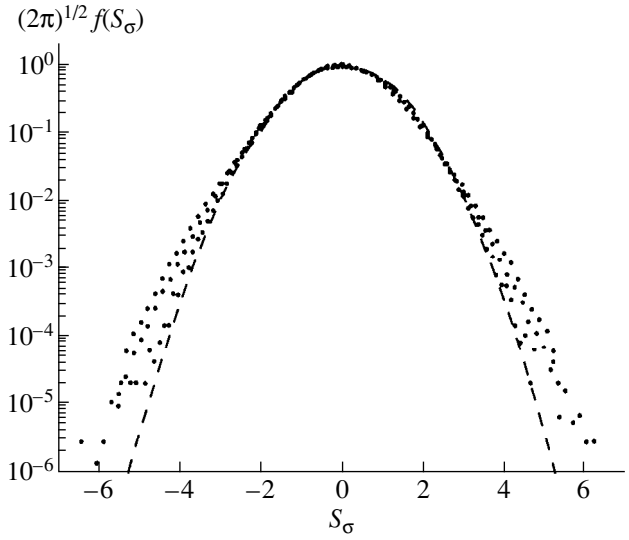

Fig. 2. Distribution function $f(S_\sigma)$ of local fluctuations in the nonequilibrium steady state with $F = 0.01$. Dashed line is the standard Gauss law (3.5); points represent the results of numerical experiments with $N = 1$ and $t_1 = 10, 25, 100$.

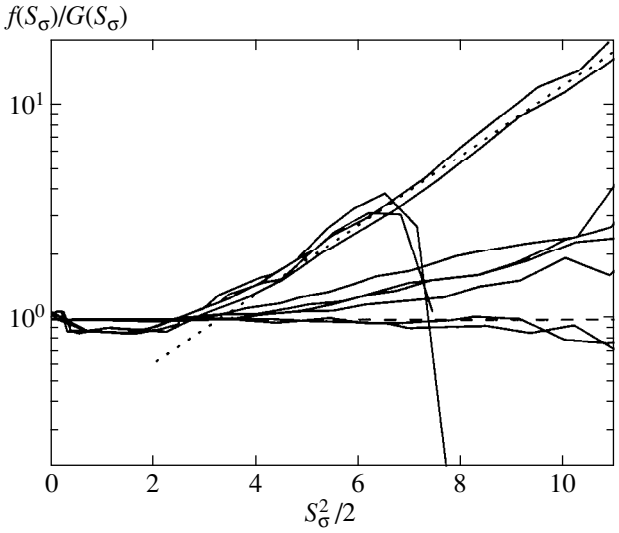

Fig. 3. The ratio of the distribution $f(S_\sigma)$ to the standard Gauss law (3.5) (broken lines). The values of the parameter N/t_1 from top to bottom are: 1/5 ($S_\sigma^2/2 < 7.5$, see text); 1/10; 1/100; 10/10, and 100/1. The oblique dotted straight line demonstrates the Gaussian shape of the tails.

An example of the actual distribution function is shown in Fig. 2 for a single trajectory with the segment length $t_1 = 10, 25, 100$ iterations, and the number of segments up to 10^7. The cap of the distribution is close to the standard Gauss form (3.5) (see also Fig. 3) but both tails clearly show a considerable enhancement of fluctuations depending on both t_1 and N (in other examples, see below).

The shape of the tails is also Gaussian but the width is larger the smaller t_1 and N. This is especially clear in a different representation of the data in Fig. 3, where the ratio of the empirical distribution to the standard Gauss one is plotted as a function of the Gaussian variable $S_G = S_\sigma^2/2$. Each run with particular values of N and t_1 is represented by two slightly different lines for both signs of S_σ. In addition to fluctuations, the difference apparently involves some asymmetry of the distribution with respect to $S_\sigma = 0$. The origin of this asymmetry is not completely clear as yet. A sharp crossover between the two Gaussian distributions at $S_G \approx 3$ is nearly independent of the parameters N and t_1, as is the top distribution below crossover. On the contrary, the tail distribution essentially depends on both parameters in a rather complicated way. The origin of the difference between the two Gaussian distributions apparently lies in dynamical correlations. In spite of a fast decay (see Section 1), the correlation in Arnold map (1.1) does affect somehow the big entropy fluctuations except in the limiting case $N \gg t_1$ (two lower lines in Fig. 3), where the correlations vanish because of random and statistically independent initial conditions of many trajectories.

For any fixed parameters N and t_1, the fluctuations are bounded ($F \ll 1$),

$$|S_\sigma| < \sqrt{3Nt_1}, \tag{3.6}$$

which follows from Eqs. (2.6), (3.3), and (3.4). This is clearly seen in Fig. 3 for minimal $Nt_1 = 5$. If only force F is fixed instead, the relative entropy fluctuations

$$\frac{S_i}{\langle S_i \rangle} \approx \pm\frac{1}{F} \tag{3.7}$$

are also restricted but can be arbitrarily large for small F and, moreover, can have either sign. This implies a nonmonotonic growth of the entropy at the expense of the segments with $S_i < 0$.

The probability (in the number of trajectory segments) of extremely large fluctuations, Eqs. (3.6) and (3.7), is exponentially small (see Eq. (3.5) and below). However, the probability of the fluctuations with a negative entropy change ($S_i < 0$) (without time reversal!) is generally not small at all, reaching 50% as $\tau \longrightarrow 0$ (for arbitrary N and t_1). In principle, this is known, at least for the systems with an equilibrium steady state (Section 1). Nevertheless, the first, to my knowledge, direct observation of this phenomenon in a nonequilibrium steady state [27] has so much staggered the authors that they even entitled the paper "Probability of Second Law violations in Shearing Steady State". In fact, this is simply a sort of peculiar fluctuations that are big not so much with respect to their size but primarily to their probability (cf. discussion in Section 1). However, the important point is that all those negative entropy fluctuations (transforming the heat into work) are randomly

scattered among the others of positive entropy, and for making any use of the former a Maxwell's demon is required who is known by now to be well in a "peaceful coexistence" with the Second Law.

A Gaussian distribution of the entropy fluctuations shifted with respect to $S_i = 0$ in a nonequilibrium steady state first observed in [27] was also theoretically explained there in terms of the Lyapunov exponents (see Eq. (8) in [27]). This was the first form of what is now called the "Fluctuation Theorem" (see, e.g., D. Ruelle in [6, p. 540]). In my opinion, a more physical representation of this theorem would be the ratio of the two moments in Eq. (3.4). In any representation, the theorem essentially depends on both the underlying dynamics and the type of fluctuations considered (see Sections 4 and 5).

Another interesting limit is $t_1 \longrightarrow t_f \longrightarrow \infty$ (a single segment) [27] with $\tau \longrightarrow 0$, which is possible if $F \longrightarrow 0$ too. In this case, the probability of zero entropy change in the entire motion also approaches 50%. However, the probability of any negative entropy fluctuation vanishes (see Eq. (3.3)). An interesting question is whether there exists some intermediate region of parameters where the latter probability remains finite. In other words, are the Poincaré recurrences to negative entropy change $S_i < 0$ possible in a nonequilibrium steady state as these are in the equilibrium (Section 1)? The answer to this question is given by the statistics of the global fluctuations.

4. NONMONOTONIC ENTROPY PRODUCTION: GLOBAL FLUCTUATIONS

The definition of the global fluctuations is similar to, yet essentially different from that of the local fluctuations in the previous section. Namely (cf. Eqs. (3.3) and (3.4)), the principal dimensionless random variable $S_\sigma(t)$ now explicitly depends on time,

$$S_\sigma(t) = \frac{S(t) - \langle S(t)\rangle}{\sigma} = \frac{S(t) - \tau}{\sigma}, \qquad (4.1)$$

where $S(t)$ is calculated from Eq. (2.6), $S(0) = 0$, $\langle S(t)\rangle = NtF^2 \equiv \tau$ (see Eq. (3.2)), and the rms fluctuation σ is given by the same relation (3.4) with a new time variable τ,

$$\sigma^2 = \frac{\tau}{3}. \qquad (4.2)$$

In other words, the global fluctuations are described as a diffusion with the constant rate

$$D = \frac{\sigma^2}{\tau} = \frac{1}{3}. \qquad (4.3)$$

The global fluctuations can also be viewed as a continuous time-dependent deviation of the entropy from its average growth unlike the local fluctuations in the ensemble of fixed trajectory segments (Section 3).

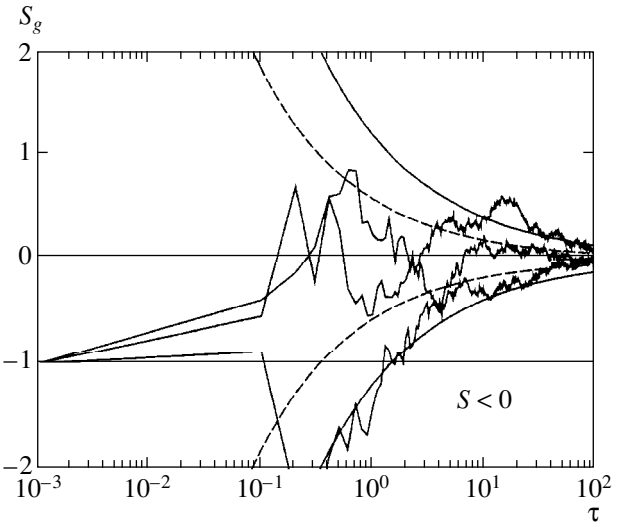

Fig. 4. Time dependence of the reduced global fluctuations $S_g(\tau)$, Eq. (4.4): three sets by $N = 10$ trajectories with different initial conditions but the same initial entropy $S(0) = 0$ and $F = 0.01$. Horizontal solid line $S_g = 0$ represents the average entropy growth. The lower solid line $S = 0$ is the border between positive and negative entropy. A pair of dashed curves corresponds to the standard rms fluctuation σ, Eq. (4.2), and two solid curves represent the maximum diffusion fluctuations σ_b, Eq. (4.5).

Now, the primary goal is to find whether the entropy can reach negative values $S(t) < 0$ as $t \longrightarrow \infty$. As was discussed in the previous section, this is possible at some finite segments of the trajectory with the probability rapidly decreasing (but always finite) as the segment length grows.

In Fig. 4, three examples of global fluctuations are shown in a slightly different representation (cf. Eq. (4.1))

$$S_g(\tau) = \frac{S(\tau)}{\tau} - 1 \qquad (4.4)$$

chosen in order to always keep the most important border $S(\tau) = 0$ in front of one's eyes (with $S_g(\tau) = -1$, the horizontal line in Fig. 4). Eventually, all trajectories converge to the average entropy growth (the horizontal line $S_g = 0$ in Fig. 4). During the initial stage of diffusion, the probability of negative entropy is roughly 50%, similar to the local fluctuations (Section 3). However, the situation cardinally changes at $t \gtrsim 1$, with all the trajectories moving away from the border $S = 0$. Moreover, the relative distance to the border with respect to the fluctuation size increases indefinitely.

The fluctuation size is characterized by two parameters. The first one is the well-known rms dispersion σ, Eq. (4.2) (two dashed curves in Fig 4), which estimates the fluctuation distribution width. In the problem under consideration, the most important is the second characteristic, σ_b (two solid curves in Fig. 4), which sets the maximum size (the upper bound) of the diffusion fluctuations, and therefore insures against the recurrence

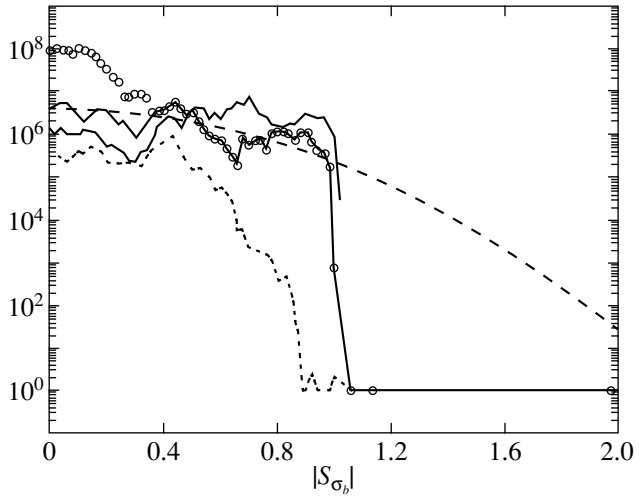

Fig. 5. Histogram of the global fluctuations in the number of entries per bin of the width 0.02: $F = 0.01$; $N = 100$; $R_\sigma \approx 3$. From bottom to top in the left-most part of figure: $\tau = 10^5$ (dashed line); 10^6 (two solid lines, different initial conditions); 10^7 (circles); the total motion time $t = 100\tau$ iterations. For comparison, the smooth dashed line shows unbounded Gaussian distribution (4.7) for $\tau = 10^6$.

into the region $S < 0$ in a sufficiently long time. The ratio of the two sizes

$$R_\sigma(\tau) = \frac{\sigma_b}{\sigma} = \sqrt{2\ln\ln(A\tau)} \qquad (4.5)$$

is given by the famous Khinchin law of iterated logarithms [28].

We emphasize again that the principal peculiarity and importance of the border σ_b is that it characterizes a sharp drop of the fluctuation probability down to zero (in the limit as $\tau \longrightarrow \infty$). In other words, almost any trajectory approaches infinitely many times arbitrarily close to this border from below, but the number of border crossings remains finite. In Fig. 4, this corresponds to the eternal confinement of trajectories in the gap between the two solid curves.

This surprising behavior of random trajectories is well known to mathematicians but, apparently, not to physicists. In Fig. 5 several examples of the fluctuation distributions are shown for illustration of that impenetrable border.

In the Khinchin theorem, factor A in Eq. (4.5) is irrelevant and is set to $A = 1$. This is because the theorem can be proved in the formal limit as $\tau \longrightarrow \infty$, only as most theorems in the probability theory (as well as in the ergodic theory, by the way). However, in numerical experiments on a finite time, even if arbitrarily large, one needs a correction to the limit expression. In addition, it would be desirable to look at the border over the whole motion down to the dynamical time scale determined by the correlation decay. In the model under consideration. it is of the order of the relaxation time $t_r \sim 1$ (see Section 1). The additional parameter A can be fixed by the condition

$$\sigma_b(\tau_1) = \sigma(\tau_1), \quad \tau_1 = NF^2, \qquad (4.6)$$

for minimal $t = 1$ on the dynamical time scale of the diffusion. It then follows from Eq. (4.5) that

$$A\tau_1 = 5.2,$$

which is used in Figs. 4 and 5. The condition assumed is, of course, somewhat arbitrary but the dependence on A remains extremely weak provided $\tau_1 \ll 1$.

The histogram in Fig. 5 is given in the absolute numbers of trajectory entries into bins in order to graphically demonstrate a negligible number of exceptional crossings of the border. The exact formulation of the Khinchin theorem admits a finite number of crossings in infinite time. Actually, all those "exceptions" are concentrated within a relatively short initial time interval $\tau \sim 1$ (for the accepted A value, see Fig. 4).

The distribution of entropy fluctuations between the borders is characterized by its own big fluctuations due to a large time interval ($\sim\tau$) required for crossing the distribution region (see Eq. (4.3)). The spectacular precipice of many orders of magnitude is reminiscent of a diffusion "shock wave" cutting away the Gaussian tail. The unbounded Gauss curve is also shown in Fig. 5 by the smooth dashed line.

In terms of the variable $S_{\sigma_b} = S_\sigma/R_\sigma$, the standard Gauss law is no longer a stationary distribution (cf. Eq. (3.5)),

$$\sqrt{2\pi}G(S_{\sigma_b}) = R_\sigma(\tau)\exp\left(-\frac{S_{\sigma_b}^2}{2}R_\sigma^2(\tau)\right). \qquad (4.7)$$

Both the probability density at the border $|S_{\sigma_b}| = 1$ and the integral probability beyond that are slowly decreasing $\propto 1/\ln(A\tau)$. The "shock wave" decays but still continues to "hold back" the trajectories.

Thus, unlike unrestricted entropy fluctuations out of the equilibrium steady state (Section 1), the strictly restricted fluctuations in the nonequilibrium steady state are well separated, in a short time, from the negative-entropy region, separated in a large excess. that grows in time. In other words, the Poincaré recurrences to any negative entropy quickly and completely disappear leaving the system with ever increasing, even if nonmonotonically, entropy.

As the nonequilibrium steady state involves a heat bath of the infinite phase-space volume (or its nice substitute, the Gauss heat bath), the Poincaré recurrence theorem is not applicable. However, the "anti-recurrence" theorem is not generally true either. For example, the entropy repeatedly crosses the line $S = \tau$ of the average growth in spite of the infinite heat bath, yet it does not do so for the line $S = 0$ of the initial entropy.

We note that the new ratio $\sigma_b^2 \langle S(t)\rangle$ (cf. Eq. (3.4)) represents another "Fluctuation Theorem" as compared to the known one mentioned in Section 3.

5. BIG ENTROPY FLUCTUATIONS IN CRITICAL DYNAMICS

The strict restriction of the global entropy fluctuations in a nonequilibrium steady state considered in the previous section is a result of the "normal," Gaussian, diffusion of the entropy with a constant rate (4.3) and with the surprising impenetrable border (4.5). In turn, this is related to a particular underlying dynamics of model (1.1) with very strong statistical properties. We note that the border (4.5) has a statistical nature because it is much less than the maximum dynamical fluctuation (3.7).

However, it is well known by now that the homogeneous diffusion can in general be "abnormal" in the sense that the diffusion rate depends on time,

$$D(t) \propto t^{c_D}, \quad -1 \le c_D \le 1, \tag{5.1}$$

where c_D is the so-called critical diffusion exponent. The term "critical" refers to a particular class of such systems with a very intricate and specific structure of the phase space (see, e.g., [29] and references therein). The "normal" diffusion corresponds to $c_D = 0$, while a positive $c_D > 0$ represents a superfast diffusion with the upper bound $c_D = +1$, the maximum diffusion rate possible for a homogeneous diffusion. The latter is, of course, the most interesting case for the problem under consideration here. A superslow diffusion for a negative $c_D < 0$ is also possible with the limit $c_D = -1$, which means the absence of any diffusion for $c_D < -1$. An interesting example of a superslow diffusion with $c_D = -1/2$ was considered in [30]. Besides a particular application to the plasma confinement in magnetic field, the example is of a special interest because this slow diffusion is the result of the time-reversible diffusion of particles in a chaotic magnetic field. For other examples and various discussions of abnormal diffusion, see [31].

A number of dynamical models exhibiting the superfast diffusion are known including the limiting case $c_D = 1$ [29, 32]. Interestingly, a simple simulation of the abnormal diffusion is possible by a minor modification of the model under consideration. It concerns the additional 1D map (2.6) only, which now becomes

$$\bar{S} = S + (2p_1F + F^2)t_s, \tag{5.2}$$

where the new variable t_s is defined by a simple relation

$$t_s = s^{-c_s}, \quad s = 1 - 2|p_1|, \tag{5.3}$$

with s being the distance from any of the two borders $p_1 = \pm 0.5$ homogeneously distributed within the interval $(0 < s < 1)$. The quantity $t_s > 1$ describes the sticking of a trajectory in the "critical structure" concentrated near $s = 0$. Actually, the model does not involve this structure, however its effect is simulated by the "sticking time" t_s that enhances both the fluctuations and the average entropy (5.2). In a sense, this simulation is similar in spirit to that of the Gauss heat bath. All the properties of that sticking are described by a single parameter c_s, the critical sticking exponent $(0 \le c_s \le 1)$. In particular, it is directly related to the diffusion exponent c_D (see below).

The statistical properties of the abnormal diffusion in this model are determined by the first two moments of the t_s distribution, which can be directly evaluated from the above relations as follows. For the first moment, we have

$$\langle t_s\rangle = \int_0^1 t_s(s)ds = \frac{1}{1-c_s}, \quad c_s < 1, \tag{5.4a}$$

and

$$\langle t_s\rangle \approx \ln\frac{1}{s_1} \approx \ln t, \quad c_s = 1. \tag{5.4b}$$

In the latter case the integral diverges and is determined by the minimum $s \approx s_1 \sim 1/t$ reached over time t that is the total motion time in the iterations of the map. It must be distinguished from the "physical time" in a true model of the critical structure,

$$\tilde{t} \approx t\langle t_s\rangle \approx \begin{cases} \dfrac{t}{1-c_s}, & c_s < 1, \\ t\ln t, & c_s = 1. \end{cases} \tag{5.5}$$

Similarly, the second moment is given by three relations:

$$\langle t_s^2\rangle = \frac{1}{1-2c_s}, \quad c_s < \frac{1}{2}, \tag{5.6a}$$

for the normal diffusion,

$$\langle t_s^2\rangle \approx \ln\frac{1}{s_1} \approx \ln t, \quad c_s = \frac{1}{2}, \tag{5.6b}$$

in the critical case, and

$$\langle t_s^2\rangle \approx \frac{s_1^{1-2c_s}}{2c_s-1} \approx \frac{t^{2c_s-1}}{2c_s-1}, \quad \frac{1}{2} < c_s \le 1, \tag{5.6c}$$

for the superfast diffusion.

The average entropy production is found from Eq. (5.2) as

$$\langle S(t)\rangle = NF^2t\langle t_s\rangle = NF^2\tilde{t} \equiv \tau, \tag{5.7}$$

with the redefined time variable τ (cf. Eq. (3.3)). In this section, we only consider the simplest case of a single trajectory ($N = 1$).

Evaluating the superfast diffusion requires a slightly different averaging $\langle(2p_1t_s)^2\rangle$ (see Eq. (5.2)). However,

it is easily verified that asymptotically as $\tau \longrightarrow \infty$, the difference with respect to Eq. (5.6c) vanishes, and one arrives at the following estimate for the critical rms dispersion σ_{cr}:

$$\frac{\sigma_{cr}^2(\tau)}{B^2} = \tilde{t}D(\tilde{t}) = F^2\langle t_s^2\rangle t = \frac{(1-s_s)^{2c_s}}{2c_s-1}\frac{\tau^{2c_s}}{F^{4c_s-2}} \quad (5.8a)$$

if $1/2 < c_s < 1$ (5.6c), and

$$\frac{\sigma_{cr}(\tau)}{B} = \frac{\tau}{F\ln(\tau/F^2)} \quad (5.8b)$$

in the most interesting limiting case where $c_s = 1$. The empirical factor $B \sim 1$ accounts for all the approximations in the above relations.

The limit as $c_s \longrightarrow 1$ in Eq. (5.8a) crucially differs from the limiting relation (5.8b). The origin of this discrepancy is Eq. (5.4a). A more accurate evaluation for $c_s \approx 1$ reads

$$\langle t_s\rangle = \int_{s_1}^{1} t_s(s)ds = \frac{1-s_1^{1-c_s}}{1-c_s} = \frac{1-\exp[(1-c_s)\ln s_1]}{1-c_s}, \quad (5.9)$$

where $s_1 \sim 1$ is the minimum s over t iterations of the map (cf. Eq. (5.4b)). Relation (5.4a) is therefore valid under the condition $\epsilon \ln t > 1$ only (with $\epsilon = 1 - c_s$), while in the opposite limit, we have $\langle t_s\rangle \approx \ln t$ as for $c_s = 1$, Eq. (5.4b). The crossover between the two scalings occurs at

$$t_{cro} \sim e^{1/\epsilon}, \quad \tau_{cro} \sim \frac{e^{1/\epsilon}}{\epsilon}F^2. \quad (5.10)$$

The deviation from Eq. (5.8a) is essential for a sufficiently small ϵ only.

The ratio of fluctuations to the average entropy production is given by the reduced entropy (see Eq. (4.4))

$$S_g = \pm\frac{\sigma_{cr}}{\tau} \approx \frac{B}{F\ln(\tau/F^2)}, \quad (5.11)$$

where the latter expression is estimate (5.8b) for the rms fluctuations. They are slowly decreasing with time, and at

$$\tau \gtrsim \tau_0 = F^2\exp(1/F),$$

the rms line crosses the border $S_g = -1$ of zero entropy. Afterwards, the entropy remains mainly positive. To be more precise, the probability for a trajectory to enter into the negative-entropy region is systematically decreasing with time, although rather slowly. This must be compared with the F-independent crossover $\tau_0 = 1/3$ and a rapid drop of the probability to return to $S < 0$ for the normal diffusion (Section 4).

However, there exists another mechanism of big fluctuations, specific for the critical dynamics. Namely, a separated individual fluctuation can be produced as the result of a single extremely big sticking time t_s over the total motion up to the moment the fluctuation springs up in a single map iteration. We recall that in the present model, each sticking corresponds to just one map iteration. The increments of dynamical variables in this jump are obtained from Eq. (5.2) as

$$\Delta S = \pm Ft_s, \quad \Delta\tau = F^2t_s, \quad (5.12)$$

where $t_s \gg 1$ (with $2p_1 \approx 1$) is assumed (a big fluctuation). The reduced fluctuation is then given by

$$S_g \approx \frac{S}{\tau} = \pm\frac{Ft_s}{\tau + F^2t_s} \approx \pm\frac{1/F}{1+\tau/\Delta\tau}. \quad (5.13)$$

The maximum single sticking time over the motion time t is, on the average,

$$\langle t_s\rangle \approx t\ln t = \frac{\tau}{F^2}. \quad (5.14)$$

Therefore, a single fluctuation (5.13) has the upper bound

$$|S_g| \lesssim \frac{A}{F}, \quad (5.15)$$

where an empirical factor $A \sim 1$ is introduced similarly to Eq. (5.8b).

The border (5.15) considerably exceeds the rms diffusion fluctuation (5.11) and, even more importantly, the former never crosses the zero-entropy line $S_g = -1$. Therefore, the critical fluctuations repeatedly bring the system into the negative-entropy region. This is because the upper bound (5.15) does not depend on time τ provided that $\Delta\tau \gtrsim \tau$ in Eq. (5.13). However, in a chain of successive fluctuations, the values of τ in Eqs. (5.13) and (5.14) are not generally equal. While in the former relation it is always the total motion time as assumed above, it must be the preceeding period of fluctuations in Eq. (5.14): $\tau_n \longrightarrow P_n < \tau_n$, where n is the serial number of fluctuations. Hence, the approach to the upper bound (5.15) is only possible under the condition $P_n \gg P_{n-1}$, which implies $P_n \approx \tau_n$. Thus, the fluctuations become more and more rare with the period growing exponentially in time. In other words, the fluctuations are stationary in $\ln\tau$ with a sufficiently big mean period $\langle\ln P\rangle \approx 5$ (see Fig. 6).

In Fig. 6, an example of several big critical fluctuations in the limiting case $c_s = 1$ is presented for five single sufficiently long trajectories with different initial conditions and the motion time up to $\tau \approx 5 \times 10^9$ and $t = 10^{10}$ iterations. To achieve such a long time, the force was increased up to $F = 0.1$ (see Eq. (5.14)).

Unlike a similar Fig. 4 for the normal diffusion, only several big fluctuations with $F|S_g| > 0.3$ are presented in Fig. 6. For the full picture of critical fluctuations, the

required output becomes formidably long. The distribution of all fluctuations, independent of time, is shown in Fig. 7.

Each fluctuation in Fig. 6 is presented by a pair of FS_g values connected by the straight line: one at a map iteration just before the fluctuation (circles), and the other (stars) at the next iteration when the fluctuation springs up (see above). Both are plotted at the same, latter, τ to follow the pairs. This slightly shifts the circles to the right.

The most important, if only preliminary, result of numerical experiments is the confirmation of the fluctuation upper bound (5.15) that is independent of time. As expected, the circles represent considerably smaller $F|S_g|$ values, roughly following the diffusive scaling (5.11).

The border (5.15) qualitatively reminds the strict upper bound for the normal diffusion (Section 4), including a logarithmic ratio with respect to the rms size (4.5), as compared to the ratio

$$R_{cr}(\tau) \approx \ln(\tau/F^2) \tag{5.16}$$

in the critical diffusion. An interesting question whether the new, critical, border is also as strict as the old one in the normal diffusion remains, to my knowledge, open, at least for the physical model under consideration where the superdiffusion is caused by a strong long-term correlation of successive entropy changes due to the sticking of trajectory.

However, for a much simpler problem of statistically independent changes, various generalizations of Khinchin theorem to the abnormal diffusion were proved by many mathematicians (see, e.g., [33]). In the present model, this is precisely the case for description in the map time t with statistically independent iterations. The most general and complete result was recently obtained by Borovkov [34]. In the present notation, it can be approximately represented in a very simple form for the ratio

$$R_{cr} = \frac{\sigma_b}{\sigma} \sim (\ln t)^{c_s} \tag{5.17}$$

in the entire superdiffusion interval ($1/2 < c_s \le 1$). For the most important reduced fluctuation (5.13), we then arrive at the two relations

$$|S_g| \lesssim \frac{\sigma_b}{\tau} \sim \frac{\tau^{c_s-1}}{F^{2c_s-1}}\left(\ln\frac{\tau}{F^2}\right)^{c_s} \tag{5.18a}$$

for $c_s < 1$ and

$$|S_g| \lesssim \frac{\sigma_b}{\tau} \sim \frac{1}{F} \tag{5.18b}$$

in the limiting case $c_s = 1$. The latter confirms estimate (5.15), which, in turn, is in a good agreement with the empirical data in Fig. 6. In any event, a simple physical estimate (5.15) seems to provide an efficient description of the fluctuation upper bound.

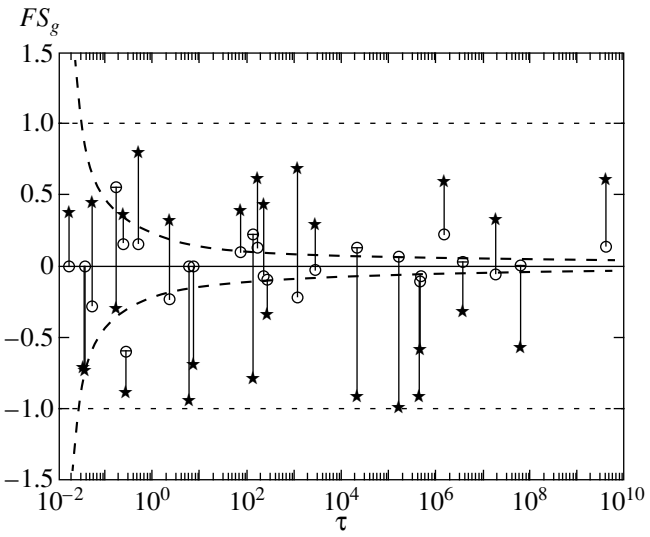

Fig. 6. Time dependence of 26 big fluctuations in critical dynamics: 5 single trajectories up to 10^{10} iterations, $c_s = 1$, $F = 0.1$. Only fluctuations with $F|S_g| > 0.3$ are shown, each by a pair of points connected by the straight line: the big fluctuation itself (stars) and at the preceding map iteration (circles, see text). Two dashed curves show the rms fluctuations of $F|S_g|$, Eq. (5.11), with $B = 1$. Horizontal dotted lines mark the upper bound, Eq. (5.15), with $A = 1$.

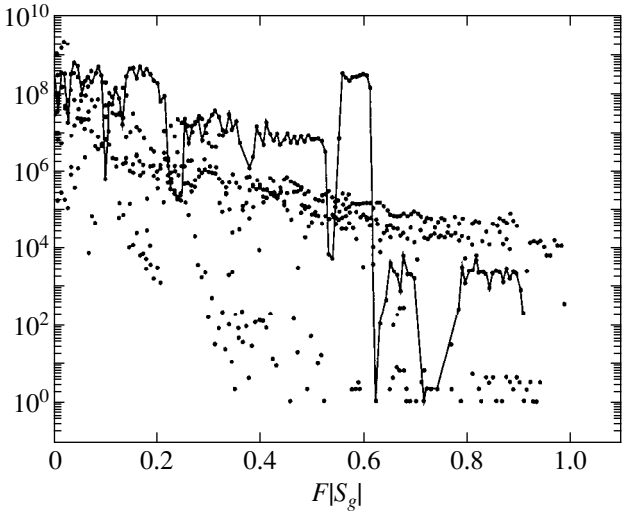

Fig. 7. Histogram of critical fluctuations in the number of entries per bin of width 0.007 for the data in Fig. 6. The border $S = 0$ corresponds to $F|S_g| = -F = -0.1$. The points for the longest trajectory are connected by line.

In Fig. 7, an example of all (at each map's iteration) fluctuations is shown for the data from the same runs as in Fig. 6. In addition to very large overall distribution fluctuations, a sharp drop by about four orders of magnitude is clearly seen near the expected upper

bound (5.15). It is similar to the drop in Fig. 5 for the normal diffusion.

Thus, the critical diffusion results in infinitely many recurrences far into the negative-entropy region $S < 0$ (for $F \ll 1$), the sojourn time in that region being comparable to the total motion time. Of course, the former is less than 50% on the average, so that asymptotically in time the entropy is always growing. In this respect, the global critical fluctuations are similar to the local ones in the normal diffusion (Section 3).

We note, however, that the upper bound $\sigma_b/\tau \sim 1/F$ (5.18b) is permanent in the strict limit $c_s = 1$ only. For any deviation from the limit $\epsilon = 1 - c_s > 0$, this bound lasts a finite time determined by the crossover (5.10) ($\tau \lesssim F^2\exp(1/\epsilon)/\epsilon$) to decreasing $\sigma_b/\tau \longrightarrow 0$, Eq. (5.18a). Another interesting representation of this intermediate behavior is the crossover in the sticking exponent,

$$\epsilon \lesssim \frac{1}{\ln(\tau/F^2)} \approx F|S_g|, \tag{5.19}$$

which is actually shown in Fig. 6 by the upper dashed line. For the longest $\tau = 5 \times 10^9$, the latter crossover is $\epsilon_{cro} \approx 0.037$.

Again, the new cardinally different critical ratio $\sigma_b^2/\langle S(t)\rangle$ and the distribution of entropy fluctuations lead to yet another "Fluctuation Theorem" as compared to the two previous ones mentioned in Sections 3 and 4.

6. DISCUSSION AND CONCLUSIONS

In the present paper, the results of extensive numerical experiments on big entropy fluctuations in a nonequilibrium steady state of classical dynamical systems are presented and their peculiarities are analyzed and discussed. For comparison, some similar results for the equilibrium steady state are briefly described in the Introduction (they will be published in detail elsewhere [10]).

All numerical experiments have been carried out on the basis of a very simple model, the Arnold cat map (1.1) on a unit torus, with only three minor, but important, modifications that allowed comprising all the problems under consideration. The modifications are:

(1) The expansion of the torus in the p direction (1.2), which allows more impressive diffusive fluctuations out of the equilibrium steady state (Fig. 1 in Section 1).

(2) The addition of 1D map (2.5) with the constant driving force F and with an ingenious time-reversible friction force that represents the so-called Gauss heat bath and which allows modeling a physical thermostat of infinitely many degrees of freedom [20, 21]. This is the principal modification in the present studies of fluctuations in a nonequilibrium steady state (Sections 3–5).

(3) The addition of a new parameter t_s, Eq. (5.3), in map (5.2) which allows for the study of very unusual fluctuations of an "abnormal," critical, dynamical diffusion (Section 5).

Big fluctuations in the equilibrium steady state are briefly considered in Section 1. The simplest one of this class, which we call the Boltzmann fluctuation, is shown in Fig. 1. It is obviously symmetric under time reversal, and at least in this case, therefore, there is no physical reason for the notorious "time arrow" concept. Nevertheless, a related concept, for example, the thermodynamic arrow, pointing in the direction of the average increase of entropy, makes sense in spite of the time symmetry. The point is that the relaxation time of the fluctuation is determined by the model parameter C only, and does not depend on the fluctuation itself. On the contrary, the expectation time for a given fluctuation, or the mean period between successive fluctuations, rapidly grows with the fluctuation size and with the number of trajectories (or degrees of freedom).

Besides the simplest Boltzmann fluctuation, various others are also possible, typically with a much smaller probability. One of those—the two correlated Boltzmann fluctuations, which we call the Schulman fluctuation—was recently described in [36] using the same Arnold cat map. However, this model is not related to cosmology as was speculated in [36]. At least, the Universe and most of the macroscopic phenomena therein require qualitatively different models, ones without an equilibrium steady state. These structures do appear (with a probability of 1) as a result of certain regular collective processes that lead to very complicated nonequilibrium and inhomogeneous states with ever increasing entropy. This is in contrast with a constant, on the average, entropy in ES systems.

A nonequilibrium steady state, the main subject of this paper, is but a little, characteristic though, piece of the chaotic collective processes. In model (2.5), the driving force F represents a result of some preceding collective processes, the spring of free energy, and the Gauss friction does so for an infinite environment around, the sink of the energy, converting the work into heat, on the average. An interesting peculiarity of these systems is that the big fluctuations can, and under certain conditions, do the opposite, converting some heat back into the work.

Two types of fluctuations were studied:

(i) the local ones on a set of trajectory segments of length-t_1 iterations and of the entropy change S_i (Section 3), and

(ii) ones of the global entropy $S(t)$ along a trajectory with respect to the initial entropy set to zero, $S(0) = 0$ (Sections 4 and 5).

The former were found to have a stationary unrestricted distribution close to the standard Gauss law with some enhancement of an unknown mechanism for large fluctuations. The study of the latter effect will be continued. The distribution is symmetric with respect

to the average entropy, growing in proportion to time in agreement with previous studies on a more complicated (and more realistic) model [27]. Even though the distribution is asymmetric with respect to zero entropy change, the probability of negative $S_i < 0$ is generally not small provided $F^2 Nt_1 \lesssim 1$. This phenomenon, apparently a new one in the nonequilibrium steady state, was first observed in [27] but has been interpreted there as a violation of the Second Law. It seems to be the reflection of a common, but wrong in my opinion, understanding of the Second Law as a monotonic growth of entropy, neglecting all the fluctuations including the large ones. The nonmonotonic rise of entropy is clearly seen, for instance, in Fig. 4, and discussed in detail in Sections 3 and 4.

The behavior of global entropy is completely different as the data in the same Fig. 4 demonstrate (Section 4). Although the entropy evolution remains nonmonotonic, it quickly crosses the line of the initial zero entropy and does not return into the negative entropy region $S < 0$. This is insured by the famous Khinchin theorem about the strict upper bound for the diffusion process. At least for physicists, this limitation of a statistical nature for a random motion is surprising and apparently less known. That unidirectional evolution is the most important distinction of the nonequilibrium steady states from the equilibrium ones. In particular, it leads to a certain asymmetry of the entropy distribution sometimes called the "Fluctuation Theorem" or "Fluctuation Law". However, one should bear in mind that this law essentially depends on the underlying dynamics as briefly discussed in Sections 3–5.

This characteristic feature of nonequilibrium steady state further justifies the concept of the thermodynamic arrow pointing to a larger, on the average, entropy. Yet, again it is not related to the properties of time. Of course, the entropy will systematically decrease upon formal time reversal, which is also the case with the model under consideration because the Gauss heat bath is time reversible. Within the steady state approximation, or rather restriction, this would be an infinitely large fluctuation that never comes to the end. However, this fluctuation would never occur either, as a result of the natural time evolution of the system, opposite to the case of equilibrium fluctuations. The ultimate origin of that crucial difference is that the former process, even asymptotically in time, is a tiny little part of the full underlying dynamics of an infinite system. In particular, the initial state $S(0) = 0$ is not a result of the preceding fluctuation, as is the case in ES, but has been eventually caused, for instance, by instability of the initial ES at a very remote time in the past. If one imagined the time reversal at that instant, nothing would change because the thermodynamic arrow does not depend on the direction of time provided, of course, the time reversible fundamental dynamics. Precisely this universal overall dynamics unifies the time for all the interacting objects like particles and fields throughout the Universe. In particular, it is incompatible with the two opposite time arrows (an old Boltzmann's hypothesis [2] that still has some adherents [36]).

Coming back to nonequilibrium steady states, it is worth mentioning that the regularities of the fluctuations in those, both local and global, can be applied, at least qualitatively, to a small part of a big fluctuation in a statistical equilibrium (Fig. 1) on both sides of the maximum. This interesting question will be considered in detail elsewhere [10].

Finally, some preliminary numerical experiments on the global entropy fluctuations and the theoretical analysis were carried out in a special case of the critical dynamics, which turned out to be the most interesting one for the problem in question (Section 5). The point is that the critical dynamics leads to the "abnormal" superdiffusion with the rate $D \propto \tau^{2c_s - 1}$ and the rms fluctuation size $\sigma_{cr} \propto \tau^{c_s}$, where c_s is a new parameter of the third model ($1/2 < c_s \le 1$). This implies that for $c_s \approx 1$, the reduced entropy $|S_g| \propto \tau^{c_s - 1}$ decreases very slowly compared to the normal diffusion $|S_g| \propto 1/\sqrt{\tau}$. In the limiting case where $c_s = 1$, the entropy $|S_g| \propto 1/\ln\tau$ is still decreasing. However, in addition to diffusive fluctuations, there is a set of infinitely many separated fluctuations whose size does not decrease with time (Fig. 6). In other words, these preliminary numerical experiments suggest that in the limiting case of the critical dynamics, the Poincaré recurrences to the initial state $S = 0$ and beyond repeatedly occur without limit. These are preliminary results to be confirmed and further studied in detail.

In this paper, we only considered the fluctuations in classical mechanics. In general, the quantum fluctuations must be significantly different. However, according to the Correspondence Principle, the dynamics and statistics of a quantum system in the semiclassical regime must be close to the classical ones on the appropriate, generally finite, time scales (for details, see [12, 35]). Interestingly, the computer classical dynamics (that is, the simulation of a classical dynamical system on digital computer) is of a qualitatively similar character. This is because any quantity is discrete ("overquantized") in the computer representation. As a result, the correspondence between the classical continuous dynamics and its computer representation in numerical experiments is restricted to certain finite time scales as in quantum mechanics (see the first two references in [35]).

The discreteness of computer phase space leads to another peculiar phenomenon: generally, the computer dynamics is irreversible because of the rounding-off operation unless the special algorithm is used in numerical experiments. Nevertheless, this does not affect the statistical properties of chaotic computer dynamics. In particular, the statistical laws in computer representation remain time-reversible in spite of the (nondissipative) irreversibility of the underlying dynamics. This

simple example demonstrates that contrary to a common belief, the statistical reversibility is a more general property than the dynamical one.

In the very conclusion, we briefly remark on a very difficult, complicated and vague problem, the so-called (physical) causality principle, i.e., the time-ordering of the cause and the effect. A detailed discussion of this important problem will be published elsewhere [37]. We only note the example of a simple Boltzmann fluctuation shown in Fig. 1. I adhere to the idea of statistical nature of causality. Indeed, the cause is, by definition, an "absolutely" independent event that is only possible in the chaotic dynamics. Moreover, the concept of cause loses its usual physical meaning in any purely dynamical description. For example, the initial conditions precisely determine the entire infinite trajectory ($-\infty < t < \infty$), i.e., both the future and the past of such a "cause." For a single Boltzmann fluctuation, an appropriate cause is the minimum entropy (at $t = t_i$ in Fig. 1). This was exactly the procedure used in numerical experiments for the location of a fluctuation of an approximately given size. The principal difference from the exact dynamical initial conditions is that the former cause is an approximate (e.g., average) fluctuation size, which is sufficient for the complete statistical description of the fluctuation, however it leaves enough freedom for the independence from other events, including the preceding fluctuations. However, this cause determines not only the future relaxation of the fluctuation (in agreement with the causality principle) but also the past rise of the same fluctuation, which is a violation of causality, or acausality (spontaneous rise of a fluctuation), or anti-causality, which is perhaps the most appropriate term. Upon the time reversal, the causality/anticausality exchange, which allows for the concept of the causality arrow, however this is not related to the physical time. In this philosophy, the directions of the thermodynamic and causal arrows, coincide independently of the direction of time. An important point of this philosophy is that the "arrow" concept is related to the interpretation of a physical phenomenon rather than to the phenomenon itself. In particular, the question "how to fix or maintain the arrow" [36] is up to the researcher alone. In a more complicated Schulman's double fluctuation, the causality mechanism becomes more interesting [36], and will be discussed in [37] from a different point of view.

I am grateful to Wm. Hoover for attracting my attention to a new class of highly efficient dynamical models with the Gauss heat bath and for stimulating discussions and suggestions. I very much appreciate the initial collaboration with O.V. Zhirov. I am also indebted to A.A. Borovkov for elucidation of Khinchin theorem and of its recent generalizations to the "abnormal" superdiffusion.

REFERENCES

1. D. G. Luchinsky, P. McKlintock, and M. I. Dykman, Rep. Prog. Phys. **61**, 889 (1998).
2. L. Boltzmann, *Vorlesungen über Gastheorie* (J. A. Barth, Leipzig, 1896–98); *Lectures on Gas Theory* (Univ. of California Press, Berkeley, 1964; Gostekhizdat, Moscow, 1956).
3. E. Schrödinger, *Über die Umkehrung der Naturgesetze* (Sitzungsber. Preuss. Akad. Wiss., 1931), p. 144.
4. A. N. Kolmogoroff, Math. Ann. **113**, 766 (1937); **112**, 155 (1936); *Selected Papers on Probability Theory and Mathematical Statistics*, Ed. by Yu. V. Prokhorov (Nauka, Moscow, 1986), p. 197, p. 173.
5. A. M. Yaglom, Dokl. Akad. Nauk SSSR **56**, 347 (1947); Mat. Sb. **24**, 457 (1949).
6. *Proceedings of the 20th IUPAP International Conference on Statistical Physics, Paris, 1998*, Physica A (Amsterdam) **263**, 516 (1999).
7. I. P. Kornfeld, S. V. Fomin, and Ya. G. Sinai, *Ergodic Theory* (Nauka, Moscow, 1980; Springer-Verlag, New York, 1982).
8. V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968; RCD, Izhevsk, 1999).
9. A. Lichtenberg and M. Lieberman, *Regular and Chaotic Dynamics* (Springer-Verlag, New York, 1992); *Regular and Stochastic Motion* (Springer-Verlag, New York, 1982; Mir, Moscow, 1984).
10. B. V. Chirikov and O. V. Zhirov, E-print archives nlin-CD/0010056.
11. D. V. Anosov, Dokl. Akad. Nauk SSSR **145**, 707 (1962).
12. B. V. Chirikov, in *Law and Prediction in the Light of Chaos Research*, Ed. by P. Weingartner and G. Schurz (Springer-Verlag, Berlin, 1996), p. 10; Open Syst. Inf. Dyn. **4**, 241 (1997); E-print archives chao-dyn/9705003.
13. L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Nauka, Moscow, 1995; Pergamon, Oxford, 1980), Part 1.
14. B. V. Chirikov, Wiss. Z. Humboldt Univ. Berlin, Ges.-Sprachwiss. Reihe **24**, 215 (1975).
15. J. Jeans, Philos. Trans. R. Soc. London, Ser. A **199**, 1 (1929).
16. H. Haken, *Synergetics* (Springer-Verlag, Berlin, 1978; Mir, Moscow, 1980).
17. A. Turing, Philos. Trans. R. Soc. London, Ser. B **237**, 37 (1952); G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems* (Wiley, New York, 1977).
18. A. Cottrell, in *The Encyclopedia of Ignorance*, Ed. by R. Duncan and M. Weston-Smith (Pergamon, New York, 1977), p. 129.
19. P. Glansdorf and I. Prigogine, *Thermodynamic Theory of Structure, Stability, and Fluctuations* (Wiley, New York, 1971; Mir, Moscow, 1972).
20. D. Evans and G. Morriss, *Statistical Mechanics of Nonequilibrium Liquids* (Academic, New York, 1990); Wm. G. Hoover, *Computational Statistical Mechanics* (Elsevier, New York, 1991); in *Proceedings of the Workshop on Time-Reversal Symmetry in Dynamical Systems, Warwick, 1996*; Physica D (Amsterdam) **112** (1–2), 225 (1998).

21. Wm. Hoover, *Time Reversibility, Computer Simulation, and Chaos* (World Scientific, Singapore, 1999).
22. N. I. Chernov, G. Eyink, J. Lebowitz, and Ya. G. Sinai, Phys. Rev. Lett. **70**, 2209 (1993).
23. R. Peierls, in *Theoretical Physics in Twentieth Century*, Ed. by M. Fiera and V. Weisskopf (Wiley, New York, 1961).
24. K. Gauss, J. Reine Angew. Math. **4**, 232 (1829).
25. Wm. Hoover, Phys. Lett. A **255**, 37 (1999).
26. F. Galton, *Natural Inheritance* (Macmillan, London, 1889).
27. D. Evans, E. Cohen, and G. Morriss, Phys. Rev. Lett. **71**, 2401 (1993).
28. A. Ya. Khinchin, C. R. Hebd. Seances Acad. Sci. **178**, 617 (1924); in *Selected Papers on Probability Theory*, Ed. by B. V. Gnedenko and A. M. Zubkov (TVP, Moscow, 1995), p. 10.; M. Loéve, *Probability Theory* (Van Nostrand, Princeton, 1955); A. A. Borovkov, *Probability Theory* (Nauka, Moscow, 1986).
29. B. V. Chirikov and D. L. Shepelyansky, Physica D (Amsterdam) **13**, 395 (1984); B. V. Chirikov, Chaos, Solitons and Fractals **1**, 79 (1991).
30. A. Rechester and M. Rosenbluth, Phys. Rev. Lett. **40**, 38 (1978); B. V. Chirikov, Open Syst. Inf. Dyn. **4**, 241 (1997); E-print archives chao-dyn/9705003.
31. P. Levy, *Théorie de l'Addition des Variables Eléatoires* (Gauthier-Villiers, Paris, 1937); T. Geisel, J. Nierwetberg, and A. Zacherl, Phys. Rev. Lett. **54**, 616 (1985); R. Pasmanter, Fluid Dyn. Res. **3**, 320 (1985); Y. Ichikawa *et al.*, Physica D (Amsterdam) **29**, 247 (1987); R. Voss, Physica D (Amsterdam) **38**, 362 (1989); G. M. Zaslavskiĭ, M. Yu. Zakharov, A. I. Neĭshtadt, *et al.*, Zh. Éksp. Teor. Fiz. **96**, 1563 (1989) [Sov. Phys. JETP **69**, 885 (1989)]; H. Mori *et al.*, Prog. Theor. Phys. Suppl., No. 99, 1 (1989).
32. B. V. Chirikov, Zh. Éksp. Teor. Fiz. **110**, 1174 (1996) [JETP **83**, 646 (1996)]; B. V. Chirikov and D. L. Shepelyansky, Phys. Rev. Lett. **82**, 528 (1999).
33. J. Chover, Proc. Am. Math. Soc. **17**, 441 (1966); T. Mikosh, Vestn. Leningrad. Univ., No. 13, 35 (1984); Yu. S. Khokhlov, Vestn. Mosk. Univ., No. 3, 62 (1995).
34. A. A. Borovkov, submitted to Sib. Mat. Zh. (2000).
35. B. V. Chirikov, F. M. Izrailev, and D. L. Shepelyansky, Sov. Sci. Rev., Sect. C **2**, 209 (1981); B. V. Chirikov, in *Lectures in Les Houches Summer School on Chaos and Quantum Physics, 1989* (Elsevier, Amsterdam, 1991), p. 443; G. Casati and B. V. Chirikov, in *Quantum Chaos: Between Order and Disorder*, Ed. by G. Casati and B. V. Chirikov (Cambridge Univ. Press, Cambridge, 1995), p. 3; Physica D (Amsterdam) **86**, 220 (1995); B. V. Chirikov, in *Proceedings of the International Conference on Nonlinear Dynamics, Chaotic and Complex Systems, Zakopane, 1995*, Ed. by E. Infeld, R. Zelazny, and A. Galkowski (Cambridge Univ. Press, Cambridge, 1997), p. 149; B. V. Chirikov and F. Vivaldi, Physica D (Amsterdam) **129**, 223 (1999).
36. L. S. Schulman, Phys. Rev. Lett. **83**, 5419 (1999); G. Casati, B. V. Chirikov, and O. V. Zhirov, Phys. Rev. Lett. **85**, 896 (2000); L. S. Schulman, Phys. Rev. Lett. **85**, 897 (2000).
37. B. V. Chirikov, private communication (2000).

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 20–27.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 27–34.

ATOMS, SPECTRA, RADIATION

Ponderomotive Nongradient Force Acting on a Relativistic Particle Crossing an Inhomogeneous Electromagnetic Wave

A. V. Serov

Lebedev Physical Institute, Russian Academy of Sciences, Leninskiĭ pr. 31, Moscow, 117924 Russia

Received May 17, 2000

Abstract—The ponderomotive force acting on a relativistic charged particle crossing an inhomogeneous electromagnetic wave is investigated numerically and analytically. The initial velocity of the particle is perpendicular to the electric field vector of the wave and to the direction of its propagation. The wave has zero gradient in the direction of propagation and is inhomogeneous in both transverse directions. It is shown that the ponderomotive force acting on the particle is parallel to the wave vector. The magnitude of the force is determined not only by the extent of wave inhomogeneity in the direction of the translational motion of particle, but also by its inhomogeneity in the transverse direction. It is found that the trajectory of a particle is determined by the action of ponderomotive forces as well as by its drift in a nonuniform field. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The introduction of the concept of ponderomotive forces simplifies the description of particle dynamics in rapidly oscillating nonuniform fields [1, 2]. For example, ponderomotive forces are of potential nature in many cases in spite of the fact that electric and magnetic fields acting on a particle in an electromagnetic wave contain only vortex components. This makes it possible to reduce the order of equations describing the motion of particles and to obtain an approximate solution for a number of problems [3, 4]. In such problems, the term ponderomotive forces denotes time-averaged forces acting on an individual particle.

In the method of time-averaging the motion equations for particles, which is widely used for describing the dynamics of charged particles in an external spatially inhomogeneous high-frequency electromagnetic field, the motion of an individual charged particle is presented as the sum of a smooth translation $\mathbf{r}_0(t)$ and a rapid oscillatory motion $\mathbf{r}_\sim(t)$ at the frequency of the external field. It is assumed that the value of $\mathbf{r}_\sim$ averaged over the period T vanishes, while the value of $\mathbf{r}_0$ varies insignificantly over this time period. The variation of the smooth component of the particle momentum is interpreted as the result of the action of ponderomotive forces.

The authors of earlier publications devoted to an analysis of the ponderomotive forces acting on a particle in a spatially inhomogeneous electromagnetic field (see, for example, [2]) assumed that the initial velocity $\dot{\mathbf{r}}_0(0)$ of the particle, as well as the oscillatory velocity $\dot{\mathbf{r}}_\sim(t)$ acquired by the particle under the action of the field of the wave, are much smaller than the velocity of light. The field of the wave was regarded as weakly inhomogeneous, and the number of periods of particle vibrations during the time of its flight through the region of the inhomogeneous field was assumed to be large. It was proved that under these conditions, the ponderomotive force is given by [2]

$$\mathbf{f_p} = -\frac{e^2}{4m\omega^2}\nabla\overline{E^2}, \qquad (1)$$

where e and m are the charge and mass of the particle, and ω and E are the frequency and strength of the external electromagnetic field. The bar indicates averaging over the field period.

It follows from Eq. (1) that the force is directed along the gradient of average intensity of the field. On account of the direction of its action, this force is often referred to as the gradient force. Another term used for this force in the literature is the Gaponov–Miller force. Subsequently, ponderomotive forces acting on a particle in strong fields were investigated [5, 6] in the case when the velocity $\dot{\mathbf{r}}_\sim(t)$ of the oscillatory motion becomes relativistic.

It was proved by us earlier [7] that when the initial translational velocity $\dot{\mathbf{r}}_0$ is comparable with the velocity of light, the dynamics of particles can be noticeably affected by another ponderomotive force, which differs considerably from the gradient force. The features of this force were studied in [8, 9], where the dynamics of relativistic particles crossing a linearly polarized wave H_{10} propagating along a waveguide with a constant rectangular cross section. Particles were injected from a point on the waveguide wall at right angles to the wave vector and to the electric vector of the wave. The wave was inhomogeneous in the direction of the translational motion of the particles and homogeneous in the

transverse directions. There are no gradient forces acting on a particle interacting with wave H_{10} in the direction of the wave vector. It was proved that particles crossing such a wave experience the action of a ponderomotive force having a component parallel to the wave vector. The value of this force is proportional to the initial velocity v_0 and to the fourth power E^4 of the electric field strength.

The interaction of relativistic electrons with an inhomogeneous electromagnetic wave was considered by Bituk and Fedorov [10] who separated the motion into fast and slow components in the second order of the perturbation theory and averaged the equation of motion of a particle in the given electromagnetic field over fast oscillations. They derived expressions for relativistic ponderomotive forces in the general case and considered a specific realization of an inhomogeneous field in the form of a steady-state plane-focused beam. It was proved that apart from gradient forces, there also emerge force components parallel to the wave vector and to the direction of field polarization. The force acting along the field polarization vector emerges even when the field gradient in this direction is equal to zero. The field of the electromagnetic wave was regarded in [10] as quite weak.

The structure of ponderomotive forces acting on an electron in strong fields was considered by Taranukhin [11]. He obtained an expression for the ponderomotive force acting on a classical electron in a weakly inhomogeneous field of arbitrary intensity by averaging over the proper time of the leading center of the electron. Scattering of relativistic electrons from a focused laser pulse was studied by Narozhnyĭ and Fofanov [12] who proposed, among other things, a three-dimensional model of the field, which is an exact solution of Maxwell's equations. The equations of averaged motion of an electron in the field of a laser pulse were derived by using the standard method of separating motion into smooth and rapidly oscillating components.

In the present paper, we consider the action of a ponderomotive nongradient force acting parallel to the wave vector. The ponderomotive force acting on a relativistic particle crossing a wave inhomogeneous in both transverse directions, i.e., inhomogeneous in the cross section of the waveguide, is studied numerically and analytically. The main result obtained by us here is that the magnitude of the ponderomotive force is determined not only by the extent of the inhomogeneity of the wave in the direction of the translational motion of particles, but also by the wave inhomogeneity in a transverse direction. The ponderomotive force directed parallel to the wave vector is proportional to E^2 only when the wave is inhomogeneous in both transverse directions. If the wave is inhomogeneous only in the direction of translational motion of particles, the force is proportional to E^4 [7–9]. An analysis of the dynamics carried out by the method of iterations in a field leads to results which agree with the results of numerical calculations.

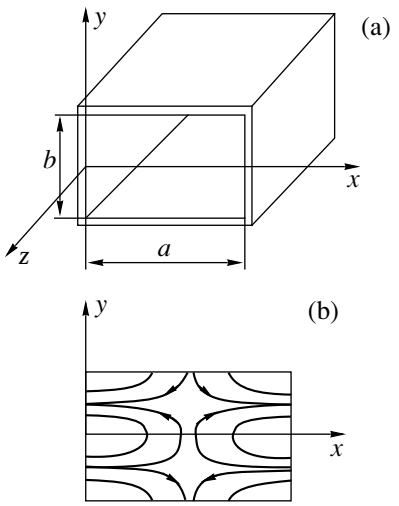

Fig. 1. Geometry of the problem (a) and electric field distribution over the waveguide cross section (b).

2. NUMERICAL ANALYSIS OF PARTICLE DYNAMICS

Expressions describing the field distribution of an electromagnetic beam in empty space are usually derived using certain approximations. It is difficult to estimate the influence of these approximations on the accuracy of the calculation of particle vibrations, and the more so, on the smooth motion of particles. This is especially pertinent while estimating the accuracy of calculations for particles moving in regions far away from the beam axis and crossing the beam boundaries, i.e., for particles entering the wave and leaving it. As a rule, the deviations from Maxwell's equations in approximate expressions describing the fields of electromagnetic beams become significant with increasing distance from the axis. Since we study here the action of averaged forces, which are of the difference type and have second or higher order in a field, it is necessary that the electric and magnetic fields used in numerical calculations satisfy Maxwell's equations on the entire trajectory of the particle. For this reason, we consider here the motion of particles in the field of an electromagnetic wave propagating in a waveguide, for which the expressions describing the spatial field distribution satisfy Maxwell's equations at all points in the space. Besides, in contrast to a wave in empty space, a wave in a waveguide does not diverge, and hence a particle in the waveguide is subjected neither to a gradient force, nor to any other forces associated with a change in the wave cross section along the particle.

The geometry of the problem is shown in Fig. 1. Particles interact with an electromagnetic wave H_{12}

propagating along a rectangular waveguide. The field of this wave is described by the formulas

$$\begin{aligned} E_x &= E\left(\frac{2a}{b}\right)\cos\frac{\pi x}{a}\sin\frac{2\pi y}{b}\sin\varphi, \\ E_y &= -E\sin\frac{\pi x}{a}\cos\frac{2\pi y}{b}\sin\varphi, \\ E_z &= 0, \\ H_x &= E\chi\sin\frac{\pi x}{a}\cos\frac{2\pi y}{b}\sin\varphi, \\ H_y &= E\chi\left(\frac{2a}{b}\right)\cos\frac{\pi x}{a}\sin\frac{2\pi y}{b}\sin\varphi, \\ H_z &= -E\alpha\left(\frac{\lambda}{2a}\right)\cos\frac{\pi x}{a}\cos\frac{2\pi y}{b}\cos\varphi, \end{aligned} \tag{2}$$

where $\varphi = \omega t - k_z z$, $\omega = 2\pi c/\lambda$, λ being the wavelength, $\alpha = (b^2 + 4a^2)/b^2$, $\chi = [1 - \alpha(\lambda/2a)^2]^{1/2}$, $k_z = k\chi = (2\pi/\lambda)[1 - \alpha(\lambda/2a)^2]^{1/2}$, a and b being the waveguide dimensions along the x- and y-axes. The spatial distribution of the electric field of the H_{12} mode is presented in Fig. 1b. It can be seen that the field of the wave is inhomogeneous in both transverse directions (along the x-axis as well as along the y-axis).

Since we analyze a nongradient ponderomotive force, the conditions of particle injection and the beam cross section were chosen so that the effect of gradient forces on the particle dynamics was reduced to the minimum. It follows from Eqs. (2) that the wave is homogeneous in the direction of its propagation at all points of the space, and hence no gradient force is acting along the z-axis. The width of the bunch (the size along the z-axis) can be arbitrary since the z_0 value of the initial coordinate does not affect the nature of the ponderomotive forces acting on a particle. When the particle moves in the xz plane, the gradient force along y is also equal to zero. It emerges only on segments of the particle trajectory lying above or below this plane. In order to reduce the effect of this force, the height of the bunch (the size along the y-axis) must be such that the change in the field strength over the beam cross section is much smaller than the field strength. This condition sets a limit on the height of the bunch: it must be much smaller than the transverse size of the wave.

The calculations were made under the following initial conditions. Particles were injected from points located at a waveguide wall symmetrical relative to the xz plane and had a velocity v_0 directed along the x-axis. At the initial moment of time, the particles moved at right angles to the direction of wave propagation (z-axis) and to the electric field vector of the wave (y-axis). The case when particles are injected in the symmetry plane xz at a certain angle to the z-axis can be reduced to the case considered by us here. For this purpose, we must go over to a reference frame moving at a constant velocity along the z-axis and carry out the Lorentz transformations of the wave field.

It follows from Fig. 1b and expressions (2) that under the given conditions of injection, the field of the H_{12} mode along an unperturbed trajectory is close to the field of a linearly polarized wave. On this trajectory, the electric field vector has a large E_y component which is a function of the transverse coordinate x. In the xz plane, the gradient force has an x component only.

The numerical method was used to solve the equation of motion [13] of a particle in the given electromagnetic field (2):

$$\ddot{\mathbf{r}} = \frac{e}{m\gamma}\left\{\mathbf{E} + \frac{1}{c}[\dot{\mathbf{r}}\times\mathbf{H}] - \frac{1}{c^2}\dot{\mathbf{r}}(\dot{\mathbf{r}}\cdot\mathbf{E})\right\}, \tag{3}$$

where $\gamma = (1 - (v/c)^2)^{-1/2}$ is the relative energy of the particle.

Particles were injected from five points with different y coordinates ($y_0 = 0, \pm 0.25y_b, \pm 0.5y_b$) and identical x and z coordinates ($x_0 = z_0 = 0$). The motion of one hundred particles injected successively with a time interval $dt = 0.01T$ across the wave was analyzed for each value of the initial transverse coordinate y_n. Under such conditions, the dynamics of a strip bunch with a vertical size y_b and length $L = vT$, which was injected into the wave during the time equal to a period T, was simulated in our calculations. We calculated the position and the velocity of the bunch during its motion in the wave and at the exit from the wave. The position and velocity of the bunch at a given instant were determined as the values of coordinates $\langle x\rangle$, $\langle y\rangle$, $\langle z\rangle$ and velocities $\langle v_x\rangle$, $\langle v_y\rangle$, $\langle v_z\rangle$ averaged over all the particles, the angle brackets indicating averaging over all the particles. The calculations were made for various values of the waveguide height b, wavelength λ, initial electron velocity v_0, and electric field E of the wave.

Figure 2 shows the trajectory of the center of gravity of the bunch, obtained as a result of calculations. The calculations were made for the vertical size $y_b = 0.3$ mm of the bunch, waveguide parameters $a = 10$ mm and $b = 25$ mm, and wavelength $\lambda = 1$ mm. A wave with such a relation between the wavelength and the transverse dimensions can be regarded as weakly inhomogeneous since a particle crossing the wave performed more than ten oscillations. The electric field of the wave was $E = 3 \times 10^6$ V/cm. Particles were injected at a velocity $v_0 = 2.75 \times 10^{10}$ cm/s, which corresponds to the relative velocity $\beta = v/c = 0.92$. For such parameters of the wave and initial velocity of the particle, the relative velocity of oscillatory motion was $\beta_\sim = v_\sim/c = 0.1$. It can be seen from the figure that the amplitude of oscillations increases as the bunch moves to the region with a high field strength and decreases as it leaves this region. Besides, the average velocity of the bunch along the x axis varies smoothly, and the bunch is gradually displaced along the z axis. The change in the velocity

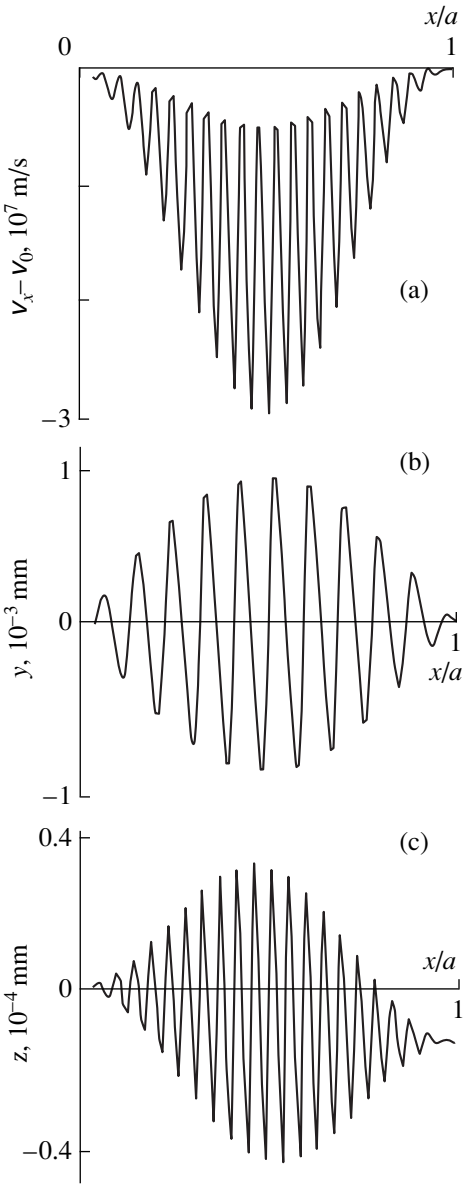

Fig. 2. Variation of the velocity of a bunch in the direction of injection (a) and bunch coordinates in the directions of the electric field vector (b) and the wave vector (c): $a = 10$ mm, $b = 25$ mm, $\lambda = 1$ mm, $y_b = 0.3$ mm, $E = 3 \times 10^6$ V/cm, $v_0 = 2.75 \times 10^{10}$ cm/s.

along the *x*-axis can be attributed to the action of a gradient ponderomotive force. This force decelerates the bunch during its motion in an increasing field from the point of injection to the middle of the waveguide and accelerates it during its motion in a decreasing field from the middle of the waveguide to the exit from the wave.

The displacement of the bunch along the *z*-axis indicates the existence of another ponderomotive force directed parallel to the wave vector. This force is of a nongradient type since the field of the H_{12} wave has

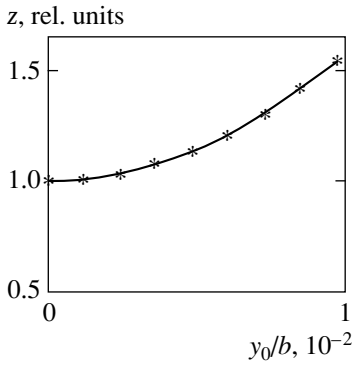

Fig. 3. Displacement of particles along the *z*-axis as a function of the initial transverse coordinate y_0: $a = 10$ mm, $b = 25$ mm, $\lambda = 1$ mm, $E = 3 \times 10^6$ V/cm, $v_0 = 2.75 \times 10^{10}$ cm/s.

zero gradient along the *z*-axis. In order to determine the order of magnitude of this ponderomotive force, we analyzed the dependence of the longitudinal coordinate *z* of the bunch at the exit from the wave as a function of the wave field. Calculations show that for small values of the field, the change in the bunch coordinate is proportional to E^2.

According to calculations, the value of the ponderomotive force increases with the distance from the point of injection to the *xz* plane. The change in the force does not depend on the sign of the initial transverse coordinate y_0 of a particle and is a function of its absolute value only. Figure 3 shows the dependence of the displacement of particles along the *z*-axis on the initial transverse coordinate y_0. The displacement was calculated as the mean displacement over all particles injected from the given point. It can be seen that the increment of force is proportional to the square of coordinate y_0 since $\Delta z \sim (y_0/b)^2$. The value of the force acting on the particles at a distance $y_0 = 0.01$b from the *xz* plane is 1.5 times larger than the value of the force in the plane.

According to the results of calculations, the longitudinal force F_z is sensitive to the spatial inhomogeneity of the field not only in the direction of the translational motion of the particle (along the *x*-axis), but also to the field inhomogeneity in the transverse direction (along the *y*-axis). In order to study the effect of this inhomogeneity on force F_z, we analyzed the particle dynamics for various values of the waveguide height *b*. It follows from Eqs. (2) that the field strength E_y on the *x*-axis remains unchanged upon a variation of the size of the electromagnetic beam along the *y*-axis (wave height), while the gradient of field E_y along the *y*-axis changes as well as the magnitudes of fields E_x, H_y, and H_z. If we make the waveguide height tend to infinity, fields E_x and H_y, and hence the gradient of field E_y along the *y*-axis, will tend to zero. Wave H_{12} degenerates into

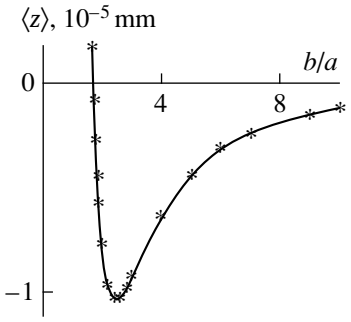

Fig. 4. Displacement of a bunch along the *z*-axis as a function of the ratio of the wave height *b* to the wave width *a*: $a = 10$ mm, $\lambda = 1$ mm, $E = 3 \times 10^6$ V/cm, $v_0 = 2.75 \times 10^{10}$ cm/s.

wave H_{10} whose electric field is homogeneous along the *y*-axis. The dynamics of the bunch in the field of the H_{10} wave was considered in [7–9], where is was proved that the ponderomotive longitudinal force acting on a particle in such a wave is proportional to E^4. In the given wave H_{12}, however, the force is proportional to E^2. Thus, the field inhomogeneity in the direction perpendicular to the translational motion of particles induces a longitudinal ponderomotive force of a lower (second) order.

The results of the calculation of the displacement of the bunch along the *z*-axis as a function of the ratio *b*/*a* of the wave dimensions in the transverse directions are presented in Fig. 4. It can be seen that with increasing *b*/*a*, the displacement of the bunch in the positive direction becomes smaller and is transformed into displacement in the negative direction. For a certain ratio of the height of the wave to its width, the displacement passes through its maximum value. A further increase in the value of *b*/*a* leads to a decrease in the displacement. Our calculations proved that the values of the wave height-to-width ratio for which the dependence $z = f(b/a)$ passes through zero and through its maximum value do not change upon a variation of the initial velocity of the particles, the field strength, and the wave frequency.

3. ANALYTIC CALCULATION OF PONDEROMOTIVE FORCES

In order to derive expressions describing ponderomotive forces, we carried out analytical calculations of the motion of particles through a given electromagnetic wave. We analyzed the dynamics in a weakly inhomogeneous field; a particle crossing this field performed a large number of oscillations. Besides, we assumed that the velocity of oscillatory motion associated with the interaction of a particle with the wave is much smaller than the velocity of light. The equation of motion (3) of particles taking into account the fields (2) acting on a particle can be written in the form

$$\frac{d\beta_x}{dt} = \frac{e}{mc\gamma} \times \{E_x + \beta_y H_z - \beta_z H_y - \beta_x(\beta_x E_x + \beta_y E_y)\}, \quad (4)$$

$$\frac{d\beta_y}{dt} = \frac{e}{mc\gamma} \times \{E_y + \beta_z H_x - \beta_x H_z - \beta_y(\beta_x E_x + \beta_y E_y)\}, \quad (5)$$

$$\frac{d\beta_z}{dt} = \frac{e}{mc\gamma} \times \{\beta_x H_y - \beta_y H_x - \beta_z(\beta_x E_x + \beta_y E_y)\}, \quad (6)$$

were $\beta_i = v_i/c$ are relative velocities of particles.

The solution of this system of equations was obtained by the method of successive approximations. Using the quantity $\eta = v_\sim/c = (eE/mc\gamma\omega)$ as a smallness parameter, we can present the velocities and displacements of particles in the form of series:

$$\beta_x = \beta_x^{(0)} + \beta_x^{(1)} + \ldots, \quad \beta_y = \beta_y^{(0)} + \beta_y^{(1)} + \ldots,$$

$$\beta_z = \beta_z^{(0)} + \beta_z^{(1)} + \ldots, \quad x = x^{(0)} + x^{(1)} + \ldots,$$

$$y = y^{(0)} + y^{(1)} + \ldots, \quad z = z^{(0)} + z^{(1)} + \ldots .$$

In the case under investigation, a particle at the initial instant has a constant velocity β_0 directed along the *x*-axis, and hence $\beta_y^{(0)} = \beta_z^{(0)} = 0$ and $x^{(0)} = c\beta_0 t$. Taking these expressions into account and integrating Eq. (5), we obtain the following expression for the first approximation of the velocity β_y:

$$\beta_y^{(1)} = \frac{1}{1-\xi^2}\left(\frac{eE}{mc\gamma\omega}\right) \times [\xi(\alpha-1)(\sin\psi\cos\Omega t - \sin\varphi_0) + (1-\alpha\xi^2)\cos\psi\sin\Omega t], \quad (7)$$

where $\xi = \Omega/\omega = \beta_0\lambda/2a$, $\Omega = \pi c\beta_0/a$, $\alpha = (b^2 + 4a^2)/b^2$, and $\psi = \omega t - kz + \varphi_0$, φ_0 being the phase of the wave at the instant of injection.

For a relativistic initial velocity, the value of ξ, which is proportional to the ratio of the wavelength λ to the transverse size *a* of the wave, characterizes the wave inhomogeneity in the direction of particle injection. In the case under investigation, the wave was weakly inhomogeneous, and $\xi = 0.04 \ll 1$. The value of α is determined by the electromagnetic wave mode. A transition from the fields of the H_{12} mode to the fields of the H_{10} mode in Eqs. (2) occurs as the waveguide height *b* tends to infinity. As we go over in Eq. (7) to H_{10} mode (whose field is homogeneous along the *y*-axis), the value of H_{10} must be taken to be equal to unity. In this case, the first term on the right-hand side vanishes,

and the transverse velocity oscillates with a phase shift $\pi/2$ relative to the transverse components of the wave field [7]. Thus, the inhomogeneity of the field in the y direction leads to the emergence of additional terms in the expression for velocity β_y, which describe the velocity oscillations in the phase coinciding with the phase of transverse fields of the wave.

The integration constant in Eq. (7) was determined from the condition that the transverse velocity $\beta_y^{(1)}$ is equal to zero at the initial instant of time. Carrying out integration and taking into account the initial condition $y^{(1)}(0) = 0$, we obtain the following expression for the displacement along the y-axis:

$$y^{(1)} = \frac{1}{(1-\xi^2)^2}\frac{c}{\omega}\left(\frac{eE}{mc\gamma\omega}\right) \times [\xi(2-\alpha-\alpha\xi^2)(\cos\psi\cos\Omega t - \cos\varphi_0) + \sin\psi\sin\Omega t] - \frac{c}{1-\xi^2}\left(\frac{eE}{mc\gamma\omega}\right)\xi(\alpha-1)t\sin\varphi_0. \quad (8)$$

Let us consider Eq. (6) describing the acceleration component along the z-axis. In the first approximation, the acceleration $d\beta_z^{(1)}/dt$ is equal to zero since the first term on the right-hand side is equal to zero because $H_y(y = 0) = 0$, while the remaining terms are of a higher order. It follows from the initial conditions that $\beta_z^{(1)}(0) = 0$ and $z^{(1)}(0) = 0$. We assume that the quantity $\xi = v_\sim/c$ is a smallness parameter, and hence the amplitude of oscillations is much smaller than the transverse dimensions of the waveguide. Consequently, the function $\sin(2\pi y/b)$ in the expression for the magnetic field H_y can be replaced by the first term of the expansion. Taking into account this fact, we substitute Eqs. (7) and (8) into Eq. (6). Since we are determining the force acting on the bunch, we must carry out averaging over all the particles in the bunch, i.e., over initial phases φ_0 of the injection. After averaging, we obtain the following expression for the acceleration of the bunch along the z-axis:

$$\left\langle\frac{d\beta_z^{(2)}}{dt}\right\rangle_\varphi = -\frac{\chi\Omega}{2(1-\xi^2)^2}\frac{a^2}{b^2}\left(\frac{eE}{mc\gamma\omega}\right)^2 \times [\xi^2(\alpha-1)\sin 2\Omega t + (1-\xi^2)\sin\Omega t\cos\omega t - \xi(2-\alpha-\xi^2\alpha)\cos\Omega t\sin\omega t] - \frac{\chi\Omega^2}{(1-\xi^2)}\frac{a^2}{b^2}\left(\frac{eE}{mc\gamma\omega}\right)^2(\alpha-1)t\cos\Omega t\cos\omega t, \quad (9)$$

where $\langle\ldots\rangle_\varphi$ denotes averaging over the initial phases of injection.

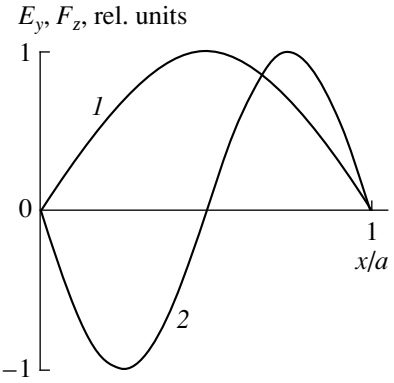

Fig. 5. Distribution of the transverse electric field strength E_y (curve *1*) and the longitudinal force F_z (curve *2*) along the x-axis.

The right-hand side of this expression contains terms of different origins: the first term describes only the smooth variation of acceleration over time periods of the order of a/c, while the second and third terms describe, in addition to the smooth variation, the oscillations with the frequency of the external field. After averaging over the period of oscillations, we obtain the following expression for the ponderomotive force F_z:

$$F_z \sim \overline{\left\langle\frac{d\beta_z^{(2)}}{dt}\right\rangle_\varphi} = -\frac{\pi c\chi}{2}\frac{\lambda^2\beta_0^3 a}{b^4}\left(\frac{eE}{mc\gamma\omega}\right)^2\sin\left(\frac{2\pi x}{a}\right). \quad (10)$$

Since we assumed that a particle performs a large number of oscillations during the time of its crossing the wave, the quantities $\sin\Omega t$ and $\cos\Omega t$ were regarded as constants during averaging over the period of oscillations.

It follows from Eq. (10) that the ponderomotive force is proportional to E^2 and is opposite to the direction of wave propagation when a particle enters the region of a strong field, while for a particle moving into the region of a weak field, the direction of the force coincides with the direction of wave propagation. The force is observed only for a wave inhomogeneous in both transverse directions. As the wave size b perpendicular to the velocity of the particle increases, the field gradient in the direction of the y axis decreases, and the force decreases in proportion to $1/b^4$. For a wave homogeneous in y (for $b = \infty$), the force, which is proportional to E^2, vanishes and the particle will experience only the action of a ponderomotive force proportional to E^4 and directed along the z-axis [7–9]. Figure 5 shows the field strength E_y and force F_z as functions of the longitudinal coordinate x.

A comparison of the results of numerical calculations (see Fig. 4) with the structure of expression (10)

and Fig. 5 shows that the trajectory of the cluster is determined not only by the ponderomotive force F_z. Indeed, the sign of the ponderomotive force does not change upon variation of the transverse dimensions of the wave, while the displacement of the bunch in the z direction at the exit from the wave changes its direction. In order to describe the particle trajectory, we must take into account the drift of the particles during their motion in an inhomogeneous field. After integrating Eq. (9) and averaging over the period of oscillations, we obtain the following expression for the average velocity:

$$\overline{\langle \beta_z \rangle_\varphi} = \frac{\chi a^4}{4b^4}\left(\frac{\lambda\beta_0}{2a}\right)^2\left(\frac{eE}{mc\gamma\omega}\right)^2 \times\left[\frac{4a^2}{b^2}(\cos 2\Omega t - 1) - 4\left(\frac{2a^2 - b^2}{a^2}\right) + 2(1 + \cos\Omega t)\right]. \quad (11)$$

In this expression, only the first term is associated with the action of a ponderomotive force, while the second and third terms describe the drift velocity of the bunch in an inhomogeneous field. Integrating Eq. (9) twice and taking into account the initial conditions, we obtain the displacement in the direction of wave propagation. After averaging, the coordinate z of the bunch at the exit from the wave is described by the relation

$$\overline{\langle z \rangle_\varphi} = \frac{2\chi\beta_0 a\lambda^2}{\pi}\left(\frac{eE}{mc\gamma\omega}\right)^2\frac{(3a^2 - b^2)}{b^4}, \quad (12)$$

which shows that the displacement is zero for $b = \sqrt{3}\,a$ and passes through a peak for $b = \sqrt{6}\,a$. Such a dependence agrees with the results of the numerical analysis of particle dynamics presented in Fig. 2.

The emergence of a nongradient ponderomotive force can be explained on a qualitative level. The longitudinal component of the force appears as a result of averaging the corresponding component of the Lorentz force. It is well known [13] that for a particle moving in a plane wave, the phases of magnetic field oscillations and the velocity component of the particle perpendicular to the magnetic field are shifted by $\pi/2$. Consequently, averaging over the wave period gives zero. In an inhomogeneous wave, a small longitudinal magnetic field component H_z appears along with the large components E_y and H_x. The magnitude of field H_z is determined by the inhomogeneity of the wave in the direction of particle injection (along the x-axis) and in the transverse direction (along the y-axis). If the wave is inhomogeneous in y, the transverse velocity β_y and the displacement y acquire components coinciding in phase with magnetic field oscillations. Consequently, averaging over the period of oscillations gives a nonzero ponderomotive force.

4. CONCLUSION

We investigated the motion of a bunch in an inhomogeneous electromagnetic wave by using numerical methods. Calculations show that in the case of a directional motion of particles through an inhomogeneous wave, they experience the action of a force differing significantly from the gradient force. For example, particles intersecting the wave at right angles to the direction of its propagation and to the electric vector are acted upon by a force F_k parallel to the wave vector. The force is proportional to E^2 if the wave is inhomogeneous in both transverse directions and to E^4 if the wave is inhomogeneous only in the direction of the translational motion of the particles.

We derived the expressions for the ponderomotive force and the velocity associated with the action of force and particle drift in an inhomogeneous field. The expression obtained for the displacement of the bunch in the direction of the wave vector agrees with the results of numerical calculations.

The nongradient ponderomotive force can be investigated experimentally by passing an electron beam through an electromagnetic wave. In such an experiment, the position of the center of gravity of the electron beam and its cross section must be recorded. The most favorable conditions for the injection are created when electrons intersect the electromagnetic wave at right angles. In this case, the contribution of gradient forces directed along the direction of wave propagation to the particle dynamics is minimal. After the interaction with the wave, first, the transverse dimensions of the electron beam increase, and second, the center of gravity is displaced in the direction opposite to the direction of wave propagation.

It should be noted that the results obtained here contradict the conclusions drawn in [10], according to which the component F_z of the ponderomotive force must be equal to zero when particles are injected at right angles to the electric vector of the wave and to the wave vector. A possible reason behind this discrepancy are the conditions under which relativistic ponderomotive forces were investigated in [10]. Bituk and Fedorov [10] confined their analysis to the lowest (linear) approximation in gradients, while in our case the ponderomotive forces are proportional to $1/b^4$, i.e., higher-order gradient terms are taken into account.

ACKNOWLEDGMENTS

The author is grateful to M.V. Fedorov and B.M. Bolotovskiĭ for fruitful discussions of the obtained results and to V.P. Yakovlev for critical remarks stimulating additional investigations. Numerical calculations were partly carried out during the author's work at the Research Institute for Scientific Measurements at Tohoku University, Japan. Thanks are due to Prof. M. Ikezawa and Dr. Y. Shibata for their kind invitation and hospitality.

This research was financed by the Russian Foundation for Basic Research (projects nos. 99-02-16271 and 99-02-18183).

REFERENCES

1. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 1: *Mechanics* (Nauka, Moscow, 1965; Pergamon, New York, 1988).
2. A. V. Gaponov and M. A. Miller, Zh. Éksp. Teor. Fiz. **34**, 242 (1958) [Sov. Phys. JETP **7**, 168 (1958)].
3. I. R. Gekker, *Interaction of Strong Electromagnetic Fields with Plasmas* (Atomizdat, Moscow, 1978; Clarendon Press, Oxrord, 1982).
4. M. V. Fedorov, *Electron in a Strong Optical Field* (Nauka, Moscow, 1991).
5. T. W. Kibble, Phys. Rev. **150**, 1060 (1966).
6. S. P. Goreslavsky, N. B. Narozhny, O. V. Shcherbachev, and V. P. Yakovlev, Laser Phys. **3**, 418 (1993).
7. A. V. Serov, Laser Phys. **7**, 1190 (1997).
8. A. V. Serov, Kvantovaya Élektron. (Moscow) **25**, 197 (1998).
9. A. V. Serov, Kvantovaya Élektron. (Moscow) **26**, 179 (1999).
10. D. R. Bituk and M. V. Fedorov, Zh. Éksp. Teor. Fiz. **116**, 1198 (1999) [JETP **89**, 640 (1999)].
11. V. D. Taranukhin, Zh. Éksp. Teor. Fiz. **117**, 511 (2000) [JETP **90**, 447 (2000)].
12. N. B. Narozhnyĭ and M. S. Fofanov, Zh. Éksp. Teor. Fiz. **117**, 867 (2000) [JETP **90**, 753 (2000)].
13. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Nauka, Moscow, 1988; Pergamon, Oxford, 1975).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 28–36.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 35–44.

ATOMS, SPECTRA, RADIATION

Dynamics of Mass Transfer Caused by the Photoinduced Spatially Inhomogeneous Modulation of Mobility in a Multicomponent Medium

R. Caputo[a], A. V. Sukhov[b,] *, C. Umeton[a], and R. F. Ushakov[b]

[a]*Dipartimento di Fisica, Università della Calabria, 87036, Arcavacata di Rende, CS, Italia*
[b]*Institute for Problems of Mechanics, Russian Academy of Sciences, pr. Vernadskogo 101, Moscow, 117526 Russia*
** e-mail: sukhov@ipmnet.ru; soukhov@fis.unical.it*

Received June 26, 2000

Abstract—The reasons behind the three-dimensional mass transfer of the inert components of a mixture in which photoinduced spatially inhomogeneous polymerization takes place are discussed. Such a mass transfer is responsible for the laser-induced formation of controllable diffraction gratings in polymer composites containing liquid crystals [R. L. Sutherland, V. P. Tondiglia, L. V. Natarajan, and T. J. Bunning, Chem. Mater. **5**, 1533 (1993); R. L. Sutherland, V. P. Tondiglia, and L. V. Natarajan, Appl. Phys. Lett. **64**, 1074 (1994); R. L. Sutherland, V. P. Tondiglia, L. V. Natarajan, *et al.*, J. Nonlinear Opt. Phys. Mater. **5**, 89 (1996); R. Caputo, A. V. Sukhov, and C. Umeton, Mol. Mater. **12**, 192 (1999)]. The semiphenomenological model proposed in this work is based on the assumption that mass transfer results from conventional Fickian diffusion which, however, does not occur over the entire volume of the sample, but only in regions free from clusters of long polymer chains. It is shown that such a "restriction of the active volume" is responsible for the conventional diffusion of inert components of liquid crystals as well as reagents and low-molecular products (short chains), in spite of the initial spatial homogeneity of concentrations in the sample. The qualitative predictions of the model coincide with the experimental results [R. Caputo, A. V. Sukhov, and C. Umeton, Mol. Mater. **12**, 192 (1999)].

1. INTRODUCTION

The photoinduced formation of three-dimensional periodic gratings in composites containing liquid crystals (LC) has attracted considerable attention during the last decade (see, for example, [1–3]). The reasons behind such an interest are obvious and are mainly associated with the possibility of obtaining cheap holographic elements, which can be switched (or controlled) by moderate electric voltages (e.g., 10–100 V, see [1–3]). The traditional approach to the problem of producing such gratings [1, 2] involves the induction of a spatially inhomogeneous phase separation of the LC component of the mixture. This is usually achieved as a result of the photoinitiation of the polymerization process in an LC mixture containing initial reagents by the interference pattern of laser radiation of the appropriate frequency (blue–green or near-UV part of the spectrum). Such a process, which is usually called "laser curing," leads to the formation of a three-dimensional periodic structure (grating) consisting of alternating polymer strips and strips of a polymer-disperse liquid crystal (PDLC).

Although the diffraction efficiency of such gratings is quite high [2], the spatial inhomogeneity of the structure of PDLC strips results, as a rule, in considerable losses due to the disordered scattering of the radiation used for the grating readout. Attempts to prevent such losses by reducing the average size of the encapsulated LC drops to much below the wavelength of the radiation result in a considerable increase in the required values of the controlling electric field strength (10 V/μm, [5]).

The possibility of a basically new approach towards the problem of creating such gratings was experimentally demonstrated recently [4]. It was found that if the process of polymerization occurs in a mixture of LC and reagents which initially dissolve quite easily in each other, the phase separation as such can be avoided and a spatial modulation of the LC-component concentration can be attained. If such a modulation is quite deep, the liquid undergoes a phase transition into LC at the modulation peaks. Consequently, the grating obtained in this way is formed by strips of a uniformly oriented LC phase separated by polymer walls [4] rather than the above-mentioned alternating strips of the polymer and PDLC strips. Since such a grating does not reveal inhomogeneities within a strip, there are practically no scattering losses, and the required controlling field is found to be quite weak (3–8 V/μm [4]).

The dependence of the diffraction efficiency of the gratings obtained in this way on the laser radiation intensity (i.e., on the reaction rate) and the three-dimensional grating interval (i.e., on the characteristic diffusion

time within one interval) has the following distinguishing features.

(1) The dependence of the diffraction efficiency on the intensity of laser radiation displays one peak, the optimal value of the intensity increasing sharply with decreasing grating interval.

(2) Each specific combination of the mixture components corresponds to a certain limiting value of the grating interval (usually of the order of 0.2 μm) for which recording of the grating is possible; for shorter grating intervals, the diffraction efficiency is extremely low for laser radiation of any intensity.

The present paper aims at developing a model of the processes responsible for the formation of a grating under the given conditions, including an analysis of real chemical transformations and mass transfer processes, which is capable of explaining these qualitative features.

2. QUALITATIVE DESCRIPTION OF THE MODEL

To begin with, we assume that conventional Fickian diffusion is responsible for the change in the LC concentration distribution from the initially uniform distribution. However, this diffusion occurs not in the entire volume of the sample, but only in its "active part" which has not yet been occupied by immobile polymer clusters (this part of the volume will be referred to as "passive" volume in the following analysis). The passive volume contains only the residual concentration of LC molecules, determined approximately by its solubility in the corresponding polymer. The remaining LC molecules are expelled from the passive part of the volume to the active part and thus the concentration of LC molecules in the active volume is found to be higher than the mean concentration in the entire volume and increases upon an increase in the passive volume. Considering that the rate of growth of the passive volume is determined by the local intensity of radiation and is therefore nonuniform over a grating interval, it turns out that the LC concentration in the active volume is modulated within a grating interval in contrast to the mean concentration over the entire volume, which has so far remained unchanged and uniform. Consequently, conventional Fickian diffusion takes place in the active volume and tends to level out the concentration in the active volume over a grating interval, thus modulating the mean LC concentration which is the parameter determining the local variation of the refractive index of the medium.

At the same time, the molecules of the monomer start diffusing in the opposite direction, and also strive to level out its concentration in the active volume within a grating interval. Thus, the LC molecules diffuse to the regions with the smallest fraction of the passive volume, while the monomer molecules diffuse in the opposite direction.

The latter process can occur in two possible regimes. If the reaction is quite slow (for a given grating interval), the diffusion of the monomer has time to occur before it ultimately perishes in the immobile polymer chains. For a quite fast reaction, on the other hand, there is no time for the monomer diffusion to occur. We shall call these regimes respectively the "slow curing" and "fast curing" modes.

It is important to specify here the reaction products that are attributed to the passive volume. Short polymer chains are definitely mobile and hence must belong to the active volume. We introduce a certain critical length N_0 of a polymer chain starting from which the chain can be treated as immobile (diffusion in a condensed medium is an activation process, hence the diffusion coefficient D depends quite strongly on the molecular mass, and we can confine the analysis, at least at the present stage, to model concepts about N_0 without resorting to the more complex dependences of D on N).

Note (see, for example, [6]) that during the polymerization of radicals, long chains appear just for slow reaction rates, i.e., in the vicinity of the interference minima in our case.

For low radiation intensities, mainly long chains (longer than N_0) must also appear at the interference minima. Consequently, the passive volume is found to be large, but almost uniformly distributed over the grating interval, and hence the modulation is weak. Conversely, only short chains are formed almost everywhere for high intensities; i.e., the passive volume practically does not exist. In this case also, the modulation is weak. The formation of long chains at the minima and short chains at the peaks of the interference pattern (and hence a strong modulation of the passive volume) can be expected only in a certain region of intermediate intensities. Qualitatively, this situation corresponds to the experimentally observed peak in the dependence of the diffraction efficiency of the grating on the radiation intensity.

It follows easily from all that has been said above that the choice of the fast or slow mode is determined by the ratio of two time constants, viz., the characteristic reaction time $\tau_R \propto W_0^{-1/2}$ (W_0 is the radiation intensity, see below) and the characteristic time of diffusion over a grating interval $\tau_d \propto \Lambda^2$ (Λ is the three-dimensional grating interval). Whichever regime is found to be optimal according to the results of subsequent analysis, it is characterized by a certain relation between these times, and hence the optimal value of W_0 must increase with decreasing Λ. This is also in accord with the experimental observations [4].

Thus it can be seen that the model proposed here leads to qualitative conclusions that are in agreement with the experimental results, and hence merits a quantitative analysis in spite of the fact that the mathematical apparatus used for this purpose is apparently quite cumbersome.

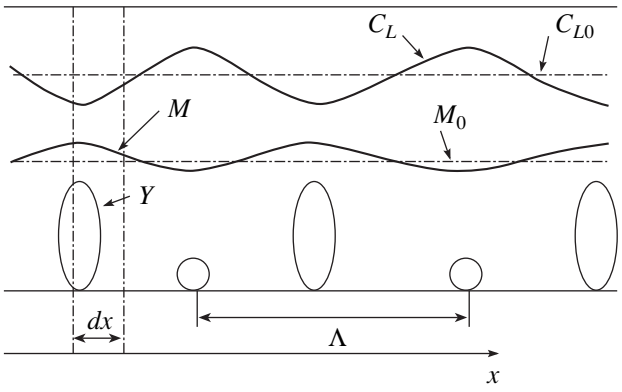

Fig. 1. Illustrating the formation of passive volume.

3. DERIVATION OF THE SYSTEM OF BASIC EQUATIONS

Let us consider a thin plane-parallel layer of the initial mixture (Fig. 1) in a field of spatially inhomogeneous radiation in coordinate x:

$$W = W_0(1 + m\sin(q_0 x)).$$

Here, $q_0 = 2\pi/\Lambda$ is the wave number of the interference pattern and m is its modulation depth. The rate of polymerization as well as the spectrum of the lengths of the chains formed are different in different volume elements dV corresponding to the elements dx for various values of x. Hence the concentration Y [cm^{-3}] of monomer elements included in long chains and thus constituting the passive volume is found to depend on x. This results in a modulation of the concentrations C_L and M of LC and monomer molecules, respectively.

For the subsequent analysis, we make the following assumptions.

(1) The LC molecules are partially expelled from the passive volume in which only a small residual concentration C_{NS} is retained. This concentration corresponds to a saturated solution of LCs in the given polymer, which is an easily measurable parameter in the experiments. In this work, we do not set out to make a quantitative comparison of our results with the experimental data and assume that $C_{NS} = 0$. This does not make any fundamental changes in the analysis, but considerably simplifies the calculations.

(2) All monomer molecules that do not participate in a reaction remain in the active volume according to its definition.

(3) All chemical transformations and mass transfer occur in the active volume.

(4) The spatial redistribution of the components follows the conventional Fickian diffusion which, however, occurs only in the active volume. The equilibrium distribution corresponds to spatially homogeneous concentrations of the mobile components in the active volume, but not to homogeneous mean concentrations.

(5) The radicals participating in a reaction are considered to be immobile from the diffusion point of view. This assumption is justified in view of their short lifetimes.

(6) The concentration of the photoinitiator (see [6]) is homogeneous and remains constant in the course of a reaction. This means, above all, that the initial concentration C_I of the initiator is assumed to be surplus. The validity of such an assumption depends on the mean chain length over a grating interval, and hence on the radiation intensity. Nevertheless, this assumption is always valid in a successful experiment on the recording of gratings for the following reason. If the process occurs with a deficit of the initiator, the "uncured" monomer remains in the reaction product. However, all the monomers used in the experiments contain admixtures of inhibitors for preventing spontaneous polymerization during storage. Consequently, the radicals trapped temporarily by the inhibitor will be released at the end of the controlled photoinduced process, which is followed by uncontrolled polymerization (postcuring, see [1–3]). This process, which lasts from a few hours to several days, inevitably spoils the morphology of the recorded grating. Hence real experiments are always carried out with mixtures containing an excess of the inhibitor. Moreover, it follows from our assumption that the initiator molecules are easily trapped in the passive volume and are not expelled into the active volume. This is not quite obvious, but can be expected in view of the small size of the initiator molecules (usually diphenyl ketone and its derivatives). In any case, this assumption is required for the purpose of analysis (see below) and can be treated as a sort of "zeroth approximation".

(7) Finally, we disregard the volume occupied by radicals and initiator molecules, since their concentrations are quite low as compared to M and C_N. Although this assumption is not of fundamental importance, it considerably simplifies the calculations.

In the light of the above assumptions, we can easily establish a relation between the mean concentrations of the components and their concentration in the active volume (marked by the additional superscript "f") and also write the corresponding relations for the reaction rates and diffusion fluxes:

$$C_L^f = C_L\frac{1}{1 - Y/Z},$$
$$M^f = M\frac{1}{1 - Y/Z}, \quad C_I^f = C_I. \tag{1}$$

Here Z is the total concentration of molecules:

$$Z = C_L^0 + M^0$$

(the superscript "0" indicates the initial conditions before the onset of the reaction). For the sake of simplicity, the ratio of the molecular volumes of the monomer (uncured or polymerized) and the nematic in Eqs. (1) is put equal

to unity, which is nearly true for real materials. This assumption is also not fundamental, but simplifies the calculations. The expression for the diffusion flux across a unit surface S (see Fig. 1) has the form

$$W_D = -D\left(1-\frac{Y}{Z}\right)^{2/3}\frac{\partial}{\partial x}\left(\frac{M}{1-Y/Z}\right). \quad (2)$$

Formula (2) is written for a monomer, but the form of the expression for flux remains the same for any other mobile component. Here also, the fact that both the active volume and the active part of surface S vary upon an increase in Y is taken into consideration, and hence the effective value of the diffusion coefficient (or mobility) depends on Y; D is the true diffusion coefficient in the active volume, and the expression in the parenthesis following D describes the modulation of the effective value. Since we are speaking of diffusion in the active volume which does not contain long polymerized chains, it can be assumed in the model under consideration that the diffusion coefficient D is independent of the degree of polymerization. In the passive volume, the value of D is determined by the degree of polymerization and is equal to zero.

For the process of radical polymerization, we use the following well-established classical scheme (see, for example, [6]):

(1) $I + h\nu \longrightarrow I^*$,

(2) $I^* \longrightarrow 2\dot{R}$,

(3) $\dot{R} + M \xrightarrow{k_3} \dot{P}_1$,

(4) $\dot{P}_n + M \xrightarrow{k_4^n} \dot{P}_{n+1}$, $1 < n < \infty$,

(5) $\dot{R} + \dot{R} \xrightarrow{k_5} I$,

(6) $\dot{R} + \dot{P}_N \xrightarrow{k_6^N} D_N$, $1 < N < \infty$,

(7) $\dot{P}_N + \dot{P}_M \xrightarrow{k_6^{NM}} D_{N+M}$, $1 < N + M < \infty$.

The first two stages describe the photoexcitation of the initiator and its dissociation into controlling radicals (chain initiation), stages 3 and 4 describe the growth of polymer chains (continuation of the chain), while stages 5–7 describe the extinction of controlling radicals as a result of their combination with one another and with polymer chains (rupture of the chain). It should be noted that in contrast to the ordinary chain reactions, the end product in our case emerges not at the stage of chain continuation, but at the stage of its rupture (reactions 6 and 7).

For the purpose of analysis, we traditionally assume [6] that the constants in the stages of continuation and rupture of chains are independent of the length of the reacting radical chains. It is true that this assumption is not quite satisfactory (especially for reactions 5 and 7) for a detailed prediction of the results in the case of certain specific reagents, but it is nevertheless treated as a traditional assumption in analysis. Thus we obtain

$$k_4 = k_3 = k_4^n,$$

for all n, and

$$k_7 = k_5 = k_6^N = k_7^{NM}$$

for all N and M.

In this case, we can write the following expression in the conventional approximation of the quasi-stationary concentrations of radicals [6]:

$$0 = \frac{d\dot{R}}{dt} = gWC_I - k_4\dot{R}C_M - k_7\dot{R}\left(R + \sum_1^\infty \dot{P}_m\right),$$

$$0 = \frac{d\dot{P}_n}{dt} = k_4C_M(\dot{P}_{n-1} - \dot{P}_n) - k_7\left(\dot{R} + \sum_1^\infty \dot{P}_m\right)\dot{P}_n, \quad (3)$$

$$n = 1, \ldots, \infty, \text{ for } n = 1 \quad P_{n-1} = P_0 \equiv \dot{R}.$$

Here g is the effective constant of photoinduced initiation of a chain. Summing Eqs. (3) and considering that $P_\infty = 0$, we obtain the expression

$$gWC_I = k_7X^2 \longrightarrow X = \frac{1}{k_7}\sqrt{k_7gW(x)C_I}, \quad (4)$$

for the sum of concentrations of radicals

$$X = \dot{R} + \sum_1^\infty \dot{P}_m.$$

It follows from formula (4) that X depends only on the parameters whose values in the active volume coincide with their mean values. Thus, using the standard equation

$$\frac{dM_f}{dt} = -k_4M_f\left(\dot{R} + \sum_1^\infty \dot{P}_n\right)$$

$$= -k_4XM_f = -\frac{k_4}{k_7}\sqrt{k_7gW(x)C_I}M_f,$$

for the rate of depletion of the monomer, we can easily obtain an expression for the number of elementary acts in a reaction involving this monomer in the active volume $V - \Delta V$ (and hence in the entire volume V depicted in Fig. 1):

$$dN_M^{ch} = -\frac{k_4}{k_7}\sqrt{k_7gWC_I}\frac{N_M}{V-\Delta V}dt(V-\Delta V).$$

Here N_M is the total number of monomer molecules in the volume element V. In terms of the mean concentrations, we can easily write an expression of the same kind:

$$\frac{dM}{dt} = -\frac{k_4}{k_7}\sqrt{k_7gWC_I}M. \quad (5)$$

It now remains for us to formulate an equation for the dynamics of growth of the "passive volume concentration" *Y*. The latter quantity is the sum of concentrations D_N of "*N*-mers" taken from the smallest "immobile" chain length N_0, multiplied by the corresponding chain lengths *N*:

$$Y(x, t) = \sum_{N_0}^{\infty} N D_N,$$

as described above. The rate of production of such an *N*-mer is

$$\frac{\partial D_N}{\partial t} = \frac{k_7}{2}\sum_{0}^{N} P_m P_{N-m},$$
$$\frac{P_m}{P_{m-1}} = \frac{1}{1 + k_7 X/k_4 M} = q, \quad (6)$$
$$P_m = X(1-q)q^m.$$

Here we have taken into consideration reactions 6 and 7 by simply taking the sum from $m = 0$ instead of $m = 1$, and putting $P_0 = R$ by definition. The ratio of the concentrations of radicals with two successive *m*'s is the definition of parameter *q* and can be obtained easily from the system of Eqs. (3). Although this ratio depends on time, it is independent of *m*. Hence at any given instant of time, P_m's form a decreasing geometric progression. Thus, introducing the relative concentrations of components

$$\sigma = C_L/Z, \quad \mu = M/Z, \quad \nu = Y/Z,$$

as well as the dimensionless time and coordinate

$$a_0\tau = t, \quad q_0 x = \xi,$$

$\alpha_0 = \sqrt{gW_0C_I/k_7}$ is the reciprocal characteristic reaction time), we arrive at the following system:

$$\frac{\partial \mu}{\partial \tau} - B\frac{\partial}{\partial \xi}\left[(1-\nu)^{2/3}\frac{\partial}{\partial \xi}\frac{\mu}{1-\nu}\right] + \sqrt{1 + m\sin\xi}\,\mu = 0,$$
$$\frac{\partial \nu}{\partial \tau} = \frac{G}{2}(1 + m\sin\xi)$$
$$\times\left[N_0^2(1-q) + N_0(1+q) + \frac{2q}{1-q}\right]q^{N_0}, \quad (7)$$
$$q = \left[1 + \frac{G\sqrt{(1 + m\sin\xi)}}{\mu}\right]^{-1}.$$

Here

$$B = \frac{D(q_0)^2}{\alpha_0} = \frac{4\pi D\sqrt{k_7}}{\sqrt{gW_0C_I}\Lambda^2},$$

$$G = \frac{k_7\alpha_0}{k_4^2 Z} = \frac{\sqrt{k_7 g W_0 C_I}}{k_4^2 Z}.$$

The initial conditions have the form

$$\mu(0, \xi) = \mu_0 = M_0/Z, \quad \nu(0, \xi) = 0. \quad (8)$$

We have intentionally omitted the material balance equation for the LC concentration σ from the system (7), (8) for the following reasons. First, the system of Eqs. (7), (8) is self-consistent and can be solved autonomously. Second, in order to analyze the evolution of the concentration σ with time, we must consider the dynamics of production and diffusion of short mobile chains in accordance with the condition of incompressibility, which makes the problem extremely complicated. On the other hand, we need to find only the final steady-state distribution σ(ξ) which is defined completely by the final distribution ν(ξ) through the condition of homogeneity of the equilibrium concentration in the active volume:

$$\sigma = a(1 - \nu(\xi)), \quad \int_0^1 \sigma(\xi)d\xi = \sigma_0 = C_L^0/Z. \quad (9)$$

The normalized integral in (9) is required for defining *a* for a given final profile σ(ξ).

Obviously, the nonlinear system of Eqs. (7), (8) cannot be integrated analytically in the general case. Hence we will consider possible limiting cases.

4. POSSIBLE REGIMES FOR THE PROCESS

There are two basic parameters *G* and *B* governing the evolution of the concentration distribution. An additional parameter N_0 is determined by the invariable properties of the given medium, which cannot be affected by the conditions of the experiment. It can be used as a fitting parameter for a given initial composition of the mixture (we mean the set of components and not their initial concentrations).

Parameter *G*, which is related to the radiation intensity, controls the rate of formation of the passive volume as well as the length of the chain averaged over the grating interval (through the initial value q_0). The term in the brackets in the expression for *q* is just the "reciprocal kinetic length of the chain," i.e., the ratio of the rates of continuation and rupture of the chain. The smaller this ratio, the larger the value of *q*, and hence the concentration P_m of radicals for large *m*. Consequently, neither too small (say, of the order of 0.01), nor too large (say, more than 10) initial values of this parameter are acceptable in view of the considerations formulated above during a qualitative analysis of the problem.

Parameter *B*, which is associated with the intensity as well as the three-dimensional grating interval controls the ability of the monomer to diffuse over the grating interval during the reaction time, i.e., controls the

switching of the process from the "fast" mode to the "slow" mode. Obviously, the "boundary" of these modes lies in the vicinity of $B = 1$.

Another circumstance is worth noting. It was mentioned above that a large value of B (i.e., a small value of the characteristic diffusion time) may activate the diffusion of radicals, leading to a "blurring" of the grating. Let us estimate the value of B corresponding to this situation. For this purpose, we carry out the summation of the system of Eqs. (3) once again, putting $M = M_0$ instead of assuming that the right-hand sides are equal to zero (we consider the beginning of the process of initiation, when the monomer is practically unused). In this case, we arrive at the following expression for the sum of concentrations X of the radicals:

$$\frac{dX}{dt} = k_7X_0^2 - k_7X^2 \longrightarrow X(t) = X_0\frac{1-\exp(-\beta t)}{1+\exp(-\beta t)}, \quad (10)$$

where

$$\beta = \frac{1}{2}\sqrt{k_7gWC_I} = \frac{k_7}{2k_4}\alpha_0\sqrt{1+m\sin\xi}$$

and

$$X_0 = \sqrt{\frac{gWC_I}{k_7}}$$

is the standard expression for the steady-state value of X. It should be observed that the expression for β coincides with the expression for the reciprocal characteristic time of the reaction except for the large factor $k_7/2k_4$ ($k_7 \gg k_4$, since the chain rupture reactions are practically activationless, while the continuation reactions presume the rupture of a double bond). It can easily be concluded from here that the relaxation time for the concentration of radicals considered here becomes commensurate with the diffusion time for

$$B = k_7/2k_4.$$

We now observe that B is constant for a constant value of $\Lambda^2\sqrt{W_0}$, and assume that a decrease in Λ for a constant optimal value of the intensity leads, for example, to the value

$$B = 10k_7/2k_4.$$

In this case, the radicals diffuse actively and obstruct the grating recording. In order to overcome this difficulty for a given value of Λ, i.e., in order to attain the value

$$B \approx 0.1k_7/2k_4,$$

we must increase W_0 by four orders of magnitude. However, this is in contradiction to the requirement of not-too-large values of G (see above). Thus, a combination of the conditions $B < k_7/2k_4$ and $G < 10$ (approximately) leads to the following constraint:

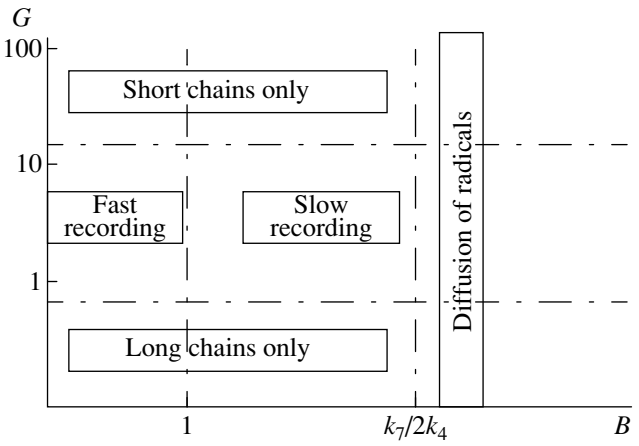

Fig. 2. Schematic diagram showing the possible lattice recording modes.

$$\Lambda \geq \sqrt{\frac{8\pi^2D}{k_4Z}} \approx \Lambda_C, \quad (11)$$

where Λ_C is the minimum possible value of the grating interval observed experimentally for the given mixture, below which the gratings cannot be recorded for any radiation intensity.

Figure 2 shows schematically the regions of the parametric plane BG corresponding to the above analysis. In spite of the fact that this is just an approximate schematic diagram, it can be useful for choosing the experimental conditions for recording of gratings.

Let us now consider the limiting cases in which the system of Eqs. (7), (8) can be solved analytically.

4.1. "Fast" Mode

In this case, we assume that $B \ll 1$ and disregard the diffusion term in the constituent Eq. (7) for the monomer. The equation then assumes a simplified form:

$$\frac{\partial\mu}{\partial\tau} + \sqrt{1+m\sin\xi}\,\mu = 0, \quad (12)$$

whence we obtain the following expression for the monomer concentration:

$$\mu = \mu_0\exp(-a\tau), \quad a = a(\xi) = \sqrt{1+m\sin\xi}. \quad (13)$$

We can now explicitly integrate the second equation in (7). Taking into account the explicit form of

$$q(\tau) = [1 + Ga\exp(a\tau)/\mu_0]^{-1},$$

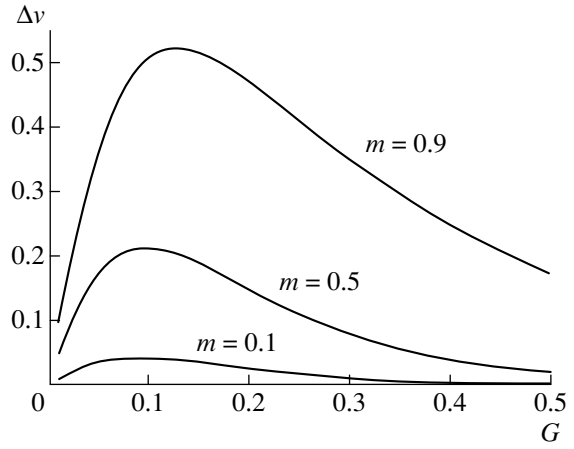

Fig. 3. Dependence of the amplitude $\Delta\nu$ of concentration modulation on the dimensionless intensity G for the "fast" mode. The initial monomer concentration is $\mu_0 = 0.5$. The curves correspond to three different values of the contrast m of the interference pattern.

we can present it as follows:

$$\frac{d\nu}{dq} = \frac{Ga^2}{2}\left[N_0^2 q^{N_0-1} + N_0\frac{1+q}{1-q}q^{N_0-1} + \frac{2q^{N_0}}{(1-q)^2}\right].$$

As a result, we arrive at the following expression for the final distribution of ν:

$$\nu_\infty = \frac{\mu_0}{2}q_0^{N_0-1}(N_0 + q_0 + 1 - N_0 q_0), \quad q(\tau = 0) = q_0, \quad q(\infty) = 0. \tag{14}$$

For a quite small radiation intensity ($G/\mu_0 \ll 1$), we obtain

$$\nu_\infty \simeq \mu_0\left[1 - (N_0 - 1)\frac{Ga}{\mu_0}\right].$$

Thus, in accordance with the qualitative analysis presented above, the modulation of the monomer concentration $Y(x)$ is found to be weak for low radiation intensities. For large radiation intensities ($G/\mu_0 \gg 1$), the modulation is also weak in view of a small value of the concentration $Y(x)$ itself:

$$\nu_\infty \simeq \frac{\mu_0}{2}(N_0 - 1)\left(\frac{\mu_0}{Ga}\right)^{N_0-1}.$$

Figure 3 shows the dependences of the amplitude of the concentration modulation $\Delta\nu$ on parameter G for $\mu_0 = 0.5$ (a typical experimental value [4]) and $N_0 = 10$ (which seems to be a realistic value for the loss of mobility in a polymer chain), calculated by using the formula

$$\Delta\nu = \frac{\nu_{\max} - \mu_{\min}}{2} = \nu_\infty(G, \mu_0, a = \sqrt{1+m}) - \nu_\infty(G, \mu_0, a = \sqrt{1-m}). \tag{15}$$

Strictly speaking, this amplitude does not determine the diffraction efficiency K of the obtained grating. The parameter which actually determines the diffraction efficiency is the first spatial Fourier component $\nu_{\sin}$:

$$\nu_{\sin} = \frac{1}{\pi}\int_{-\pi}^{\pi}\nu(\xi)\sin\xi\, d\xi,$$

$$K \simeq \sin^2\left(\frac{\pi(n_L - n_M)\nu_{\sin}L}{\lambda}\right).$$

Here, n_L and n_M are the refractive indices of the pure nematic and monomer respectively. However, the corresponding integrals cannot be evaluated analytically, and formula (15) describes quite accurately the qualitative form of the dependence $K(G)$.

Thus, it can be seen that the qualitative conclusions for the "fast" mode are in satisfactory agreement with the experimental results.

In this work, we shall not delve into the details of the obtained results since a quantitative comparison with the experimental data requires not only more detailed computations of the dependences $\nu_\infty(G, B, \mu_0, m)$, but also a large number of additional experimental measurements.

4.2. "Slow" Mode

Unfortunately, an analysis (direct solution of the system of Eqs. (7)) of this mode is not possible in the general case. We shall consider the simplest case for determining the quantitative nature of the effect of diffusion on the resulting diffraction efficiency of the grating (see the qualitative analysis above). Such an approach is possible only in the framework of perturbation theory. In other words, we assume that

$$B \ll 1, \quad \mu = \mu_F + \mu_1, \quad \nu = \nu_F + \nu_1,$$

the subscript "F" refers to the solution in the "fast" mode ($B = 0$), while μ_1 and $\nu_1 = O(B)$ are the first-order corrections in perturbation theory. Linearizing system (7) in μ_1, ν_1 and B, we can easily obtain the following equations:

$$\frac{\partial\mu_1}{\partial\tau} + a(\xi)\mu_1 = \hat{D}(\mu_F), \quad \frac{\partial\nu_1}{\partial\tau} = \left.\frac{\delta F}{\delta\mu}\right|_{\mu_F}\mu_1 = f(\mu_F)\mu_1. \tag{16}$$

Here, $\hat{D}$ is the diffusion operator from the first equation in system (7), acting on the solutions $\mu(\tau, \xi)$ and $\nu(\tau, \xi)$ in the "fast" mode, and F stands for the right-hand side of the second equation in (7). The general solution of the system of Eqs. (16) can obviously be written in the form

$$\nu_1 = \int_0^\tau d\tau_1 f(\mu_F)\exp(-a\tau_1)\int_0^{\tau_1} d\tau_2 \hat{D}(\mu_F)\exp(\tau_2).$$

Considering that

$$\nu_{1\infty} = \int_0^\infty d\tau \frac{\partial F}{\partial \tau}\int_0^\tau d\tau_1 \hat{D}(\mu_F(\tau_1))\frac{1}{a\mu_F(\tau_1)},$$

we obtain for the final modulation

$$f(\mu_F)\exp(-a\tau) = \frac{\delta F}{\delta \mu}\bigg|_{\mu_F}\frac{d\mu_F}{d\tau}\frac{1}{a\mu_0}.$$

Integrating this expression by parts (taking into account the explicit form $F(\mu_F) = \partial\nu_F/\partial\tau$ and $F(\infty) = 0$), we obtain

$$\nu_{1\infty} = \int_0^\infty d\tau \frac{\partial \nu_F}{d\tau}\frac{\hat{D}(\mu_F)}{a\mu_F}. \quad (17)$$

An analytic evaluation of this integral may not be possible in the general case because of the complex form of $\hat{D}(\mu_F)$. For large intensities ($G/\mu_0 \gg 1$), however, q is found to be small ($q \ll 1$) from the very beginning of the process, and hence $\nu_\infty \ll \mu_0$. In this case, the expression for $\hat{D}(\mu_F)$ is simplified considerably:

$$\hat{D}(\mu_F) \approx B\left[\left(\tau\frac{\partial a}{\partial \xi}\right)^2 - \tau\frac{\partial^2 a}{\partial \xi^2}\right],$$

which gives

$$\nu_{1\infty} = -\frac{1}{a}\frac{\partial^2 a}{\partial \xi^2}\int_0^\infty \frac{\partial \nu_F}{\partial \tau}\tau d\tau + \frac{1}{a}\left(\frac{\partial a}{\partial \xi}\right)^2\int_0^\infty \frac{\partial \nu_F}{\partial \tau}\tau^2 d\tau. \quad (18)$$

Substituting into this equation the explicit form of $\partial\nu_F/\partial\tau$, we can easily obtain the expression

$$\frac{\nu_{F\infty}}{\nu_{1\infty}} \simeq -\frac{B}{N_0 a^2}\frac{\partial^2 a}{\partial \xi^2}. \quad (19)$$

The second integral in (18) is a quantity of the order of $1/N_0$ as compared to the first integral and can therefore be disregarded while writing expression (19).

Let us now assume that the modulation depth of the interference pattern is small, i.e., $m \ll 1$. In this case, the corresponding Fourier components can be calculated easily:

$$(\nu_{F\infty})_{\sin} \simeq (N_0^2 - 1)\frac{\mu_0 m}{4}\left(\frac{\mu_0}{G}\right)^{N_0 - 1},$$
$$(\nu_{1\infty})_{\sin} \simeq -\frac{B}{N_0}(\nu_{F\infty})_{\sin}. \quad (20)$$

Here the subscript "sin" indicates the first spatial Fourier component of the quantity in the parentheses preceding it.

Thus, we arrive at the conclusion (which is certainly not rigorous in view of a large number of assumptions made in the above analysis) that the "initiation" of diffusion of the monomer upon approaching the "slow" mode decreases the diffraction efficiency of the grating (the quantities $(\nu_{F\infty})_{\sin}$ and $(\nu_{1\infty})_{\sin}$ in (20) have opposite signs).

Let us now consider the results of numerical computations for arbitrary values of m and B.

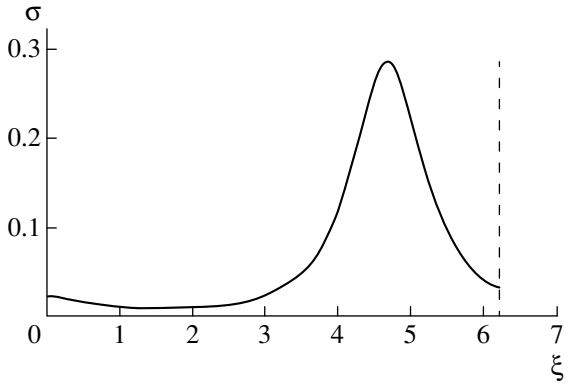

Fig. 4. A typical $\sigma(\xi)$ profile obtained as a result of a numerical solution of the system (7), (8).

4.3. Numerical Simulation of the Process

The system of Eqs. (7) and (8) was solved numerically for the values of the initial monomer concentration $\mu_0 = 0.5$–0.9 and the modulation depth $m = 0.9$, which are close to the experimental values (the value of m in actual experiments was practically equal to unity, but such a contrast of the pattern is "unacceptable" for numerical methods on account of the singularities at the points with zero intensity). Implicit and explicit four-point diagrams were used for a numerical solution of the system of equations. Figure 4 shows a typical resultant $\sigma(\xi)$ profile. As expected, this profile is considerably nonsinusoidal, and hence we shall operate with the amplitude F and phase φ of its first Fourier component. It is this amplitude that determines the diffraction efficiency of the grating, while the phase describes the possible spatial grating shift of the refractive index relative to the interference pattern. The results are presented in Fig. 5 in the form of dependences of amplitude F and phase φ on parameter B for

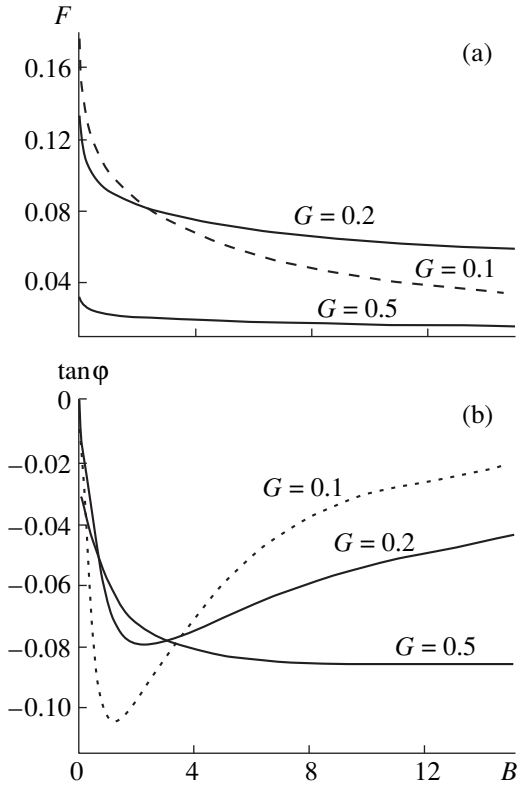

Fig. 5. Dependence of amplitude F (a) and phase φ (b) on parameter B for various values of parameter G.

various values of parameter G. In accordance with the preliminary conclusions drawn in the preceding section, the dependence $F(B)$ is found to be a decreasing function. It should be noted in particular that the higher the value of G, the sharper the decrease in the dependence. Let us compare the curves corresponding to $G = 0.1$ and $G = 0.2$ in Fig. 5a. These curves intersect for $B \approx 2$, which leads to the following qualitative conclusion: while the lower of the two intensities (corresponding to the above-mentioned values of G) is preferable from the point of view of diffraction efficiency for small values of B (for large grating intervals), the higher intensity is preferable for large B. This conclusion is in agreement with the above-mentioned experimental results concerning the increase in optimal intensity upon a decrease in the grating interval.

As regards the dependence $\varphi(B)$, the nonzero values of the directed phase shift of the grating relative to the interference pattern appear rather strange from general considerations, since the initial formulation of the problem was symmetric and did not contain polar axes. However, it must be borne in mind that, first, the system of Eqs. (7), (8) (whose stability is yet to be tested) is considerably nonlinear and, second, a directional energy-exchange between the recording waves was observed experimentally [7] in the course of polymerization. It is well known that such an energy exchange is typical of photorefraction (see, for example, [8]), and can occur only for nonzero values of φ. Hence the qualitative physical reasons behind nonzero grating shifts are not clear at present, and further investigations must be carried out to resolve this issue.

5. CONCLUSION

Thus, it can be stated that the model of volume-restricted diffusion proposed by us is qualitatively in agreement with the observed results. A preliminary analysis reveals that this model is capable of explaining all of the main qualitative features of nontrivial experimental dependences of the diffraction efficiency on the intensity of the recording radiation and grating interval obtained for polymer composites containing LCs.

Further quantitative comparisons of the predictions of the model with the experimental results are being made at present, but additional measurements are required before specific conclusions can be drawn. It should be remarked that the model described here is based on the real kinetics of molecular transformations and mass transfer, and practically does not contain any abstract fitting parameters. Hence it can be used effectively not only for explaining the obtained results, but also for predicting the optimal conditions for recording of gratings in specific materials.

ACKNOWLEDGMENTS

The authors are grateful to D. Duka, V. Lazarev, S. Nersesyan, and G. Cipparonne for fruitful discussions.

This research was partially supported by the European Foundation for Regional Development under the project “Progetto Sud INFM”, subproject FESR-UME.

REFERENCES

1. R. L. Sutherland, V. P. Tondiglia, L. V. Natarajan, and T. J. Bunning, Chem. Mater. **5**, 1533 (1993).
2. R. L. Sutherland, V. P. Tondiglia, and L. V. Natarajan, Appl. Phys. Lett. **64**, 1074 (1994).
3. R. L. Sutherland, V. P. Tondiglia, L. V. Natarajan, *et al.*, J. Nonlinear Opt. Phys. Mater. **5**, 89 (1996).
4. R. Caputo, A. V. Sukhov, and C. Umeton, Mol. Mater. **12**, 192 (1999).
5. P. S. Drzaich, *Liquid Crystal Dispersions* (World Scientific, Singapore, 1995).
6. P. W. Atkins, *Physical Chemistry* (Oxford Univ. Press, Oxford, 1987), p. 717.
7. A. V. Sukhov, N. V. Tabirian, and C. Umeton, private communication.
8. P. Gunter and J. P. Huignard, *Photorefractive Materials and Their Applications* (Springer-Verlag, Berlin, 1989).

Translated by Ram Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 37–60.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 45–70.

ATOMS, SPECTRA, RADIATION

Generalized Sturm Expansions of the Coulomb Green's Function and Two-Photon Gordon Formulas

A. A. Krylovetsky*, N. L. Manakov, and S. I. Marmo**

Voronezh State University, Universitetskaya pl. 1, Voronezh, 394693 Russia
**e-mail: kryl@kryl.vsu.ru*
***e-mail: manakov@thp.vsu.ru*

Received August 8, 2000

Abstract—The radial component of the Coulomb Green's function (CGF) is written in the form of a double series in Laguerre polynomials (Sturm's functions in the Coulomb problem), which contains two free parameters α and α'. The obtained result is applicable both in the nonrelativistic case and for the CGF of the squared Dirac equation with a Coulomb potential. The CGF is decomposed into the resonance and potential components (the latter is a smooth function of energy) for $\alpha = \alpha'$. In the momentum representation, the CGF with the free parameters is written in the form of an expansion in four-dimensional spherical functions. The choice of the parameters α and α' in accordance with the specific features of the given problem radically simplifies the calculation of the composite matrix elements for electromagnetic transitions. Closed analytic expressions (in terms of hypergeometric functions) are obtained for the amplitudes of bound–bound and bound–free two-photon transitions in the hydrogen atom from an arbitrary initial state $|nl\rangle$, which generalize the known (one-photon) Gordon formulas. The dynamic polarizability tensor components $\alpha_{nlm}(\omega)$ for an arbitrary n are expressed in terms of the hypergeometric function ${}_2F_1$ depending only on l and $\tilde{\omega}$ and through the polynomial functions $f_{nl}(\tilde{\omega})$ of frequency $\tilde{\omega} = \hbar\omega/|E_n|$. The Rydberg ($n \gg 1$) and threshold ($\hbar\omega \sim |E_n|$) asymptotic forms of polarizabilities are investigated. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The probabilities of one- and many-photon processes in the framework of perturbation theory in the interaction with laser fields can be expressed in terms of the matrix elements of the corresponding transitions between the initial and final states of an atom. In the case of many-electron atoms, such matrix elements can be calculated only approximately, but for hydrogen-like systems, such calculations can be carried out exactly. In most problems, the interaction with the field can be regarded as a dipole interaction. In this case, the radial matrix elements $\langle n'l' = l \pm 1|r|nl\rangle$ of one-photon transitions between the initial $|nl\rangle$ and the final $|n'l'\rangle$ state in the discrete spectrum of the hydrogen atom can be calculated analytically in the form of a combination of two hypergeometric functions ${}_2F_1(a, b; c; z)$ with integral negative parameters a and b (hypergeometric polynomials) and simple algebraic factors (the well-known Gordon formulas [1]):

$$\begin{aligned}\langle n'l-1|r|nl\rangle &= (-1)^{n'+1}\\ &\times\frac{[(n-l)_{2l+1}(n'-l+1)_{2l-1}]^{1/2}}{4z(2l-1)!}\frac{\xi^{l+1}}{(1-\xi)^{(n+n')/2}}\\ &\times[{}_2F_1(-n+l+1,-n'+l;\,2l;\,\xi)\\ &-(1-\xi){}_2F_1(-n+l-1,-n'+l;\,2l;\,\xi)],\end{aligned}\qquad(1)$$

where

$$\xi = -4nn'/(n-n')^2,$$

and $(a)_k = \Gamma(a+k)/\Gamma(a)$ is the Pochhammer symbol. Here and below, the formulas are written in atomic units.

Gordon's formulas completely describe the emission and absorption of a photon by an electron bound in the Coulomb potential, and an analytic continuation of formula (1) in n' (in the given case, the substitution $n' \longrightarrow i/\sqrt{2E}$) gives the amplitudes $\langle El' = l \pm 1|r|nl\rangle$ of bound–free transitions (photoionization and recombination), which are also polynomials. The expressions for the amplitudes $\langle E'l' = l \pm 1|r|El\rangle$ of free–free transitions (of the bremsstrahlung type) are also well known, in which an additional singular term with a delta function $\delta(E - E')$ emerges along with the term obtained by analytic continuation in n and n' and having the form of a combination of two complete hypergeometric functions ${}_2F_1$. Such analytical formulas in the theory of atomic photoinduced processes are of the reference type: tabulated results as well as numerous approximations for particular values of n, l, and/or energy E are widely used in the problem of classical (one-photon) optical spectroscopy.

The introduction of lasers into optical spectroscopic technique and the experimental investigations of many-

photon processes stimulated analytical calculations of the cross sections for many-photon (primarily, two-photon) transitions in a Coulomb field. Even for the radial matrix element

$$\mathcal{M}^{L=l\pm1}_{l;\,l'=l,l\pm2}(n, n', E = E_n \pm \omega) = \sum_k \frac{\langle n'l'|r|kL\rangle\langle kL|r|nl\rangle}{E_k - E_n \mp \omega - i0} \equiv \langle n'l'|rg_L(E; r, r')r'|nl\rangle, \quad (2)$$

of a two-photon transition, the problem is obviously complicated basically since the calculation of spectral sums is necessary in this case. It is clear that even when calculated in closed form, the matrix element (2) must have a more complex functional form than expression (1) since, in addition to the dependence on the quantum numbers n, n', and l, it depends essentially (resonantly) on the continuous parameter, viz., the frequency ω of the external monochromatic perturbation. The history of the analytic calculations of amplitudes of type (2) for a Coulomb potential runs into nearly 40 (see, e.g., [2]) and includes about one hundred publications. A number of alternative approaches were used (including various modifications of the method for integrating inhomogeneous differential equations for the first-order correction function in transient perturbation theory and algebraic approaches based on $O(4)$ symmetry of the Coulomb problem), but the most effective method for calculating spectral sums is that using the explicit expression for Green's function

$$G_E = \sum_{klm} \frac{|klm\rangle\langle klm|}{E_k - E - i0} = \frac{1}{\hat{H} - E - i0}, \quad (3)$$

and the Schrödinger equation with the Coulomb Hamiltonian. For transitions between the initial and/or final states with a fixed orbital angular momentum, it is most expedient to use the multipole expansion of G_E:

$$G_E(\mathbf{r}, \mathbf{r}') = \sum_{lm} g_l(E; r, r')Y_{lm}(\hat{\mathbf{r}})Y^*_{lm}(\hat{\mathbf{r}}'), \quad (4)$$

where $Y_{lm}(\hat{\mathbf{r}})$ is a spherical function. In this case, the problem is reduced to the calculation of matrix elements of type (2) with a radial Green's function $g_l(E; r, r')$.

The effectiveness of the application of the Coulomb Green's function (CGF) in the theory of two-photon processes was demonstrated for the first time in an analysis of the dynamic polarizability $\alpha_{1s}(\omega)$ of the ground state of hydrogen in a compact analytic form [3] (see also [4]):

$$\alpha_{1s}(\omega) = \frac{1}{\omega^2}[T(E_{1s} + \omega) + T(E_{1s} - \omega) - 1], \quad (5)$$

where

$$T(E) = \frac{\eta}{2(2-\eta)}{}_2F_1(1, 4; 3-\eta; x),$$

$$x = -\frac{(1-\eta)^2}{4\eta}, \quad \eta = \frac{Z}{\sqrt{-2E - i0}}.$$

Thus, in the simplest case when $n = n' = 1$, the matrix elements (2) can be reduced to the complete hypergeometric function ${}_2F_1$ with two integral first parameters, one of which is equal to unity (incomplete beta-function [5]). Similar analytical results were also obtained for transitions from the ground state to lower excited states with $n = 2, 3$ [6] and also for transitions between excited energy levels [6, 7] up to $n = 4$ [8]. In all cases, the results can be expressed in the form of a combination of the functions ${}_2F_1$, but the number of functions increases rapidly upon an increase in the principal quantum numbers n and n'. It should be noted that such functions can in principle be reduced to one of such functions through recurrence relations in view of above-mentioned integral values of the parameters. However, the corresponding algorithm is quite cumbersome and can be realized only by computer methods [9]. Along with the calculation of the amplitudes of two-photon transitions with preset values of n and n', some authors analyzed the matrix elements (2) for the general case of arbitrary n and/or n'. In this case, the results have a simple form only for elastic ($n' = n$) processes. A compact expression was obtained, for example, in [10] for the scalar component of the polarizability of the state $|nl\rangle$, emerging in the course of the computation of the Bethe logarithm for the Lamb shift (see also [8] for $n \le 4$ and [11] for the ns-states). In [12], three independent components of the polarizability tensor for an arbitrary n as well as the amplitude of the "nondiagonal" transition $|nl\rangle \longrightarrow |nl' = l \pm 2\rangle$ are expressed in terms of the linear combination $\sim(n_r = n - l - 1)$ of the functions ${}_2F_1$. The expression for the amplitude of inelastic transitions from the ground state, $|1s\rangle \longrightarrow |nl = 0, 2\rangle$, was obtained similarly, in the form of a combination of n hypergeometric functions [13]. The matrix elements (2) for the general case when $n \neq n'$ were analyzed in [14, 15], but no closed expression was obtained in terms of the known special functions. The method of the Sturm expansion of the CGF used in [14] makes it possible to present the result in the form of an infinite series of the products of two hypergeometric polynomials (similar to those appearing in the Gordon formula for photoionization), which in addition is found to be diverging at frequencies exceeding the ionization potential $|E_n|$ of the initial state $|nl\rangle$. In [16, 17], the result is written in the form of integrals that cannot be reduced to known special functions, while Marian [15], who claims to have obtained the result in a compact analytic form, presented it as a cumbersome sum of several dozens of functions of six parameters and four arguments of the hypergeometric type, which were specially introduced for the given problem. These functions are equivalent to a triple sum including ${}_2F_1$ functions.

An analysis of the matrix element (2) for bound–free transitions, $|nl\rangle \longrightarrow |El'\rangle$, is complicated by the presence of the Coulomb function of the continuous spectrum, which itself is a degenerate hypergeometric series. Nevertheless, the amplitude of two-photon ionization from the ground state can be calculated in a closed analytic form, but now in terms of complete hypergeometric functions of two variables, viz., Appel's functions F_1 [18, 19] (see also similar results obtained by Gavrila [20] for Compton's scattering with ionization). Although these calculations can easily be generalized to the case of low-lying excited levels also, no closed results for the amplitude were obtained in the general case, and various versions of numerical algorithms for an analytic continuation of the hypergeometric-type series (see, e.g., [21–23, 10]) are used in specific calculations for large n (especially for above-threshold ionization, when $\omega > |E_n|$). It should also be noted that matrix elements of type (2) (describing, e.g., double bremsstrahlung or Rutherford scattering in the presence of a light wave) for two-photon transitions between states of the continuous spectrum with fixed values of the orbital angular momentum can be expressed in terms of the integral of the complete function ${}_2F_1$ [24].

In the present work, we derive two-photon Gordon's formulas, viz., compact analytic expressions for matrix elements of type (2) with arbitrary n and n' and for bound–free transitions from the $|nl\rangle$ state in a canonically simple form (to the extent possible in such a general case) in terms of the known special functions. All analytic results obtained earlier for two-photon dipole transitions in the hydrogen atom are just special cases of these formulas. It was found that a more complex form of the results for arbitrary n and n' as compared to the ground state is associated not with the increase in the number of hypergeometric functions ${}_2F_1$ or Appel's function F_1, as could be expected from the well-known results for low- lying levels, but with the fact that the amplitudes can be expressed in the general case in the form of a simple linear combination of the products of two functions of the hypergeometric type. One of these functions is a hypergeometrical polynomial of an order not exceeding $n_< = \min\{n_r, n_r'\}$, which is similar to polynomials in the classical Gordon formula, while the other is Appel's function F_1. In the case of bound–free transitions, the latter function is similar to those determining the amplitude of two-photon ionization from the ground state. For bound–bound transitions, the above-mentioned Appel's function has an integral negative parameter and is equivalent to a finite sum of functions ${}_2F_1$ with a unit first parameter, which are similar to those appearing in Eq. (5). Besides, in both cases the result also contains polynomial terms with products of hypergeometric polynomials of one (${}_2F_1$) and two (F_1) variables. The results are simplified radically for "diagonal" matrix elements (2) with $n' = n$: in this case, polynomial terms are missing, and $\mathcal{M}^L_{l;\,l'}(n, n, E = E_n \pm \omega)$ can be written as a simple product of the sum of two hypergeometric polynomials and the sum of two complete hypergeometric functions ${}_2F_1$.

We succeeded in generalizing Gordon's formulas to the case of two-photon processes by using the effective computational method based on a new representation (generalized Sturm expansion) of the radial CGF $g_l(E; r, r')$ in the form of a double series in Laguerre polynomials (or Sturm's functions $S_{nl}(x)$ in Eq. (8) from the Coulomb problem), which contains two arbitrary (free) parameters α and α':

$$g_l(E; r, r') = \sum_{k,\,k'=0}^{\infty} g^l_{kk'}(\nu; \alpha, \alpha') S_{kl}\left(\frac{2r}{\alpha}\right) S_{k'l}\left(\frac{2r'}{\alpha'}\right). \quad (6)$$

The basic circumstance that renders considerable flexibility to the generalized Sturm expansion in various applications is the factorized dependence of the terms of series (6) on r and r' and the energy parameter $\nu = 1/\sqrt{-2E - i0}$. The entire dependence on energy E is contained in the kernel $g^l_{kk'}(\nu; \alpha, \alpha')$ which is independent of radial variables and can be expressed in terms of hypergeometric functions. A rational choice of parameters α and α' in accordance with the specific features of a given problem (e.g., $\alpha = n$ and $\alpha' = n'$ in an analysis of two-photon bound–bound transitions) makes it possible in some cases to radically simplify the computation procedure for matrix elements with $g_l(E; r, r')$. In particular, two-photon Gordon's formulas can be presented in the closed analytic form described above in terms of four functions $g^l_{kk'}$ with different k and k'. Thus, the amplitudes of two-photon processes are as if contained even in representation (6) itself for $g_l(E; r, r')$. It is essential that expansion (6) is valid for nonintegral values of the parameter $l = \gamma$ also. This allows us to obtain a two-parametric expansion for a nonrelativistic CGF also in terms of Sturm's functions of the squared Dirac equation with the Coulomb potential. For integral l and $\alpha' = \alpha$, expansion (6) is transformed into a one-parametric representation of Green's function for the nonrelativistic Coulomb problem, which was obtained earlier by Heller [25] (see also [10]).

In Section 2, we will derive expansion (6) and analyze some properties of the new representation of the CGF. Among other things, we will obtain a two-parametric representation of the CGF $G_E(\mathbf{p}, \mathbf{p}')$ in the momentum space, which generalizes the well-known result obtained by Schwinger, as well as the decomposition of $G_E(\mathbf{r}, \mathbf{r}')$ into the "resonance" component (containing poles for $E = E_n$) and the "potential" component (which is a smooth function of energy and is real-valued for real values of E). In Sections 3 and 4, expansion (6) will be used for deriving two-photon Gordon's for-

mulas for the amplitudes of inelastic bound–bound (Section 3) and bound–free (Section 4) transitions. In Section 5, the case when $n' = n$ will be analyzed in detail: the Stark matrix elements will be calculated for an energy level with an arbitrary n, and simple formulas will be derived for the scalar, vector, and tensor-type polarizability, which are the most natural generalizations of expression (5). The Rydberg and threshold asymptotic forms of polarizabilities will also be investigated.

Some of the results which will be presented here were briefly described in [26].

2. GENERALIZED STURM EXPANSION OF THE CGF

The existence of a convenient representation of Green's function suitable for a given problem is often a necessary condition for successful analytical or numerical calculations. For this reason, various versions of such representations have been developed starting from the middle of the 60s, when intense studies of the effect of laser radiation on atoms began, to nowadays. We will briefly consider two such versions (a detailed analysis of various CGF representations including the relativistic case and peculiarities of their application can be found, for example, in [27–29]).

In many-photon calculations, the most popular is the expansion of $g_l(E; r, r')$ into series in Sturm's function of the Coulomb problem [30]:

$$g_l(E; r, r') = \nu \sum_{k=0}^{\infty} \frac{k! S_{kl}(2r/\nu) S_{kl}(2r'/\nu)}{\Gamma(k+2l+2)(k+l+1-\eta)}, \quad (7)$$

where

$$S_{kl}(2r/\nu) = \frac{2}{\nu}(2r/\nu)^l \exp(-r/\nu) L_k^{2l+1}(2r/\nu), \quad (8)$$

L_n^{α} is a generalized Laguerre polynomial [31], $\nu = (-2E - i0)^{-1/2}$, and $\eta = Z\nu$. The Sturm expansion is useful for analytical transformations, but it is most suitable for direct numerical calculations of radial composite matrix elements $M_{fi}^{(N)}$ in higher orders of perturbation theory. In this case, the results can be presented in the form of multiple series of hypergeometric polynomials rapidly converging for subthreshold ($E < 0$) values of energy of intermediate states (Green's functions (3) and (4)). For above-threshold energies, the series for matrix elements become diverging since the series in Eq. (7) diverges for $E > 0$ (for imaginary ν). However, using an appropriate transformations of (7), we can extend the technique of computations with Sturm's functions to the given class of problems also [32]. Another method of the summation of series for matrix elements in the case of above-threshold energies involves the application of the Padé approximation techniques [33] (see also [34]). The Sturm expansion of a radial CGF (with an appropriate modification of the angular component in Eq. (4)) is successfully used in relativistic calculations also. In order to obtain Green's function $g_\gamma(E; r, r')$ of the squared Dirac equation from Eq. (7) (or (6)) [35], we must make the substitution $l \longrightarrow \gamma$ and put

$$\nu = \frac{\alpha}{\sqrt{1-\epsilon^2}}, \quad \epsilon = \frac{E}{mc^2}, \quad \alpha = \frac{e^2}{\hbar c},$$

$$\eta = \epsilon Z\nu, \quad \gamma = \sqrt{\left(j+\frac{1}{2}\right)^2 - (\alpha Z)^2} + \frac{s-1}{2},$$

$$s = \pm 1 \text{ at } j = l \mp \frac{1}{2}$$

(where j and l are the total and orbital angular momenta for electrons, respectively). Green's function (7) with a nonintegral parameter l emerges also in calculations with the model Füss potential (Coulomb potential with an additional centrifugal term proportional to $\propto r^{-2}$ [28]).

Along with representations (4) and (7), the CGF representations in the momentum space are also used widely. The most popular is the expansion of $G_E(\mathbf{p}, \mathbf{p}')$ with $E < 0$ in the discrete basis of four-dimensional spherical functions $Y_{klm}(\xi)$ defined on the unit hypersphere S ("Fock's sphere") in $\mathbf{R}^{(4)}$ [36]:

$$G_E(\mathbf{p}, \mathbf{p}') = \frac{16\nu^5}{[(1+\nu^2\mathbf{p}^2)(1+\nu^2\mathbf{p}'^2)]^2} \times \sum_{lm} \sum_{k=l+1}^{\infty} \frac{Y_{klm}(\xi) Y_{klm}^*(\xi')}{1-\eta/k}, \quad (9)$$

where ξ is the unit vector in $\mathbf{R}^{(4)}$:

$$\xi = \{\cos\varphi, \sin\varphi\hat{\mathbf{p}}\}, \quad \cos\varphi = \frac{1-\nu^2\mathbf{p}^2}{1+\nu^2\mathbf{p}^2}.$$

The harmonics $Y_{nlm}(\xi)$ orthonormal on the sphere S have the form

$$Y_{nlm}(\xi) = 2^{l+1} l! \left[\frac{n(n-l-1)!}{(n+l)! 2\pi}\right]^{1/2} \times (\sin\varphi)^l C_{n-l-1}^{l+1}(\cos\varphi) Y_{lm}(\hat{\mathbf{p}}), \quad (10)$$

where $C_m^{l+1}(\cos\varphi)$ is the Gegenbauer polynomial [31]. It can easily be verified that representation (9) is a momentum analog of expansions (4) and (7) in the coordinate space (see, for example, [27]) Representation (9) leads to a compact integral representation of the CGF [36, 37] which is widely used in computations. It turned out in practice that the effectiveness of coordinate and momentum representations of the CGF is approximately the same: the application of either of these representations leads in most cases to results of equivalent complexity (or just identical results).

2.1. Expression for the Kernel $g^l_{kk'}$

Our aim is to derive expansion (6) for $g_l(E; r, r')$ in generalized Sturm functions and arguments containing, instead of the energy parameter ν, the free parameters α and α' which can be chosen appropriately in each specific problem. It should be noted that the idea to introduce a free parameter in the CGF was successfully employed earlier. Manakov *et al.* [32] used a special form of the one-parametric Sturm expansion of $g_l(E; r, r')$ to continue the matrix elements analytically to the region of continuous spectrum ($E > 0$ in the nonrelativistic case and $|E| > mc^2$ in the relativistic case). Heller [25] derived a one-parametric expansion of $g^l_{kk'}(\nu; \alpha, \alpha)$ for the nonrelativistic case (see relation (17) below) on the basis of an alternative technique of trinomial recurrent relations for the reciprocal matrix of the Coulomb Hamiltonian on the quadratically integrable (L^2) basis $S_{kl}(2r/\alpha)$. The high effectiveness of this expansion in an analysis of two-photon matrix elements between identical hydrogen-like states was demonstrated by Broad [10].

In order to find the kernel $g^l_{kk'}$ of representation (6), we will use the formal reexpansion of Sturm's functions on (7) into a series in the complete system of these functions with another value of the argument containing the free parameter α:

$$S_{kl}\left(\frac{2r}{\nu}\right) = \sum_{n=0}^{\infty} c_{nk}(\alpha) S_{nl}\left(\frac{2r}{\alpha}\right). \tag{11}$$

Using the completeness of the functions $S_{nl}(2r/\alpha)$ and the well-known integral of the product of Laguerre polynomials, we can express coefficients c_{nk} in terms of the hypergeometric function with integral negative upper parameters (hypergeometric polynomial):

$$c_{nk}(\alpha) = (-1)^k \frac{(2l+2)_k}{k!}\left(\frac{4\alpha\nu}{(\alpha+\nu)^2}\right)^{l+1} \times \left(\frac{\alpha-\nu}{\alpha+\nu}\right)^{n+k} {}_2F_1(-n, -k; 2l+2; z).$$

Here and below, we use the notation

$$z = -\frac{4\alpha\nu}{(\alpha-\nu)^2}, \quad z' = -\frac{4\alpha'\nu}{(\alpha'-\nu)^2}. \tag{12}$$

Using expansion (11) with the free parameters α and α' as well as the coefficients $c_{nk}(\alpha)$ and $c_{n'k}(\alpha')$ for $S_{kl}(2r/\nu)$ and $S_{kl}(2r'/\nu)$ in expansion (7), we arrive at the following identity:

$$g^l_{nn'}(\nu; \alpha, \alpha') = f(\alpha, \alpha') \sum_{k=0}^{\infty} [(2l+2)_k \times {}_2F_1(-n, -k; 2l+2; z)\,{}_2F_1(-n', -k; 2l+2; z')] \times [k!(k+l+1-\eta)[(1-z)(1-z')]^{k/2}]^{-1}, \tag{13}$$

where the factor f has the form

$$f(\alpha, \alpha') = \nu \frac{(4\nu\sqrt{\alpha\alpha'})^{2l+2}}{\Gamma(2l+2)} \times \frac{(\alpha-\nu)^n}{(\alpha+\nu)^{n+2l+2}} \frac{(\alpha'-\nu)^{n'}}{(\alpha'+\nu)^{n'+2l+2}}.$$

Presenting the hypergeometric polynomial with the parameter $-n$ in (13) in differential form [5],

$${}_2F_1(-n, -k; 2l+2; z) = \frac{(1-z)^{k+n+2l+2}}{(2l+2)_n z^{2l+1}} \times \frac{d^n}{dz^n}[z^{n+2l+1}(1-z)^{-k-2l-2}], \tag{14}$$

and using the elementary identity

$$(k+l+1-n)^{-1} = \int_0^1 dt\, t^{k+l-\eta},$$

we can sum up series (13) with the help of the generating function [5] for ${}_2F_1(-n', -k; 2l+2; z')$. After a series of transformations, we obtain

$$g^l_{nn'}(\nu; \alpha, \alpha') = f(\alpha, \alpha') \frac{d^n}{dz^n}\{\varphi_n(z) I_{n'}(z)\}\Big|_{z_0 = z},$$

where

$$\varphi_n(z) = \frac{[z(1-z_0)]^n}{(2l+2)_n}\left(\frac{z}{z_0}\right)^{2l+1}\left(\frac{1-z_0}{1-z}\right)^{l+1+\eta},$$

$$I_{n'}(z) = \int_0^{(1-z_0)/(1-z)} dt \frac{t^{l-\eta}(1-yt)^{n'}}{(1-y't)^{n'+2l+2}}, \tag{15}$$

$$y = \sqrt{\frac{1-z'}{1-z_0}}, \quad y' = \frac{1}{\sqrt{(1-z')(1-z_0)}}.$$

The differentiation of $\varphi_n(z)$ again leads the hypergeometric polynomial

$$\frac{d^{n-p}}{dz^{n-p}}\varphi_n(z)\Big|_{z_0 = z} = \frac{[z(1-z)]^p}{(2l+2)_p} \times {}_2F_1(-n+p, l+1-\eta+p; 2l+2+p; z).$$

The integral $I_{n'}(z)$ for $z_0 = z$ gives the Appel function F_1 [5]:

$$I_{n'} \equiv I_{n'}(z)\big|_{z_0 = z} = \frac{1}{l+1-\eta} \times F_1(l+1-\eta; -n', n'+2l+2; l+2-\eta; y, y'). \tag{16}$$

It can be seen from (15) that for coinciding free parameters $\alpha' = \alpha$, all the derivatives $I_{n'}(z)$ with respect to z having an order not higher than n' vanish if we put

$z_0 = z$. Considering also that the argument $y = 1$ for $\alpha' = \alpha$ in (16), so that Appel's function $I_{n'}$ is transformed into ${}_2F_1$, and introducing the notation $n_< = \min\{n, n'\}$, $n_> = \max\{n, n'\}$, we can write $g^l_{nn'}(\nu; \alpha, \alpha)$ in the form

$$g^l_{nn'}(\nu; \alpha, \alpha) = \frac{\nu}{\Gamma(2l+2)}\left(\frac{\alpha-\nu}{\alpha+\nu}\right)^{n_<}$$

$$\times {}_2F_1(-n_<, l+1-\eta; 2l+2; z)\left(\frac{\alpha^2-\nu^2}{4\alpha\nu}\right)^{n_>} \qquad (17)$$

$$\times \frac{n_>!\,{}_2F_1(n_>+1, n_>+2l+2; n_>+l+2-\eta; z^{-1})}{(l+1+\eta)_{n_>+1}}.$$

For $\alpha' \neq \alpha$, the derivatives of $I_{n'}(z)$ can be calculated using the generating function for Appel's functions F_1 [38]. The final result has the form

$$g^l_{nn'}(\nu; \alpha, \alpha') = f(\alpha, \alpha')\left[\frac{{}_2F_1(-n, l+1-\eta; 2l+2; z)}{l+1-\eta}\right.$$

$$\times F_1(l+1-\eta; -n', n'+2l+2; l+2-\eta; y, y')$$

$$+ \sum_{p=1}^{n} C_n^p(-z)^p \qquad (18)$$

$$\times [{}_2F_1(-n+p, l+1-\eta+p; 2l+2+p; z)]$$

$$\left.\times \{(2l+2)_p\}^{-1}\Phi_p^{l,n'}\right].$$

Here,

$$\Phi_p^{l,n'} = -\frac{(l+2+\eta-p)_{p-1}(1-y)^{n'}}{(1-y')^{n'+2l+2}}$$

$$\times F_1(-p+1; -n', n'+2l+2;$$

$$l+2+\eta-p; 1/(1-y), 1/(1-y')),$$

$$y = \frac{\alpha-\nu}{\alpha+\nu}\frac{\alpha'+\nu}{\alpha'-\nu}, \quad y' = \frac{\alpha-\nu}{\alpha+\nu}\frac{\alpha'-\nu}{\alpha'+\nu},$$

and C_n^p is the binomial coefficient. Appel's function in $\Phi_p^{l,n'}$ is a finite polynomial in both arguments. Consequently, expansion (18) contains two essentially different groups of terms: Appel's function in the first term has an integral negative parameter $-n'$ and is equivalent to a linear combination of $(n'+1)$ hypergeometric functions ${}_2F_1$ that cannot be reduced to polynomials: the terms with $\Phi_p^{l,n'}$ are the products of hypergeometric polynomials in one variable (${}_2F_1$) and two variables (F_1). It can easily be verified that for $\alpha' = \alpha$, expression (18) is transformed into (17). In the above derivation, we did not use the fact that parameter l is an integer. Consequently, as was noted above (see also [26]), all the results are also valid for a radial CGF of the squared Dirac equation with a nonintegral $l = \gamma$. For integral l, expression (17) coincides with the result obtained by Heller [25].

It should be noted that the derivation of expressions for $g^l_{kk'}(\nu; \alpha, \alpha')$ can be carried out by transforming the second function in (13) instead of the first function ${}_2F_1$. In this case, we obtain for $g^l_{kk'}(\nu; \alpha, \alpha')$ an expression that follows from (18) after the substitution $(k, \alpha) \rightleftarrows (k', \alpha')$. This result corresponds to the symmetry condition

$$g^l_{kk'}(\nu; \alpha, \alpha') = g^l_{k'k}(\nu; \alpha', \alpha), \qquad (19)$$

which obviously follows already from the initial expansion (6). Pay attention to the fact that the symmetry condition (19) and the explicit expression (18) for $g^l_{kk'}(\nu; \alpha, \alpha')$ determine the nontrivial identity connecting bilinear forms containing the functions ${}_2F_1$ and F_1 (see formula (A.1) in Appendix A). This identity can be helpful in transforming the matrix elements calculated using function (6) (see Subsection 5.1 below).

2.2. Decomposition of the CGF into the Resonant and Potential Components

The double series for $g_l(E; r, r')$ derived above have a more complex structure as compared to the standard Sturm expansion (7). Consequently, it would be interesting to verify the fulfillment of the general properties of Green's functions for them. On the other hand, formulas (6), (17), and (18) allow us to obtain new results, in particular, the decomposition of $g_l(E; r, r')$ into the resonant component (containing poles for $E = E_n$) and the potential component (which is a smooth function of E):

$$g_l(E; r, r') = g_l^{\mathrm{res}}(E; r, r') + g_l^{\mathrm{pot}}(E; r, r'). \qquad (20)$$

In the simplest case when $\alpha' = \alpha$, the kernels of the Sturm expansion (6) for $g_l^{\mathrm{res}}(E; r, r')$ and $g_l^{\mathrm{pot}}(E; r, r')$ have the form (see Appendix A)

$$g^{l,\mathrm{res}}_{kk'}(\nu; \alpha) = \frac{1}{\Gamma(2l+2)}\frac{\nu}{l+1-\eta}\left(\frac{\alpha-\nu}{\alpha+\nu}\right)^{k+k'}$$

$$\times {}_2F_1(1, 2l+2; l+2-\eta; z^{-1}) \qquad (21)$$

$$\times {}_2F_1(-k, l+1-\eta; 2l+2; z)$$

$$\times {}_2F_1(-k', l+1-\eta; 2l+2; z),$$

$$g^{l,\mathrm{pot}}_{kk'}(\nu; \alpha) = \frac{-\nu}{\Gamma(2l+2)}\left(\frac{\alpha-\nu}{\alpha+\nu}\right)^{k+k'}$$

$$\times {}_2F_1(-k_<, l+1-\eta; 2l+2; z)$$
$$\times \sum_{p=1}^{k_>} \frac{C_{k_>}^p (-z)^p (\eta - l)_{p-1}}{(k_> + 2l + 2 - p)_p} \quad (22)$$
$$\times {}_2F_1(-k_> + p, l+1-\eta; 2l+2; z)$$
$$\times {}_2F_1(-p+1, 2l+2-p; l+2-\eta-p; z^{-1}).$$

It can easily be verified that $g_{kk'}^{l,\mathrm{pot}}$ is a smooth function of the energy E (of parameter η), while $g_{kk'}^{l,\mathrm{res}}$ has poles at $l+1-\eta = -n_r = 0, -1, -2. \ldots$, corresponding to resonances at the energy levels of the discrete spectrum with $E = E_n$.

The resonant term (21) is symmetric and factorized in k, k', and hence the double sum in an expression of the type (6) for g_l^{res} can be calculated in closed form using relation (A.7) from Appendix A:

$$g_l^{\mathrm{res}}(\nu; r, r') = \frac{1}{(2l+1)!}\frac{\nu}{l+1-\eta}\left(\frac{\nu^2-\alpha^2}{4\alpha\nu}\right)^{2l+2}$$
$$\times \left(\frac{\nu+\alpha}{\nu-\alpha}\right)^{2\eta} {}_2F_1(1, 2l+2; l+2-\eta; z^{-1}) \quad (23)$$
$$\times \frac{1}{rr'} M_{\eta, l+1/2}(2r/\nu) M_{\eta, l+1/2}(2r'/\nu).$$

Proceeding to the limit $\eta \longrightarrow l+1+n_r$, where $n_r = 0, 1, \ldots$, we can verify that the residues at the poles for $E = E_n$ are equal to the product of the corresponding eigenfunctions $R_{nl}(r)$ and $R_{nl}(r')$.

Proceeding from formula (22), we can prove by direct verification that the potential term $g_l^{\mathrm{pot}}(E; r, r')$ for real-valued α is real for all real-valued E, but this circumstance can be verified more easily on the basis of the results obtained by Broad [10]. Let us write expression (17) for $g_{kk'}^l(\nu; \alpha, \alpha)$ in terms of the notation proposed in [25, 10]:

$$g_{kk'}^l(E, \alpha, \alpha) = \frac{2\alpha p_{k_<}^l(x, \alpha) q_{k_>}^{+l}(E, \alpha)}{(1+2E\alpha^2)(k+1)_{2l+1}(k'+1)_{2l+1}}, \quad (24)$$

where

$$p_k^l(x, \alpha) = \frac{(2l+2)_k}{k!}\left(\frac{\alpha-\nu}{\alpha+\nu}\right)^k \times {}_2F_1(-k, l+1-\eta; 2l+2; z), \quad (25)$$

$$q_k^{+l}(E, \alpha) = -2\frac{\Gamma(k+2l+2)}{(l+1-\eta)_{k+1}}\left(\frac{\alpha^2-\nu^2}{4\alpha\nu}\right)^{k+1} \times {}_2F_1(k+1, k+2l+2; k+l+2-\eta; z^{-1}). \quad (26)$$

The functions p_k^l introduced above and proportional to the kth degree hypergeometric polynomials ${}_2F_1$ in z are also kth degree polynomials in

$$x = \frac{2E\alpha^2 - 1}{2E\alpha^2 + 1}$$

(Pollaczek polynomials [31]). These polynomials satisfy the trinomial recurrence relation

$$(k+1)p_{k+1}^l - 2((k+l+1-\alpha Z)x + \alpha Z)p_k^l + (k+2l+1)p_{k-1}^l = 0 \quad (27)$$

with the initial conditions

$$p_0^l = 1, \quad p_{-1}^l = 0$$

and belong to the class of so-called nonclassical orthogonal polynomials [31]. The functions q_k^{+l} are referred to in [25] as Pollaczek functions and can be treated as the second solution of the recurrence relation (27), which is linearly independent of p_k^l. It should be noted that, in accordance with (26), $q_{-1}^{+l} = -2(2l)!$, while the hypergeometric function appearing in (23) is proportional to q_0^{+l}. We are using here the Pollaczek polynomials and functions as a convenient notation for writing Green's function for $\alpha' = \alpha$ in compact form. When we solved the Coulomb problem by introducing the quadratically integrable basis $S_{kl}(2r/\alpha)$, the recurrence relation (27) was an analog of the Schrödinger equation, and Green's function was constructed using two linearly independent solutions of (27) [25] (see also [39]). In the notation p_k^l, q_k^{+l}, relation (A.6) from Appendix A can be written as follows [10]:

$$q_k^{+l}(E, \alpha) = \bar{q}_k^l(x, \alpha) + p_k^l(x, \alpha) q_0^{+l}(E, \alpha). \quad (28)$$

The functions $\bar{q}_k^l(x, \alpha)$ introduced here satisfy relation (27) with the initial conditions $\bar{q}_0^l = 0$, $\bar{q}_{-1}^l = -2(2l)!$ (which obviously follow from the initial conditions for q_k^{+l} and p_k^l) and are therefore polynomials in x. Although this fact was established in [10], the polynomial $\bar{q}_k^l(x, \alpha)$ was not obtained in an explicit form. Formula (A.6) makes it possible to find a closed (although rather cumbersome) expression for $\bar{q}_k^l$ in the form of a bilinear sum of hypergeometric polynomials.

Bearing in mind the application of Green's functions in the form of (18) or (17), (24) to the calculation of matrix elements of two-photon transitions, we put the free parameter $\alpha = n/Z$ and write the energy parameter of Green's function in the form $E = E_n \pm \omega$. Then $x = 1 \mp 2/\tilde{\omega}$, where $\tilde{\omega} = 2Z^{-2}n^2\omega$ is the dimensionless frequency measured in the units of the ionization poten-

tial of the state $|nl\rangle$ under investigation. It can now be easily seen from (27) that p_k^l and $\bar{q}_k^l$ are real-valued polynomials in $1/\tilde{\omega}$. In this way, we verify the real-valuedness of the potential term (22) in Green's function, which is proportional to the product $p_k^l(x, \alpha)\bar{q}_k^l(x, \alpha)$, and specify the frequency dependence of these polynomials.

The resonant terms (21) and (23) are complex-valued for $E > 0$. Consequently, in accordance with the general properties of Green's function, for a positive energy, the following relation must hold:

$$\begin{aligned} \mathrm{Im}g_l^{\mathrm{res}}(E>0; r, r') &= \mathrm{Im}g_l(E>0; r, r') \\ &= i\pi R_{El}(r)R_{El}(r'), \end{aligned} \tag{29}$$

where $R_{El}(r)$ are the wave functions of the continuous spectrum, normalized to energy:

$$\begin{aligned} R_{El}(r) &= \frac{i^{l+1}}{(2l+1)!}\frac{C_{El}}{2p}\frac{1}{r}M_{ia,\,l+1/2}(-2ipr), \\ C_{El} &= \sqrt{\frac{2p}{\pi}}e^{\pi a/2}|\Gamma(l+1+ia)|, \end{aligned} \tag{30}$$

where $a = Z/p$, $p = \sqrt{2E}$. The validity of relation (29) can be verified directly by separating the imaginary component of g_l^{res} in (23). It should be borne in mind in this case that for real α, the product of Whittaker's functions is real-valued, and the imaginary component of complex-valued terms can be determined using relation (A.8) in Appendix A.

Concluding the subsection, we note that a three-dimensional CGF can also be presented as the sum of the resonant and potential terms:

$$G(E; \mathbf{r}, \mathbf{r}') = G^{\mathrm{res}}(E; \mathbf{r}, \mathbf{r}') + G^{\mathrm{pot}}(E; \mathbf{r}, \mathbf{r}'), \tag{31}$$

where the expressions for G^{res} and G^{pot} obviously follow from (4), (6), (21), and (22). The decomposition of Green's functions into the resonant and potential components is convenient for some applications, for example, in collision problems (see, e.g., [40, 41]). The decomposition of the meromorphic function G_E into the "pole" and "smooth" components which are complex functions of energy is ambiguous, and expression (31) gives only one of such representations (see [41, 42] for details). The expression for the resonant component of the CGF $g_l(E; r, r')$ in terms of the Whittaker functions appearing in (23) was derived in [40] (see also [42]) by using the standard representation of $g_l(E; r, r')$ in terms of Whittaker's functions with the arguments $r_>$, $r_<$, where $r_>(r_<) = \max\{r, r'\}$ ($\min\{r, r'\}$).

2.3. Two-Parametric Momentum Representation of CGF

Free parameters can also be introduced in the momentum representation of the CGF in the form of (9). It can be done most simply by calculating the Fourier transform (in the variables $\mathbf{r}$ and $\mathbf{r}'$) of expression (4) using expansion (6) for $g_l(E; r, r')$. Evaluating the integral of the product of the Laguerre polynomials $L_n^{\alpha}(x)$ and the spherical Bessel function $j_l(pr)$ (emerging in the Fourier transformation) using the generating functions for $L_n^{\alpha}(x)$ as well as the Gegenbauer polynomials $C_k^{\beta}(\cos\varphi)$, and also taking into account the definition (10) of the hyperspherical harmonic Y_{nlm}, we obtain

$$\begin{aligned} G_E(\mathbf{p}, \mathbf{p}') &= \sum_{lm}\sum_{k,\,k'=l+1}^{\infty} G_{kk'}^l(\nu; \alpha, \alpha') \\ &\times Y_{klm}(\xi)Y_{k'lm}^*(\xi'), \end{aligned} \tag{32}$$

where

$$\xi = \{\cos\varphi, \sin\varphi\hat{\mathbf{p}}\}, \quad \cos\varphi = \frac{1-\alpha^2\mathbf{p}^2}{1+\alpha^2\mathbf{p}^2},$$

and ξ' is determined similarly in terms of $\mathbf{p}'$ and α'. The kernel of expansion (32) has the form

$$\begin{aligned} G_{kk'}^l(\nu; \alpha, \alpha') &= (-1)^{k+k'}\left[\frac{4\alpha\alpha'}{(1+\boldsymbol{\alpha}^2\mathbf{p}^2)(1+\boldsymbol{\alpha}'^2\mathbf{p}'^2)}\right]^2 \\ &\times\sqrt{\frac{k(k+l)!}{(k-l-1)!}}\sqrt{\frac{k'(k'+l)!}{(k'-l-1)!}}g_{k-l-1,\,k'-l-1}^l(\nu; \alpha, \alpha'), \end{aligned} \tag{33}$$

where $g_{nn'}^l(\nu; \alpha, \alpha')$ is defined by relations (17) or (18). For $\alpha' = \alpha = \nu$, expression (32) is transformed into (9).

3. MATRIX ELEMENTS OF INELASTIC TWO-PHOTON TRANSITIONS

The matrix elements of two-photon transitions between the states $|nl\rangle$ and $|n'l'\rangle$ of the discrete spectrum with $n' \neq n$ must be calculated while analyzing real two-photon processes (Raman scattering of light by an atom, two-photon excitation, or the decay of atomic states) as well as in the case of many-photon transitions in the presence of two-photon resonances at intermediate levels (ionization and the generation of higher harmonics of laser radiation). For example, two-photon resonance plays a decisive role in experiments on the generation of mixed frequencies in various schemes of 4-wave mixing under the conditions of the self-induced transparency of a gaseous medium (see, e.g., [43] as well as the publication by Zhang *et al.* [44] on the resonant generation of UV radiation according to the scheme

$$E_{1s} + \omega + \omega + \omega' = E_{np}$$

in the transitions $|np\rangle \longrightarrow |1s\rangle$ with $n = 3$ to 8 in atomic hydrogen).

Experience shows that it is most convenient to use in analytical calculations of two-photon matrix elements the operator of electron interaction with the field in the "form of velocity"

$$V = \frac{i}{c}\mathbf{A}\cdot\nabla + \frac{1}{2c^2}\mathbf{A}^2,$$

where $\mathbf{A}$ is the vector potential of the field of the wave with the electric vector

$$\begin{aligned}\mathbf{F}(\mathbf{r}, t) &= F\mathrm{Re}\{\mathbf{e}\exp(-i\omega t + i\mathbf{k}\cdot\mathbf{r})\},\\ |\mathbf{k}| &= \omega/c, \quad \mathbf{e}\cdot\mathbf{k} = 0, \quad \mathbf{e}\cdot\mathbf{e}^* = 1.\end{aligned} \tag{34}$$

In the dipole approximation, the term with $\mathbf{A}^2$ makes zero contribution to the amplitudes of inelastic transitions. After the integration over angular variables, these amplitudes can be expressed in terms of radial matrix elements of the form

$$\mathrm{M}^L_{ll'}(n, n', E) = \langle n'l'|\hat{D}(l', L)g_L(E)\hat{D}(L, l)|nl\rangle, \tag{35}$$

where

$$\hat{D}(l_1, l_2) = \frac{d}{dr} + \frac{\mathrm{sgn}(l_2 - l_1)\max(l_1, l_2) + 1}{r}.$$

Note that in the case of inelastic transitions, matrix elements in the form of "velocity" (35) are connected with the matrix elements in the form of "length" (2) through the simple relation

$$\begin{aligned}&\langle n'l'|\hat{D}(l', L)g_L(E)\hat{D}(L, l)|nl\rangle\\ &= (E_{n'} - E)(E - E_n)\langle n'l'|rg_L(E)r|nl\rangle\end{aligned}$$

(cf. formula (55) below for inelastic transition), which can easily be verified by using the commutation relations between the momentum and coordinate operators.

Using expansion (6) for g_L, we can obtain closed analytic expressions for $\mathrm{M}^L_{ll'}(n, n', E)$ which are valid for any n and n'. For this purpose, we write the result of the action of operators on the wave functions (35) in the form

$$\begin{aligned}\hat{D}(l+1, l)R_{nl}(r) &= \frac{2Z^{5/2}}{n^3[(d)_{2l+1}]^{1/2}}\rho^l e^{-\rho/2}\\ &\times (L^{2l+3}_{d-3}(\rho) - L^{2l+3}_{d-1}(\rho)),\\ \hat{D}(l+1, l+2)R_{nl+2}(r) &= \frac{2Z^{5/2}}{n^3[(d-3)_{2l+5}]^{1/2}}\rho^l e^{-\rho/2}\\ &\times ((s+1)(s+2)L^{2l+3}_{d-3}(\rho) - (d-1)(d-2)L^{2l+3}_{d-1}(\rho)),\end{aligned} \tag{36}$$

where

$$d = n - l, \quad s = n + l, \quad \rho = 2Zr/n,$$

$$R_{nl}(r) = 2Z^{3/2}n^{-2}[(d)_{2l+1}]^{-1/2}\rho^l e^{-\rho/2}L^{2l+1}_{d-1}(\rho)$$

is the radial wave function of the state $|nl\rangle$, and choose the following values of the free parameters in (6):

$$\alpha = n/Z, \quad \alpha' = n'/Z. \tag{37}$$

As a result, the integration with respect to r and r' in (35) can be reduced to direct application of the orthogonality condition for Laguerre polynomials

$$\int_0^\infty e^{-\rho}\rho^\beta L^\beta_n(\rho)L^\beta_m(\rho)d\rho = \frac{\Gamma(\beta + n + 1)}{n!}\delta_{mn}, \tag{38}$$

the sums in expansion (6) become finite, and $\mathrm{M}^L_{ll'}(n, n', E)$ can be expressed in the form of a simple linear combination of the four quantities $g^L_{kk'} = g^L_{kk'}(\nu; n/Z, n'/Z)$:

$$\begin{aligned}\mathrm{M}^{l+1}_{ll}(n, n', E) &= -Z\frac{[(d)_{2l+1}(d')_{2l+1}]^{1/2}}{4nn'}\\ &\times ((d-1)(d-2)(d'-1)(d'-2)g^{l+1}_{d-3, d'-3}\\ &-(s+1)(s+2)(d'-1)(d'-2)g^{l+1}_{d-1, d'-3}\\ &-(d-1)(d-2)(s'+1)(s'+2)g^{l+1}_{d-3, d'-1}\\ &+(s+1)(s+2)(s'+1)(s'+2)g^{l+1}_{d-1, d'-1}),\end{aligned} \tag{39}$$

$$\begin{aligned}\mathrm{M}^{l-1}_{ll}(n, n', E) &= -Z\frac{[(d)_{2l+1}(d')_{2l+1}]^{1/2}}{4nn'}\\ &\times (g^{l-1}_{d-1, d'-1} - g^{l-1}_{d+1, d'-1} - g^{l-1}_{d-1, d'+1} + g^{l-1}_{d+1, d'+1}),\end{aligned} \tag{40}$$

$$\begin{aligned}\mathrm{M}^{l+1}_{ll+2}(n, n', E) &= -Z\frac{[(d)_{2l+1}(d'-2)_{2l+5}]^{1/2}}{4nn'}\\ &\times ((d-1)(d-2)g^{l+1}_{d-3, d'-3} - (s+1)(s+2)g^{l+1}_{d-1, d'-3}\\ &-(d-1)(d-2)g^{l+1}_{d-3, d'-1} + (s+1)(s+2)g^{l-1}_{d-1, d'-1}).\end{aligned} \tag{41}$$

We have introduced the following notation for the difference and sum of n' and l:

$$d' = n' - l, \quad s' = n' + l.$$

One more matrix element allowed by the dipole selection rules can be obtained from (41) as a result of the substitution of indices:

$$\mathrm{M}^{l-1}_{ll-2}(n, n', E) = M^{l+1}_{ll+2}(n', n, E)\big|_{l\to l-2}. \tag{42}$$

The same method can be used for calculating matrix elements of the type

$$\mathrm{M}^{m, m'} = \langle n'l'|r^{m'}q_L(E)r^m|nl\rangle,$$

emerging when the operator of dipole interaction is taken in the form of "length" ($m, m' = 1$) and when the

quadrupole, etc. terms (m, $m' \geq 2$) are taken into account in the interaction with the field. When free parameters are chosen in accordance with (37), the matrix elements $M^{m,m'}$ can be expressed through the integrals of the products of two Laguerre polynomials with identical arguments. The elementary evaluation of these polynomials with the help of recurrence relations again boils down to the application of the orthogonality condition (38). As a result, the explicit expression for $M^{m,m'}$ is a linear combination of the functions $g^l_{kk'}$ as before. Pay attention to the peculiarities of calculations based on the generalized Sturm expansion. If free parameters are chosen in accordance with (37), the entire energy dependence of Green's function (6) (and hence, the frequency dependence of the matrix elements) is contained in the functions $g^l_{kk'}$, which are known beforehand and are defined by formula (18). However, the integration with respect to radial variables makes it possible to determine only the rational coefficients of $g^l_{kk'}$, which are functions of the quantum numbers n, l, n', and l', in formulas of the type (39)–(41).

Formulas (39)–(42) are two-photon analogs of Gordon's formulas (1) since they provide exact expressions for the matrix elements of two-photon dipole transitions in terms of hypergeometric functions. For obvious reasons, two-photon Gordon's formulas are much more complicated than one-photon formulas. First, they cannot be reduced to polynomials any longer, but contain Appel's function F_1 with an integral negative superscript, which is equivalent to a finite linear combination of the hypergeometric functions ${}_2F_1$. Besides, $\mathrm{M}^L_{ll'}$ can be expressed in terms of the products of hypergeometric functions and hypergeometric polynomials. Nevertheless, formulas (39)–(42) give the simplest expressions for two-photon dipole matrix elements. It should be noted that earlier [15], matrix elements of type (35) were expressed in terms of cumbersome hypergeometric-type functions of four parameters and three arguments, which were specially introduced for this problem and which are considerably more complicated than formulas (39)–(42).

General formulas can be simplified significantly for the maximum values of the orbital angular momentum of the initial and/or final states. In this case, the terms containing $g^L_{kk'}$ with negative subscripts are missing in (39)–(42) since the factors of these terms vanish (the fact that the matrix elements cannot be expressed in terms of $g^L_{kk'}$ with $k, k' < 0$ obviously follows even from the initial expansion (6)). The simplest case is that of transition $|ns\rangle \longrightarrow |n's\rangle$ with $n = 1, 2$, and $n' > n$. The matrix elements of such transitions can be written in the form of a universal expression

$$\mathrm{M}^1_{00}(n, n', E) = \frac{3}{2}\left(\frac{n}{n'}\right)^{1/2}((n'-1)(n'-2)g^1_{0,n'-3} - (n'+1)(n'+2)g^1_{0,n'-1}) + \delta_{n,2}\frac{2^{13/2}n'^{5/2}\nu^2}{3(\nu^2-4)(n'^2-4)^2}\left(\frac{n'-2}{n'+2}\right)^{n'}, \quad (43)$$

where $n = 1$ or 2. The functions $g^L_{kk'}$ with zero subscript k or k' can be written in a compact form:

$$g^1_{0k} = \frac{\nu}{6(2-\eta)}\left(\frac{16nn'\eta^2}{(n+\eta)^2(n'+\eta)^2}\right)^2\left(\frac{n'-\eta}{n'+\eta}\right)^k \times F_1(2-\eta; -k, k+4; 3-\eta; y, y'),$$

$$y = \frac{n-\eta}{n+\eta}\frac{n'+\eta}{n'-\eta}, \quad y' = \frac{n-\eta}{n+\eta}\frac{n'-\eta}{n'+\eta}.$$

For $n = 1$, the last term in (43) vanishes, and the result can be expressed in terms of two Appel's functions F_1. The expression for the matrix element of the transition $|1s\rangle \longrightarrow |nd\rangle$ analogous to (43) has the form

$$\mathrm{M}^1_{02}(1, n', E) = \frac{3}{2}(n')^{-1/2}(n'^2-1)^{1/2} \times (n'^2-4)^{1/2}(g^1_{0,n'-3} - g^1_{0,n'-1}).$$

The above expressions for the transitions $|1s\rangle \longrightarrow |ns\rangle$ and $|1s\rangle \longrightarrow |nd\rangle$ coincide with the results obtained by Marian [15] (formulas (34) and (35)).

The cross section of any two-photon transition between bound states can be expressed in terms of the matrix elements (35). For example, the cross section for Raman scattering involving the transition $|nl\rangle \longrightarrow |n'l+2\rangle$ contains only $\mathrm{M}^{l+1}_{ll+2}(n, n', E)$:

$$\frac{d\sigma_{nl,n'l+2}}{d\Omega_{\mathbf{k}'}} = \alpha^4\frac{\omega'}{\omega}\frac{(l+1)(l+2)}{10(2l+1)(2l+3)} \times \left|\mathrm{M}^{l+1}_{ll+2}(n, n', E_n+\omega) + \mathrm{M}^{l+1}_{ll+2}(n, n', E_n-\omega')\right|^2 \times \left(1 + |\mathbf{e}'\cdot\mathbf{e}|^2 - \frac{2}{3}|\mathbf{e}'\cdot\mathbf{e}^*|^2\right), \quad (44)$$

where α is the fine-structure constant, and $\mathbf{e}$ and ω ($\mathbf{e}'$ and ω') are the polarization vector and frequency of an incident (scattered) photon.

Table 1 contains numerical values of the matrix elements $\mathrm{M}^{l\pm1}_{ll\pm2}(n, n', E)$ and $\mathrm{M}^{l\pm1}_{ll}(n, n', E)$ for $n' > n$ and $E = (E_n + E_{n'})/2$. It can be seen from the table that the value of the relevant matrix element increases with the orbital angular momentum of the final state, which is similar to the well-known Bethe rule for one-photon matrix elements.

Table 1. Values of matrix elements $M_{ll\pm2}^{l\pm1}(n, n', E)$ and $M_{ll}^{l\pm1}(n, n', E)$ where $E = (E_n + E_{n'})/2$

n	l	n'	$M_{ll-2}^{l-1}(n, n')$	$M_{ll}^{l-1}(n, n')$	$M_{ll}^{l+1}(n, n')$	$M_{ll+2}^{l+1}(n, n')$
3	2	10	3.33409(–1)	7.26967(–1)	9.11768(–1)	5.65901(0)
3	2	15	5.64955(–2)	8.70688(–2)	2.20482(–1)	8.40251(–1)
3	2	20	2.66131(–2)	3.21271(–2)	1.20379(–1)	3.73535(–1)
10	9	15	–2.30971(–2)	–4.68375(–1)	5.97962(–2)	–4.36418(0)
10	9	20	–3.53832(–3)	–1.14158(–1)	5.93551(–2)	–5.75417(–1)
10	9	25	6.64088(–3)	2.07508(–2)	1.06041(–1)	8.80864(–1)

4. TWO-PHOTON GORDON FORMULAS FOR BOUND–FREE TRANSITIONS

A generalization of the results obtained in the previous section to the case when one of the states belongs to the continuous spectrum with energy E makes it possible to obtain closed analytic expressions for the amplitudes of Compton scattering of a rather hard photon of frequency ω, involving the ionization of the atom and the emission of a photon of frequency $\omega' = E_n + \omega - E$, as well as two-photon ionization of a state with arbitrary quantum numbers n and l. The above method of calculation of amplitudes of transitions between states of the discrete spectrum, which is actually based of the orthogonality condition for Laguerre polynomials, cannot be extended to the present case since there is no relation similar to (38) for functions of the continuous spectrum. An analysis shows, however, that in analogy with the generalization of one-photon Gordon's formulas to the case of photoionization or photorecombination, expressions (39)–(41) can be continued analytically in one of the variables n or n' to the region of continuous energy spectrum. For example, in order to obtain

$$M_{ll'}^{L}(n, E, \mathscr{E}) = \langle R_{El'}|\hat{D}(l', L)g_L(\mathscr{E})\hat{D}(L, l)|R_{nl}\rangle \tag{45}$$

from $\langle n'l'|\hat{D}(l', L)g_L(\mathscr{E})\hat{D}(L, l)|nl\rangle$, we must carry out the substitution n' $n' \longrightarrow iZ/p$, where $p = \sqrt{2E}$ and introduce an additional factor taking into account the difference in the normalization of the wave functions of the discrete and continuous spectra.

For the sake of reference, we will write here all four radial matrix elements emerging in the calculation of the amplitude of bound–free transition from the state $|nl\rangle$:

$$M_{ll}^{l+1}(n, E, \mathscr{E}) = -i^{l+1}aC_{El}\frac{[(d)_{2l+1}]^{1/2}}{8nZ^{1/2}}$$
$$\times[(d-1)(d-2)(l+1-ia)(l+2-ia)g_{d-3,\,ia-l-3}^{l+1}$$
$$-(s+1)(s+2)(l+1-ia)(l+2-ia)g_{d-1,\,ia-l-3}^{l+1} \tag{46}$$
$$-(d-1)(d-2)(l+1+ia)(l+2+ia)g_{d-3,\,ia-l-1}^{l+1}$$
$$+(s+1)(s+2)(l+1+ia)(l+2+ia)g_{d-1,\,ia-l-1}^{l+1}],$$

$$M_{ll}^{l-1}(n, E, \mathscr{E}) = -i^{l+1}aC_{El}\frac{[(d)_{2l+1}]^{1/2}}{8nZ^{1/2}}$$
$$\times[g_{d-1,\,ia-l-1}^{l-1} - g_{d+1,\,ia-l-1}^{l-1}$$
$$-g_{d-1,\,ia-l+1}^{l-1} + g_{d+1,\,ia-l+1}^{l-1}], \tag{47}$$

$$M_{ll+2}^{l+1}(n, E, \mathscr{E}) = -i^{l+3}aC_{El+2}\frac{[(d)_{2l+1}]^{1/2}}{8nZ^{1/2}}$$
$$\times[(d-1)(d-2)g_{d-3,\,ia-l-3}^{l+1}$$
$$-(s+1)(s+2)g_{d-1,\,ia-l-3}^{l+1} \tag{48}$$
$$-(d-1)(d-2)g_{d-3,\,ia-l-1}^{l+1}$$
$$+(s+1)(s+2)g_{d-1,\,ia-l-1}^{l+1}],$$

$$M_{ll-2}^{l-1}(n, E, \mathscr{E}) = -i^{l-1}aC_{El-2}\frac{[(d)_{2l+1}]^{1/2}}{8nZ^{1/2}}$$
$$\times[(l-1-ia)(l-ia)g_{d-1,\,ia-l-1}^{l-1}$$
$$-(l-1-ia)(l-ia)g_{d+1,\,ia-l-1}^{l-1} \tag{49}$$
$$-(l-1+ia)(l+ia)g_{d-1,\,ia-l+1}^{l-1}$$
$$+(l-1+ia)(l+ia)g_{d+1,\,ia-l+1}^{l-1}],$$

where

$$a = Z/p, \quad C_{El} = \sqrt{\frac{2p}{\pi}}e^{\pi a/2}|\Gamma(l+1-ia)|.$$

It can be seen that the matrix elements are expressed in terms of the kernel $g_{kk'}^{l}$ of expansion (6), which must be taken for the complex value of the second subscript, $k' = \beta$, and for the following arguments (see (18)):

$$g_{k\beta}^{l} = g_{k\beta}^{l}(\nu;\, n/Z,\, i/p).$$

In this case, the first term in (18) contains the complete Appel function F_1 which, in contrast to the case of bound–bound transitions, cannot be reduced to the combination of functions ${}_2F_1$ any longer. It is important to note that the functions $\Phi_p^{l, n' = i/p}$ in formula (18) for $g_{k\beta}^l(\nu; n/Z, i/p)$ are expressed, as before, through hypergeometric polynomials. Thus, for arbitrary n and l, each of the four partial amplitudes $M_{ll'}^L(n, E, \mathscr{E})$ is in turn a combination of four terms each of which is the sum of not more than $n_r + 2$ ($n_r = n - l - 1$) nonresonant polynomial terms (which are smooth functions of frequency) and a resonant term which is the product of a hypergeometric polynomial of the order n_r or $n_r + 2$ and the complete Appel function F_1 (the first term in (18)). The numerical calculation of these quantities obviously does not involve any technical difficulties either for $|E_n|/2 < \omega < |E_n|$ or in the above-threshold case $\omega > |E_n|$.

Strictly speaking, the analytic continuation of amplitudes from the discrete spectral region to the continuous energy range could be incorrect, but in our case no ambiguity emerges when this procedure is carried out, and the obtained expressions are in accord with the known analytical and numerical results for some special cases. In particular, two-photon Gordon's formulas (46)–(49) completely solve the problem of the two-photon ionization of the hydrogen atom from an arbitrary state $|nl\rangle$ since the total cross section, as well as the angular distribution of photoelectrons, can be expressed in terms of $M_{ll'}^L(n, E, \mathscr{E})$. For example, the two- photon ionization cross section integrated with respect to the angles of emission of a photoelectron in field (34) with a linear polarization has the form

$$\begin{aligned}\sigma_{nl} = \frac{\pi^2\alpha F^2\omega}{15(2l+1)}\Bigg(&\frac{(l+1)(4l^2+8l+5)}{(2l+1)(2l+3)}\left|M_{ll}^{l+1}\right|^2 \\ &+ \frac{l(4l^2+1)}{(2l-1)(2l+1)}\left|M_{ll}^{l-1}\right|^2 \\ &+ \frac{4l(l+1)}{2l+1}\mathrm{Re}\left((M_{ll}^{l+1})^* M_{ll}^{l-1}\right) \\ &+ \frac{2(l+1)(l+2)}{2l+3}\left|M_{ll+2}^{l+1}\right|^2 + \frac{2l(l-1)}{2l-1}\left|M_{ll-2}^{l-1}\right|^2\Bigg),\end{aligned} \tag{50}$$

where

$$M_{ll'}^L = M_{ll'}^L(n, E, \mathscr{E}), \quad E = E_n + 2\omega, \quad \mathscr{E} = E_n + \omega.$$

Similar expressions in the case of the elliptic polarization of the wave for the parameters of the angular distribution of photoelectrons are given in [45].

As in the case of bound–bound transitions, the results are simplified significantly for $l \sim n$ Indeed, if the initial state has the maximum value of the orbital angular momentum, $l = n - 1$ ($n_r = 0$), the matrix elements (46), (48), for example, assume the form

$$\begin{aligned}M_{n-1n-1}^n(n, E, \mathscr{E}) &= i(-1)^n[(2n-1)!]^{1/2}\frac{aC_{En-1}}{4Z^{1/2}} \\ &\times((n-ia)(n+1-ia)F(0) \\ &-(n+ia)(n+1+ia)F(2)),\end{aligned} \tag{51}$$

$$\begin{aligned}M_{n-1n+1}^n(n, E, \mathscr{E}) &= i(-1)^{n+1}[(2n-1)!]^{1/2} \\ &\times\frac{aC_{En+1}}{4Z^{1/2}}(F(0)-F(2)),\end{aligned}$$

and can be expressed only in terms of function F_1:

$$\begin{aligned}F(m) = \frac{\nu}{n+1-\eta}&\frac{(4\eta)^{2n+2}}{(2n)!}\frac{(an)^{n+1}}{(n+\eta)^{2n+2}} \\ &\times\frac{(ia-\eta)^{ia-n-2+m}}{(ia+\eta)^{ia+n+m}}F_1(n+1-\eta; \\ &-ia+n+2-m, ia+n+m; n+2-\eta; y, y'),\end{aligned}$$

$$y = \frac{n-\eta}{n+\eta}\frac{ia+\eta}{ia-\eta}, \quad y' = \frac{n-\eta}{n+\eta}\frac{ia-\eta}{ia+\eta}.$$

For the ground state $n = 1$, the other two matrix elements (47) and (49) do not appear in the transition amplitude, and the results described above completely coincide with those obtained in [18] (see also [19]).

Relatively simple expressions (46)–(49) for matrix elements make it possible to calculate the two-photon ionization cross sections for excited states in a wide frequency range. Figure 1 shows the frequency dependence of cross sections σ_{nl} (50) beyond the one-photon ionization threshold for states with $n = 10$, $l = 0, 5, 9$. It can be seen that the ionization cross section decreases rapidly upon an increase in ω. The same figures illustrate the accuracy of the "resonant" approximation, which implies the inclusion of only the imaginary component of matrix elements in an analysis of the two-photon ionization cross section:

$$M_{ll'}^L(n, E, \mathscr{E}) \longrightarrow \mathrm{Im}M_{ll'}^L(n, E, \mathscr{E}).$$

Our results indicate that the contribution of the imaginary component decreases upon an increase in the orbital angular momentum and the energy of the final state.

5. POLARIZABILITIES OF EXCITED STATES

5.1. Exact Analytical Results for Polarizability Tensor Components

It is convenient to analyze two-photon transitions between (degenerate) states with the same energy in the language of atomic polarizabilities. Polarizability describes the linear response of an atom to an external

monochromatic action (induced dipole moment) and determines the amplitude of elastic (Rayleigh) scattering of light as well as the shift, splitting, and ionization-induced broadening of atomic energy levels in the field of a light wave. The polarizabilities of highly excited hydrogen-like states are of special interest in certain cases, e.g., in astrophysical applications and in the spectroscopy of Rydberg levels. The polarizabilities of the hydrogen atom can be calculated analytically, and such computations were made by dozens of authors. In this section, the generalized Sturm expansion of the CGF will be used for deriving analytic expressions for the polarizabilities of arbitrary hydrogen states, which are essentially as simple as formula (5) for the polarizability of the ground state.

A general analysis of the second-order perturbation in F of the spectrum of an atom in the field of an electromagnetic wave (34) was carried out in [12]. For this reason, we will confine our analysis to the case of linear or circular polarization of the wave, for which the states $|nlm\rangle$ with different values of components of the orbital angular momentum m are not mixed by the field. A specific feature of the Coulomb problem is the degeneracy of energy levels in l, the two-photon matrix element between the states $|nlm\rangle$ and $|nl'm\rangle$ with $l' = l \pm 2$ differing from zero. For this reason, even in the case of a linear or circular polarization of the wave, the perturbation of the spectrum of excited states is determined by a secular equation with a tridiagonal matrix of rank $n - |m|$ [46]:

$$\left|Q_{mm}^{ll'} - \Delta E_n \delta_{ll'}\right| = 0,$$

where

$$Q_{mm}^{ll'} = -\frac{1}{4}F^2\langle nl'm|\{\mathbf{e}^* \cdot \mathbf{r}G_{E_n+\omega+i0}\mathbf{e}\cdot\mathbf{r} + \mathbf{e}\cdot\mathbf{r}G_{E_n-\omega+i0}\mathbf{e}^*\cdot\mathbf{r}\}|nlm\rangle. \quad (52)$$

The dependence of the Stark matrix elements $Q_{mm}^{ll'}$ on the magnetic quantum number m can easily be determined by the methods of quantum theory of angular momentum [47]. Matrix elements diagonal in l have the form

$$Q_{mm}^{ll} = -\frac{1}{4} \times F^2\left[\alpha^s(\omega) + \xi\alpha^a(\omega)\frac{m}{2l} - \alpha^t(\omega)\frac{3m^2 - l(l+1)}{2l(2l-1)}\right] \quad (53)$$

for circular polarization ($\xi = \pm 1$ for right- and left circular polarizations) and

$$Q_{mm}^{ll} = -\frac{1}{4}F^2\left[\alpha^s(\omega) + \alpha^t(\omega)\frac{3m^2 - l(l+1)}{l(2l-1)}\right] \quad (54)$$

for linear polarization. Thus, the elements Q_{mm}^{ll} are determined by the three invariant parameters $\alpha^s(\omega)$, $\alpha^a(\omega)$, and $\alpha^t(\omega)$ known as the scalar, vector (antisymmetric) and tensor (symmetric) polarizabilities. The dependence on m in nondiagonal matrix elements can be singled out in a similar way. In the case of the circular polarization of the wave, the element $Q_{mm}^{ll'=l+2}$ has the form

$$Q_{mm}^{ll'=l+2} = \frac{1}{8}F^2\beta_{n,l+2}(\omega) \times \sqrt{\frac{((l+1)^2 - m^2)((l+2)^2 - m^2)}{(2l+1)(2l+5)}},$$

while the expression for $Q_{mm}^{l,l'=l-2}$ can be obtained from this formula through the substitution $l \longrightarrow (l-2)$. For a linear polarization of $\mathbf{F}(t)$, we have

$$Q^{ll'(l)} = -2Q^{ll'(c)}.$$

It can be seen that the nondiagonal elements $Q^{l,l\pm2}$ are determined by a single invariant atomic parameter

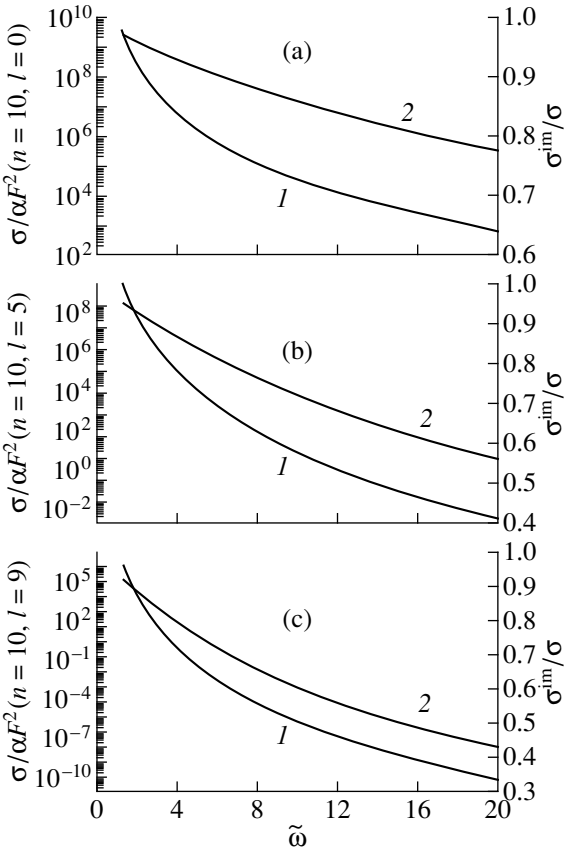

Fig. 1. Dependence of the ionization cross section (on logarithmic scale) $\sigma_{nl}/\alpha F^2$ (curve *1*) and of the ratio $\sigma_{nl}^{im}/\sigma_{nl}$ of the cross section calculated taking into account only the imaginary component of matrix elements to the exact expressions (curve *2*) on the frequency $\tilde{\omega} = 2n^2\omega$: (a) $n = 10$, $l = 0$; (b) $n = 10$, $l = 5$, and (c) $n = 10$, $l = 9$.

$\beta_{nl'=l\pm2}(\omega)$, which can be referred to as the "nondiagonal" polarizability.

Explicit expressions for the polarizabilities $\alpha^{s,a,t}$ and β can be written in terms of two-photon radial matrix elements in r (2) or in $\hat{D}$ (35), which are connected through the following relation in the case of elastic transitions:

$$\langle nl'|rg_L(E_n+\omega)r|nl\rangle = -\frac{1}{\omega}\langle nl'|r^2|nl\rangle - \frac{1}{4\omega^2}\langle nl'|nl\rangle \times[2+2L(L+1)-l(l+1)-l'(l'+1)] - \frac{1}{\omega^2}\langle nl'|\hat{D}(l',L)g_L(E_n+\omega)\hat{D}(L,l)|nl\rangle. \quad (55)$$

We will use below the following expressions for polarizabilities in terms of the matrix elements in the form of "velocity," i.e., $\mathrm{M}^L_{ll'}(n,n,E)\equiv\mathrm{M}^L_{ll'}(E)$:

$$\alpha^s_{nl}(\omega) = -\frac{1}{\omega^2}-\frac{1}{\omega^2}\frac{1}{3(2l+1)}[l\rho^{(+)}_{l-1}+(l+1)\rho^{(+)}_{l+1}],$$

$$\alpha^a_{nl}(\omega) = -\frac{1}{\omega^2}\frac{l}{2l+1}[\rho^{(-)}_{l-1}-\rho^{(-)}_{l+1}],$$

$$\alpha^t_{nl}(\omega) = \frac{1}{\omega^2}\frac{l}{3(2l+1)}\left[\rho^{(+)}_{l-1}+\frac{2l-1}{2l+3}\rho^{(+)}_{l+1}\right], \quad (56)$$

$$\beta_{nl'}(\omega) = -\frac{1}{\omega^2}\frac{1}{2l'-1}[\mathrm{M}^L_{ll'}(E_n+\omega)+\mathrm{M}^L_{ll'}(E_n-\omega)],$$

$$L = l\pm1, \quad l' = L\pm1,$$

where

$$\rho^{(\pm)}_L = M^L_{ll}(E_n+\omega)\pm M^L_{ll}(E_n-\omega). \quad (57)$$

Instead of (36), it is more convenient to use the following auxiliary relations for the action of operators $\hat{D}$ on the wave functions for calculating polarizabilities:

$$\hat{D}(l+1,l)R_{nl}(r) = -\frac{Z}{n}\left(\frac{d-1}{s+1}\right)^{1/2}R_{nl+1}(r) - \frac{4Z^{5/2}}{n^2(s+1)[(d)_{2l+1}]^{1/2}}\rho^l e^{-\rho/2}(L^{2l+3}_{d-1}(\rho)-L^{2l+3}_{d-2}(\rho)),$$

$$\hat{D}(l+1,l+2)R_{nl+2}(r) = \frac{Z}{n}\left(\frac{d-2}{s+2}\right)^{1/2}R_{nl+1}(r) + \frac{4Z^{5/2}}{n^2[(d-3)_{2l+5}]^{1/2}}\rho^l e^{-\rho/2} \times((s+1)L^{2l+3}_{d-3}(\rho)-(d-2)L^{2l+3}_{d-2}(\rho)). \quad (58)$$

Since the first terms in these formulas differ from the wave function of the state $|n, l+1\rangle$ only in the factor, the corresponding terms in $\mathrm{M}^L_{ll'}$ can easily be evaluated using the relation

$$g_l(E)|R_{nl}\rangle = \frac{1}{E_n-E}|R_{nl}\rangle. \quad (59)$$

In the case under investigation, the energies of the initial and final states in $\mathrm{M}^L_{ll'}$ are identical, and hence we can use representation (6) with identical parameters $\alpha' = \alpha = n/Z$ for Green's function, and the subsequent integration with respect to r and r' can be reduced to the trivial application of the orthogonality condition (38). As a result, we obtain, for example, the following expression for M^{l+1}_{ll}:

$$\mathrm{M}^{l+1}_{ll}(E) = \frac{Z^2}{n^2(E_n-E)}\frac{d-1}{s+1}(d)_{2l+1}((s+2)^2 g^{l+1}_{d-1,d-1} - 2(d-1)(s+2)g^{l+1}_{d-1,d-2}+(d-1)^2 g^{l+1}_{d-2,d-2}). \quad (60)$$

Using expression (24) for $g^l_{kk'}$ as well as the relation (see formula (A.3) in Appendix A)

$$q^{+l}_k p^l_{k-1} - q^{+l}_{k-1}p^l_k = 2(k+1)_{2l},$$

we can write expression (60) in the following form:

$$\mathrm{M}^{l+1}_{ll}(E_n+\omega) = \frac{2}{\tilde{\omega}} - \frac{2n}{\tilde{\omega}(s+1)^2(d)_{2l+1}} \times(p^{l+1}_{d-1}-p^{l+1}_{d-2})(q^{+(l+1)}_{d-1}-q^{+(l+1)}_{d-2}). \quad (61)$$

Similar expressions for M^{l-1}_{ll} and M^{l+1}_{ll+2} have the form

$$\mathrm{M}^{l-1}_{ll}(E_n+\omega) = \frac{2}{\tilde{\omega}} - \frac{2n}{\tilde{\omega}}\frac{[(s-1)p^{l-1}_{d-1}-dp^{l-1}_d][(s-1)q^{+(l-1)}_{d-1}-dq^{+(l-1)}_d]}{(d)_{2l+1}}, \quad (62)$$

$$\mathrm{M}^{l+1}_{ll+2}(E_n+\omega) = \frac{2}{\tilde{\omega}}\left(\frac{d-1}{s+1}\frac{d-2}{s+2}\right)^{1/2}+\frac{2n}{\tilde{\omega}} \times\frac{[(s+1)p^{l+1}_{d-3}-(d-2)p^{l+1}_{d-2}][q^{+(l+1)}_{d-1}-q^{+(l+1)}_{d-2}]}{(s+1)[(d)_{2l+1}(d-2)_{2l+5}]^{1/2}}. \quad (63)$$

The expression for M^{l-1}_{ll-2} can be obtained from (63) through substitution (42). Note that the elementary terms in $\mathrm{M}^L_{ll'}$, which are proportional to $1/\tilde{\omega}$, vanish in the course of evaluating the sum in $\rho^{(+)}$ (57) and the difference in α^a and make no contribution to the polarizabilities.

Relations (56) and (61)–(63) completely determine the invariant components of the dynamic polarizability tensor of the *n*-shell. It can be seen that the general form of the dependences $\alpha^{s,a,t}(\omega)$ and $\beta(\omega)$ on the parameters of the problem (field frequency ω, quantum

numbers n and l, and the nuclear charge Z) can be presented as

$$\alpha_{nl}(\omega) = \frac{1}{\omega^2} f(n, l, \tilde{\omega}). \tag{64}$$

Thus, the "true" frequency ω appears in the polarizability in fact only as the common factor in Eqs. (56), and the details of the frequency dependence are determined by the quantity $\tilde{\omega}$. Using relation (28) for Pollaczek functions q_k^{+L} in (61)–(63) and taking into account the fact that p_k^l and $\bar{q}_k^l$ are polynomials in $1/\tilde{\omega}$ (see Subsection 2.2), we can readily see that the scalar, vector, and tensor polarizabilities for an arbitrary n have the following universal structure (cf. Eqs. (5)):

$$\begin{aligned}
\alpha_{nl}^s &= \frac{1}{\omega^2}\left[\frac{l}{2l+1}\mathcal{T}^+_{n,l-1}(\tilde{\omega})\right. \\
&\left. + \frac{l+1}{2l+1}\mathcal{T}^+_{n,l+1}(\tilde{\omega}) + \psi^s_{nl}\left(\frac{1}{\tilde{\omega}^2}\right) - 1\right], \\
\alpha_{nl}^a &= \frac{1}{\omega^2}\left[\frac{3l}{2l+1}\right. \\
&\left. \times (\mathcal{T}^-_{n,l-1}(\tilde{\omega}) - \mathcal{T}^-_{n,l+1}(\tilde{\omega})) + \psi^a_{nl}\left(\frac{1}{\tilde{\omega}}\right)\right], \\
\alpha_{nl}^t &= -\frac{1}{\omega^2}\left[\frac{l}{2l+1}\mathcal{T}^+_{n,l-1}(\tilde{\omega})\right. \\
&\left. + \frac{l(2l-1)}{(2l+1)(2l+3)}\mathcal{T}^+_{n,l+1}(\tilde{\omega}) + \psi^t_{nl}\left(\frac{1}{\tilde{\omega}^2}\right)\right],
\end{aligned} \tag{65}$$

where

$$\mathcal{T}^\pm_{n,L}(\tilde{\omega}) = \varphi^L_{nl}(\tilde{\omega})T_L(\tilde{\omega}) \pm \varphi^L_{nl}(-\tilde{\omega})T_L(-\tilde{\omega}),$$

$$T_L(\tilde{\omega}) = \frac{\eta}{L+1-\eta}{}_2F_1(1, 2L+2; L+2-\eta; z^{-1}),$$

$$\varphi^{L=l+1}_{nl} \equiv \varphi^{(+)}_{nl} = \frac{(2l+3)!}{3(s+1)^2(d)_{2l+1}}[p^{l+1}_{d-1} - p^{l+1}_{d-2}]^2,$$

$$\varphi^{L=l-1}_{nl} \equiv \varphi^{(-)}_{nl} = \frac{(2l-1)!}{3(d)_{2l+1}}[(s-1)p^{l-1}_{d-1} - d\,p^{l-1}_{d}]^2,$$

and $\psi^{s,a,t}_{nl}$ can be expressed in terms of p^L_k and $\bar{q}^L_k$ (see Appendix B).

It follows from the results of Subsection 2.2 that the functions $\varphi^{(\pm)}_{nl}$ and $\psi^{s,a,t}_{nl}$ are real-valued polynomials in $\tilde{\omega}^{-1}$, $\psi^{s,t}$ containing only even powers of frequency and ψ^a, only odd powers. The degree of these polynomials is determined by the radial quantum number n_r, while the coefficients are determined by the values of n_r and of the orbital quantum number l. For the ground state (see (5)), we have $\varphi^{(+)}_{10} = 1/2$ and $\psi^s_{10} = 0$. Explicit expressions for $\varphi^{(\pm)}_{nl}$ and $\psi^{s,a,t}_{nl}$ for $n \le 4$ are given in Appendix B.

Formulas (65) show that the polarizabilities of states with arbitrary n and l can be expressed in terms of the universal function

$${}_2F_1(1, 2m+2; m+2-\eta; z^{-1}) \tag{66}$$

with $m = l \pm 1$, in which the principal quantum number appears only through the parameter

$$\eta = \frac{Z}{\sqrt{-2(E_n \pm \omega) - i0}} = \frac{n}{\sqrt{1 \mp \tilde{\omega} - i0}}.$$

The possibility of such a representation of polarizabilities was noted earlier [12, 29], but explicit formulas for arbitrary values of n were not derived. Function (66) possesses a number of peculiar properties that can be used in computational algorithms. It should be noted here that function (66) can be expressed in terms of an incomplete beta function (the integral on the right-hand side of (A.4) is the definition of the incomplete beta function [5]). Moreover, relation (A.8) makes it possible to separate the imaginary component of $T_L(\tilde{\omega})$, which appears for $\tilde{\omega} > 1$ and can be reduced to elementary functions,

$$\begin{aligned}
\mathrm{Im}\,T_L(\tilde{\omega}) &= |\eta|\frac{|\Gamma(L+1-\eta)|^2}{2(2l+1)!}\left(\frac{4n|\eta|}{n^2+|\eta|^2}\right)^{2l+2} \\
&\times \exp(\pi|\eta| - 4|\eta|\,\mathrm{arccot}(|\eta|/n)),
\end{aligned} \tag{67}$$

and hence the imaginary component of polarizabilities also since the remaining quantities appearing in (65) are real-valued. Since the photoionization cross section $\sigma_{nl}(\omega)$ of the state $|nl\rangle$ is connected with the imaginary component of the scalar polarizability in accordance with the optical theorem, i.e.,

$$\sigma_{nl}(\omega) = \frac{2\pi\omega}{c(2l+1)}\mathrm{Im}\,\alpha^s_{nl}(\omega), \tag{68}$$

relation (67) makes it possible to express $\sigma_{nl}(\omega)$ in terms of elementary functions:

$$\begin{aligned}
\sigma_{nl}(\omega) &= \frac{4\pi\omega}{c(2l+1)^2} \\
&\times \mathrm{Im}(l\varphi^{(-)}_{nl}\mathrm{Im}\,T_{l-1} + (l+1)\varphi^{(+)}_{nl}\mathrm{Im}\,T_{l+1}).
\end{aligned}$$

In particular, for ns states we have

$$\begin{aligned}
\sigma_{ns}(\omega) &= \frac{a_0^2e^2}{\hbar c}\frac{2^{10}\pi^2n^4}{3Z^2\tilde{\omega}^5}(\tilde{\omega} + n^2 - 1) \\
&\times \frac{\exp(-4|\eta|\,\mathrm{arccot}(|\eta|/n))}{1 - \exp(-2\pi|\eta|)}\varphi^{(+)}_{n0}(\tilde{\omega}),
\end{aligned} \tag{69}$$

(in formulas (68) and (69), ordinary units are used, and a_0 is Bohr's radius). Thus, it is only the polynomials

$\varphi_{nl}^{(\pm)}$ that are connected with the ionization cross section. It is interesting to note that in view of the analytic energy dependence of the scattering amplitude, the polynomials $\varphi_{nl}^{(\pm)}(+\tilde{\omega})$ through which $\sigma_{nl}(\omega)$ is expressed for $\tilde{\omega} > 1$ determine $\mathcal{T}^{\pm}_{n,L}(+\tilde{\omega})$ for $\tilde{\omega} < 1$ also, while the polynomials $\varphi_{nl}^{(\pm)}(-\tilde{\omega})$ appear in the matrix elements with Green's function $g_{l\pm1}(E_n - \omega)$, which remain real-valued for any $\tilde{\omega}$.

5.2. Polarizabilities of Rydberg States

The asymptotic properties of the polarizabilities of states with arbitrary n and l were analyzed by many authors. The high-frequency asymptotic form, for example, was studied in [12, 48] (see also [49]). The low-frequency limit is also considered in [12]. Semiclassical calculations of the polarizabilities of the n shell averaged over l were calculated in [50], while some peculiarities of the polarizabilities of Rydberg s states were analyzed in [11]. In this subsection, we determine the Rydberg asymptotic form of polarizabilities (for $n \gg 1$), which holds for any frequency. In this case, the states with small and large values of the orbital angular momentum should be distinguished.

5.2.1. Small values of the orbital angular momentum: $l \ll n$. In this case, the asymptotic form of $\alpha_{nl}^{s,a,t}(\omega)$ and $\beta_{nl'}$ can be obtained by proceeding to the limit of $n \gg 1$ in formulas (61)–(63). The expression for the hypergeometric polynomial appearing in (25) for large n can be derived by the term-by-term transition to the limit

$$ {}_2F_1(-n+k, b; c; z) \longrightarrow \Phi(b, c; 4\eta), \tag{70}$$

where Φ is a degenerate hypergeometric function. In order to find the limiting expression for a hypergeometric function of the type ${}_2F_1(n+a, n+b; n+c; z^{-1})$ (see (26)), in which all the three parameters and the argument are large for $n \gg 1$, we present it in the integral form:

$$\begin{aligned} {}_2F_1(n+a, n+b; n+c; z^{-1}) &= \frac{\Gamma(n+c)}{\Gamma(n+b)\Gamma(c-b)} \\ \times\left(\frac{z}{z-1}\right)^{n+a}&\int_0^\infty t^{c-b-1}(1+t)^{a-c}\left(1+\frac{zt}{z-1}\right)^{-n-a}dt. \end{aligned} \tag{71}$$

Considering that $z \propto n^{-1}$ and proceeding to the limit in the integrand, we obtain

$$\begin{aligned} {}_2F_1(n+a, n+b; n+c; z^{-1}) &\longrightarrow \frac{\Gamma(n+c)}{\Gamma(n+b)} \\ \times\left(\frac{z}{z-1}\right)^{n+a}&\Psi(c-b, a-b+1; 4\eta), \end{aligned} \tag{72}$$

where $\Psi(a, c; x)$ is a degenerate hypergeometric function which is irregular at zero.

The substitution of the obtained limiting expressions (70), (72) into formulas (61)–(63) leads to the cancellation of terms principal in $1/n$ (the possibility of such cancellations follows even from formulas (36)). In order to avoid the calculation of the next expansion terms in $1/n$ in formulas (70) and (72), we transform expressions (61)–(63) so that the principal terms of the asymptotic form cancel out in these initial expressions. Using the Gauss relation (see formula (2.8.32) in [5]) for transforming the differences of Pollaczek polynomials and the relation

$$\begin{aligned} &(n+l+1-\eta){}_2F_1(n, n+2l+1; n+l+1-\eta; z^{-1}) \\ &= (l+1-\eta){}_2F_1(n, n+2l+2; n+l+2-\eta; z^{-1}) \\ &+ n(1-z^{-1}){}_2F_1(n+1, n+2l+2; n+l+2-\eta; z^{-1}) \end{aligned}$$

for transforming the differences of Pollaczek functions, we arrive at the following asymptotic expressions for matrix elements:

$$\begin{aligned} \tilde{M}_{ll}^{l+1}(E_n+\omega) &= \frac{Z^2}{\omega n^2} + \frac{2Z^2}{\omega n^3}e^{-4\tilde{\eta}}\frac{\Gamma(l+2-\tilde{\eta})}{\Gamma(2l+4)} \\ &\times[(l+1-\tilde{\eta})\Phi(l+2-\tilde{\eta}, 2l+4; 4\tilde{\eta}) \\ &+ (l+2-\tilde{\eta})\Phi(l+3-\tilde{\eta}, 2l+4; 4\tilde{\eta})] \\ &\times[(2l+3-2\tilde{\eta})\Psi(-l-1-\tilde{\eta}, -2l-2; 4\tilde{\eta}) \\ &- (l+2-\tilde{\eta})\Psi(-l-1-\tilde{\eta}, -2l-3; 4\tilde{\eta})], \end{aligned} \tag{73}$$

$$\begin{aligned} \tilde{M}_{ll}^{l-1}(E_n+\omega) &= \frac{Z^2}{\omega n^2} + \frac{2Z^2}{\omega n^3}e^{-4\tilde{\eta}}\frac{\Gamma(l-\tilde{\eta})}{\Gamma(2l)} \\ &\times[(l+\tilde{\eta})\Phi(l-\eta, 2l; 4\tilde{\eta}) \\ &- (l-\tilde{\eta})\Phi(l+1-\tilde{\eta}, 2l; 4\tilde{\eta})] \\ &\times[2\tilde{\eta}\Psi(-l+1-\tilde{\eta}, -2l+2; 4\tilde{\eta}) \\ &+ (l-\tilde{\eta})\Psi(-l+1-\tilde{\eta}, -2l+1; 4\tilde{\eta})], \end{aligned} \tag{74}$$

$$\begin{aligned} \tilde{M}_{ll+2}^{l+1}(E_n+\omega) &= \frac{Z^2}{\omega n^2}\left(\frac{d-1}{s+1}\frac{d-1}{s+2}\right)^{1/2} \\ &+ \frac{2Z^2}{\omega n^3}e^{-4\tilde{\eta}}\frac{\Gamma(l+2-\tilde{\eta})}{\Gamma(2l+4)} \\ &\times[(l+2+\tilde{\eta})\Phi(l+2-\tilde{\eta}, 2l+4; 4\tilde{\eta}) \\ &- (l+2-\tilde{\eta})\Phi(l+3-\tilde{\eta}, 2l+4; 4\tilde{\eta})] \\ &\times[(2l+3-2\tilde{\eta})\Psi(-l-1-\tilde{\eta}, -2l-2; 4\tilde{\eta}) \\ &- (l+2-\tilde{\eta})\Psi(-l-1-\tilde{\eta}, -2l-3; 4\tilde{\eta})], \end{aligned} \tag{75}$$

where

$$\tilde{\eta} = \frac{Z}{\sqrt{-2\omega - i0}}.$$

Formulas (73)–(75) lead to the dependence $\alpha_{nl}^{s,a,t}, \beta_{nl} \propto n^{-3}$ since the terms proportional to $\propto n^{-2}$ cancel out when we evaluate the polarizabilities. Such a dependence corresponds to the factor $n^{-3/2}$ in the normalization of the Rydberg wave functions, but the limit under consideration cannot be obtained by substituting the latter functions into (35) since this would lead to a diverging integral (this circumstance was noted in [11]).

If we proceed to the limit $\tilde{\eta} \longrightarrow m$ in Eqs. (73)–(75), the result must obviously coincide with the limit $n \longrightarrow \infty$ taken from the corresponding resonant term in the exact matrix element (35):

$$\lim_{\tilde{\eta} \to m} \tilde{M}_{ll'}^{L}(E_n - \omega) = \lim_{n \to \infty} \frac{1}{\Delta} \langle nl' | \hat{D}(l', L) | mL \rangle \langle mL | \hat{D}(L, l) | nl \rangle, \quad (76)$$

where $\Delta = E_m - E_n + \omega$. The fulfillment of relation (76) was verified for states with $l = 0, 1$ in several low-lying resonances. We have also verified the matching of the obtained formulas to the high-frequency limit: if we make $\omega \longrightarrow \infty$ in Eqs. (73)–(75), the result coincides with that obtained from $M_{ll'}^{L}$ for the inverse sequence of proceeding to the limit: first $\omega \longrightarrow \infty$, and then $n \longrightarrow \infty$. In this case, the first two terms of the expansion of Rydberg's formulas (73) and (74) for the diagonal matrix elements and in exact expressions for $M_{ll}^{l \pm 1}$ in powers of $1/\omega$ coincide:

$$M_{ll}^{l+1} = \tilde{M}_{ll}^{l+1} \approx \frac{Z^2}{\omega n^2} + \frac{2Z^4}{\omega n^3} \frac{1}{(l+1)(2l+1)}, \quad (77)$$

$$M_{ll}^{l-1} = \tilde{M}_{ll}^{l-1} \approx \frac{Z^2}{\omega n^2} - \frac{2Z^4}{\omega n^3} \frac{1}{l(2l+1)}. \quad (78)$$

Simple analytic expressions for polarizabilities in the high-frequency limit are given in [12, 48].

5.2.2. Large values of the orbital angular momentum: $n - l \ll n$. An analysis of the high-frequency asymptotic form of polarizabilities in [48] revealed that $1/\omega n^3$ is a parameter of the expansion for $l \sim n$. This means that the principal terms of the expansion of α_{nl} into a series in ω^{-1} for a given n also determine the principal terms of the expansion in n^{-1} for a fixed frequency ω. Using the results obtained in [12, 48], we get

$$\alpha_{nl}^{s} = -\frac{1}{\omega^2} - \frac{2}{\omega^6} \frac{Z^8}{n^{12}}, \quad \alpha_{nl}^{a} = \frac{18}{\omega^7} \frac{Z^{10}}{n^{15}},$$
$$\alpha_{nl}^{t} = -\frac{1}{\omega^4} \frac{Z^4}{n^6}, \quad \beta_{nl'} = -\frac{9Z^8}{\sqrt{2}\omega^6 n^{14}}. \quad (79)$$

Thus, the elements of Stark's matrix with a small l ($\sim n^{-3}$) in the Rydberg range are larger than the matrix elements with a large l ($\propto n^{-6}$), which obviously affects the form of the spectrum.

Pay attention to the peculiar form of the frequency dependence of polarizabilities in the limit $n \longrightarrow \infty$. Since in this case any fixed value of frequency ω is a above-threshold value ($\omega > |E_n|$), the matrix elements with $E = E_n + \omega$ (which determine the resonant structure of the polarizabilities of energy levels with a finite n, which condense to the threshold $\omega = |E_n|$) turn out to be nonresonant (and the corresponding parameter η is purely imaginary). The matrix elements $M_{ll}^{L}(E_n - \omega)$ have poles at frequencies corresponding to resonances at low-lying energy levels with $n' = L + 1, L + 2, \ldots, n'_{\max} = (n - 1)$. It can easily be verified that matrix elements with a large l have no resonances in the above-threshold region (see (79)), while $M_{ll}^{L}(E_n - \omega)$ for $l \ll n$ have a finite number of resonance (for a fixed n), the separation between which increases with frequency. In the Rydberg asymptotic forms (73)–(75), the resonant dependence is associated with the factors $\Gamma(l + 2 - \tilde{\eta})$ and $\Gamma(l - \tilde{\eta})$ so that the position of the poles is shifted by the binding energy $|E_n|$ relative to the true values $\omega_{\text{res}} = E_n - E_{n'}$ in accordance with the condition $n \gg 1$. By way of an illustration of the frequency dependence, Figs. 2 and 3 show the dispersion curves of polarizabilities of states with $n = 3$ and $n = 10$ for the minimum ($l = 0$) and maximum ($l = n - 1$) values of the orbital angular momentum. A peculiar feature of the frequency dependence of polarizabilities with a small l is the presence of a negative dispersion region in a narrow range near resonances. Indeed, it can be seen from Fig. 2 that as we go over through a resonance, the real component of polarizability decreases rapidly and attains its minimum value, and then increases monotonically up to the frequency range near the next resonance. As we go over through this resonance, the polarizability first decreases, attaining the minimum (negative) value, and then increases, approaching zero. Obviously, such a (quite complicated) frequency dependence cannot be approximated by elementary formulas. At the same time, the asymptotic expressions (73)–(75) lead to a satisfactory agreement with exact expressions even for low-lying levels, the matching being improved considerably with increasing n. Pay attention to the fact that "semiasymptotic" formulas which can be obtained

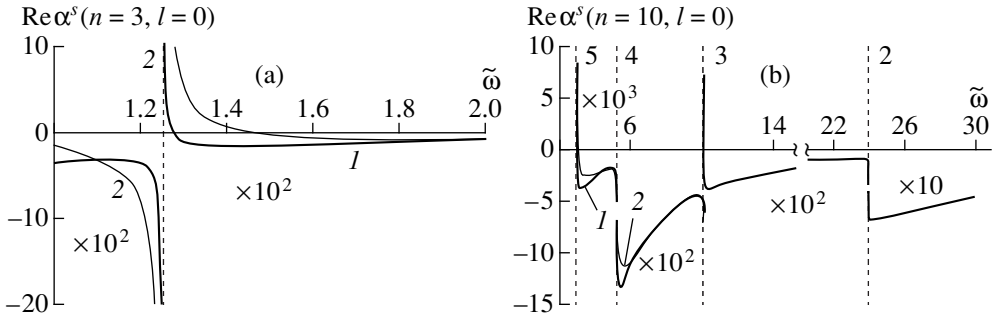

Fig. 2. Dependence of the scalar polarizability α^s_{nl} of states with (a) $n = 3$ and (b) $n = 10$, $l = 0$, calculated using the exact formula (1) and "semiasymptotic" formulas obtained from (73)–(75) as a result of the substitution $\tilde{\eta} \longrightarrow \eta$ on the frequency $\tilde{\omega} = 2n^2\omega$. Resonant frequencies are shown by vertical dashed lines on which the principal quantum number of the corresponding resonant level is indicated. The scale on the ordinate axis is indicated for each interval between resonances. In (b), curves *1* and *2* in the last two intervals between resonances coincide.

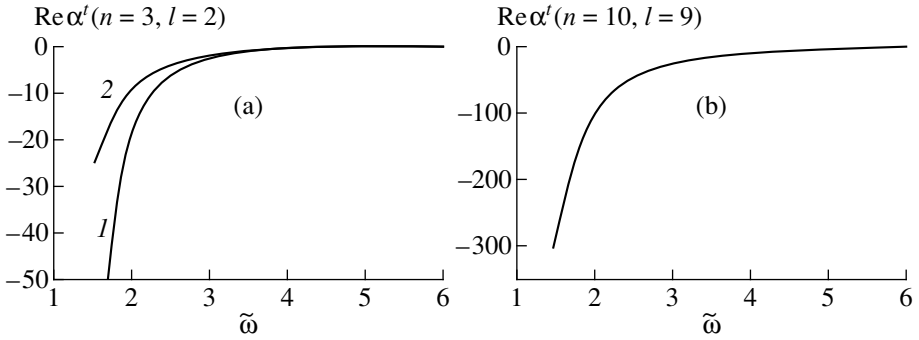

Fig. 3. Dependence of the tensor polarizability α^t_{nl} of states with (a) $n = 3$, $l = 2$ and (b) $n = 10$, $l = 9$, calculated using the exact formula (curve *1*) and the approximate formula (79) (curve *2*) on the frequency $\tilde{\omega} = 2n^2\omega$. Curves *1* and *2* in (b) coincide.

from (73)–(75) by substituting the exact value of η (depending on n) for $\tilde{\eta}$ lead to more exact results for polarizabilities, which ensures the correct positions of resonances. The dependence $\alpha^s(\omega)$ calculated in this way is also presented in Fig. 2. The polarizabilities $\alpha^{s,a,t}$ of states corresponding to large l are monotonic functions of frequency and their behavior can be correctly described by the power dependence on ω in the asymptotic forms (79). In this case, the main contribution to the scalar polarizability comes from the term $-1/\omega^2$, and the correction to it in the elements of Stark matrix (53), (54) is determined by the term with α^t. The frequency dependence of the tensor polarizability α^t for $n = 3$, $n = 10$, and $l = n - 1$, which is calculated using the exact (56) and approximate (79) formulas is presented in Fig. 3. The imaginary polarizability component is not shown in the figures for the sake of compactness. It should be noted that $\mathrm{Im}\alpha^{s,a,t}_{nl}$ is a smooth monotonically decreasing function of frequency, which is successfully approximated by the Rydberg asymptotic form for small l, while for large l, it vanishes in the limit under consideration.

5.3. Threshold Asymptotic Form of Polarizabilities

Let us now consider the asymptotic form of polarizabilities at the threshold frequencies $\omega \sim |E_n|$. For fixed values of n and frequencies $\omega \longrightarrow Z^2/2n^2$ ($\eta \longrightarrow \infty$), the matrix element $\mathrm{M}^L_{ll'}(E_n - \omega)$ is a smooth function, while $\mathrm{M}^L_{ll'}(E_n + \omega)$ has poles condensing at the Rydberg levels. Using a special representation of Green's function in the threshold region, Khristenko and Vetchinkin [51] showed that the asymptotic expression for $\mathrm{M}^L_{ll'}(E_n + \omega)$ has the form

$$\mathrm{M}^L_{ll'}(E_n + \omega) = A^L_{nll'} + \pi\cot\pi\eta B^L_{nll'}, \qquad (80)$$

and calculated coefficients A and B for the ground state $n = 1$. The coefficient B_{1s} of the resonance cotangent was expressed in terms of Laguerre polynomials, while the expression for A_{1s} was obtained in the form of a con-

Table 2. Values of coefficients A^i_{nl} and B^i_{nl} from formula (82) for states with $n \le 5$

n	l	A^s_{nl}	B^s_{nl}	A^a_{nl}	B^a_{nl}	A^t_{nl}	B^t_{nl}
1	0	–4.310(0)	–4.910(0)				
2	0	–6.045(1)	–4.605(1)				
	1	–9.827(1)	–4.221(1)	8.773(1)	5.756(1)	1.255(1)	5.372(0)
3	0	–3.473(2)	–1.765(2)				
	1	–5.122(2)	–1.825(2)	5.707(2)	2.305(2)	1.457(2)	2.690(1)
	2	–8.371(2)	–1.268(2)	1.478(3)	2.421(2)	4.159(2)	3.788(1)
4	0	–8.596(2)	–4.648(2)				
	1	–8.746(2)	–4.912(2)	–2.593(2)	5.762(2)	–1.911(2)	8.124(1)
	2	–7.937(2)	–4.457(2)	–7.571(2)	8.225(2)	–4.156(2)	1.372(2)
	3	–2.876(2)	–2.378(2)	–2.344(3)	5.206(2)	–9.699(2)	1.007(2)
5	0	–2.387(3)	–9.930(2)				
	1	–2.583(3)	–1.052(3)	5.531(2)	1.151(3)	1.744(1)	1.904(2)
	2	–2.872(3)	–1.041(3)	1.151(3)	1.852(3)	9.012(1)	3.304(2)
	3	–2.937(3)	–7.981(2)	1.104(3)	1.717(3)	–1.968(–2)	3.413(2)
	4	–2.149(3)	–3.411(2)	–1.272(3)	8.040(2)	–7.713(2)	1.750(2)

tour integral, which can be subsequently evaluated numerically. Approximate expressions for $B^L_{nll'}$ for arbitrary n, l, and l' were obtained by us earlier [12] in the form of a cumbersome combination of finite (triple) sums. Using asymptotic expressions for p^l_k and q^l_k for $\eta \longrightarrow \infty$ (we omit the details of such computations), we can derive analytic expressions for A and B for arbitrary n, l, and l', but only in the form of an infinite series for the logarithmic derivative of the Γ-function, $\psi(x)$:

$$p^l_k \longrightarrow (-1)^k L^{2l+1}_k(4n),$$

$$q^{+l}_k \longrightarrow 2(-1)^{k+1}(4n)^{2l+1}e^{-4n}L^{2l+1}_k(4n)$$
$$\times(\pi\cot\pi\eta + \ln 4n) + 2(-1)^k(4n)^{2l+1}$$
$$\times\sum_{p=0}^{\infty}\frac{(k+1)_{p+2l+1}}{p!(p+2l+1)!}(-4n)^p(\psi(p+2l+2) \qquad (81)$$
$$+\psi(p+1)-\psi(p+k+2l+2))$$
$$+2(-1)^k\sum_{p=0}^{2l}\frac{(2l-p)!(k+1)_p}{p!}(4n)^p.$$

For $k = 0$, the infinite series can be expressed through the integral exponential function $\mathrm{Ei}(x)$:

$$q^{+l}_0 \longrightarrow -2(4n)^{2l+1}e^{-4n}$$
$$\times(\pi\cot\pi\eta + \mathrm{Ei}(4n)) + 2\sum_{p=0}^{2l}(2l-p)!(4n)^p,$$

which leads to the results obtained in [3] for the ground state.

Using expressions (80) and (81), we can present the polarizabilities $\alpha^i_{nl}(\omega)$ ($i = s, a, t$) for $\omega \longrightarrow |E_n| - 0$ in the form

$$\alpha^i_{nl}(\omega) = A^i_{nl} + \cot\pi\eta B^i_{nl}. \qquad (82)$$

In order to obtain the asymptotic form of the matrix element $\mathrm{M}^L_{ll'}(E_n + \omega)$ for the frequency approaching the ionization potential from above, $\omega \longrightarrow |E_n| + 0$, we must make the substitution $\cot\pi\eta \longrightarrow -i$ in (82).

Table 2 shows, by way of an example, the values of A^i_{nl} and B^i_{nl} for states with $n \le 5$. For $n = 1$, our results for A_{1s} and B_{1s} coincide with those obtained in [51].

6. CONCLUDING REMARKS

The application of the generalized Sturm expansion of the CGF (6) has made it possible to derive closed analytic expressions (two-photon Gordon formulas) for the matrix elements of two-photon transitions from an arbitrary state $|nl\rangle$ of a hydrogen-like atom to the bound state $|n'l'\rangle$, (39)–(42) and (61)–(63), as well as the continuous spectrum, (46)–(49). In all cases, the result can be expressed through the kernel $g^l_{kk'}$ of representation (6), which contains only one complete hypergeometric function ${}_2F_1$ or Appel's function F_1. Consequently, we can hope that in spite of the cumbersome nature of the obtained results, they provide complete information on

the atomic parameters required in the problems of the two-photon spectroscopy of hydrogen-like atomic energy levels in the simplest possible analytic form and conclude the protracted history of the analytical calculations of two-phonon transitions in the Coulomb field for particular values of quantum numbers of the initial and/or final states. The application of the asymptotic expressions of hypergeometric functions for particular values of parameters and/or arguments can obviously simplify the general results in various limiting cases (like the analysis of polarizabilities in Subsections 5.2 and 5.3). It should be noted that the two-photon Gordon formulas (39)–(42) cannot be extended to the region of continuous spectrum in both variables n and n' (since the variable associated with the upper limit of summation in (18) must be an integer). Consequently, two-photon transition in the continuous spectrum requires a separate analysis.

The case of two-photon transitions considered by us here does not exhaust the possible applications of parametric representation of the CGF $G_E(\mathbf{r}, \mathbf{r}')$ and $G_E(\mathbf{p}, \mathbf{p}')$. The effectiveness of expansions of type (4), (6), and (32) is determined by the following two circumstances. First, an appropriate choice of free parameters considerably simplifies the integration procedure in some cases and makes it possible to present the results in the simplest analytical form (see above). Since the matrix elements with Green's functions for an arbitrary energy E can be expressed, as a rule, in terms of quite complicated special functions, the application of an inadequate computational algorithm either leads to much more cumbersome constructions (see, for example, [15]), or makes it impossible altogether to present the result in an analytical form. The other even more important circumstance is the possibility of using free parameters for extending the method of direct numerical calculations of a Sturm series of type (6) to matrix elements with Green's functions for $E > 0$, which emerge, for example, in collision problems, in the theory of self-ionization states, and in an analysis of above-threshold many-photon transitions. It was mentioned above that expansions (7) and (9) are inapplicable for $E > 0$, and the corresponding series for matrix elements diverge (apart from exceptional cases when the result can be expressed in terms of known special functions that can be continued analytically in E, e.g., formula (5) or two-photon Gordon formulas). The corresponding generalization of representations (9) and (7) to the region $E > 0$ were obtained in [52] and [53], and the relativistic CGF for $|E| > mc^2$ was considered in [54]. However, as a result of such a generalization, the summation over the discrete index k in (7) and (9) is replaced by the integration with respect to continuous parameters of special functions, which considerably complicates the application of such representations in actual numerical calculations (the choice of a modified Sturm basis with a discrete spectrum for an expansion of Green's functions with $E > 0$ is considered in [55]). The application of an expansion of type (6) and a rational choice of parameters α and α' can ensure the convergence of Sturm series in direct numerical calculations of the matrix elements for $E > 0$ (and also accelerate their convergence considerably for energies $E < 0$). Thus, the generalized Sturm representation makes it possible in fact to carry out a numerical analytical continuation of the results obtained by using the ordinary Sturm expansion (7) to the energy range in which series (7) diverges.

At the present time, the method of complex scaling (or complex coordinates) [56] is used as one of the most effective numerical methods in problems with a continuous spectrum and with resonance (quasi-stationary) states in atomic and molecular processes. It involves a non-Hermite extension (complex dilatation) of the initial Hamiltonian of the problem through the substitution $\mathbf{r} \longrightarrow \mathbf{r}e^{i\alpha}$ and "continuum discretization" [57] by introducing the basis of quadratically integrable functions depending on a (complex) parameter α for determining the complex eigenvalues, i.e., the energies of resonances. For example, Maquet *et al.* [58] used the system of Sturm's functions $S_{nl}(2r/\alpha)$ with a complex parameter α as the basis for the numerical calculation of complex quasi-energies of the hydrogen atom without employing perturbation theory in the interaction with a high-intensity laser field (see also the review in [59]), while Gersbacher and Broad [60] applied this method for an analysis of self-ionization resonances in the photoionization of helium. The application of generalized Sturm representations of the CGF is essentially an analytic implementation of the complex scaling and continuum discretization methods in problems with the Coulomb Green's function for a positive (or complex-valued) energy E.

The effectiveness of representation (6) in calculations of matrix elements in the higher orders of perturbation theory (when the result cannot be presented in terms of known special functions) was verified by us in numerical calculations of the hyperpolarizabilities of highly excited states of the hydrogen atom in the field of laser radiation, which contain matrix elements with three Green's functions for $E > 0$ [61]: with an appropriate choice of (complex) parameters α and α', the series for the matrix elements converge rapidly at frequencies ω exceeding the ionization potential $|E_n|$ of the state $|nl\rangle$ under investigation by more than an order of magnitude. The possibility to ensure the convergence of Sturm-type series for the composite matrix elements of electromagnetic transitions for above-threshold energies justifies the application of expansions of type (6) in relativistic problems, also, e.g., for calculating the Rayleigh and Compton scattering cross sections of X-rays and γ-radiation by multiply charged ions or inner shells of atoms with large values of Z as well as in problems of the quantum electrodynamics of bound states, which require a nonperturbative (in parameter αZ) analysis of a strong Coulomb field acting on an electron in virtual states.

ACKNOWLEDGMENTS

This work was partially financed by the joint program INTAS–RFBR (grant no. 97-673) and RFBR (grant no. 00-02-17843), as well as the Competition Center of Fundamental Science and the program "Universities of Russia: Basic Research."

APPENDIX A

1. The symmetry conditions (19) and expression (18) lead to the identity connecting the "bilinear forms" which contain functions ${}_2F_1$ and F_1:

$$\frac{{}_2F_1(-n, l+1-\eta; 2l+2; z)}{l+1-\eta}$$
$$\times F_1(l+1-\eta; -n', n'+2l+2; l+2-\eta; y, y')$$
$$-\frac{(1-y)^{n'}}{(1-y')^{n'+2l+2}}\sum_{p=1}^{n} C_n^p(-z)^p\frac{(l+2+\eta-p)_{p-1}}{(2l+2)_p}$$
$$\times {}_2F_1(-n+p, l+1-\eta+p; 2l+2+p; z)$$
$$\times F_1(-p+1; -n', n'+2l+2;$$
$$l+2+\eta-p; 1/(1-y), 1/(1-y'))$$
$$= \frac{{}_2F_1(-n', l+1-\eta; 2l+2; z')}{l+1-\eta} \qquad \text{(A.1)}$$
$$\times F_1(l+1-\eta; -n, n+2l+2; l+2-\eta; 1/y, y')$$
$$-\frac{(1-1/y)^{n}}{(1-y')^{n+2l+2}}\sum_{p=1}^{n'} C_{n'}^p(-z')^p\frac{(l+2+\eta-p)_{p-1}}{(2l+2)_p}$$
$$\times {}_2F_1(-n'+p, l+1-\eta+p; 2l+2+p; z')$$
$$\times F_1(-p+1; -n, n+2l+2;$$
$$l+2+\eta-p; 1/(1-y), 1/(1-y')).$$

A special case emerges when the free parameters coincide: $\alpha' = \alpha$. Proceeding to the limit $\alpha' \longrightarrow \alpha$ in (A.1), we arrive at a relation which has the following form in terms of Pollaczek polynomials and functions:

$$q_{n_>}^l p_{n_<}^l - q_{n_<}^l p_{n_>}^l = (-1)^{n_<}\frac{(n_<+1)_{2l+1}(n_>+1)_{2l+1}}{(2l+1)!}$$
$$\times\frac{2}{(4\alpha\nu)^{n_<+1}}\frac{(\alpha-\nu)^{n_>+n_<+1}}{(\alpha+\nu)^{n_>-n_<-1}}\sum_{p=n_<+1}^{n_>} C_{n_>}^p(-z)^p$$
$$\times\frac{(-p+1)_{n_<}(l+2+\eta+n_<-p)_{p-n_<-1}}{(2l+2)_p} \qquad \text{(A.2)}$$
$$\times {}_2F_1(-n_>+p, l+1-\eta+p; 2l+2+p; z)$$
$$\times {}_2F_1(-p+1+n_<, n_<+2l+2;$$
$$l+2+\eta+n_<-p; (\alpha+\nu)^2/4\alpha\nu),$$

where $n_< = \min\{n, n'\}$, $n_> = \max\{n, n'\}$.

In the simplest case when $n_< = n = n_> - 1$, (A.2) leads to the relation

$$q_n^{+l} p_{n-1}^l - q_{n-1}^{+l} p_n^l = 2(n+1)_{2l}, \qquad \text{(A.3)}$$

corresponding to the particular case of the symmetry condition (19):

$$g_{kk-1}^l(\nu; \alpha, \alpha) = g_{k-1k}^l(\nu; \alpha, \alpha).$$

Note that Broad [10, formula (A.18)] interpreted relation (A.3) as the "Wronskian" of the solutions of the trinomial recurrence relation (27).

2. We transform the function ${}_2F_1(k+1, k+2l+2; k+l+2-\eta; x)$ appearing in (17) by using the well-known relations [5]:

$$\frac{d^k}{dx^k}{}_2F_1(1, 2l+2; l+2-\eta; x) = \frac{k!(2l+2)_k}{(l+2-\eta)_k}$$
$$\times {}_2F_1(k+1, k+2l+2; k+l+2-\eta; x),$$
$${}_2F_1(1, 2l+2; l+2-\eta; x) = (l+1-\eta)x^{\eta-l-1} \qquad \text{(A.4)}$$
$$\times(1-x)^{-\eta-l-1}\int_0^x t^{l-\eta}(1-t)^{l+\eta}dt.$$

This gives

$$(l+1-\eta)^{-1}{}_2F_1(k+1, k+2l+2; k+l+2-\eta; x)$$
$$= \frac{(l+2-\eta)_k}{(2l+2)_k k!}\left[\sum_{p=1}^{k} C_k^p\frac{d^{k-p}}{dx^{k-p}}(x^{-l-1+\eta}(1-x)^{-l-1-\eta})\right. \qquad \text{(A.5)}$$
$$\times\frac{d^{p-1}}{dx^{p-1}}(x^{l-\eta}(1-x)^{l+\eta})$$
$$\left.+\frac{d^k}{dx^k}(x^{-l-1+\eta}(1-x)^{-l-1-\eta})\int_0^x t^{l-\eta}(1-t)^{l+\eta}dt\right].$$

Expressing the derivatives in (A.5) in terms of hypergeometric polynomials (see (14)) and going over to the reciprocal argument of these polynomials,

$${}_2F_1(-k, -k-2l-1; -k-l+\eta; x) = \frac{(2l+2)_k}{(l+1-\eta)_k}$$
$$\times(-x)^k{}_2F_1(-k, 1+l-\eta; 2l+2; x^{-1}),$$

we arrive at the following expansion:

$${}_2F_1(k+1, k+2l+2; k+l+2-\eta; x)$$
$$= \frac{(l+2-\eta)_k}{k!}(1-x)^{-k}{}_2F_1(-k, l+1-\eta; 2l+2; x^{-1})$$

$$\times {}_2F_1(1, 2l+2;\ l+2-\eta;\ x)$$
$$-\frac{(l+1-\eta)_{k+1}}{(2l+2)_k k!}(1-x)^{-k}\sum_{p=1}^{k} C_k^p (2l+2)_{k-p} \qquad \text{(A.6)}$$
$$\times (\eta-l)_{p-1}(-x)^{-p}\,{}_2F_1(-k+p,\ l+1-\eta,\ 2l+2,\ x^{-1})$$
$$\times {}_2F_1(-p+1,\ 2l+2-p;\ l+2-\eta-p;\ x).$$

Substituting (A.6) into (17), we obtain expressions (21) and (22).

3. Formula (23) can be derived using the following expansion of the Whittaker function $M_{\eta,\, l+1/2}(2r/\nu)$ in the complete system of Sturm's functions $S_{nl}(2r/\alpha)$:

$$\sum_{K=0}^{\infty}\left(\frac{\alpha-\nu}{\alpha+\nu}\right)^k {}_2F_1(-k,\ l+1-\eta;\ 2l+2;\ z)S_{kl}(2r/\alpha) \qquad \text{(A.7)}$$
$$= \left(\frac{\nu^2-\alpha^2}{4\alpha\nu}\right)^{l+1}\left(\frac{\nu+\alpha}{\nu-\alpha}\right)^{\eta}\frac{1}{r}M_{\eta,\, l+1/2}(2r/\nu),$$

which can be obtained in analogy with expansion (11). In fact, (A.7) is the expansion of the solution of the Coulomb problem with an (arbitrary) energy E in the quadratically integrable basis $S_{kl}(2r/\alpha)$, which is regular at zero. A detailed analysis of such expansions (including the case of a irregular solution) is given in [10, 62].

4. In order to separate the imaginary component of Green's function $g_l(E; r, r')$ for $E > 0$, we will use the following relation:

$$\frac{{}_2F_1(1, 2l+2, l+2-\eta;\ z^{-1})}{l+1-\eta}$$
$$+\frac{{}_2F_1(1, 2l+2, l+2+\eta;\ 1-z^{-1})}{l+1+\eta} \qquad \text{(A.8)}$$
$$= \frac{\Gamma(l+1-\eta)\Gamma(l+1+\eta)}{\Gamma(2l+2)}\frac{z^{2l+2}}{(z-1)^{\eta+l+1}}.$$

If z is determined by formula (12), α is real-valued, and ν and η are imaginary, the terms on the left-hand side are complex conjugates, which allows us to single out the imaginary component in (23).

APPENDIX B

The expressions for $\psi_{nl}^{s,a,t}$ in terms of p_k^l and $\bar{q}_k^l$ have the form

$$\psi_{nl}^s = \frac{l}{3(2l+1)}[f^{l-1}(\tilde{\omega}) + f^{l-1}(-\tilde{\omega})]$$
$$+\frac{l+1}{3(2l+1)}[f^{l+1}(\tilde{\omega}) + f^{l+1}(-\tilde{\omega})],$$

$$\psi_{nl}^a = \frac{l}{(2l+1)}[(f^{l-1}(\tilde{\omega}) + f^{l-1}(-\tilde{\omega}))$$
$$-(f^{l+1}(\tilde{\omega}) - f^{l+1}(-\tilde{\omega}))],$$

$$\psi_{nl}^t = \frac{l}{3(2l+1)}[f^{l-1}(\tilde{\omega}) + f^{l-1}(-\tilde{\omega})]$$
$$+\frac{l}{3(2l+1)}\frac{2l-1}{2l+3}[f^{l+1}(\tilde{\omega}) + f^{l+1}(-\tilde{\omega})],$$

where

$$f^{l+1}(\tilde{\omega}) = \frac{2n}{\tilde{\omega}(s+1)^2(d)_{2l+1}}$$
$$\times[p_{d-1}^{l+1} - p_{d-2}^{l+1}][\bar{q}_{d-1}^{l+1} - \bar{q}_{d-2}^{l+1}],$$

$$f^{l-1}(\tilde{\omega}) = \frac{2n}{\tilde{\omega}(d)_{2l+1}}[(s-1)p_{d-1}^{l-1} - dp_d^{l-1}]$$
$$\times[(s-1)\bar{q}_{d-1}^{l-1} - d\bar{q}_d^{l-1}].$$

The explicit form of the polynomials $\varphi_{nl}^{(\pm)}$ and $\psi_{nl}^{s,a,t}$ appearing in expression (65) for polarizabilities is given below for states with the principle quantum number $n \le 4$:

$$\varphi_{10}^{(+)} = \frac{1}{2}, \quad \psi_{10}^s = 0,$$

$$\varphi_{20}^{(+)} = 1, \quad \psi_{20}^s = 0,$$

$$\varphi_{21}^{(+)} = \frac{5}{12}, \quad \psi_{21}^s = -\psi_{21}^t = \frac{32}{27\tilde{\omega}^2},$$

$$\varphi_{21}^{(-)} = \frac{8}{9\tilde{\omega}^2}, \quad \psi_{21}^a = -\frac{4}{3\tilde{\omega}},$$

$$\varphi_{30}^{(+)} = \frac{3}{2}\left(1+\frac{4}{3\tilde{\omega}}\right)^2, \quad \psi_{30}^s = \frac{16}{\tilde{\omega}^2},$$

$$\varphi_{31}^{(+)} = \frac{5}{3}, \quad \psi_{31}^s = -\psi_{31}^t = \frac{16}{9\tilde{\omega}^2}\left(1+\frac{8}{\tilde{\omega}^2}\right),$$

$$\varphi_{31}^{(-)} = \frac{32}{9\tilde{\omega}^2}\left(1+\frac{2}{\tilde{\omega}}\right)^2, \quad \psi_{31}^a = -\frac{8}{\tilde{\omega}}\left(1-\frac{4}{\tilde{\omega}^2}\right),$$

$$\varphi_{32}^{(+)} = \frac{7}{18}, \quad \psi_{32}^s = -\psi_{32}^t = \frac{16}{25\tilde{\omega}^2},$$

$$\varphi_{32}^{(-)} = \frac{4}{15\tilde{\omega}^2}, \quad \psi_{32}^a = -\frac{8}{5\tilde{\omega}},$$

$$\varphi_{40}^{(+)} = 2\left(1+\frac{4}{\tilde{\omega}}+\frac{8}{3\tilde{\omega}^2}\right)^2,$$

$$\psi_{40}^{s} = \frac{320}{3\tilde{\omega}^2}\left(1+\frac{8}{15\tilde{\omega}^2}\right),$$

$$\varphi_{41}^{(+)} = \frac{25}{6}\left(1+\frac{4}{5\tilde{\omega}}\right)^2,$$

$$\psi_{41}^{s} = \frac{448}{27\tilde{\omega}^2}\left(1+\frac{552}{35\tilde{\omega}^2}+\frac{384}{35\tilde{\omega}^4}\right),$$

$$\varphi_{41}^{(-)} = \frac{80}{9\tilde{\omega}^2}\left(1+\frac{28}{5\tilde{\omega}}+\frac{24}{5\tilde{\omega}^2}\right)^2,$$

$$\psi_{41}^{a} = -\frac{80}{3\tilde{\omega}}\left(1-\frac{144}{25\tilde{\omega}^2}-\frac{3328}{75\tilde{\omega}^4}\right),$$

$$\psi_{41}^{t} = -\frac{512}{135\tilde{\omega}^2}\left(1+\frac{69}{\tilde{\omega}^2}+\frac{48}{\tilde{\omega}^4}\right),$$

$$\varphi_{42}^{(+)} = \frac{7}{3}, \quad \psi_{42}^{s} = -\psi_{42}^{t} = \frac{64}{75\tilde{\omega}^2}\left(1+\frac{16}{3\tilde{\omega}^2}\right),$$

$$\varphi_{42}^{(-)} = \frac{8}{5\tilde{\omega}^2}\left(1+\frac{4}{3\tilde{\omega}}\right)^2,$$

$$\psi_{42}^{a} = -\frac{64}{5\tilde{\omega}}\left(1-\frac{16}{15\tilde{\omega}^2}\right),$$

$$\varphi_{43}^{(+)} = \frac{3}{8}, \quad \psi_{43}^{s} = -\psi_{43}^{t} = \frac{64}{147\tilde{\omega}^2},$$

$$\varphi_{43}^{(-)} = \frac{8}{63\tilde{\omega}^2}, \quad \psi_{43}^{a} = -\frac{12}{7\tilde{\omega}}.$$

REFERENCES

1. W. Gordon, Ann. Phys. (Leipzig) **2**, 1031 (1929); V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics* (Nauka, Moscow, 1980; Pergamon, Oxford, 1982), Parag. 52; H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Academic, New York, 1957; Fizmatgiz, Moscow, 1960), Parag. 63.
2. C. Schwartz and T. J. Tiemann, Ann. Phys. (N.Y.) **6**, 178 (1959); M. H. Mittleman and F. Wolfe, Phys. Rev. **128**, 2686 (1962).
3. M. Gavrila, Preprint IFA FI-60 (Inst. of Atomic Physics, Bucharest, 1966); Phys. Rev. **163**, 147 (1967).
4. S. I. Vetchinkin and S. V. Khristenko, Chem. Phys. Lett. **1**, 437 (1967); S. V. Khristenko and S. I. Vetchinkin, Opt. Spektrosk. **25**, 650 (1968).
5. *Higher Transcendental Functions (Bateman Manuscript Project)*, Ed. by A. Erdelyi (McGraw-Hill, New York, 1953; Nauka, Moscow, 1973), Vol. 1.
6. B. A. Zon, N. L. Manakov, and L. P. Rapoport, Zh. Éksp. Teor. Fiz. **55**, 924 (1968) [Sov. Phys. JETP **28**, 480 (1969)].
7. M. Gavrila, Z. Phys. A **293**, 269 (1979); V. Florescu, Phys. Rev. A **30**, 2441 (1984); V. Florescu and A. Cionga, Z. Phys. A **321**, 187 (1985).
8. A. Maquet, Phys. Rev. A **15**, 1088 (1977).
9. M. A. Preobrazhenskiĭ, Zh. Éksp. Teor. Fiz. **111**, 816 (1997) [JETP **84**, 448 (1997)].
10. J. T. Broad, Phys. Rev. A **31**, 1494 (1985).
11. V. Yakhontov and K. Jungmann, Z. Phys. D **38**, 141 (1996).
12. A. A. Krylovetsky, N. L. Manakov, and S. I. Marmo, Laser Phys. **7**, 817 (1997).
13. A. Costescu, I. Brändus, and N. Mezincescu, J. Phys. B **18**, L11 (1985); V. Florescu, S. Pätrascu, and O. Stoican, Phys. Rev. A **36**, 2155 (1987).
14. E. Karule, J. Phys. B **4**, L67 (1971); A. Maquet, Phys. Lett. A **48A**, 199 (1974).
15. T. A. Marian, Phys. Rev. A **39**, 3816 (1989).
16. A. I. Ignat'ev, Zh. Éksp. Teor. Fiz. **70**, 484 (1976) [Sov. Phys. JETP **43**, 250 (1976)].
17. J. P. Gazeau, J. Math. Phys. **19**, 1041 (1978).
18. L. P. Rapoport, B. A. Zon, and N. L. Manakov, Zh. Éksp. Teor. Fiz. **56**, 400 (1969) [Sov. Phys. JETP **29**, 220 (1969)].
19. S. Klarsfeld, Lett. Nuovo Cimento **1**, 682 (1969); **2**, 548 (1969); **3**, 395 (1970); V. G. Gorshkov and V. S. Polikanov, Pis'ma Zh. Éksp. Teor. Fiz. **9**, 464 (1969) [JETP Lett. **9**, 279 (1969)].
20. M. Gavrila, Lett. Nuovo Cimento **2**, 180 (1969); Phys. Rev. A **6**, 1360 (1972); Rev. Roum. Phys. **19**, 483 (1974).
21. E. Karule, J. Phys. B **11**, 441 (1978); E. Karule and R. H. Pratt, J. Phys. B **24**, 1585 (1991); E. Karule, J. Phys. B **18**, 220 (1985).
22. S. Klarsfeld and A. Maquet, J. Phys. B **12**, L553 (1979); Phys. Lett. A **78A**, 40 (1980); J. Phys. B **73**, 100 (1979).
23. A. G. Fainshtein, N. L. Manakov, and S. I. Marmo, Phys. Lett. A **104A**, 347 (1984).
24. N. L. Manakov, S. I. Marmo, and A. V. Shaposhnikov, in *Atoms and Molecules in Strong Field of Laser Radiation*, Ed. by F. V. Bunkin and I. I. Tugov (Nauka, Moscow, 1992), p. 87.
25. E. J. Heller, Phys. Rev. A **12**, 1222 (1975).
26. N. L. Manakov, A. Maquet, S. I. Marmo, and C. Szymanowski, Phys. Lett. A **237**, 234 (1998).
27. S. A. Zapryagaev, N. L. Manakov, and V. G. Pal'chikov, *Theory of Multiply Charged Ions with One and Two Electrons* (Énergoatomizdat, Moscow, 1985), Chap. 2.
28. N. L. Manakov, V. D. Ovsiannikov, and L. P. Rapoport, Phys. Rep. **141**, 319 (1986).
29. A. Maquet, V. Véniard, and T. A. Marian, J. Phys. B **31**, 3743 (1998).
30. L. Hostler, J. Math. Phys. **11**, 2966 (1970).
31. *Higher Transcendental Functions (Bateman Manuscript Project)*, Ed. by A. Erdelyi (McGraw-Hill, New York, 1953; Nauka, Moscow, 1974), Vol. 2.
32. N. L. Manakov, S. I. Marmo, and A. G. Fainshtein, Teor. Mat. Fiz. **59**, 49 (1984).
33. E. Arnous, J. Bastian, and A. Maquet, Phys. Rev. A **27**, 977 (1983).
34. R. Shakeshaft, J. Phys. B **18**, L611 (1985); Phys. Rev. A **34**, 244 (1986).

35. B. A. Zon, N. L. Manakov, and L. P. Rapoport, Yad. Fiz. **15**, 508 (1972) [Sov. J. Nucl. Phys. **15**, 282 (1972)]; N. L. Manakov, L. P. Rapoport, and S. A. Zapryagaev, Phys. Lett. A **43A**, 139 (1973).
36. J. Schwinger, J. Math. Phys. **5**, 1606 (1964).
37. L. C. Hostler, J. Math. Phys. **5**, 1235 (1964).
38. *Higher Transcendental Functions (Bateman Manuscript Project)*, Ed. by A. Erdelyi (McGraw-Hill, New York, 1955; Nauka, Moscow, 1974), Vol. 3.
39. R. Papp, J. Phys. A **20**, 153 (1987).
40. L. P. Presnyakov and A. M. Urnov, Zh. Éksp. Teor. Fiz. **68**, 61 (1975) [Sov. Phys. JETP **41**, 31 (1975)].
41. C. Greene, U. Fano, and G. Strinati, Phys. Rev. A **19**, 1485 (1979).
42. V. E. Chernov, N. L. Manakov, and A. F. Starace, Eur. Phys. J. D **8**, 347 (2000).
43. C. Dorman, I. Kucukkara, and J. P. Marangos, Phys. Rev. A **61**, 013802 (2000).
44. G. Z. Zhang, D. W. Tokaryk, and B. P. Stoicheff, Phys. Rev. A **56**, 813 (1997).
45. N. L. Manakov, A. Maquet, S. I. Marmo, *et al.*, J. Phys. B **32**, 3747 (1999).
46. B. A. Zon, N. L. Manakov, and L. P. Rapoport, Opt. Spektrosk. **38**, 13 (1975) [Opt. Spectrosc. **38**, 6 (1975)].
47. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Scientific, Singapore, 1988).
48. N. L. Manakov, V. A. Sviridov, and A. G. Fainshtein, Zh. Éksp. Teor. Fiz. **95**, 790 (1989) [Sov. Phys. JETP **68**, 451 (1989)].
49. V. P. Gavrilenko, Zh. Éksp. Teor. Fiz. **90**, 857 (1986) [Sov. Phys. JETP **63**, 500 (1986)].
50. I. L. Beigman, Zh. Éksp. Teor. Fiz. **100**, 125 (1991) [Sov. Phys. JETP **73**, 68 (1991)]; I. L. Beigman, L. A. Bureeva, and R. H. Pratt, J. Phys. B **27**, 5833 (1994).
51. S. V. Khristenko and S. I. Vetchinkin, Opt. Spektrosk. **26**, 310 (1969).
52. A. M. Perelomov and V. S. Popov, Zh. Éksp. Teor. Fiz. **50**, 179 (1966) [Sov. Phys. JETP **23**, 118 (1966)].
53. N. E. Firsova and A. I. Sherstyuk, Teor. Mat. Fiz. **81**, 59 (1989).
54. R. Szmytkowski, J. Phys. A **31**, 4963 (1998); **31**, 7415 (1998).
55. A. I. Sherstyuk, Opt. Spektrosk. **87**, 765 (1999) [Opt. Spectrosc. **87**, 696 (1999)].
56. B. R. Junker, Adv. At. Mol. Phys. **18**, 208 (1982); N. Moiseyev, Phys. Rep. **302**, 211 (1998).
57. H. A. Yamani and W. P. Reinhardt, Phys. Rev. A **11**, 1144 (1975).
58. A. Maquet, S. I. Chu, and W. P. Reinhardt, Phys. Rev. A **27**, 2946 (1983).
59. R. M. Potvliege and R. Shakeshaft, in *Atoms in Intense Laser Fields*, Ed. by M. Gavrila (Academic, New York, 1992), p. 373.
60. R. Gersbacher and J. T. Broad, J. Phys. B **23**, 365 (1990).
61. N. L. Manakov, S. I. Marmo, E. A. Pronin, and A. F. Starace, in *Abstracts of Annual Conference of American Physical Society, DAMOP'2000, Storrs: Connecticut, USA, 14–17 June, 2000* [Bull. Amer. Phys. Soc. **45**, 118 (2000)].
62. L. J. Dubé and J. T. Broad, J. Phys. B **23**, 1711 (1990).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 61–68.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 71–79.

ATOMS, SPECTRA, RADIATION

Spontaneous Multipole Radiation in a Condensed Medium

E. V. Tkalya

Skobel'tsyn Research Institute of Nuclear Physics, Moscow State University, Vorob'evy gory, Moscow, 119899 Russia
e-mail: tkalya@ibrae.ac.ru

Received August 28, 2000

Abstract—The probability of the spontaneous emission of electromagnetic radiation of an arbitrary multipolarity in an infinite homogeneous isotropic nonabsorbent condensed medium with constant permittivity and permeability is determined using the perturbation theory for quantum electrodynamics. The local field inside a sphere is calculated for fields of arbitrary configuration in the medium with the help of the real and virtual cavity models. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The probability of the spontaneous emission of dipole radiation by atoms and molecules depends on the electronic properties of the surrounding medium. It has been proved experimentally that an infinite homogeneous nonabsorbent dielectric medium with refractive index n at the radiation frequency accelerates the electric dipole (E1) transitions approximately by a factor of n [1, 2] and the magnetic dipole (M1) transitions by a factor of n^3 [1] in comparison with the analogous transitions in vacuum.

It was also found that in the case of $E1$-radiation, the local electric field $\mathbf{E}_{\text{loc}}$ acting on an atom emitting radiation slightly differs from the mean value of the electric component of the macroscopic electric field $\mathbf{E}_m$ in the medium and is connected with it through the relation $\mathbf{E}_{\text{loc}} = f(n)\mathbf{E}_m$. The function $f(n)$ obtained in the framework of the "real" or "empty" cavity model studied in detail by Glauber and Lewenstein [3] is in good agreement with the experimental data presented in [1, 2]. Another popular model of "virtual" or Lorentz cavity [4] (see Chapter 2 in [5] for details) has also received a certain experimental substantiation [6]. A fairly complete description of both approaches to the problem of local field and the corresponding lists of references can be found in [7–10].

The dependence of the probability of spontaneous dipole transitions on the dielectric properties of the medium, which was predicted earlier in [3, 11], was confirmed by the experimental results obtained in [1, 2]. A large number of theoretical works [3, 7–25] published in recent years were devoted to various aspects of this problem, including the quantization of an electromagnetic field in absorbing and nonabsorbing insulators, the spontaneous decay of excited states, the dipole–dipole interaction, Einstein's coefficients, etc. Significant results were also obtained from the investigations of the effect of the periodic structure of an unbounded dielectric on its optical properties. Such materials, which were termed as "photon crystals", are described in detail in [26–33]. Finally, the effect of a homogeneous isotropic dielectric medium on the probability of spontaneous nuclear emission in the optical range as a result of the decay of the anomalously low-lying level $3/2^+$ (3.5 ± 1.0 eV) in the ^{229}Th nucleus was studied theoretically in our recent publications [34, 35].

Optical transitions $E2$ [36] and even $E3$ [37, 38] were detected quite recently in singly charged ions. The present paper is devoted to a derivation of the formulas for the probability of the emission of radiation of an arbitrary multipolarity in a nonabsorbing homogeneous isotropic medium with permittivity ϵ and permeability μ. A relation between the mean field of an arbitrary configuration in a medium and the local field inside a sphere with permittivity and permeability differing from the values ϵ and μ for the medium is derived for the real and virtual cavity models in the near radiation band. In this case, the medium is assumed to be unbounded, and the separation between the atoms of the substance is much smaller than the radiation wavelength. In the system of units adopted in this work, $\hbar = c = 1$.

2. NONINTERACTING FIELDS APPROXIMATION

In order to calculate the probability of spontaneous emission, we make use of the perturbation theory for quantum electrodynamics (QED), developed in the approximation of noninteracting fields.

The first-order S-matrix element corresponding to the process of photon emission from a bounded system (atom, ion, molecule, nucleus, etc.) has the form

$$S^{(1)} = -i\int H_{\text{int}}(t)dt, \qquad (1)$$

where $H_{\text{int}}(t)$ is the interaction Hamiltonian in the interaction representation. It is connected with the density of the Lagrangian function $\mathcal{L}_{\text{int}}(x)$, where $x = (t, \mathbf{r})$, describ-

ing the interaction between the electron current $j^{\nu}_{fi}(x)$ of a transition and the photon field $A_{\nu}(x)$, i.e.,

$$\mathcal{L}_{\text{int}}(x) = j^{\nu}_{fi}(x)A_{\nu}(x), \tag{2}$$

through the relation (see, for example, [39, 40])

$$H_{\text{int}}(t) = -\int \mathcal{L}_{\text{int}}(t, \mathbf{r})d^3r. \tag{3}$$

When speaking of the interaction of a photon with an electron of an atom or an ion, we can take the current density operator in the interaction representation from Eq. (2), for example, in the one-particle approximation, and write it in the form

$$j^{\nu}_{fi}(x) = e\bar{\psi}_f(x)\gamma^{\nu}\psi_i(x),$$

where e is the electron charge, γ^{ν} are the Dirac matrices, and $\psi(x)$ are the operators of Dirac spinors. On the other hand, this can also be the nuclear current describing a certain collective transition. In any case, $j^{\nu}_{fi}(x)$ is an extraneous current [41] whose properties do not depend significantly on the parameters of the medium in the given approximation.

In the interaction representation, the operators $\psi(x)$ and $A_{\nu}(x)$ of the electron and photon fields appearing in (2) satisfy the same equations of motion and the same permutation relations as the operators of free (not interacting with one another) electron and photon fields in the Heisenberg representation [39, 40]. Hence we follow the standard procedure for deriving an equation for the electromagnetic field operator to be used for constructing the Lagrangian $\mathcal{L}_{\text{int}}(x)$ in (2). We first use the Maxwell equation to derive the classical equation of motion for the vector potential of an electromagnetic field in a medium in the presence of extraneous currents, and construct the Lagrangian of the system. After this, we obtain a homogeneous equation for the operator $A_{\nu}(x)$ using the formal coincidence of the classical equations with the equations of motion for the field operators in the Heisenberg representation. The solution of this equation will depend on the given properties of the spatial region of field propagation. This could be the above-mentioned periodicity, resulting in the formation of an optical band structure [26, 27], or the boundaries of the region at which a certain relation between the wavelength and the size of the region can lead to effects such as a sharp deceleration of the decay or a considerable increase in the probability of spontaneous emission [42–44], as well as some other effects. In the present work, we will be interested in the dependence of $A_{\nu}(x)$ on ϵ and μ, i.e., on the electronic properties of the medium itself.

The evolution of the vector of state of the system in the interaction representation is determined by the S-matrix. When written in the general form in terms of the T-exponent [39, 40], i.e.,

$$S = T\exp\left(i\int \mathcal{L}_{\text{int}}(x)d^4x\right),$$

the S-matrix can be expanded into a series in perturbation theory. The first-order element (1) of the S-matrix describes the emission (absorption) of photons by a bounded system.

The probability of the emission of a photon per unit time can be calculated by using the formula [45]

$$W = \frac{1}{2J_i + 1}\sum_{M_i, M_f, \lambda}\int \frac{|S^{(1)}|^2}{t}\frac{d^3k}{(2\pi)^3}, \tag{4}$$

where we have carried out averaging and summation respectively over the initial and final states of the emitting system, as well as summation over polarizations and integration with respect to the photon momentum. In the above equation, t is a large but finite time interval. It cancels out with the analogous time interval appearing in the numerator during the computation of $|S^{(1)}|^2$ [45].

Thus, the electromagnetic field, which is connected with the transition current in the initial equations of motion in the Heisenberg representation, is transformed in the interaction representation into the free field $A_{\nu}(x)$ satisfying the homogeneous equation of motion. Here, the field $A_{\nu}(x)$ carries complete information about the electronic properties of the medium. Moreover, the emission probability is also found to depend on these properties through formulas (1)–(4).

The above explanation is probably the simplest of all those proposed so far in the publications. Nevertheless, this explanation is quite sufficient to reveal, even before the derivation of any formulas, that the spontaneous emission probability depends on ϵ and μ. Almost every textbook on quantum electrodynamics contains such an explanation in its introductory chapters in one form or another. The quantitative laws governing the process of photon emission by a bounded system were calculated using the approximation of noninteracting fields even in the first order of the perturbation theory for QED. On the other hand, it has been well known for a long time that the components of a free electromagnetic field in a condensed medium differ from those in a vacuum [41]. Hence, in view of formulas (2) and (4), we can expect the dependence of the emission probability on ϵ and μ beforehand, even as a result of calculations of the first-order diagram. Higher orders of perturbation theory lead to corrections that are important for studying various types of fine effects, but are insignificant for the case being considered here.

3. ELECTROMAGNETIC FIELD IN A MEDIUM

The equation

$$\Delta\mathbf{A}(t, \mathbf{r}) - \epsilon\mu\partial_t^2\mathbf{A}(t, \mathbf{r}) = -\mu\mathbf{j} \tag{5}$$

for the vector part of the electromagnetic field potential $\mathbf{A}(t, \mathbf{r})$ follows directly from Maxwell's equations for the electromagnetic filed in a homogeneous medium with permittivity ϵ and permeability μ in the presence of a current $\mathbf{j}$ that is extraneous with respect to the medium [41]. Equation (5) was derived by using the standard definition of the electric field and magnetic induction in terms of $A_\nu = (A_0, \mathbf{A})$:

$$\mathbf{E} = -\partial_t \mathbf{A} - \mathrm{grad} A_0, \quad \mathbf{B} = \mathrm{curl}\mathbf{A}, \tag{6}$$

under the condition $A_0 = 0$ and the Coulomb calibration $\mathrm{div}\,\mathbf{A} = 0$. Here the electric and magnetic inductions $\mathbf{D}$ and $\mathbf{B}$ are defined as $\mathbf{D} = \epsilon\mathbf{E}$ and $\mathbf{B} = \mu\mathbf{H}$.

It can be verified easily as a result of direct computation that Eq. (5) follows from the Lagrangian (see, for example, [23])

$$\mathscr{L} = \frac{1}{2}\left[\epsilon(\partial_t \mathbf{A})^2 - \frac{1}{\mu}(\mathrm{curl}\mathbf{A})^2\right] + \mathbf{j}\cdot\mathbf{A}.$$

The term $\mathbf{j}\cdot\mathbf{A}$ describes the interaction of the extraneous current with the electromagnetic field in the medium. In the potential calibration used here and in the absence of extraneous charges, this expression coincides with the interaction Lagrangian in formula (2).

Pursuing the approach outlined above, we now consider the equation for a free electromagnetic field in a medium. This equation coincides with Eq. (5) except for the right-hand side. The classical solution of such a homogeneous equation in the form of an expansion in plane waves is quantized by replacing the Fourier coefficients by the photon creation and annihilation operators $\hat{a}^+_{\mathbf{k},\lambda}$ and $\hat{a}_{\mathbf{k},\lambda}$, for which the standard permutation relations must be obeyed. As a result, we can present the vector potential in the form

$$\hat{\mathbf{A}}(\mathbf{r}, t) = \sum_{\mathbf{k}}\sum_{\lambda=1,2}[\hat{a}_{\mathbf{k},\lambda}\mathbf{A}_{\mathbf{k},\lambda}(\mathbf{r})e^{-i\omega t} + \hat{a}^+_{\mathbf{k},\lambda}\mathbf{A}^*_{\mathbf{k},\lambda}(\mathbf{r})e^{-i\omega t}], \tag{7}$$

where the functions

$$\mathbf{A}_{\mathbf{k},\lambda}(\mathbf{r}) = \mathbf{e}_{\mathbf{k},\lambda}\sqrt{\frac{2\pi}{\epsilon\omega}}e^{i\mathbf{k}\cdot\mathbf{r}} \tag{8}$$

constitute the fundamental set of solutions of the Helmholtz vector equation

$$\Delta\mathbf{A}(\mathbf{r}) + k^2\mathbf{A}(\mathbf{r}) = 0 \tag{9}$$

with the supplementary condition $\mathrm{div}\,\mathbf{A} = 0$. In Eq. (9), we have introduced the notation

$$k^2 = \epsilon\mu\omega^2. \tag{10}$$

Momentum $\mathbf{k}$, used above for describing plane waves, is associated with parameter $\mathbf{k}$ introduced here through the relation $\mathbf{k} = k\mathbf{n}_\mathbf{k}$, where $\mathbf{n}_\mathbf{k}$ is the unit vector along the direction of $\mathbf{k}$. In Eq. (8), $\mathbf{e}_{\mathbf{k},\lambda}$ is the unit polarization vector of the plane wave, and $\mathbf{n}_\mathbf{k}\cdot\mathbf{e}_{\mathbf{k},\lambda} = 0$. In Eq. (7), $\sum_{\lambda=1,2}$ indicates summation over two photon polarizations. The normalized volume is taken as equal to unity.

The factor $\sqrt{2\pi/\epsilon\omega}$ in the plane wave (8) can be obtained during the calculation of the electromagnetic field energy

$$\mathscr{E} = \frac{1}{8\pi}\int(\mathbf{E}\mathbf{D} + \mathbf{H}\mathbf{B})d^3r$$

in the medium with the help of formula (6) and the classical solution of Eq. (5) with the right-hand side equal to zero.

The emitter wave functions in the expression for the transition current $j^\nu_{fi}(x)$ in formula (2) are bound states. Such wave functions are not eigenfunctions of the momentum operator. The quantum number being conserved is the square of the angular momentum. Hence, in order to calculate the corresponding matrix elements, we expand the plane wave (8) in multipoles [46]:

$$\mathbf{A}_{\mathbf{k},\lambda}(\mathbf{r}) = \sqrt{\frac{2\pi}{\epsilon\omega}}\lambda\sum_{Lm}\sqrt{2\pi(2L+1)}i^L \times D^L_{m\lambda}(\varphi_\mathbf{k}, \vartheta_\mathbf{k}, 0)[\mathbf{A}^M_{Lm}(k, \mathbf{r}) + i\lambda\mathbf{A}^E_{Lm}(k, \mathbf{r})]. \tag{11}$$

The functions $\mathbf{A}^E_{Lm}(k, \mathbf{r})$ and $\mathbf{A}^M_{Lm}(k, \mathbf{r})$ are called the electric and magnetic multipoles respectively and, together with the longitudinal multipole $\mathbf{A}^Y_{Lm}(k, \mathbf{r})$, constitute another fundamental set of solutions of Eq. (9). In Eq. (11), $D^L_{m\lambda}(\varphi_\mathbf{k}, \vartheta_\mathbf{k}, 0)$ is Wigner's D-function, and $\vartheta_\mathbf{k}$ and $\varphi_\mathbf{k}$ are the angles defining the direction of $\mathbf{k}$ in the given coordinate system.

The explicit form of the fields $\mathbf{A}^{E,M,Y}_{Lm}(k, \mathbf{r})$ is known and can be written as follows:

$$\mathbf{A}^M_{Lm}(k, \mathbf{r}) = j_L(kr)\mathbf{Y}_{LL;m}(\mathbf{n}_\mathbf{r}), \tag{12}$$

$$\mathbf{A}^E_{Lm}(k, \mathbf{r}) = \sqrt{\frac{L+1}{2L+1}}j_{L-1}(kr)\mathbf{Y}_{LL-1;m}(\mathbf{n}_\mathbf{r}) - \sqrt{\frac{L}{2L+1}}j_{L+1}(kr)\mathbf{Y}_{LL+1;m}(\mathbf{n}_\mathbf{r}), \tag{13}$$

$$\mathbf{A}^Y_{Lm}(k, \mathbf{r}) = \sqrt{\frac{L}{2L+1}}j_{L-1}(kr)\mathbf{Y}_{LL-1;m}(\mathbf{n}_\mathbf{r}) + \sqrt{\frac{L+1}{2L+1}}j_{L+1}(kr)\mathbf{Y}_{LL+1;m}(\mathbf{n}_\mathbf{r}). \tag{14}$$

Here, $j_L(kr)$ are spherical Bessel functions [47], and $\mathbf{Y}_{JL;\,m}(\mathbf{n_r})$ are spherical vector harmonics defined by the relations [46]

$$\mathbf{Y}_{JL;\,m}(\mathbf{n_r}) = \sum_{m'\lambda} C^{Jm}_{Lm'1\lambda} Y_{Lm'}(\mathbf{n_r})\mathbf{e}_\lambda,$$

where $C^{Jm}_{Lm'1\lambda}$ are the Clebsch–Gordan coefficients and $Y_{Lm}(\mathbf{n_r})$ are spherical functions.

The potentials $\mathbf{A}^{E,M}_{Lm}(k, \mathbf{r})$ appearing in the expansion (11) of a plane wave satisfy the condition [46]

$$\mathrm{div}\mathbf{A}^{E,M}_{Lm}(k, \mathbf{r}) = 0.$$

The longitudinal potential $\mathbf{A}^{Y}_{Lm}(k, \mathbf{r})$ in (14) is calculated with the help of the gradient formula [46] proceeding from the definition

$$\mathbf{A}^{Y}_{Lm}(k, \mathbf{r}) = \frac{1}{k}\mathrm{grad}[j_L(kr)Y_{Lm}(\mathbf{n_r})]. \tag{15}$$

Hence the transversality condition is not satisfied for this potential [46].

In the problem under consideration, the properties of the extraneous current $j^{\nu}_{fi}(t, \mathbf{r})$ depend weakly on the parameters of the medium. Hence, as will be shown below, the corresponding matrix elements in the long-wave approximation can be expressed in terms of the matrix elements of the transition in a vacuum. It is not necessary to specify the origin (atomic, nuclear, etc.) of the transition current inducing the emission of radiation. The only important thing is that, like the electromagnetic current, this current must also satisfy the continuity equation

$$\mathrm{div}\,\mathbf{j}_{fi}(\mathbf{r}) = i\omega_{fi} j^0_{fi}(\mathbf{r}), \tag{16}$$

where ω_{fi} is the transition energy.

4. ELECTRIC-TYPE RADIATION

The interaction Hamiltonian for *EL*-radiation can be obtained from formulas (2), (3) by substituting expressions (7) and (11) into them:

$$H^{EL}_{\mathrm{int}}(t) = e^{i(\omega-\omega_{fi})t}\sqrt{\frac{2\pi}{\epsilon\omega}}\sqrt{2\pi(2L+1)}(-i)^{L+1} \times \sum_m D^{L*}_{m\lambda}(\varphi_{\mathbf{k}}, \vartheta_{\mathbf{k}}, 0)\int \mathbf{A}^{E*}_{Lm}(k, \mathbf{r})\mathbf{j}_{fi}(\mathbf{r})d^3r. \tag{17}$$

Using the long-wave approximation $kr \ll 1$, which is obviously valid in the optical region of photon energy, we transform the part of the Hamiltonian in the integrand of (17). For this purpose, we discard in formulas (13) and (14) the terms containing $j_{L+1}(kr)$, which are small compared to the terms with $j_{L-1}(kr)$ in view of the familiar behavior of the Bessel functions for $kr \ll 1$ [47]. Further, taking into account formula (15), we obtain for the electric potential the expression

$$\mathbf{A}^{E}_{Lm}(k, \mathbf{r}) \approx \sqrt{\frac{L+1}{L}}\mathbf{A}^{Y}_{Lm}(k, \mathbf{r}) = \sqrt{\frac{L+1}{L}}\frac{1}{k}\mathrm{grad}[j_L(kr)Y_{Lm}(\mathbf{n_r})]. \tag{18}$$

Using expressions (18) and (16) and integrating by parts, we can easily transform the part of the Hamiltonian under consideration to:

$$\int \mathbf{A}^{E*}_{Lm}(k, \mathbf{r})\mathbf{j}_{fi}(\mathbf{r})d^3r \approx -i\frac{\omega}{k}\sqrt{\frac{L+1}{L}}\int j_L(kr)Y_{Lm}(\mathbf{n_r})j^0_{fi}(\mathbf{r})d^3r. \tag{19}$$

Further, expanding the Bessel function in (19) [47]

$$j_L(kr) \approx \frac{(kr)^L}{(2L+1)!!}$$

and substituting the obtained result into (17), we obtain a relation between the interaction Hamiltonian $(H^{EL}_{\mathrm{int}})_m$ in the medium and the interaction Hamiltonian $(H^{EL}_{\mathrm{int}})_{\mathrm{vac}}$ in a vacuum. Taking formula (10) into account, we obtain

$$(H^{EL}_{\mathrm{int}})_m = \epsilon^{(L-3)/2}\mu^{(L-1)/2}(H^{EL}_{\mathrm{int}})_{\mathrm{vac}}.$$

The phase volume also depends on ϵ and μ. The expression for d^3k in Eq. (4) leads to the multiplier $\epsilon^{3/2}\mu^{3/2}$ in contrast to the analogous expression for a vacuum.

Let us formally take into account the effect of the local field mentioned in the Introduction. For this purpose, we express the local electric field $(\mathbf{E}_L)_{\mathrm{loc}}$ acting on the emitting object in terms of the mean field $(\mathbf{E}_L)_m$ in the medium through the relation

$$(\mathbf{E}_L)_{\mathrm{loc}} = f_L(\epsilon)(\mathbf{E}_L)_m, \tag{20}$$

without specifying the explicit form of the function $f_L(\epsilon)$ for the time being (see below for the exact derivation of this function).

Let us substitute the obtained expression into (4). The final formula connecting the probability of spontaneous decay with the emission of an *EL*-type photon in the medium and in a vacuum can be presented in the form

$$W^{EL}_m = f^2_L(\epsilon)\epsilon^{L-1/2}\mu^{L+1/2}W^{EL}_{\mathrm{vac}}. \tag{21}$$

5. MAGNETIC-TYPE RADIATION

The interaction Hamiltonian for the *ML*-radiation can be constructed in analogy with $H_{\rm int}^{EL}(t)$. Carrying out appropriate substitutions, we obtain

$$H_{\rm int}^{ML}(t) = \exp[i(\omega-\omega_{fi})t]\sqrt{\frac{2\pi}{\epsilon\omega}}\sqrt{2\pi(2L+1)}$$

$$\times\lambda(-i)^L\sum_m D_{m\lambda}^{L*}(\varphi_{\mathbf{k}},\vartheta_{\mathbf{k}},0)\int\mathbf{A}_{Lm}^{M*}(k,\mathbf{r})\mathbf{j}_{fi}(\mathbf{r})d^3r.$$

We use formula (12) for the vector potential $\mathbf{A}_{Lm}^{M}(k,\mathbf{r})$. Expanding the Bessel function $j_L(kr)$ in the small argument, we obtain in the long-wave approximation a relation between the interaction Hamiltonians in the medium and in a vacuum:

$$(H_{\rm int}^{ML})_m = \epsilon^{(L-1)/2}\mu^{L/2}(H_{\rm int}^{ML})_{\rm vac}.$$

Further, using formula (4) for the emission probability, we express the probability of spontaneous decay with the emission of an *ML*-type photon in the medium in terms of the probability of emission in a vacuum:

$$W_m^{ML} = f_L^2(\mu)\epsilon^{L+1/2}\mu^{L+3/2}W_{\rm vac}^{ML}. \qquad (22)$$

The function $f_L(\mu)$ for the magnetic medium introduced in this expression connects the local magnetic field with the mean magnetic field in the medium, and is in complete analogy with the function $f_L(\epsilon)$ for a dielectric.

Formulas (21) and (22) are not symmetric relative to the substitution $\epsilon \longleftrightarrow \mu$ and to the multipolarities of the transition $E \longleftrightarrow M$. This is due to the well-known fact that the magnetic induction $\mathbf{B}_m$, and not the magnetic field strength $\mathbf{H}_m$, is the analog of the electric field strength $\mathbf{E}_m$ in the medium [41].

The expression obtained from (22) for the probability of the spontaneous emission of magnetic dipole radiation coincides with the results obtained in [11], and is transformed into the expression for *M*1-radiation in a dielectric medium derived in [34], but differs from the expression obtained by Glauber and Lewenstein [3]. The reasons behind the discrepancy with the results of [3], where the dependence $W_m^{M1}/W_{\rm vac}^{M1} = \epsilon^{1/2}$ instead of $\epsilon^{3/2}$ was obtained for a dielectric, are discussed in detail in [34].

6. LOCAL FIELD IN THE REAL AND VIRTUAL CAVITY MODELS

The local field, i.e., the field acting on the emitting object in a condensed medium, differs in magnitude from the field of an electromagnetic wave propagating in the given medium [1, 2]. The concept of the local field and its connection with the electric field in a dielectric was touched upon to a certain extent, or served as the main subject of investigations, in a number of theoretical works {1–5, 7–10, 13–15, 18, 22, 48, 49].

In the computations of transition probabilities or dipole–dipole interaction energies in a medium, the effect of the local field is usually taken into consideration through formula (20). Recent experimental investigations [1, 2] have shown that the results of the measurements of *E*1-transition probabilities in atoms and ions can be best described through the function

$$f(\epsilon) = 3\epsilon/(2\epsilon+1).$$

An exactly identical function $f(\epsilon)$ is obtained in the real cavity model [3, 41]. It was found that $f(\epsilon) = 1$ for *M*1-transition in a dielectric [1]. This result is quite natural: the magnetic component of the electromagnetic field is not renormalized by the dielectric medium. In this connection, it should be remarked that the function $f_L(\epsilon)$ for electric-type transitions may not depend on the permeability μ of the medium. Conversely, the function $f_L(\mu)$ for magnetic-type transitions may not depend on ϵ.

Thus, let us assume that the emitting object is located inside a real spherical cavity of a small radius, built in a dielectric. The inside of the cavity is evacuated, and the radius of the sphere is large compared to the linear size of the emitting object, but much smaller than the radiation wavelength.

In the real cavity model, it is assumed that electromagnetic fields satisfying Maxwell's equations and quantized in accordance with the general rules exist inside and outside the cavity. The solution of the local field problem is obtained by joining the corresponding components of the field at the boundary of the cavity and the medium. The resulting steady-state field in the medium differs from the mean electromagnetic field existing before the formation of the cavity (i.e., the initial field in the medium varies in the course of cavity formation).

Let us define the function $f_L(\epsilon)$ for electric-type radiation of an arbitrary polarity. The solution of the Helmholtz equation (9) outside the spherical cavity can be written in the form

$$\mathbf{A}(\mathbf{r}) = \mathbf{A}_{Lm}^{E}(k,\mathbf{r}) + b\mathbf{B}_{Lm}^{E}(k,\mathbf{r}),$$

where the vector potential $\mathbf{A}_{Lm}^{E}(k,\mathbf{r})$ corresponds to the initial unperturbed field in the medium, and the potential $\mathbf{B}_{Lm}^{E}(k,\mathbf{r})$ describes the variation introduced by the small cavity. The vector potential $\mathbf{B}_{Lm}^{E}(k,\mathbf{r})$ differs from $\mathbf{A}_{Lm}^{E}(k,\mathbf{r})$ in Eq. (13) in the replacement of the spherical Bessel function $j_L(kr)$ of the first kind by the spherical Neumann function $n_L(kr)$ (which is a Bessel function of the second kind) [47]. The momentum k is defined, as before, by formula (10).

We shall seek the solution of Eq. (9) inside the evacuated spherical cavity in the form

$$\mathbf{A}(\mathbf{r}) = a\mathbf{A}_{Lm}^{E}(p,\mathbf{r}),$$

where the photon momentum $|\mathbf{p}| \equiv p = \omega$.

The electric field of a multipole is defined in accordance with Eqs. (6) as

$$\mathbf{E}^E_{Lm}(t, k, \mathbf{r}) = -\partial_t \exp(-i\omega t)\mathbf{A}^E_{Lm}(k, \mathbf{r}),$$

which leads to the expression [46]

$$\mathbf{E}^E_{Lm}(k, \mathbf{r}) = i\omega \mathbf{A}^E_{Lm}(k, \mathbf{r}). \tag{23}$$

According to the conditions of the problem, the entire spherical cavity is situated in the near (or static) radiation band. Taking formulas (18) and (23) into consideration, we can introduce the potential $\varphi^E_{Lm}(k, \mathbf{r})$ so that the following condition is satisfied:

$$\mathbf{E}^E_{Lm}(k, \mathbf{r}) = -\mathrm{grad}\varphi^E_{Lm}(k, \mathbf{r}).$$

The potentials inside and outside the cavity have the form

$$\varphi^E_{Lm}(k, \mathbf{r}) = -i\omega\sqrt{\frac{L+1}{L}}\frac{1}{k}Y_{Lm}(\mathbf{n_r})[j_L(kr) + bn_L(kr)], \quad r \geq R,$$

$$\varphi^E_{Lm}(p, \mathbf{r}) = -i\omega\sqrt{\frac{L+1}{L}}\frac{1}{p}Y_{Lm}(\mathbf{n_r})[aj_L(pr)], \quad r \leq R.$$

The first equation for coefficients a and b follows from the condition of the continuity of potential at the boundary of the region [41]:

$$\varphi^E_{Lm}(k, R) = \varphi^E_{Lm}(p, R).$$

Expanding the Bessel and Neumann functions in the small parameter, we obtain

$$a = (\epsilon\mu)^{(L-1)/2} - \frac{b}{(\omega R)^{2L+1}}\frac{(2L-1)!!(2L+1)!!}{(\epsilon\mu)^{(L+2)/2}}. \tag{24}$$

Apart from the potential, the radial component of the electric induction vector is also continuous at the boundary [41]:

$$(\mathbf{D}^E_{Lm}(k, \mathbf{r}))_r = \epsilon(\mathbf{E}^E_{Lm}(k, \mathbf{r}))_r = -\epsilon\partial_r\varphi^E_{Lm}(k, \mathbf{r}).$$

Using the recurrent relations for the derivatives of the spherical Bessel functions [47], i.e.,

$$(2L+1)\frac{d\phi_L(x)}{dx} = L\phi_{L-1}(x) - (L+1)\phi_{L+1}(x),$$

and neglecting $j_{L+1}(x)$ and $n_{L-1}(x)$ in comparison with $j_{L-1}(x)$ and $n_{L+1}(x)$ respectively, we arrive at the following expressions for the radial components of the electric induction vector:

$$(\mathbf{D}^E_{Lm}(k, \mathbf{r}))_r = i\epsilon\omega\sqrt{\frac{L+1}{L}}Y_{Lm}(\mathbf{n_r}) \times\left[\frac{L}{2L+1}j_{L-1}(kr) - b\frac{L+1}{2L+1}n_{L+1}(kr)\right], \quad r \geq R, \tag{25}$$

$$(\mathbf{D}^E_{Lm}(p, \mathbf{r}))_r = i\omega\sqrt{\frac{L+1}{L}}Y_{Lm}(\mathbf{n_r})\left[a\frac{L}{2L+1}j_{L-1}(pr)\right], \quad r \geq R. \tag{26}$$

Resorting once again to the expansion of $j_{L-1}(x)$ and $n_{L+1}(x)$ in (25) and (26), we obtain the second equation for coefficients a and b:

$$a = \frac{1}{\mu}\left[(\epsilon\mu)^{(L+1)/2} + \frac{b}{(\omega R)^{2L+1}} \times\frac{L+1}{L}\frac{(2L-1)!!(2L+1)!!}{(\epsilon\mu)^{L/2}}\right]. \tag{27}$$

The solution of the joining equations (24) and (27) gives

$$a = (\epsilon\mu)^{(L-1)/2}\frac{\epsilon(2L+1)}{\epsilon(L+1)+L}. \tag{28}$$

In the near radiation band, the potential and the electric field strength of the unperturbed field (i.e., in the absence of a cavity) in the medium have the form

$$\tilde{\varphi}^E_{Lm}(k, \mathbf{r}) = -i\sqrt{\frac{L+1}{L}}Y_{Lm}(\mathbf{n_r})\frac{(\omega r)^L}{(2L+1)!!}(\epsilon\mu)^{(L-1)/2},$$

$$\tilde{\mathbf{E}}^E_{Lm}(k, \mathbf{r}) = i\omega^L\sqrt{\frac{L+1}{L}} \times\frac{(\epsilon\mu)^{(L-1)/2}}{(2L+1)!!}\mathrm{grad}[r^L Y_{Lm}(\mathbf{n_r})]. \tag{29}$$

Comparing formula (29) for $\tilde{\mathbf{E}}^E_{Lm}(k, \mathbf{r})$ with the expression for the electric field in the cavity

$$\mathbf{E}^E_{Lm}(p, \mathbf{r}) = i\omega^L\sqrt{\frac{L+1}{L}}\frac{a}{(2L+1)!!}\mathrm{grad}[r^L Y_{Lm}(\mathbf{n_r})],$$

we obtain the following relation connecting these two quantities:

$$\mathbf{E}^E_{Lm}(p, \mathbf{r}) = \frac{a}{(\epsilon\mu)^{(L-1)/2}}\tilde{\mathbf{E}}^E_{Lm}(k, \mathbf{r}).$$

Using (28), we obtain from here the relation

$$\mathbf{E}^E_{Lm}(p, \mathbf{r}) = f_L(\epsilon)\tilde{\mathbf{E}}^E_{Lm}(k, \mathbf{r}),$$

where

$$f_L(\epsilon) = \frac{\epsilon(2L+1)}{\epsilon(L+1)+L}. \tag{30}$$

Obviously, the field $\mathbf{E}^E_{Lm}(p, \mathbf{r})$ inside the sphere is just the local field interacting with the transition current. It can easily be seen that the function $f_1(\epsilon)$ from Eq. (30) coincides with the expression for $f(\epsilon)$ which was presented at the beginning of this section and whose validity has been confirmed experimentally [1, 2]. In conclusion, let us make a remark concerning the derivation of formula (30). While calculating the fields, we used the continuity of the potential and not of the tangential components of the electric field at the boundary between the medium and the region. This was done intentionally. The problem of a dielectric sphere in a constant external electric field was discussed in [41] (see Chapter 2). The result (30) can be treated as a generalization of the above-mentioned problem to a field of an arbitrary configuration. (Indeed, solutions (12)–(14) form a complete set and can be used to simulate any field.) Hence we followed a line of reasoning as close as possible to that adopted by Landau and Lifshitz [41]. It should also be noted that if the sphere were filled with a dielectric medium of permittivity ϵ', formula (30) would assume the form

$$f_L(\epsilon, \epsilon') = \frac{\epsilon(2L+1)}{\epsilon(L+1)+\epsilon' L}.$$

The function $f_L(\mu)$, which connects the magnetic field in the medium having a permeability μ differing from unity with the field in a nonmagnetic real spherical cavity, is calculated in a manner analogous with that described in [41] and has the form

$$f_L(\mu) = \frac{\mu(2L+1)}{\mu(L+1)+L}. \quad (31)$$

Let us now briefly consider the model of a "virtual" cavity. This model differs from the one considered above in that the formation of a cavity in a dielectric medium in which a mean field $\mathbf{E}_m$ already exists does not lead to a change in the field $\mathbf{E}_m$ outside the cavity.

Various methods are known for calculating the function $f_1(\epsilon)$ in the framework of the virtual cavity model [5, 7]. For example, the method of "auxiliary" configuration is used in [5] (see Chapter 2): the field inside the cavity is calculated with the help of an auxiliary potential generated in a vacuum by a polarized dielectric sphere. The expression for the local field in the case of a uniform polarization $\mathbf{P}_m$ presented in [5] has the form

$$\mathbf{E}_{\text{loc}} = \mathbf{E}_m + (4\pi/3)\mathbf{P}_m.$$

This expression can be easily generalized to the case of an arbitrary polarization by using the method described in [5]. In terms of the function $f^{\text{virt}}_L(\epsilon)$ sought by us, we can write

$$f^{\text{virt}}_L(\epsilon) = \frac{\epsilon L + L + 1}{2L+1}.$$

Apparently, this formula is transformed into the familiar expression for the function $f^{\text{virt}}_1(\epsilon)$ in the virtual cavity model [7–10].

7. CONCLUSION

It is well known that $E1$-transitions and, in particular, electric dipole radiation, dominate in atoms, ions and molecules. $M1$-transitions are rarely encountered in such systems, and considerable efforts are needed for their investigation. It is even more difficult to create conditions conducive for observing, say, $E2$- and $E3$-transitions. However, such studies are being carried out at present. An $E2$-transition with a wavelength of 411 nm was observed in $^{172}Yb^+$ ions in 1995 [36]. A few years later, an $E3$-transition with a wavelength of 467 nm was observed in $^{171, 172}Yb^+$ ions [37, 38]. Hence, even though a verification of expressions (21) and (22) for the EL- and ML-transition widths, as well as formulas (30) and (31) for the functions $f_L(\epsilon)$ and $f_L(\mu)$, may not be an easy task, it cannot be given up as futile. An investigation of optical transitions (including spontaneous emission) in magnetic materials and, in particular, in magnetic dielectrics, may turn out to be another prospective direction.

Multipole radiation exists in atomic nuclei. However, the energies of nuclear transitions are as a rule so high that it is impossible to establish a relation between the decay probability and the electronic properties of the medium. However, there does exist an exception, viz., the optical transition in the nucleus ^{229}Th. The energy of isomeric transition between the first excited level and the ground state of the nucleus ^{229}Th lies in the range 3.5 ± 1.0 eV [50], and the dominant multipolarity of radiation is $M1$ (the probability of $E2$-radiation is about eleven orders of magnitude lower). The possibility of detecting the n^3-dependence of the decay probability of the above-mentioned low-lying isomer $^{229}Th^m$ ($3/2^+$, 3.5 ± 1.0 eV) in the dielectric $^{229}ThO_2$ was discussed in [34, 35]. The refractive index of thorium dioxide is $n = 2$ [51] for photons with energies $\omega = 3.1$ eV. The anticipated half-life of the isometric state lies in the interval between 10 min and one hour, depending on the wavelength [34, 35]. In the framework of the approximations used in this work, there is apparently no fundamental difference between this radiation and the radiation emitted by an atom in a transparent dielectric medium. Nevertheless, a comparison with between the local fields acting on an atom and a nucleus may even lead to unexpected results.

ACKNOWLEDGMENTS

The author is grateful to A.O. Barvinskiĭ, A.M. Dykhne, A.N. Zherikhin, and M.A. Listengarten for fruitful consultations and helpful remarks.

This research was partially supported by the Russian Foundation for Basic Research (project no. 98-02-

16070a) and a grant supporting the leading research schools (grant no. 00-15-96651).

REFERENCES

1. G. L. J. A. Rikken and Y. A. R. R. Kessener, Phys. Rev. Lett. **74**, 880 (1995).
2. F. J. P. Schuurmans, D. T. N. de Lang, G. H. Wegdam, *et al.*, Phys. Rev. Lett. **80**, 5077 (1998).
3. R. J. Glauber and M. Lewenstein, Phys. Rev. A **43**, 467 (1991).
4. H. A. Lorentz, *The Theory of Electrons and Its Applications to the Phenomena of Light and Radiant Heat* (Dover, New York, 1952; Gostekhizdat, Moscow, 1953).
5. M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford, 1969; Nauka, Moscow 1973).
6. J. J. Maki, M. S. Malcuit, J. E. Sipe, and R. W. Boyd, Phys. Rev. Lett. **67**, 972 (1991).
7. S. M. Barnett, B. Huttner, R. Loudon, and R. Matloob, J. Phys. B **29**, 3763 (1996).
8. P. de Vries and A. Lagendijk, Phys. Rev. Lett. **81**, 1381 (1998).
9. S. Schell, L. Knöll, and D.-G. Welsch, Phys. Rev. A **60**, 4094 (1999).
10. S. Schell, L. Knöll, and D.-G. Welsch, Phys. Rev. A **60**, 1590 (1999).
11. G. Nienhuis and C. Th. J. Alkemade, Physica C (Amsterdam) **81**, 181 (1976).
12. L. A. Dissado, J. Phys. C **3**, 94 (1970).
13. E. Yablonovitch, T. J. Gmitter, and R. Bhat, Phys. Rev. Lett. **61**, 2546 (1988).
14. J. Knoester and S. Mukamel, Phys. Rev. A **40**, 7065 (1989).
15. S. M. Barnett, B. Huttner, and R. Loudon, Phys. Rev. Lett. **68**, 3698 (1992).
16. B. Huttner and S. M. Barnett, Phys. Rev. A **46**, 4306 (1992).
17. S. T. Ho and P. Kumar, J. Opt. Soc. Am. B **10**, 1620 (1993).
18. G. Juzeliūnas and D. L. Andrews, Phys. Rev. B **49**, 8751 (1994).
19. R. Matloob, R. Loudon, S. M. Barnett, and J. Jeffers, Phys. Rev. A **52**, 4823 (1995).
20. T. Gruner and D.-G. Welsch, Phys. Rev. A **53**, 1818 (1996).
21. R. Matloob and R. Loudon, Phys. Rev. A **53**, 4567 (1996).
22. G. Juzeliūnas, Phys. Rev. A **55**, R4015 (1997).
23. A. Tip, Phys. Rev. A **56**, 5022 (1997).
24. H. T. Dung, L. Knöll, and D.-G. Welsch, Phys. Rev. A **57**, 3931 (1998).
25. S. Schell, L. Knöll, and D.-G. Welsch, Phys. Rev. A **58**, 700 (1998).
26. E. Yablonovitch, Phys. Rev. Lett. **58**, 2059 (1987).
27. S. John, Phys. Rev. Lett. **58**, 2486 (1987).
28. E. Yablonovitch and T. J. Gmitter, Phys. Rev. Lett. **63**, 1950 (1989).
29. G. Kurizki, Phys. Rev. A **42**, 2915 (1990).
30. E. Yablonovitch, T. J. Gmitter, and K. M. Leung, Phys. Rev. Lett. **67**, 2295 (1991).
31. E. P. Petrov, V. N. Bogomolov, I. I. Kalosha, and S. V. Gaponenko, Phys. Rev. Lett. **81**, 77 (1998).
32. S.-Y. Zhu, Y. Yang, H. Chen, *et al.*, Phys. Rev. Lett. **84**, 2136 (2000).
33. Z.-Y. Li, L.-L. Lin, and Z.-Q. Zhang, Phys. Rev. Lett. **84**, 4341 (2000).
34. E. V. Tkalya, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 449 (2000) [JETP Lett. **71**, 311 (2000)].
35. E. V. Tkalya, A. N. Zherikhin, and V. I. Zhudov, Phys. Rev. C **61**, 064308 (2000).
36. P. Gill, H. A. Klein, A. P. Levick, *et al.*, Phys. Rev. A **52**, R909 (1995).
37. M. Roberts, P. Taylor, G. P. Barwood, *et al.*, Phys. Rev. Lett. **78**, 1876 (1997).
38. M. Roberts, P. Taylor, G. P. Barwood, *et al.*, Phys. Rev. A **62**, 020501 (2000).
39. A. I. Akhiezer and V. B. Berestetskii, *Quantum Electrodynamics* (Nauka, Moscow, 1969; Wiley, New York, 1965).
40. N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Nauka, Moscow, 1976; Wiley, New York, 1980).
41. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media* (Nauka, Moscow, 1982; Pergamon, New York, 1984).
42. E. M. Purcell, Phys. Rev. **69**, 681 (1946).
43. P. Goy, J. M. Raimond, M. Gross, and S. Haroche, Phys. Rev. Lett. **50**, 1903 (1983).
44. R. G. Hulet, E. S. Hilfer, and D. Kleppner, Phys. Rev. Lett. **55**, 2137 (1985).
45. V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics* (Nauka, Moscow, 1980; Pergamon, Oxford, 1982).
46. J. M. Eisenberg and W. Greiner, *Nuclear Theory*, Vol. 2: *Excitation Mechanisms of the Nucleus* (North-Holland, Amsterdam, 1970; Atomizdat, Moscow, 1973).
47. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun (Dover, New York, 1971; Nauka, Moscow, 1979).
48. D. L. Dexter, Phys. Rev. **101**, 48 (1956).
49. V. M. Agranovich and M. D. Galanin, *Electronic Excitation Energy Transfer in Condensed Matter* (Nauka, Moscow, 1978; North-Holland, Amsterdam, 1982).
50. R. G. Helmer and C. W. Reich, Phys. Rev. C **49**, 1845 (1994).
51. A. I. Sviridova and I. V. Suĭkovskaya, Opt. Spektrosk. **22**, 940 (1967).

Translated by Ram Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 69–77.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 80–90.

PLASMA, GASES

On the Possibility of Target Plasma Ignition under the Conditions of Inertial Nuclear Fusion

A. A. Andreev[a], S. Yu. Gus'kov[b], D. V. Il'in[c], A. A. Levkovskiĭ[c, *], V. B. Rozanov[b], and V. E. Sherman[c]

[a] *Institute of Laser Physics, Vavilov State Optical Institute, Birzhevaya liniya 12, St. Petersburg, 199034 Russia*
[b] *Lebedev Physical Institute, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 117924 Russia*
[c] *St. Petersburg Institute of Machine Building (Leningrad Machine Building Plant), Engineering College, St. Petersburg, 195108 Russia*
**e-mail: Levkovsk@chat.ru*
Received May 23, 2000

Abstract—The thermonuclear gain G for bulk and spark ignitions are calculated using a mathematical simulation of thermonuclear combustion in a DT plasma of laser targets for various parameters of the target plasma and (isobaric and isochoric) ignitors. The critical parameters of ignitors at which an effective nuclear burst occurs with $G \sim 100$ are calculated. It is shown that a further increase in the temperature and size of the ignitors virtually does not affect the efficiency of DT fuel burnup. Irrespective of the ignition technique, the value of G can be estimated with the help of a simple asymptotic formula. At the same time, the critical parameters of ignitors are determined to a considerable extent by the mode of ignition and by the target parameters. Spark ignition with an isochoric ignitor corresponding to the fast ignition mode is considered in detail. It is shown that the main critical parameter for optimal isochoric ignitors is their thermal energy liberated upon absorption of an auxiliary ultrashort laser pulse. The critical values of this energy are calculated. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The effectiveness of thermonuclear combustion in setups with the inertial confinement of plasma is determined primarily by the values of the temperature T and density ρ attained in the course of target compression. An important role in this case is played not only by the average values of these parameters, but also the temperature and density distribution in the target plasma at the instant of its maximum compression. High-intensity thermonuclear combustion emerging in a comparatively small high-temperature region (ignitor) can heat a region of cold plasma and involve it in the combustion process. The thermonuclear combustion wave generated in the course of this process may lead to thermonuclear burst of the plasma with a relatively low average temperature. This forms the basis of the spark ignition mechanism [1, 2]. The present work aims at the mathematical simulation, classification, and analysis of the effectiveness of plasma burnup in laser targets with various distributions of densities and temperatures at the instant of maximum compression. The results of this research can be used to formulate the requirements imposed on the parameters of laser-induced fusion (LIF) setups and to estimate the applicability and the effectiveness of various methods of target plasma ignition.

In our previous publications [3, 4], the relation between the parameters of the ignitor and the efficiency of combustion of the target was studied using semianalytical methods. We obtained qualitative estimates for the critical parameters of ignitors at which a thermonuclear burst occurs. For the ignition criterion, we used the instantaneous jump in the ignitor temperature T_f [3] or the increase in the ratio T_f/T_0 of the ignitor and the surrounding plasma temperatures [4]. (Here and below, the subscript f marks the parameters of the ignitor and 0 the parameters of the surrounding DT fuel at the instant of maximum compression.)

It was found later that this criterion is unreasonably stringent and reduces significantly the range of the initial states of a plasma leading to an effective thermonuclear burst. The mathematical simulation of the generation of a thermonuclear combustion wave in a plasma, which was carried out by us earlier [5, 6], proved that the evolution of thermonuclear combustion can proceed in two stages. The first, "subsonic" stage is characterized by a decreasing or slowly increasing temperature of the initial ignition region. The ignitor "glows," increasing slowly in size and preparing the initial conditions for the second, "ultrasonic" stage of intense combustion leading to an almost instantaneous thermonuclear burst in the surrounding plasma. For a spark ignition, it suffices to have an ignitor with parameters ensuring that the duration of the subsonic stage is much shorter than the time of thermal disintegration of the plasma. An analysis of the potentialities of such a method of ignition involves the mathematical simula-

tion of the thermonuclear combustion processes in a plasma of laser targets and is one of the goals of this research.

Taking into account the increasing interest in the fast ignition of targets and the potentialities of the available experimental laser fusion setups, we consider here various versions of spark ignition of spherically symmetric targets with a central ignitor. It is well known that the ignitor in a fast ignition mode is created with the help of an auxiliary ultrashort laser pulse over a time shorter than the characteristic time of hydrodynamic processes [7]. The pressure in this case has no time to level out, and the plasma density remains virtually unchanged in the course of creation of the high-temperature region; i.e., the isochoric conditions are met. Here, we consider two limiting cases of spark ignition: with an isobaric ignitor ($p_f = p_0$) emerging in the course of direct hydrodynamic compression of the target [1, 2], and with an isochoric ignitor ($\rho_f = \rho_0$) corresponding to the fast ignition mode.

The calculations of the gain factors for laser targets were made on the basis of the mathematical simulation of thermonuclear combustion from the instant of maximum compression to disintegration with the help of the software package TEPA [5, 6]. Its distinguishing feature is that the method of direct statistical simulation of the kinetics of transfer of energy, momentum, and mass by fast particles, which are the products of primary and secondary nuclear fusion reaction, and recoil nuclei from elastic neutron scattering, as well as by thermal x-rays, are used at each time step of the transient difference algorithm for the solution of hydrodynamic equations for two-temperature plasmas.

The main difficulty encountered in an analysis and classification of the results of calculations of gain factors is associated with a large number of initial parameters characterizing the state of the target plasma. An analysis of characteristic features of the thermonuclear combustion processes in plasmas and the evaluation of the most important variables close to scaling parameters determining the combustion efficiency make it possible to present the result in a generalized form by reducing successively the number of parameters under investigation.

Isobaric (as well as isochoric) ignitors can be characterized by two parameters: the thickness ρR_f and the initial temperature T_f. (Henceforth, the product ρR will be treated as a single physical parameter characterizing the thickness of the spherical target plasma as a measure of its transparency for fusion product particles. In this notation, $\rho R_f = (\rho R)_f$). The results of calculations show that if an effective thermonuclear burst has occurred in the plasma of the DT target, the extent of burnup weakly depends on the ignition method. Thus, the analysis of the combustion efficiency boils down to determining the critical (minimal) values of the ignitor parameters ensuring stable ignition of the target plasma. In the present work, the critical curves determining the boundaries of stable spark ignition regions have been determined for the first time. Calculations for the fast ignition mode demonstrated that the relation between the critical parameters ρR_f and T_f for small isochoric ignitors can be substantiated from the physical point of view and can be expressed in analytic form. In fact, we are dealing with only one critical parameter, viz., the thermal energy of the ignitor acquired as a result of the absorption of an auxiliary laser pulse. The values of the critical energies of ignitors are obtained for target plasmas with various values of ρR_0 and T_0.

It is shown that spark ignition is possible only in targets with a sufficiently large and relatively hot region of the main plasma at the instant of maximum compression. It can be explained by the fact that a comparatively low energy liberated by the ignitor must be sufficient for heating the neighboring layers of the main plasma of the target to fusion temperatures over the time much shorter than the time of the target disintegration. We also determine the boundaries of admissible values of ρR_0 and T_0 of the main plasma at the instant of maximum compression for carrying out isobaric and isochoric spark ignitions.

2. PHYSICAL AND MATHEMATICAL MODEL

2.1. System of Equations

The mathematical model of thermonuclear combustion of an inhomogeneous spherically symmetric plasma is described by a system of equations of continuity, motion, energy transfer, and state of the plasma together with the kinetic equations for fast thermonuclear particles. In order to describe hydro- and thermodynamics, we use the approximation of a one-liquid two-temperature (T_e, T_i) plasma, taking into account the electronic and ionic thermal conductivities and electron–ion energy transfer in the ideal gas approximation.

In terms of the standard variables of the Lagrangean system of coordinate

$$dm = \rho(r)r^2 dr,$$

where $\rho(r)$ is the density of the plasma and $V = 1/\rho$ is its specific volume, the generalized system of equations has the form [5, 6]

$$\frac{\partial V}{\partial t} = \frac{\partial}{\partial m}(r^2 u) + V^2 S,$$

$$\frac{\partial u}{\partial t} = -r^2 \frac{\partial P}{\partial m} + VF,$$

$$c_e \frac{\partial T_e}{\partial t} + P_e\left(\frac{\partial V}{\partial t} - V^2 S\right) + \frac{\partial q_e}{\partial m} + \frac{T_e - T_i}{\rho\tau} = Q_e + Q_f, \quad (1)$$

$$c_i \frac{\partial T_i}{\partial t} + P_i\left(\frac{\partial V}{\partial t} - V^2 S\right) + \frac{\partial q_i}{\partial m} - \frac{T_e - T_i}{\rho\tau} = Q_i,$$

$$v\left(\mu\frac{\partial}{\partial r}+\frac{1-\mu^2}{r}\frac{\partial}{\partial\mu}\right)f_j+\frac{\partial}{v^2\partial v}(v^2 a f_j) = W_j.$$

Here, $u(r, t)$ is the local flow rate of the substance, $P(r, t) = P_i + P_e$ is the total pressure of ions and electrons, $F(r, t)$ is the force of pressure exerted by fast thermonuclear particles, $S(r, t)$ describes the balance of the substance due to burnup and thermalization of thermonuclear particles, c_i and c_e are the specific heats for ions and electrons, q_e and q_i are the electronic and ionic thermal conductivity fluxes, $Q(r, T_e, T_i)$ is the specific energy transferred by ions and electrons to the plasma, $\tau(r, T_e)$ is the characteristic time of energy transfer between ions and electrons, f_j and W_j are the distribution functions and the generalized source of fast particles of the jth species, μ is the cosine of the angle between the radius vector and the velocity of a particle v, and a is the Coulomb drag coefficients for a fast particle.

A more detailed description of each equation from this system, explaining the physical and mathematical meaning of all the quantities appearing in it, is given in our previous article [6]. It was proved that the processes of transport of energy, momentum, and mass, which make the main contribution to the propagation of the combustion wave, are characterized by high gradients of density and temperature over the mean free path of fast thermonuclear particles, by the spatial anisotropy of distribution functions for these particles, by a complex energy dependence of Coulomb drag losses, and by the presence of a number of coupled channels of thermonuclear reactions.

Under these conditions, the Monte Carlo method is most adequate for simulating the kinetics of fast thermonuclear particles. Since the time of flight for rapid thermonuclear particles is considerably shorter than the characteristic time of variation of hydrodynamic parameters of the plasma, we can solve quasistationary kinetic equations at each time step of a time-dependent system of difference equations for a continuous medium used for a stochastic simulation of the kinetics. This algorithm of joint solution of hydrodynamic and kinetic equations is employed in a modified package of programs TEPA, which is used by us here.

2.2. Initial Conditions

It should be noted at the very outset that the package of programs TEPA can be used to simulate the evolution of a spherical laser-induced plasma with any radial temperature and density distributions at the instant of maximum compression. However, it is expedient to carry out a series of computations for simplified model configurations of a laser plasma in order to reveal general regularities and to provide limiting relations for rough estimates. For example, in model calculations, we assume that at the instant of maximum compression the plasma is inhomogeneous except for a small high-temperature region at the center, i.e., the ignitor. We consider homogeneous isobaric (ρT = const) and isochoric (ρ = const) ignitors corresponding to direct hydrodynamic compression and fast ignition, respectively, in the limiting cases. With such a formulation, the isobaric as well as isochoric ignitors possess only two independent parameters: the radius R_f and the temperature T_f. Henceforth, we will use the relative quantities $\Delta_t = T_f/T_0$ and $\Delta_r = R_f/R_0$ characterizing the extent of target inhomogeneity as the main parameters of the ignitor.

In our previous publication [4], we took into account the fact that possible values of temperature inhomogeneity Δ_t for an isobaric ignitor belong to the interval $2 < \Delta_t < 10$. It is difficult to attain higher values under direct hydrodynamic compression. For lower values, the target is virtually homogeneous and we are dealing with a bulk ignition rather than a spark ignition. According to similar considerations, the relative size of the ignitor is limited to the interval $0.1 < \Delta_r < 0.5$.

A higher degree of thermal inhomogeneity can also be attained in the course of fast ignition, but a further increase in temperature inhomogeneity weakly affects the combustion efficiency of targets with an isochoric ignitor (see below). The ignitor size increases rapidly due to thermal conduction, and the temperature decreases until the inhomogeneity attains values which are optimal from the point of view of spark ignition and lie in the above-mentioned intervals.

The range of initial parameters (ρR_0, T_0) of the main target plasma is chosen so that no thermonuclear burst occurs in a homogeneous target. If self-sustained thermonuclear combustion takes place even in the absence of an ignitor, such a case corresponds to bulk ignition, and the concept of spark ignition loses its physical meaning. Thus, the upper boundary of the temperatures we are interested in and the sizes of laser targets can be established from preliminary calculations aimed at the simulation of thermonuclear combustion of plasma in various homogeneous targets without an ignitor and at establishing critical minimal temperatures of bulk ignition for various values of ρR_0.

3. RESULTS OF MATHEMATICAL SIMULATION

3.1. Thermonuclear Combustion of Homogeneous Confined Plasma

The starting point of our analysis is the study of thermonuclear combustion of the plasma of homogeneous targets. It was mentioned in the previous section that simulation of combustion under homogeneous conditions allows us to set a limit on the initial parameters of the plasma in the case of spark ignition. Besides, calculations show that plasma parameters in the course of combustion in any effectively burning target after a thermonuclear burst approach the combustion parameters for a homogeneous plasma. Consequently, the calculations of the combustion efficiency of plasmas in homogeneous targets may form the basis

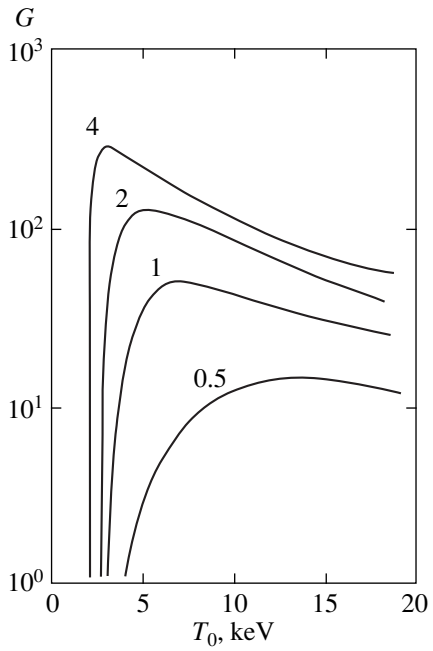

Fig. 1. Dependence of the gain factor G for a homogeneous DT plasma of laser targets on the initial temperature T_0 for various values of ρR_0 in g/cm^2 (figures on the curves) at the instant of maximum compression.

for a comparative analysis of the combustion efficiency of the plasma for various initial radial distributions of plasma parameters.

We will use the thermonuclear fusion gain factor $G = E_{TN}/E_0$ as the principal measure of the combustion efficiency, where E_{TN} is the thermonuclear energy liberated before the plasma burst and E_0 is the internal thermal energy of the plasma at the instant of maximum compression. Figure 1 presents the results of a series of calculations based on Eq. (1) which establish the dependence of the gain factor G on the initial temperature T_0 of the plasma of homogeneous DT targets with various values of ρR_0 ($0.5\ \text{g/cm}^2 \le \rho R_0 \le 4\ \text{g/cm}^2$) at the instant of maximum compression. It can be seen from the figure that for each value of ρR_0 of a homogeneous plasma in the range of relatively low temperatures, there exists an optimal (from the point of view of the combustion efficiency) initial temperature T_0 corresponding to the peak on the curve. The presence of a clearly manifested peak on each curve in this temperature range is associated with the dependence of the total liberated energy E_{TN} on T_0. The dynamics of energy liberation is determined by a complex combination of two opposite factors: a thermodynamic burst resulting in plasma cooling and thermonuclear combustion leading to plasma heating and partial burnup. As we approach the peak, the increase in E_{TN} is much faster than the increase in E_0 owing to an increasing role of self-heating, and the value of E_{TN} in the peak region attains saturation to a certain extent.

The most interesting result of calculations is that large maximum values of $G > 100$ for targets with $\rho R_0 \gtrsim 1$ g/cm^2 are attained at temperatures $T \sim 3$ to 5 keV, which are much lower than the temperatures corresponding to the peak of the DT fusion reaction rate ($T \sim 20$ keV). This is due to the fact that, at the initial stage of combustion, such targets are opaque for fast charged particles, and a considerable fraction of the liberated thermonuclear energy remains in the target, heating it to high temperatures corresponding to large values of the DT fusion reaction rates. It can also be seen that, for targets with $\rho R_0 > 1$ g/cm^2, the gain factor increases abruptly from $G \sim 1$ to $G_{\max}$ in a narrow range of initial temperatures. The sharp steepening of the temperature dependence of G with increasing ρR_0 is a consequence of the increasing role of nonlinear processes in the evolution of a thermonuclear burst. This narrow interval of initial temperatures will be referred to as the unstable ignition zone. The instability should be interpreted in the sense that insignificant fluctuations of the initial temperature may lead to radical changes in the combustion efficiency. All the initial temperatures T_0 of the plasma lead to the formation of thermonuclear bursts starting from weak bursts terminating with glow to effective bursts, at which the self-heating temperature has time to attain values close to optimal values for DT fusion. A further increase in T_0 for the plasmas of such targets leads to self-sustained bulk combustion. According to calculations, in the case of an effective thermonuclear burst, the plasma parameters in the course of combustion and, hence, the energy yield E_{TN} are determined by the dynamics of the burst itself and are virtually independent of the initial conditions. Thus, uniform heating of plasma to higher temperatures ($T_0 \sim$ 15–20 keV) directly in the course of implosion leads to a decrease in the gain factor:

$$G \sim 1/T_0.$$

In targets with $\rho R_0 < 1$ g/cm^2, the heating by charged thermonuclear particles is ineffective, which in turn leads to relatively low ($G \le 10$) values of the gain factor and its smooth dependence on T_0 in a wide range of initial temperatures. The emerging property of smooth controllability of the gain factor G in the case of bulk ignition of such targets might prove helpful for developing hybrid subcritical nuclear reactors with laser initiation [8, 9].

The behavior of gain factors in the region of unstable ignition can hardly be described analytically. At the same time, a qualitative description of the results of our calculations of thermonuclear combustion in a homogeneous spherical plasma in the case of bulk combustion at temperatures above the optimal (critical) temperature can be generalized to obtain results convenient for practical applications and simplified analytical estimates of the gain factor G.

Let us consider the dependence of the thermonuclear gain factor on the initial values ρR_0 and T_0 for a homogeneous plasma, i.e., at the instant of maximum compression. By definition, the expression for G can be written in the form

$$G = \frac{E_{TN}}{E_0} \sim \frac{\langle \sigma v \rangle \rho \Delta t}{T_0}, \quad (2)$$

where $\langle \sigma v \rangle$ is the characteristic of the thermonuclear reaction rate per unit volume averaged over a burst and Δt is the characteristic time of fusion plasma burst.

The quantity $\langle \sigma v \rangle$ averaged according to Maxwell at a fixed temperature has a clearly pronounced peak at $T \sim 20$ keV and varies insignificantly as the temperature increases subsequently to $T \sim 1000$ keV [10]. Consequently, if a thermonuclear burst has occurred, the averaged value $\langle \sigma v \rangle$ weakly depends on the initial conditions and is nearly constant. The burst time Δt can be presented as the ratio of the radius to the sound velocity in the plasma averaged over the burst:

$$\Delta t \sim \frac{R_0}{v_{\text{sound}}} \sim \frac{R_0}{\langle T \rangle^{1/2}}.$$

According to the results of our calculations, the temperature in targets with $\rho R_0 \gtrsim 1$ g/cm^2 during combustion attains hundreds of kilolectronvolts due to self-heating. This leads to a decrease in the combustion time and, hence, to a slower increase in the gain factor G with increasing ρR_0 (this factor was usually discarded). Since the thermonuclear energy liberated in such targets is much higher than the initial internal energy and its considerable fraction remains in the plasma, we can assume that the characteristic temperature of the plasma at the thermonuclear combustion stage is $\langle T \rangle \sim GT_0$. In this case, the thermonuclear gain factor can be presented in the form

$$G = F(\rho R_0, T_0)(\rho R)_0^{2/3}/T_0. \quad (3)$$

The function $F(\rho R_0, T_0)$, which depends on the parameters ρR_0 and T_0 only slightly when self-sustained bulk combustion is attained, will be referred to as the reduced gain factor.

The values of the reduced gain factors obtained by processing the results of calculations presented in Fig. 1 are shown in Fig. 2. It follows from Fig. 2 that the dependence $F(\rho R_0, T_0)$ on the initial parameters of target plasmas with $\rho R_0 \gtrsim 1$ g/cm^2 in the region of their effective bulk ignition is weak and is close to the scaling dependence: $F(\rho R_0, T_0) = 370$ to within ~10%. Consequently, the thermonuclear gain factor for a homogeneous plasma of thick targets can be estimated using the following simple expression:

$$G = 370(\rho R)_0^{2/3}/T_0, \quad (4)$$

where $(\rho R)_0$ is measured in g/cm^2 and T_0 in keV.

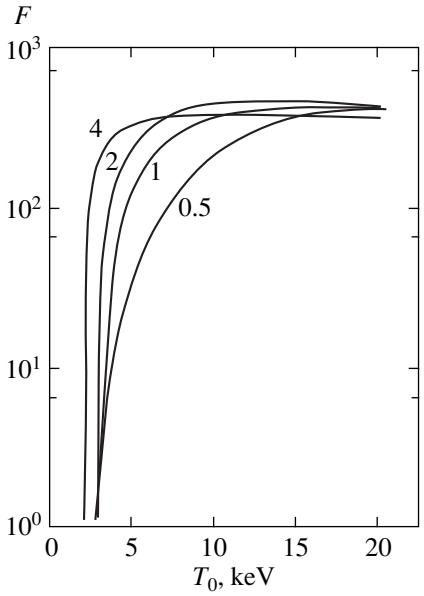

Fig. 2. Dependence of the reduced gain factor $F = GT_0/(\rho R)_0^{2/3}$ for a homogeneous DT plasma on the initial temperature T_0 for various values of $(\rho R)_0$ in g/cm^2 (figures on the curves) at the instant of maximum compression.

For targets with $\rho R_0 < 1$ g/cm^2, plasma is not heated during combustion and the characteristic plasma temperature in the course of combustion is independent of G. The same considerations as those used in the derivation of formula (3) lead to the following dependence for such targets:

$$G \sim \rho R_0/T_0^{3/2}.$$

The numerical simulation of plasma combustion in thin targets in the vicinity of the peak temperature ($T \sim 20$ keV), which was carried out by using the TEPA software package, gives the values of $G \sim 20\rho R_0$ for the maximum gain factors of these targets. These values are in good accord with the well-known approximate estimate of the maximum burnup fraction for the DT fuel [7]:

$$\Delta m/m \sim \rho R_0/7;$$

in the intermediate region of $\rho R_0 \sim 1$ g/cm^2, these values are close to that specified by Eq. (4).

3.2. Thermonuclear Combustion of Inhomogeneous Plasma in Laser Targets under Hydrodynamic and Fast Spark Ignition

It was shown above that irrespective of the initial temperature, no self-heating takes place during thermonuclear combustion of plasma in homogeneous targets with $\rho R_0 < 1$ g/cm^2 since almost the entire thermonuclear energy liberated in the plasma is carried away from the target by fast particles and X-rays. Consequently, spark

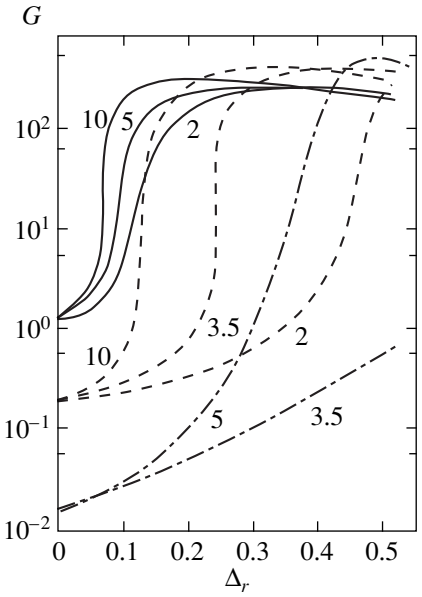

Fig. 3. Dependence of the fusion gain factor G on the relative size Δ_r of the ignitor for various values of the parameter Δ_t (figures on the curves) for an isochoric target with $(\rho R)_0$ = 2 g/cm^2. Solid, dashed, and dot-and-dash curves correspond to targets with the initial temperature of the main plasma T_0 = 3, 2, and 1 keV, respectively.

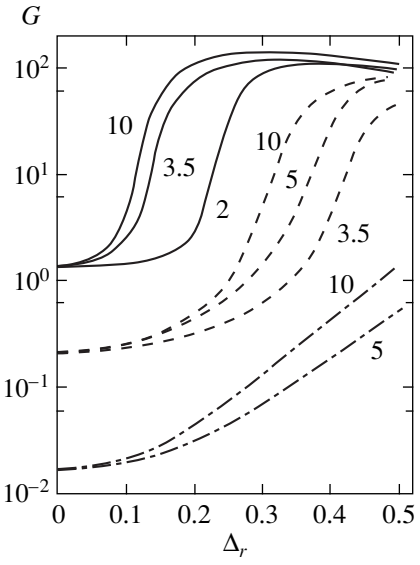

Fig. 4. Dependence of the thermonuclear gain factor G of the target on the relative size Δ_r of the ignitor for various values of the parameter Δ_t (figures on the curves) for an isobaric target with $(\rho R)_0$ = 2 g/cm^2. Solid, dashed, and dot-and-dash curves correspond to targets with the initial temperature of the main plasma T_0 = 3, 2, and 1 keV, respectively.

ignition is also impossible for such targets. On the contrary, for targets with $\rho R_0 \gtrsim 1$ g/cm^2, we can expect that a self-sustained thermonuclear combustion wave formed in the high-temperature ignitor may cause a thermonuclear burst in a plasma with an initial temperature T_0 which is considerably lower than the critical temperature of bulk ignition.

In order to study the effect of ignitor parameters on the efficiency of plasma combustion in the case of spark ignition, we studied a number of targets whose parameters are in the interval $\rho R_0 \gtrsim 1$ g/cm^2. Mathematical simulation of thermonuclear combustion was carried out for each fixed value of ρR_0 for several values of the initial temperature T_0 below the critical temperature for bulk ignition. For each chosen temperature T_0, series of model configurations with isobaric and isochoric ignitors with the parameters in the intervals $0.1 \le \Delta_r \le 0.5$ and $2 \le \Delta_t \le 10$ were considered. The evolution of target plasma up to the burst was simulated for each configuration in accordance with the system of Eqs. (1), and the gain factor G was calculated. In order to obtain reliable dependences, we investigated more than 400 various configurations of target plasmas, for which thermonuclear combustion was simulated.

We will consider the results of calculation of the gain factor G for a typical target with ρR_0 = 2 g/cm^2. Figure 3 shows the values of G calculated for isochoric conditions of ignition as a function of the relative size Δ_r of the ignitor for various initial distributions of temperature (the values of Δ_t are presented as parameters of the family of curves) and for various initial temperatures T_0 of the main plasma (various types of curves). The values of all parameters are given most completely in figure captions. Figure 4 shows for comparison the results of similar calculations for isobaric targets.

It can be seen from these figures that, for various values of the temperature inhomogeneity $\Delta_t = T_f/T_0$, there exist critical values of ignitor sizes, $\Delta_r = R_f/R_0$, near which the efficiency of thermonuclear combustion increases abruptly both for isobaric and isochoric ignitions. The narrow interval Δ_r in which the value of G increases abruptly to the maximum is the ignition instability interval. The position and width of this interval are determined to a considerable extent by the initial temperature of the main plasma. A similar form of the dependences is also observed for other values of ρR_0.

The physical nature of this phenomenon is that the time of propagation of a thermonuclear combustion wave to the outer boundary of the plasma approximately coincides with the time of target burst for values of parameters close to critical. For smaller values of Δ_r, the plasma has time to disintegrate without a nuclear burst. As Δ_r and Δ_t pass through the critical values, a nuclear burst of the plasma occurs, which is accompanied by high-efficiency thermonuclear burnup with $G \sim 100$. If the effective thermonuclear burst has occurred for any target configuration, G attains approximately the same maximum value. In other words, the past history of the process weakly affects the burnup efficiency.

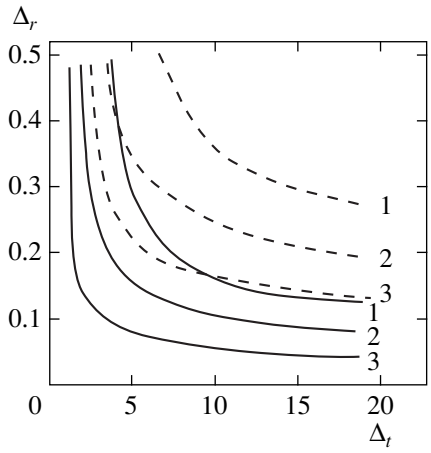

Fig. 5. Boundaries for the parameters of an isochoric ignitor, above which a thermonuclear burst of targets with various parameters T_0 and $(\rho R)_0$ of the main plasma at the instant of maximum compression is possible. The curves are marked by the values of T_0 in keV. Solid and dashed curves correspond to $(\rho R)_0 = 2$ and 1 g/cm^2, respectively.

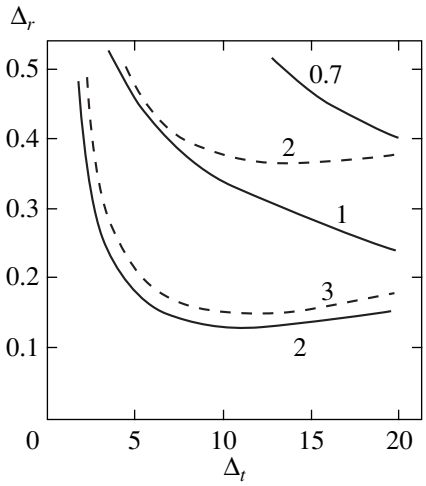

Fig. 6. Boundaries for the parameters of an isobaric ignitor, above which a thermonuclear burst of targets with various parameters T_0 and $(\rho R)_0$ of the main plasma at the instant of maximum compression is possible. The curves are marked by the values of T_0 in keV. Solid and dashed curves correspond to $(\rho R)_0 = 4$ and 2 g/cm^2, respectively.

According to the results of quantitative analysis, the values of the gain factor G can be approximately described by formula (4) in the limit of large values of Δ_r and Δ_t. The intermediate values of $1 < G < 100$ are observed in a narrow interval of ignitor parameters near the critical values. The higher the values of ρR_0 for the main plasma of the target, the narrower the unstable ignition interval, and the more effective the plasma combustion.

Having chosen the attainment of the gain factor $G = 10$ as the criterion of spark ignition, we obtain the boundaries of the regions in which an isochoric spark ignition of the plasma may be realized from stripped targets with various values of ρR_0 and T_0 on the plane of the ignitor parameters Δ_r and Δ_t. These boundaries are shown in Fig. 5. Similar dependences for the critical parameters of isobaric ignitors are shown in Fig. 6. It should be noted that the choice of the criterion $G = 10$ is conditional, but in view of the sharp increase in the gain factor in the vicinity of the critical parameters of ignitors, the specific choice of the criterion weakly affects the position of the boundaries of the spark ignition region.

A comparison of Figs. 5 and 6 shows that for identical initial parameters of the target plasma, the critical size of isobaric ignitors is several time larger than that for isochoric ignitors. Moreover, the range of admissible parameters of the targets with isochoric ignitors in which spark ignition is possible in principle is much broader than for targets with isobaric ignitors. For example, calculations show that isobaric ignition of a target with $\rho R_0 = 1$ g/cm^2 is impossible for the initial temperature $T_0 < 10$ keV, while in the case of an isochoric target, effective combustion can be realized for $T_0 = 1$ keV. Consequently, the ignition according to the fast ignition mode is more rational. This is a natural consequence of the fact that the internal energy of an isochoric ignitor is Δ_t times higher than the energy of an isobaric ignitor for the same parameters Δ_r and Δ_t.

Since we studied the isobaric spark ignition earlier [11, 12], we will consider in greater detail the combustion of a plasma with an isochoric ignitor. In contrast to the isobaric case, the energy density in the isochoric ignitor is much higher than in the remaining plasma. Consequently, instead of the target inhomogeneity parameters Δ_r and Δ_t, which were introduced by us earlier, it is more expedient to use another pair of dimensionless parameters for describing isochoric ignition, viz., the relative mass $\Delta_m = m_f/(m_0 + m_f)$ of the ignitor and the ratio $\Delta_E = E_f/E_0$ of the supplementary thermal energy of the isochoric ignitor to the total energy stored in the homogeneous plasma of the target. The latter parameter is directly connected with the energy of the supplementary laser pulse. The parameters Δ_m and Δ_E can be easily recalculated in terms of the target inhomogeneity parameters Δ_r and Δ_t:

$$\Delta_E = \Delta_r^3(\Delta_t - 1), \quad \Delta_m = \Delta_r^3. \tag{5}$$

Figure 7 shows the boundaries of the regions of isochoric spark ignition for various targets (see Fig. 5) in terms of the variables Δ_m and Δ_E. It can be seen from the figure that the critical value of the parameter Δ_E for small ignitors with a mass less than 1% of the total mass of the target is virtually independent of the size of the ignitor and is determined only by the size and temperature of the target plasma. In other words, in the case of fast ignition, there exists a minimum value of supple-

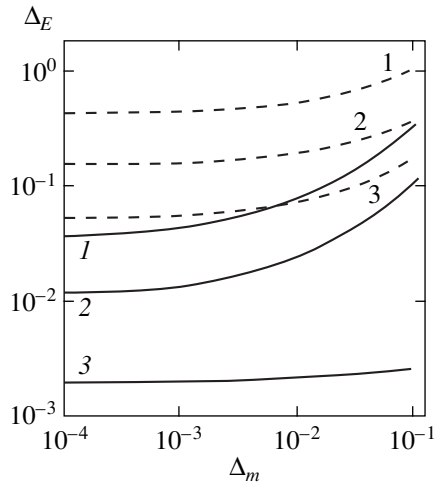

Fig. 7. Dependence of the relative value of the supplementary thermal energy $\Delta_E = E_f/E_0$ corresponding to critical values of the parameters of isochoric ignitor, its relative mass $\Delta_m = m_f/(m_0 + m_f)$. The curves are marked by the values of T_0 in keV.

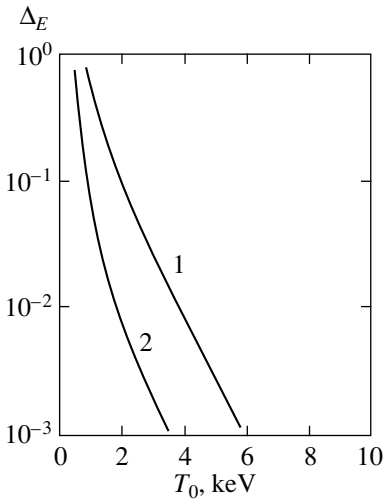

Fig. 8. Dependence of the minimal values of the relative supplementary thermal energy $\Delta_E = E_f/E_0$ of an isochoric ignitor on the plasma temperature at the instant of maximum compression for different targets. The curves are marked by the values of $(\rho R)_0$ of the target plasma in g/cm^2.

mentary absorbed laser energy corresponding to an optimal ignitor. The optimal ignitor is found to be quite small in size. As the size of the ignitor increases, its heating requires a large amount of supplementary energy, while for decreasing the size of the ignitor the supplementary energy is rapidly smeared due to thermal conduction over the volume corresponding to the optimal size of the ignitor.

Quantitative dependences of the minimum value of the fraction of the supplementary energy Δ_E required for isochoric ignition on the temperature of a plasma with two different values of ρR_0 are presented in Fig. 8. The obtained curves make it possible to estimate the minimal energy of an auxiliary laser pulse, which is required for fast ignition of the plasma of specific targets.

4. CONCLUSION

The results of our investigations allow us to formulate the following characteristic features of thermonuclear combustion of plasma in stripped laser targets.

The simulation of plasma combustion in homogeneous laser targets of a fixed thickness with various initial temperatures shows that bulk ignition becomes optimal in a very narrow range of initial temperatures. If the effective thermonuclear burst has occurred, the liberated energy is practically independent of the initial temperature, and the gain factor G decreases in proportion to $1/T_0$ upon subsequent heating (see Fig. 1). The values of the initial parameters of the target plasma, which correspond to the maximum gain factor G, form on the $(\rho R_0, T_0)$ plane a line bounding the region of bulk ignition. This boundary is presented in Fig. 9 (line *1*).

For the targets whose parameters lie outside the region of bulk ignition, we can consider the process of spark ignition. For fixed parameters of the target, there exist critical parameters of the ignitor at which a thermonuclear combustion wave can be formed during a time much shorter than the burst time. Only in this case is the thermonuclear combustion highly effective. For stripped targets with various values of ρR_0 and T_0, the corresponding boundaries of the critical parameters Δ_r and Δ_t of the ignitor are presented in Figs. 5 and 6 for isochoric and isobaric ignitors, respectively.

Since there exist technical limitations imposed on the maximum extent of inhomogeneity attainable during direct hydrodynamic compression, an isobaric spark ignition cannot be realized if the target has a temperature too low or a thickness too small. Mathematical simulation of fusion in most inhomogeneous targets with the ignitor parameters $\Delta_t = 10$ and $\Delta_r = 0.5$ makes it possible to establish the relevant lower boundaries of plasma temperature and thickness in laser targets in the case of isobaric ignition. This boundary is also shown in Fig. 9 (line *2*). The range of target plasma parameters in which isobaric ignition is possible lies between curves *1* and *2* in Fig. 9.

For an isochoric ignitor corresponding to fast ignition mode, the dependence of the gain factor of the target on the parameters of the ignitor does not differ qualitatively from the isobaric case. However, the critical parameters of isochoric ignitors are quantitatively much lower than the corresponding parameters for isobaric ignitors. It is shown that the relative value of the supplementary thermal energy of the ignitor, which is directly related to the energy of an auxiliary laser pulse, is a convenient parameter for an analysis of isochoric

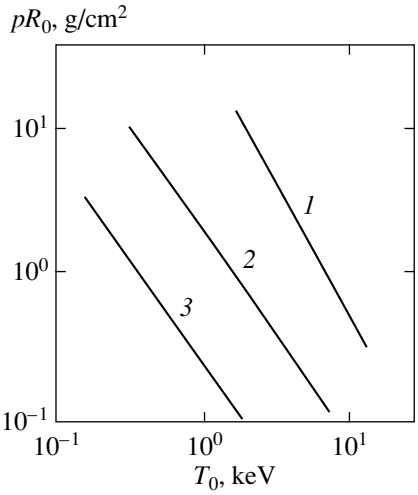

Fig. 9. Boundaries of the parameters of main target plasma at the instant of maximum compression, above which a thermonuclear burst is possible for various ignition mechanisms: bulk ignition (curve *1*), spark ignition with an isobaric ignitor (curve *2*), and spark ignition wit an isochoric ignitor (curve *3*).

ignition. According to the results of calculations, the critical (minimum) value of this quantity required for spark ignition with a small ignitor does not depend on its size and is determined only by the initial temperature and ρR_0 of the target. The corresponding quantitative dependences are presented in Fig. 8.

As in the case of isobaric ignitors, isochoric ignition cannot be realized if the target has an initial temperature too low or a thickness too small. However, the range of target plasma parameters in which isochoric ignition can be realized in principle is much wider than the corresponding range for isobaric ignition. The boundary of this region is also shown in Fig. 9 (line *3*).

Irrespective of the ignition mode, the value of the gain factor for an effectively burning target can be estimated with an admissible accuracy using the following simple relation:

$$G = 370(\rho R)_0^{2/3}/T_0.$$

Spark ignition makes it possible to extend the range of application of this expression to a wider range of initial temperatures and thicknesses of the target.

ACKNOWLEDGMENTS

This work was partly financed by the Ministry of Higher Education of Russia, grant "Thermonuclear and Nuclear Reactors-1997," and Program 015 "Fundamental High-School Studies in Natural Sciences and Humanities. Russian Universities-2000" (grant no. 990556) as well as the Russian Foundation for Basic Research (grant no. 99-02-16100).

REFERENCES

1. Yu. A. Afanas'ev, N. G. Basov, P. P. Volosevich, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **21**, 150 (1975) [JETP Lett. **21**, 68 (1975)].
2. Yu. A. Afanas'ev, N. G. Basov, P. P. Volosevich, *et al.*, Preprint No. 55, FIAN (Lebedev Institute of Physics, Academy of Sciences of USSR, Moscow, 1979).
3. O. B. Vygovskii, S. Yu. Gus'kov, A. A. Levkovskii, *et al.*, J. Russ. Laser Res. **14**, 85 (1993).
4. S. Yu. Gus'kov, A. A. Levkovskii, V. B. Rozanov, *et al.*, Proc. SPIE **2770**, 182 (1996).
5. A. A. Levkovskiĭ, Preprint No. 73, FIAN (Lebedev Institute of Physics, Academy of Sciences of USSR, Moscow, 1990).
6. S. Yu. Gus'kov, N. V. Zmitrenko, A. A. Levkovskiĭ, *et al.*, Zh. Éksp. Teor. Fiz. **106**, 1069 (1994) [JETP **79**, 581 (1994)].
7. M. Tabak, J. Hammer, M. E. Glinsky, *et al.*, Phys. Plasmas **1**, 1626 (1994).
8. L. P. Feoktistov, E. N. Avrorin, L. F. Varganova, *et al.*, Kvantovaya Élektron. (Moscow) **5**, 349 (1978).
9. N. G. Basov, A. E. Sheĭndlin, *et al.*, Izv. Akad. Nauk SSSR, Énerg. Transp. **2**, 3 (1979).
10. G. Cauglan, W. Fowler, M. Harris, and B. Zimmerman, At. Data Nucl. Data Tables **32**, 197 (1985).
11. A. A. Andreev, D. V. Il'in, A. A. Levkovskii, *et al.*, Proc. SPIE **3683**, 170 (1999).
12. A. A. Andreev, D. V. Il'in, A. A. Levkovskiĭ, *et al.*, Izv. Akad. Nauk, Ser. Fiz. **63**, 1182 (1999).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 78–85.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 91–98.

PLASMA, GASES

A Stationary Filamentary Microwave Discharge at a High Pressure of Argon

A. A. Skovoroda* and A. V. Zvonkov

Russian Research Center Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia
**e-mail: skovorod@nfi.kiae.su*

Received June 26, 2000

Abstract—A stationary discharge in the form of a filament about 1 mm in diameter is produced in a quartz tube 16 mm in diameter passed through the wider wall of a standard waveguide and field with argon at a pressure of 300 mm Hg at a microwave power of 50 W on a frequency of 7 GHz. The number of filaments increases gradually from one to seven as the power and pressure are increased. The filaments are parallel and oriented along the electric field of the wave. They are arranged symmetrically relative to the equatorial plane of the tube. The distance between the filaments decreases as their number increases. The stationary filaments arising due to ionization-overheating instability may be explained qualitatively using the interference-equilibrium model and assuming that a discharge is organized such as to provide for the maximal power absorption. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is known that a microwave discharge in gas at a high pressure is structured. A wide variety of structures is observed; however, two structures may be regarded as basic, namely, spherical and filamentary. Such structures gained wide recognition following the studies by P.L. Kapitza, who, in particular, used the capacity of microwave discharge in the atmosphere to assume a spheroidal shape for interpreting the natural phenomenon of ball lightning [1]. Allison *et al.* [2] were apparently the first to observe a filamentary structure of microwave discharge. More than ten long parallel brightly glowing filaments were observed on the walls of a spherical bulb filled with argon and placed in the microwave field of a radar. The beauty of the observed phenomenon and lack of its explanation caused Howatson to include a picture of the discharge from [2] in his monograph on the theory of gas discharge [3]. Later, filaments were observed by numerous researchers dealing with pulsed microwave discharges (see references cited in [4–6]).

The emergence of filaments may be due to two reasons. The first reason is associated with the development of ionization-overheating instability (IOI), which leads immediately to the formation of long filaments. The second reason is associated with the development of microwave streamers, when a filament is drawn gradually. The latter reason is characteristic of high-power pulsed discharges and is not associated with the stationary low-power discharges to which this paper is devoted.

The linear stage of IOI was studied theoretically by Gil'denburg and Kim [7]. They have demonstrated that small perturbations against a homogeneous background in a microwave discharge start increasing at gas pressure values at which the electromagnetic wave frequency ω become less than the collision frequency ν between electrons and neutral gas particles. The characteristic spatial scale of instability was estimated to be of the order of the geometric mean of the wavelength in vacuum and the characteristic diffusion length. The initial nonlinear stage of IOI exhibits an explosive behavior [8]. More complex and totally uninvestigated is the problem of the existence of a steady-state solution corresponding to a multifilament structure. The results of experiments described in [2] and discussed by us demonstrate the existence of such solutions.

Section 2 contains a description of the experimental facility and diagnostics employed. In Section 3, experimental results are given; in our opinion, the most interesting results include the facts of obtaining a stationary solitary filament and of observing an abrupt increase in the number of stationary filaments with increasing argon pressure and microwave power. In so doing, the filaments are almost parallel and oriented along the electric field of the wave. The arrangement of filaments relative to the equatorial plane of the discharge tube (see below) proves to be symmetric. In Section 4, the obtained results are discussed using the equilibrium theory of microwave discharges in the atmosphere and invoking the hypothesis of self-organization. In reality, the plasma in the filaments is nonequilibrium, but we assume that some qualitative features of the experiment may be interpreted using the equilibrium model.

2. EXPERIMENTAL FACILITY

Figure 1 illustrates the waveguide scheme of the facility and the design of the discharge chamber. The steady-state microwave radiation with a frequency of 7 GHz

and maximal power of 0.5 kW was delivered to the discharge chamber via a standard waveguide of 15×35 mm^2. The discharge chamber was made as a hermetically sealed quartz tube with an inside diameter of 16 mm, wall thickness of 1.5 mm, and length of 160 mm and was passed through the wider wall of the waveguide at equal distances from the narrow walls in the direction of the electric field of the wave. Outside of the waveguide, the quartz tube was placed in a water-cooled copper tube with an inside diameter of 19.2 mm. Numerous small holes were provided in the copper tube for visual observation of plasma glow outside of the waveguide. Airtight optical ports were provided at the ends of the quartz tube for photographing the discharge, as well as openings for pumping out and bleeding in of gas. An intense flow of air was delivered into the waveguide for cooling down the quartz tube section located inside the waveguide. The discharge was effected in still argon at a pressure from 0.05 to 1 atm.

The experiments described herein involved the measurements of the incident, absorbed, and reflected microwave power using directional couplers and M3-47 and M3-71 calorimetric microwave-power meters. The glow was recorded by a general-purpose videocamera with subsequent frame-by-frame digitizing of the obtained images. The filming was performed with 1/2000 exposure and a high degree of stopping to eliminate flare spots and increase the depth of focus. The temporal behavior of the discharge was monitored with the aid of a crystal detector and a photomultiplier.

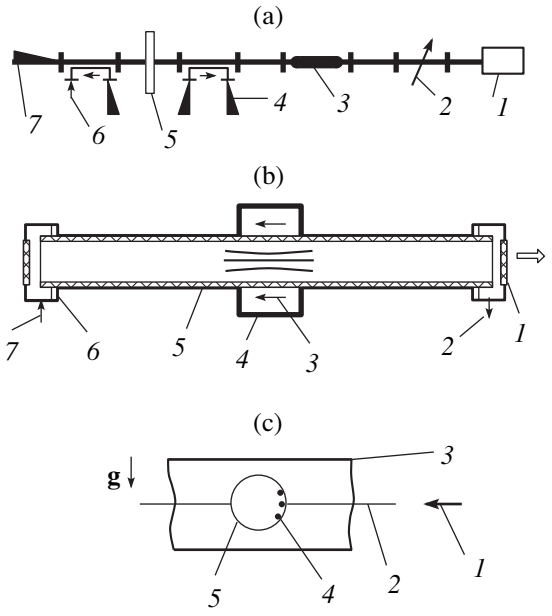

Fig. 1. (a) The waveguide scheme of the facility: (*1*) klystron, (*2*) attenuator, (*3*) ferrite circulator, (*4*) reverse coupler with a calorimetric power meter, (*5*) discharge chamber, (*6*) transmitted-wave coupler with a crystal detector, (*7*) calorimetric meter of transmitted power. (b) The waveguide scheme of the discharge chamber: (*1*) end port and direction of video filming, (*2*) pumping out, (*3*) direction of the electric field of the wave, (*4*) standard waveguide, (*5*) copper tube, (*6*) hermetically sealed quartz tube, (*7*) bleeding in of gas. A view is given of a discharge containing three filaments. (c) A view of the end port: (*1*) direction of propagation of microwave power from the klystron, (*2*) equatorial plane of the discharge chamber, (*3*) waveguide, (*4*) three filaments of the discharge, (*5*) quartz tube; *g*, gravity direction.

3. EXPERIMENTAL RESULTS

Figure 2 gives the power absorbed in the plasma and the coefficient of reflection as functions of argon pressure. The numerals indicate the moments of emergence of the respective number of filaments. The frame-by-frame scanning of continuous video records has revealed that, within one or two frames (20 to 50 ms), the number of filaments always remains constant. As the pressure increases, the number of filaments in different frames may be different. The presence in Fig. 2 of two numerals placed side by side indicates that, at different moments of time, different numbers of filaments are observed with approximately equal probability. Given in brackets is the number of filaments that is seldom observed.

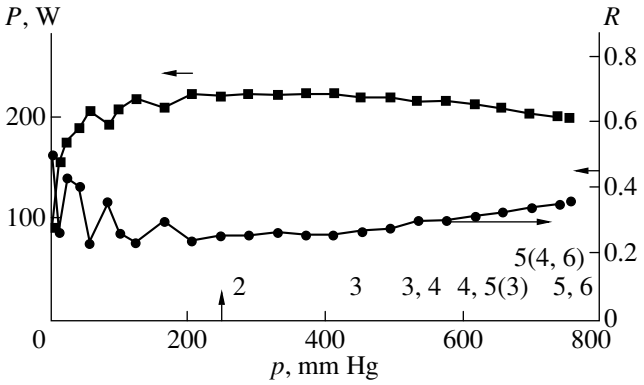

Fig. 2. The absorbed power *P* and the coefficient of reflection *R* as functions of argon pressure. The klystron power, 325 W. The arrows on the axes indicate the moment of emergence of filaments and the coefficient of reflection without plasma. The numerals at the bottom indicate the number of filaments at respective values of pressure.

The moment of emergence of filaments with rising pressure is in good agreement with the prediction of the IOI theory, $\omega \sim \nu$, if one takes into account the decrease in the density of gas inside a filament due to heating. The number of filaments increases with pressure. In so doing, the absorption and reflection of waves vary little. Note that the measurements of transmitted power using a fast crystal detector failed to reveal the disappearance of absorption during temporal "jumpovers" from one number of filaments to another.

Figure 3 gives the absorbed power, the coefficient of reflection, and the number of filaments as functions of the klystron power. For the maximal power, the maximal number of seven filaments was observed.

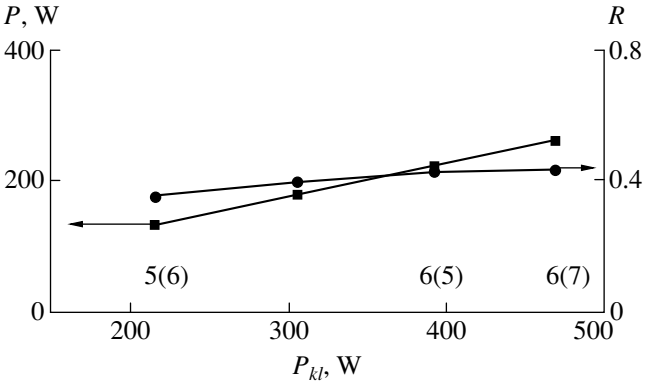

Fig. 3. The absorbed power P and the coefficient of reflection R as functions of the klystron power P_{kl}. The argon pressure, 660 mm Hg.

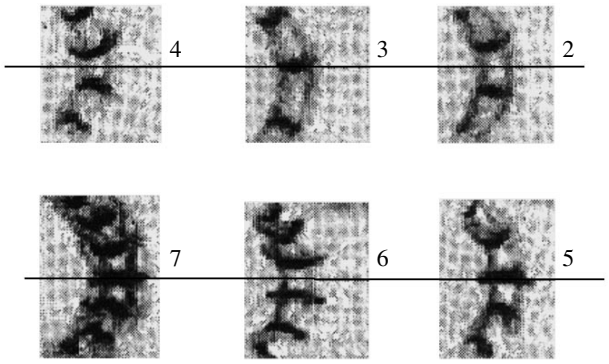

Fig. 4. Individual frames showing filaments. The filming was performed at an angle of 10° to the direction shown in Fig. 1c; the straight line indicates the equatorial plane of the waveguide and discharge chamber, and two vertical arcs indicate the edge of the waveguide.

Figure 4 is a collection of individual frames of video filming with the number of filaments ranging from two to seven. The filming was performed via optical port (see Fig. 1b) at an angle of 10° to the direction shown in Fig. 1c. One can see that the filaments are arranged symmetrically relative to the equatorial plane shown in Figs. 1c and 4; in the case of an odd number of filaments, the central filament is always located on the equatorial plane. The filaments go a little beyond the waveguide limits and disappear while diverging somewhat. As the number of filaments increases, they come to be arranged ever more closely to one another over the perimeter of the tube side facing the incident microwave power. Figure 5 is a frame with six filaments, filmed at an angle of 0°, i.e., strictly along the tube axis. The radiation is incident on the plasma from the right. The doubling of filament image is caused by the effect of reflection from the walls of the quartz tube. On analyzing Figs. 4 and 5, one can obtain the following characteristic geometric parameters of filaments: diameter, 1 to 1.5 mm; length, 15 to 25 mm; distance between filaments, 3 to 10 mm; the filaments are oriented along the electric field and spaced from the wall at a distance of the order of the filament diameter.

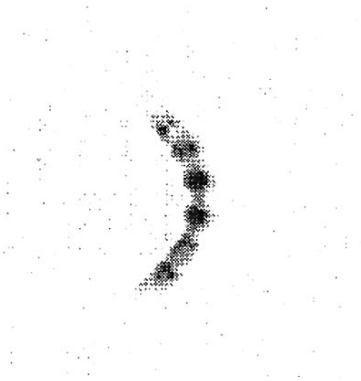

Fig. 5. A frame filmed at an angle of 0° (see Fig. 1c). The curvature radius is equal to the tube radius of 8 mm.

A special experiment was performed in order to obtain a single filament, because a discharge was required at the threshold of extinction under conditions of low argon pressure and low klystron power. The experimental procedure was as follows. A discharge was ignited at a pressure of 300 mm Hg, and the klystron power was reduced to 49 W. After that, the gas was slowly pumped out and the recording was performed continuously until the extinction of the discharge. Figure 6 illustrates the result of such an experiment. For a period of 20 s, one could observe a solitary filament located in the equatorial plane. In so doing, one can observe an almost complete agreement between the discharge chamber and incident microwave radiation and a stepped rise of absorption at the moment of doubling of filaments. The increase in absorption during the tripling of filaments is

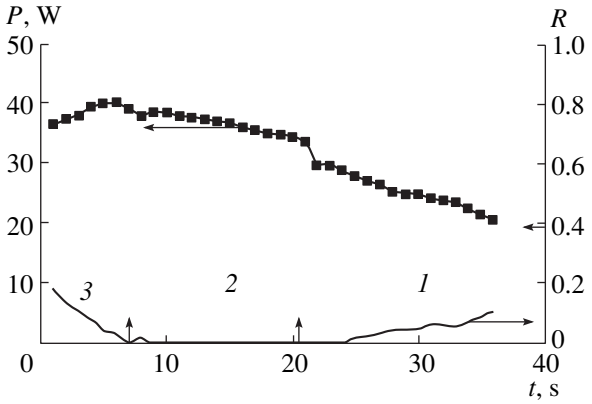

Fig. 6. The variation of the absorbed power, the coefficient of reflection, and the number of filaments during the time of experiment. The arrows indicate the moments of variation of the number of filaments and the coefficient of reflection from the discharge chamber without plasma. Argon, 300 mm Hg, 50 W.

pronounced less clearly. An absorbed microwave power of 20–30 W is sufficient for the existence of a single filament. An enlarged photograph of a single filament is shown in Fig. 7. The results of measurement of the temporal dynamics of intensity of glow of a single filament with the aid of a photomultiplier demonstrated a complete stationarity of the filament.

4. DISCUSSION

The experimental results prove the existence of a stationary discharge in the form of thin filaments extended along the electric field **E** of the wave and located in the vicinity of the wall onto which the microwave is incident. The IOI is known as the original cause of the emergence of these filaments. However, one needs to explain the reason for the stability of filaments, their visually observed geometric characteristics, i.e., to describe, at least qualitatively, a stationary filamentary microwave discharge.

The strong spatial nonuniformity of the parameters of the plasma of filamentary microwave discharge gives no grounds to apply the equilibrium theory of microwave discharges; nevertheless, we will try to derive from this latter theory a qualitative explanation of the observed phenomena while leaving the exact quantitative analysis to the nonequilibrium theory.

The orientation of filaments along the electric field is explained by the need for the flow of the filament-heating current. In our experiment, the current at the ends of filaments is closed by the bias currents to the waveguide walls via capacitors formed by the quartz tube. Therefore, the electric field intensity and the current along the filament are constant. Indicative of this is the constancy of the glow intensity along the filament within the waveguide. Therefore, we use in discussion the model of an infinite filament in the field of a plane microwave with linear polarization along the filaments.

4.1. Homogeneous Cylinder in the Field of a Plane Wave

Is the equilibrium model of discharge capable of predicting any characteristic dimension for a filament on the threshold of extinction? In order to answer this question, we will analyze the problem of heating and cooling down of an infinite and homogeneous cylinder in the field of a plane electromagnetic wave polarized along the cylinder.

The heating may be analyzed using the long-solved classical problem of scattering from an infinite cylinder (see, for example, monographs [9–11]). The specific nature of our problem is characterized only by the properties of the medium, that is, the plasma which forms the cylinder. The complex refractive index is fully defined by the equilibrium temperature of the plasma cylinder [12].

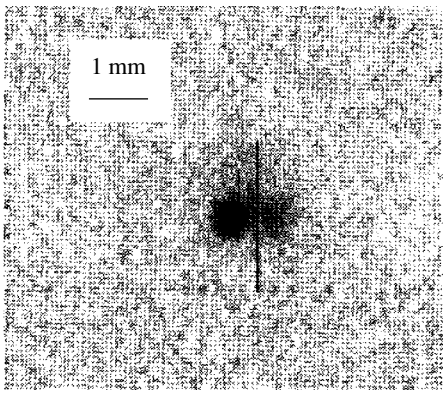

Fig. 7. An enlarged image of a solitary filament at an angle of 0°. The line indicates the position of the internal wall of the quartz tube. The halo on the right is the effect of reflection from the walls. The parameters are as in Fig. 6.

We will represent the power of heating the cylinder per unit length as

$$P = (\sigma_t - \sigma_s)S, \tag{1}$$

where S is the flux of incident power in the plane wave, $(\sigma_t - \sigma_s)$ is the absorption cross section, and σ_s is the scattering cross section. The following expressions were derived for the cross sections:

$$\sigma_t = \frac{4}{k}\sum_{n=-\infty}^{\infty} \mathrm{Re}\, c_n, \quad \sigma_s = \frac{4}{k}\sum_{n=-\infty}^{\infty} |c_n|^2, \tag{2}$$

where k is the vacuum wave vector and

$$c_n = \left[1 + i\frac{J_n(N\xi)N_n'(\xi) - NN_n(\xi)J_n'(N\xi)}{J_n(N\xi)J_n'(\xi) - NJ_n(\xi)J_n'(N\xi)}\right]^{-1}. \tag{3}$$

In (3), N is the refractive index; J_n and N_n denote the Bessel and Neumann functions, respectively; $\xi = ka$; a is the cylinder radius; and the prime indicates the derivative with respect to argument.

We will apply the derived formulas to the calculation of heating of argon plasma. We will use for this purpose the following equilibrium correlations required to derive N: the frequency of electron-neutral collisions

$$\nu[\mathrm{s}^{-1}] = 5.74\times 10^{11}\frac{T_0}{T}\sqrt{T}\frac{p}{760},$$

the cylinder temperature T and the temperature outside the cylinder T_0 in 10^3 K, the gas pressure p in mm Hg, and the electron concentration [12]

$$n_e[\mathrm{cm}^{-3}] = 1.54\times 10^{20}\sqrt{\frac{p}{760}}\exp\left(-\frac{92.88}{T}\right).$$

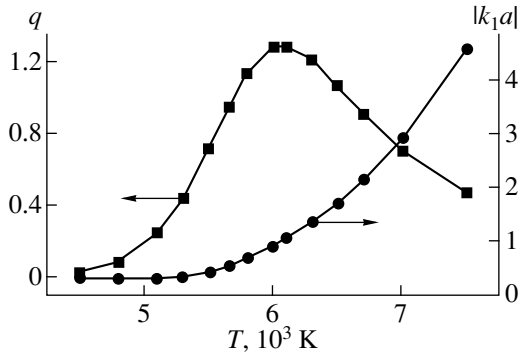

Fig. 8. The calculated dependence of the dimensionless cross section of power absorption in a cylinder and of the parameter of field penetration in this cylinder on the temperature with the cylinder radius of 0.2 cm; Ar, 300 mm Hg, 7 GHz.

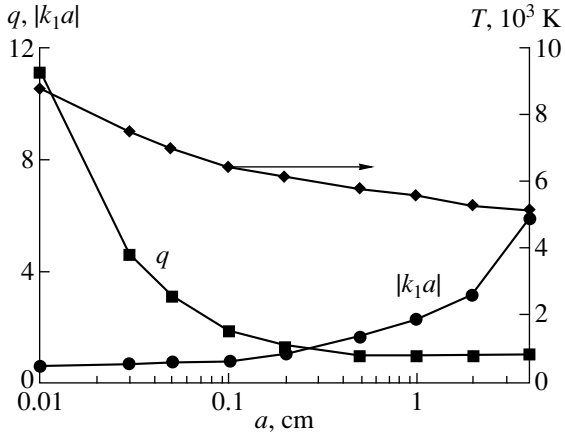

Fig. 9. The calculated dependence of the maximal absorption cross section q on the cylinder radius a. The temperature and the skinning parameter are given, which correspond to the maximal value of q. The parameters are as in Fig. 8.

Figure 8 gives the calculated temperature dependence of the dimensionless absorption cross section

$$q = \frac{\sigma_t - \sigma_s}{2a}$$

and that of the parameter $|Nka| = |k_1 a|$, which characterizes the penetration of the field deep into the cylinder, with the fixed value of the cylinder radius $a = 2$ mm for the wavelength of 4.2 cm used in the experiment (frequency of 7 GHz). One can see that the heating maximum is attained at a temperature at which the penetration parameter is of the order of unity; i.e., the distance of field decrease inside the cylinder is equal to the radius. We determine the maximal value of q and the respective value of temperature at every value of the cylinder radius a to construct the dependence given in Fig. 9. The same figure gives the values of the penetration parameter for the maximal values of q. One can see that, as the radius decreases, the temperature and the absorption cross section increase, with the penetration parameter remaining close to unity.

The flux density of heat from a cylinder surrounded by a wall of radius R_0 with a fixed temperature T_0 is

$$Q = \frac{\lambda(T - T_0)}{a \ln(R_0/a)},$$

where $\lambda[\text{erg/cm s K}] = 4 \times 10^{-4}\sqrt{T/T_0}$ is the thermal conductivity coefficient of argon. It is clear that a filament may be in thermal equilibrium $qS = Q$ only in the vicinity of the wall, when an adequate heat sink is possible. As the power S of incident wave decreases, i.e., as the threshold of extinction is approached, the heat balance may be satisfied by way of reducing the cylinder radius. In so doing, no restrictions on the filament diameter arise in the model being treated. This situation differs radically from the situation in the case of spherical microwave discharge, where such a model brings about a characteristic radius of plasma formation [13] coinciding with that predicted previously by Kapitza [1].

4.2. Inhomogeneous Cylinder in the Field of a Cylindrical Wave

Self-consistent solutions of the thermal and electrodynamic problems are required for determining the minimal radius of cylinder during extinction of a discharge. In this case, one can make use of the fact that, for a thin cylinder ($\xi \ll 1$), it is sufficient to include only zero harmonic ($n = 0$) in (3), i.e., solve a problem with cylindrical symmetry.

The set of equations has a standard form

$$\begin{aligned} \frac{1}{r}\frac{d}{dr}\left(r\lambda\frac{dT}{dr}\right) + \sigma E^2 &= 0, \\ \frac{1}{r}\frac{d}{dr}\left(r\frac{dE}{dr}\right) + \frac{\omega^2}{c^2}\varepsilon E &= 0, \end{aligned} \tag{4}$$

where

$$\sigma(T) = \frac{1}{4\pi}\frac{\omega_{pe}^2(T)\nu(T)}{\omega^2 + \nu^2(T)}$$

is the plasma conductivity,

$$\varepsilon = 1 - \frac{\omega_{pe}^2}{\omega(\omega + i\nu)}$$

is the permittivity, and ω_{pe} is the plasma frequency. It is assumed that the electric field is directed along the discharge. The heat equation may be replaced by the condition of cancellation of the heat flux and of the electromagnetic energy flux defined by Poynting's vector,

$$-\lambda\nabla_r T + S_r = 0,$$

where λ is the thermal conductivity, r is the radial coordinate, and S_r is the radial energy flux. The boundary conditions for the set of equations (4) are provided by the temperature T_0 on the wall of radius R_0 and the flow of cylindrical electromagnetic wave incident from infinity [14]. On the discharge axis, $\nabla_r T = S_r = 0$.

Figure 10 illustrates the result of characteristic calculation in the vicinity of the threshold of extinction, which demonstrates the existence of steady-state solutions of the problem with a filament less than a millimeter thick. Figure 11 gives the filament radius and the coefficient of reflection from the filament as functions of incident power at different distances from the wall. One can see the existence of the threshold of extinction, where the filament radius may be close to that which is observed. The level of threshold power likewise proves to be close to the experimentally obtained value. As the threshold is approached, the agreement between plasma and incident power is improved, which is also observed experimentally. Therefore, the simplest model predicts fairly well the qualitative characteristics of a discharge with a single filament provided the distance from the filament to the wall is preassigned. Apparently, the question of self-consistent setting of this distance cannot be resolved within the framework of equilibrium theory.

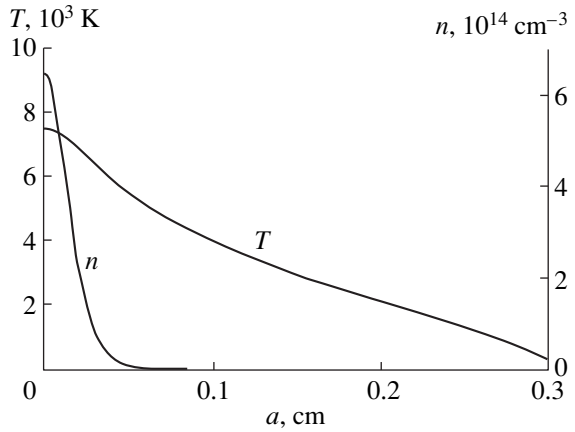

Fig. 10. The predicted profiles of the plasma electron density n and temperature T in the space from the axis to the wall (R_0 = 0.3 cm) at room temperature; Ar, 760 mm Hg, incident power per unit filament length 100 W/cm, 7 GHz.

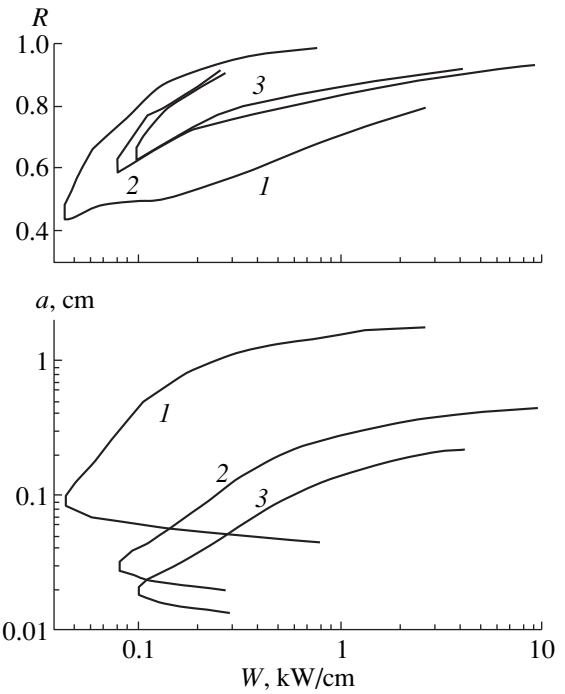

Fig. 11. The coefficient of reflection R and the filament radius a as functions of incident power W per unit filament length, for three values of the wall radius: (*1*) 2, (*2*) 0.5, (*3*) 0.3 cm. The rest of the parameters are as in Fig. 10.

4.3. Multiplicity of Filaments

The most complicated question is what causes the variation of the number of filaments. We believe that this effect is caused by the interference of electromagnetic fields scattered by the filaments.

We will treat an example with two filaments. Unlike the situation with one filament, each of two filaments is heated both by the incident wave and by the wave scattered from the neighboring filament. In so doing, the variation of the distance between the filaments leads to a variation of the phase with which these two waves interfere on the filament. Therefore, a distance must exist between the filaments, at which the power absorbed by two filaments proves to be maximal. One can estimate this distance fairly easily. In view of the fact that the necessary phase difference must be 2π and that, during scattering, a phase variation by approximately π occurs, the distance between the filaments must be half the wavelength in vacuum. It is important to note that, in this case, the absorption power exceeds the power that would result from a simple doubling of the power absorbed by a single filament. Therefore, given a certain level of microwave power input, the state with two filaments proves to be more advantageous energetically than the state with one filament. We will demonstrate that the same situation is observed for a larger number of filaments.

The diffraction of a linearly polarized plane electromagnetic wave on homogeneous infinite cylinders of radius a, spaced at equal distances L from one another in a plane perpendicular to the incident wave, is well studied [10]. We will use a simplified expression which may be derived from general formulas in the approximation of thin cylinders ($ka \ll 1$) spaced at fairly large distances from one another ($a \ll L$). In so doing, we will take into account only the zeroth harmonic ($n = 0$ in formula (3)) and the first order of diffraction; i.e., we will ignore the scattering of scattered waves. We derive the following relation for the coefficient K by which the dimensionless absorption cross section q for single

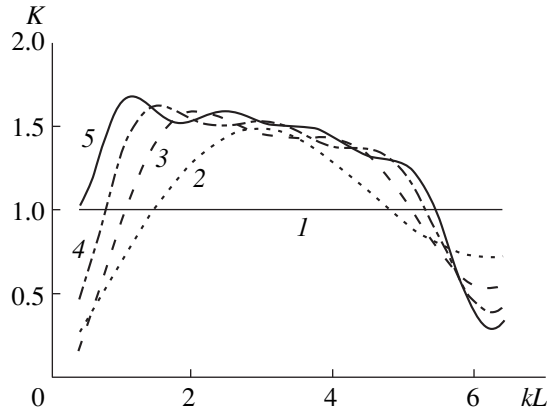

Fig. 12. The coefficient K (5) as a function of distance between filaments: the number of filaments, from one to five; Ar, 300 mm Hg, 7 GHz, T = 6500 K, a = 0.2 cm.

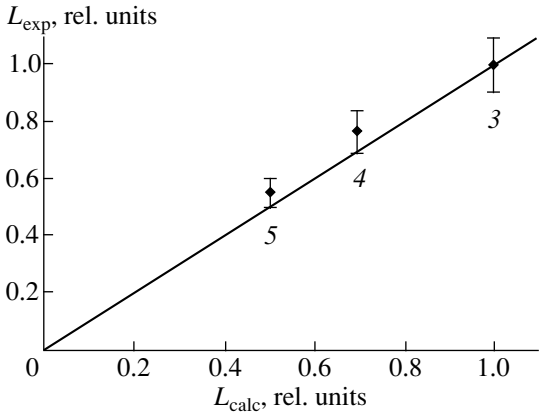

Fig. 13. Comparison of the calculated and measured values of the relative variation of the distance between filaments. The numerals indicate the number of filaments. The normalization is performed for three filaments.

filament must be multiplied:

$$K \approx \frac{1}{m}\sum_{j=1}^{m}\left|1 + {}^{1}c_0\sum_{i=1}^{m}\varepsilon H_0^{(1)}(|i-j|kL)\right|^2,$$
$$(5)$$
$${}^{1}c_0 = -\left\{1 + \frac{2i}{\pi}\left[\frac{J_0(N\xi)}{N\xi J_1(N\xi)} + \ln\left(\frac{\xi}{2}\right)\right]\right\}^{-1}.$$

Here, $\varepsilon = 0$ at $i = j$, and $\varepsilon = 1$ in all of the remaining cases; m is the number of filaments; and $H_0^{(1)}$ is zero-order Hankel function. Figure 12 illustrates an example of calculation of the coefficient K as a function of the distance between filaments for different number of filaments. One can see that the maximal value K_{max} is much higher than unity. The value K_{max} increases with the number of filaments. The distance between filaments at which K_{max} is realized decreases considerably as the number of filaments increases. This is in qualitative agreement with experiment, as is demonstrated by Fig. 13 which illustrates a comparison of the calculated and measured values of relative variation of the distance between filaments with increasing number of filaments. A fairly good agreement is attained in the case of three to five filaments. The agreement is disturbed in the case of two filaments and in the case of a large number of filaments, because, in the latter case, the calculation by formula (5) is less accurate due to violation of the $a \ll L$ approximation.

Figure 12 enables one to produce a qualitative interpretation of the observed phenomena. As the energy input to the plasma increases with increasing incident microwave power or with decreasing conduction (increasing gas pressure), the increase in the number of filaments accompanied with the reduction of the distance between them results in the possibility of absorbing more power. Because, given a sufficiently large number of filaments, the difference in the "advantageousness" for discharges with close numbers of filaments proves to be insignificant (the curves in Fig. 12 intersect), "jumpovers" may be observed.

In order to obtain the quantitative characteristics, one must completely solve the electrodynamic and thermal problems with due regard for the nonequilibrium state of plasma and for the geometry of the discharge chamber, i.e., for the presence of a waveguide and a fairly thick quartz tube. We hope that this will help us explain the substantial difference between the distance between two filaments predicted by formula (5) (approximately 2 cm) and the distance observed experimentally (about 0.8 cm). The mechanisms that cause the plasma front to propagate in the direction of the power source are well known [12]. In view of this, it is natural that the filaments concentrate in the vicinity of the surface which obstructs their motion. A complete solution of the problem for nonequilibrium discharge will enable one to self-consistently determine the distance from the filament to the wall.

5. CONCLUSION

A stationary microwave discharge in argon in a quartz tube passing through a waveguide assumes the shape of parallel filaments oriented along the electric field of the wave. Stationary filaments are obtained in the still atmosphere of argon in the vicinity of the wall facing the source of microwave power. In the case of convective motion of gas, a discharge of the "ball of filaments" type is formed inside the discharge tube.

A stationary filament has a diameter of about 1 mm. The number of stationary filaments increases from one to seven with increasing pressure (starting with 100 mm Hg) and increasing incident power (starting with 20 W). As the number of filaments increases, the distance between them decreases. The filaments are arranged symmetrically relative to the equatorial plane of the discharge tube. For rough estimation of the power absorbed by the

plasma, the number of observed filaments may be multiplied by 30 W. Experiments in the air proved to be impossible under conditions of high gas pressure because of the melting of the quartz tube.

Some properties of a stationary filamentary discharge in argon at high pressure may be explained qualitatively using the interference-equilibrium model and invoking the hypothesis that a discharge is organized such as to provide for the maximal power absorption. In order to obtain a quantitative agreement between the model and experiment, one must solve completely the self-consistent electrodynamic and thermal problems with due regard for the nonequilibrium state of the discharge and for the geometry of the waveguide.

ACKNOWLEDGMENTS

We are grateful to G.M. Batanov, V.B. Gil'denburg, and I.V. Sokolov for their interest in our study and discussion. P.M. Kosarev rendered great assistance to us in performing the experiments.

REFERENCES

1. P. L. Kapitsa, Dokl. Akad. Nauk SSSR **101**, 245 (1955).
2. J. Allison, A. L. Gullen, and A. Zavody, Nature **193**, 156 (1962).
3. A. M. Howatson, *An Introduction to Gas Discharges* (Pergamon, Oxford, 1976; Atomizdat, Moscow, 1980).
4. G. V. Bogomolov, Yu. D. Dubrovskiĭ, A. A. Letunov, and V. D. Peskov, Zh. Éksp. Teor. Fiz. **93**, 519 (1987) [Sov. Phys. JETP **66**, 295 (1987)].
5. A. L. Vikharev, V. B. Gil'denburg, S. V. Golubev, *et al.*, Zh. Éksp. Teor. Fiz. **94** (4), 136 (1988) [Sov. Phys. JETP **67**, 724 (1988)].
6. A. L. Vikharev, A. M. Gorbachev, A. V. Kim, *et al.*, Fiz. Plazmy **18**, 1064 (1992) [Sov. J. Plasma Phys. **18**, 554 (1992)].
7. V. B. Gil'denburg and A. V. Kim, Fiz. Plazmy **6**, 904 (1980) [Sov. J. Plasma Phys. **6**, 496 (1980)].
8. A. V. Kim and G. M. Fraĭman, Fiz. Plazmy **9**, 613 (1983) [Sov. J. Plasma Phys. **9**, 358 (1983)]
9. R. King and Tai Tsun Wu, *Scattering and Diffraction of Waves* (Harvard Univ. Press, Cambridge, 1959; Inostrannaya Literatura, Moscow, 1962).
10. E. A. Ivanov, *Diffraction of Electromagnetic Waves on Two Bodies* (Nauka i Tekhnika, Minsk, 1968).
11. V. V. Nikol'skiĭ and T. I. Nikol'skaya, *Electrodynamics and Transmission of Radiowaves* (Nauka, Moscow, 1989).
12. Yu. P. Raĭzer, *The Physics of Gas Discharge* (Nauka, Moscow, 1987).
13. A. A. Skovoroda, Zh. Éksp. Teor. Fiz. **112**, 877 (1997) [JETP **85**, 474 (1997)].
14. A. V. Timofeev, Fiz. Plazmy **23**, 176 (1997) [Plasma Phys. Rep. **23**, 158 (1997)].

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 86–92.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 99–106.

PLASMA, GASES

The Effect of the Gas Temperature Gradient on Dust Structures in a Glow-Discharge Plasma

V. V. Balabanov, L. M. Vasilyak*, S. P. Vetchinin, A. P. Nefedov, D. N. Polyakov, and V. E. Fortov

Institute of High Energy Density, IVTAN (Institute of High Temperatures) Scientific Association, Russian Academy of Sciences, Moscow, 127412 Russia

**e-mail: lab852@hedric.msk.su*

Received June 30, 2000

Abstract—An experimental investigation is performed of the effect of the neutral gas temperature gradient on plasma-dust formations in the positive column of a glow discharge. It is demonstrated that the thermophoretic forces arising due to the temperature gradient are comparable with radial electric fields and define the condition of formation and different shapes of plasma-dust structures, in particular, the formation of rings in the vicinity of tube walls. A model description of this effect is given. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The formation of ordered structures of micron-sized dust particles was observed under different conditions in a low-pressure plasma and in radio-frequency and glow discharges [1, 4–6], as well as in thermal and nuclear-excited plasmas [1–3]. In so doing, microparticles acquire a high negative charge $q = (10^4–10^5)e$ (where e is the electron charge), corresponding to the floating plasma potential, and the structures proper resemble Coulomb crystals [1–4]. The possibility of formation and the stability of ordered dust structures in a plasma depend on temperature. The nonideality parameter γ is proportional to the square of charge on particles,

$$\gamma = q^2 n_d^{1/3}/kT,$$

where n_d is the concentration of dust particles, and defines the temperatures at which a stable crystal structure of charged dust particles will be formed, the temperature at which structures of the liquid type will be formed, and the temperature at which no structures will be formed [1]. In addition to the value of the temperature proper, the formation of ordered dust structures must be further affected by forces associated with the temperature gradient. These forces are also capable of having decisive effect both on the conditions of emergence of ordered structures and on their geometric shape and arrangement in space. A plasma system is always characterized by the presence of sources of energy release and by the presence of boundaries; therefore, there always exist both the temperature gradient and the forces it causes. Up to now, the forces associated with fluxes and with interaction of charged components of plasma were largely included in the analyses and investigations of ordered dust formations [2, 3, 7, 8], while the forces due to the temperature gradient were disregarded.

Ordered plasma-dust formations in a dc glow discharge are formed in strata in which a fairly strong longitudinal electric field exists, which that makes it possible to contain particles in the field of gravity [1, 4–6]. Experimental results demonstrate [6] that plasma-dust structures of different shapes may form in a glow discharge. The shape and structure of dust formations depend on the conditions of equilibrium in the radial direction [6]. In the radial direction, the particles of a plasma-dust formation are acted upon by various forces directed towards the axis and the wall of the discharge tube. The force directed toward the axis is defined by the radial electric field and by the dust-particle charge. The forces urging the dust particles towards the walls may be caused by the motion of ions to the wall under conditions of ambipolar diffusion and by the neutral gas temperature gradient. The effect of temperature gradients due to Joule heating on ordered plasma structures has not been taken into account up to date. It is the objective of this study to investigate the effect of the temperature gradient on the formation of dust structures in the positive column of a glow discharge.

2. EXPERIMENT

Ordered dust structures were developed in the positive column of a glow discharge in discharge tubes 1 and 2 cm in diameter and 30 cm long. The experimental scheme is given in Fig. 1. Two metal rings *1* were glued into the tube walls at distances of 10 and 15 cm from the cathode for measuring the voltage drop in the positive column and for stabilizing the strata. Air at a pressure from 0.2 to 0.8 torr served as the working gas. The positive column was stratified, and the first stratum

emerged in the vicinity of the first ring. Introduced from above into the discharge were particles of alumina 3 to 10 μm in diameter and particles in the form of hollow glass spheres 20 to 60 μm in diameter. The discharge current varied from 0.1 to 3 mA.

The current-voltage characteristic of the positive column of a discharge in air is given in Fig. 2; also shown in this figure are the shapes assumed by dust structures at different values of current. At a low current of 0.1 to 0.3 mA, ordered filaments of particles approximately 1 cm long were observed, concentrated in the vicinity of the tube center. When the discharge current increased to 0.6–1 mA, we observed the formation of an ordered structure in the form of a cylinder with the diameter of approximately one-third or one-fourth of the tube diameter. The diameter of the ordered structure increased with current and reached two-thirds of the tube diameter at a current of 1–2 mA. The structure thickness decreased as the current increased, and, at a current from 1.5 to 2 mA, plane structures were observed consisting of several (five to ten) layers of particles. A further increase of current resulted in the formation of a ring structure whose diameter increased and the width decreased with increasing current. No particles were present in the axial region. The values of the current at which transitions occurred between different shapes of ordered structures decreased with increasing size of dust particles.

The following experiments were performed to check the effect of the temperature gradient on the structures.

1. After an ordered structure was formed, a rod *3* heated to 100°C (Fig. 1) was brought close to the side wall of the tube (opposite the ordered structure). After a period of 2–3 s, particles started to leave the structure and move away from the rod, and, after a short time, all particles from the structure moved over to the wall (Fig. 3). Approximately 10–20 s after this, if the rod was removed, the structure started to be built up gradually from the particles that came from the walls. In so doing, the center of localization of particles was first biased toward the cold wall and then shifted slowly toward the center.

2. If the wall was touched by the cooled (to –10°C) rod, the structure was set up in the form of a cone with the vertex at the point of contact between the cold rod and the wall and flowed down slowly (15 s) onto the wall (Fig. 4). After the rod was removed, the structure was built up in the same manner as that described for the case of heating.

3. In the third experiment, a heating coil *2* (Fig. 1) 0.5 cm wide was wound at a distance of 1 cm above the bottom metal ring. Under the experimental conditions, the heating largely affected the region between the metal rings. Because the heater width was small, it generated both the longitudinal and radial components of the temperature gradient. The discharge was ignited such that two strata were located between the metal rings above the heater. After switching on the heater at a power of 0.5 W (in so doing, the mean heat release in the discharge per unit length was 0.01–0.02 W/cm), the dust structure diameter in the stratum nearest to the heater decreased, and it was observed that particles

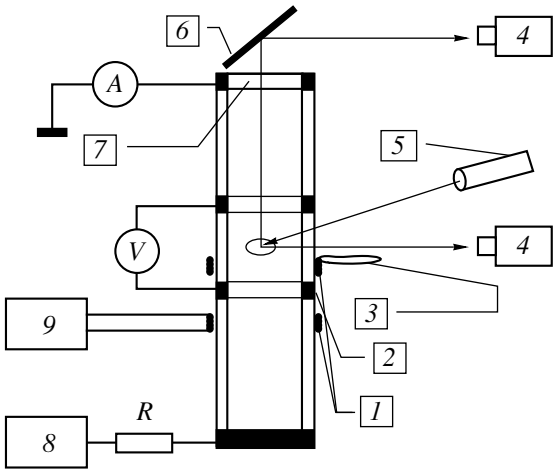

Fig. 1. Experimental scheme: (*1*) metal rings, (*2*) heater, (*3*) heated or cooled rod, (*4*) video camera, (*5*) laser, (*6*) mirror, (*7*) glass window, (*8*) discharge supply source, (*9*) heater supply source; *R*, ballast resistor.

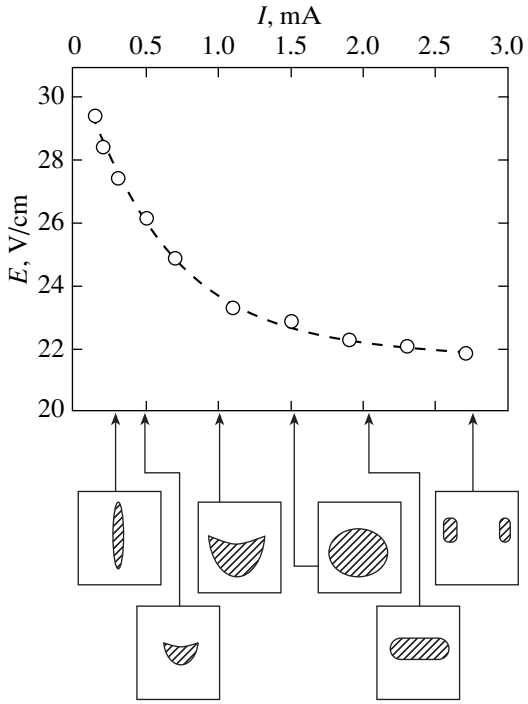

Fig. 2. The mean intensity of longitudinal electric field between rings as a function of the discharge current. Shown schematically at the bottom are the configurations of plasma-dust structures in the longitudinal cross section, corresponding to the above-identified values of current.

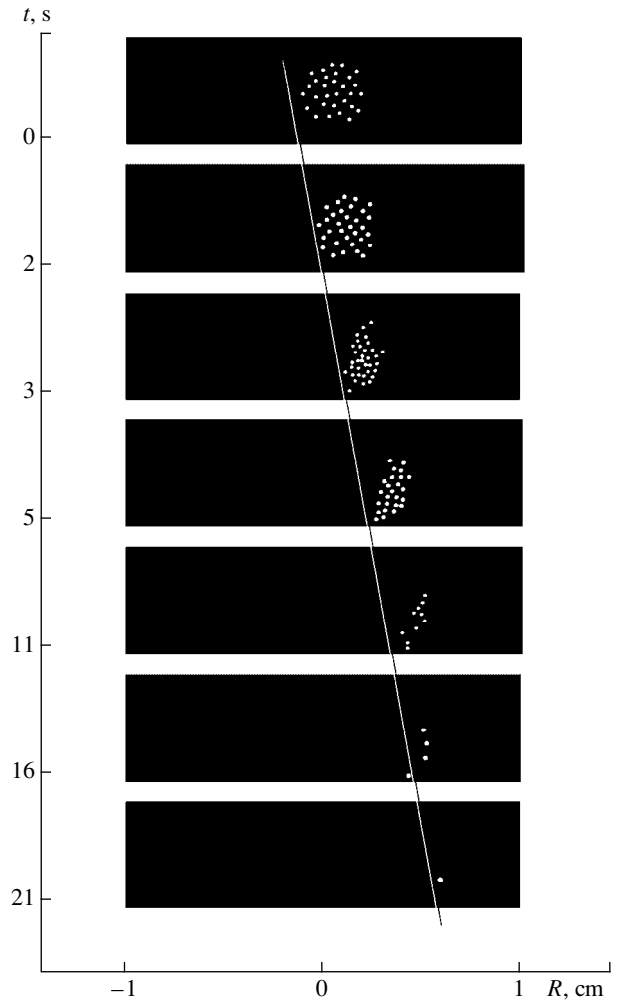

Fig. 3. Photographs of variation of the cross section of a dust structure upon contact of a heated rod with the discharge tube wall (on the left); *O*, tube center; *R*, tube radius; *T*, time of stimulation in seconds.

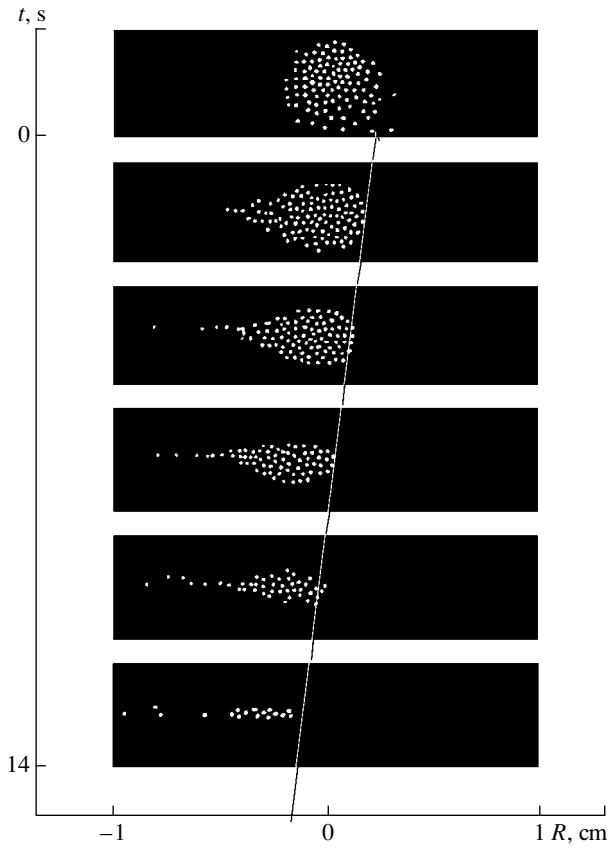

Fig. 4. Photographs of variation of the cross section of a dust structure upon contact of a cooled rod with the discharge tube wall (on the left); *O*, tube center; *R*, tube radius; *T*, time of stimulation in seconds.

occasionally flew from this stratum to the top one and back. When the heater power was increased to 1 W, particles from the bottom stratum were scattered into the region between the strata and occupied almost the entire volume. In so doing, the particles in the axial region were motionless and the distance between them increased; closer to the wall, the particles moved chaotically. Between these two regions, chaotically moving particles were observed among the motionless ones. This effect is possibly associated with the presence of particles of different sizes. In the vicinity of the wall closer to the heater, the particles performed circular motion. In the vicinity of the walls, they descended to the heater and then departed to the center, after which they ascended and moved towards the wall. This motion may be attributed to the presence of strong temperature gradients in the vicinity of the heater. When the power of heating increased to 2 W, the particles moved away from the heater and took up the top part of the region between the heater and the top ring. With a further increase of the heater power, the particles left the discharge. The effect of the rotation of particles was also observed in the case of contact between the wall tube and the cooled rod below the stratum with ordered structure. In so doing, the particles shifted towards the cooled spot (sideways and down), and the direction of particle rotation was opposite to that observed during heating.

Therefore, the presence of even insignificant temperature gradients may define the geometric shape of a dust structure, as well as fully distort or destroy the latter. The observed transition of an ordered dust structure from a disk or cylindrical shape to a ring shape with increasing discharge current may also be caused by a radial temperature gradient when the gas in the tube is heated by the flowing current.

3. CALCULATION MODEL

We will treat the steady state of a ring-shaped dust structure. The shape of the structure in the radial direction depends on the equilibrium position of dust particles, which is defined by the balance of radial forces.

The attracting force of the radial electric field of ambipolar diffusion is directed toward the center of the discharge tube, while the forces of entrainment by ambipolar ion and electron fluxes and the force caused by the temperature gradient are directed from the center to the wall of the tube. In calculating these forces, use was made of the model of the positive column of the plasma of a low-pressure glow discharge undisturbed by dust particles. The forces were calculated per single particle. Included in the calculations were the dependences of the reduced electric field intensity E/P (the values were borrowed from experiment, Fig. 2) and of the coefficients of diffusion and mobility of electrons and ions on the pressure and the tube radius [9]. In strata in the air, the longitudinal field intensity is high and the electron attachment is almost fully compensated by detachment; therefore, the concentration of negative ions in the strata is low compared with that of electrons. The electron temperature was taken to be 3 eV [10], because its variation has little effect on the calculation results. As was demonstrated by Nedospasov [11], a stratified column of a glow discharge is described by the ionization-diffusion model, and the radial distribution of the electron concentration in a stratum is the same as in a homogeneous column and is close to the Bessel function $J_0(r/\Lambda)$ with the boundary condition

$$n_e(R) = 0,$$

where $\Lambda = R/2.4$ and R is the tube radius. The loss of charged particles at the values of pressure $P = 0.2$–0.4 torr employed by us is largely defined by ambipolar diffusion, as a result of which a radial electric field arises. Because the drift velocity of electrons in the radial direction is low compared with their thermal velocity, the electron distribution over the discharge tube cross section satisfies the Boltzmann equation [12]:

$$e\varphi(r) = kT_e\ln(n_e(r)/n_e(0)), \quad (1)$$

where $\varphi(r)$ is the field potential at the space point being treated and, on the tube axis, $\varphi(0) = 0$. In view of the fact that in the diffusion approximation the radial distribution of electrons is described by the Bessel function, one can determine the potential at the preassigned space point,

$$e\varphi(r) = kT_e\ln(J_0(r/\Lambda)), \quad (2)$$

and then find the radial electric field $E_r = -d\varphi/dr$ and calculate the force acting on charged microparticles. This approach enables one to simplify the computation of radial electric field and avoid significant errors occurring when the field is found directly from an ambipolar flux of ions and electrons. Expression (2) is valid if the distance from the point being treated to the wall tube considerably exceeds the electron free path λ_e.

We expand the Bessel function in the vicinity of the wall at $R \gg \lambda_e$ into a Taylor series to derive from (2) the potential difference between the axis and the wall of the discharge tube,

$$e\varphi_W = kT_e\ln(R/\lambda_e). \quad (3)$$

Formula (3) is valid within a numerical factor on the order of unity under the logarithm sign.

Such distributions of the electron and ion concentration and of the radial electric field are distorted in the vicinity of the discharge tube walls where the value of excess space charge becomes comparable with that of ion concentration, as a result of which the quasi-neutrality of plasma is disturbed. Therefore, in calculations, we restricted the field in the vicinity of the tube walls where the quasi-neutrality was disturbed and allowed for the fact that the dust particles lost their charge in the vicinity of the tube walls and left the discharge.

In order to find the radial distribution of the gas temperature, we solved the heat-conduction problem for a plasma column with known sources of heat release that, in our case, were proportional to the concentration of electrons and to the longitudinal intensity of electric field.

Given below are predicted correlations for the main radial forces acting on dust particles, namely, the force of radial electric field, the force due to the temperature gradient, and the force of entrainment arising as a result of the ionic drift to the tube walls.

3.1. Force of Radial Electric Field

This force is equal to the product of particle charge by radial electric field,

$$F_E = -qE_r = qkT_e\frac{\nabla n_e}{n_e}, \quad (4)$$

where

$$q = aT_e\ln(0.4\sqrt{m_i/m_e})$$

is the equilibrium charge of a microparticle, and a is its radius. The dust particle charge is equal to $1.2 \times 10^4 e$. Figure 5 gives the radial electric force F_E as a function of distance.

3.2. Force of Temperature Gradient

If a temperature gradient is present in a gas, the body placed in this gas is acted upon by the force F_T which is proportional to the temperature gradient and caused by the sum of momenta imparted to a particle by bombarding molecules of gas. A microparticle moves along the line of temperature field toward a decrease in temperature. When the molecular free path of gas is

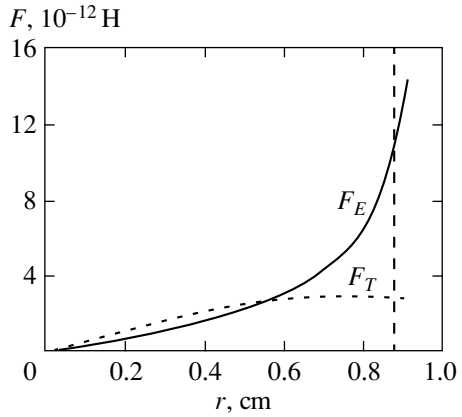

Fig. 5. The force of radial electric field and the thermophoretic force as functions of distance to the tube center at $P = 0.5$ torr and $I = 1$ mA; particle diameter, 10 μm.

much longer than the dust particle size, this thermophoretic force is defined by the formula [13]

$$F_T = -\frac{4PL}{T}a^2\frac{dT}{dr}, \tag{5}$$

where P is the gas pressure and L is the molecular free path. Given the current density and the longitudinal electric field, one can find the correlation for local heat release and then the temperature distribution and the thermophoretic force.

Because the longitudinal intensity of electric field is independent of radius, the radial heating of neutral gas is defined by the heat equation

$$\lambda\Delta T = -Q_0 J_0(r/\Lambda) \tag{6}$$

with nonuniform heat release, where the heat release on the discharge axis is $Q_0 = j(0)E$, $j(0)$ is the density of electric current on the axis, and λ is the thermal conductivity coefficient of gas. Convective heat transfer is negligibly small, because the Rayleigh criterion is invalid due to the low density of gas and insignificant temperature gradients. The solution of Eq. (6) may be written as

$$T(r) \approx \frac{1}{8}\frac{iE}{\lambda}J_0(r/\Lambda) + T_w, \tag{7}$$

where i is the total current of the discharge and T_w is the wall temperature. For the conditions of our experiment, the temperature on the axis exceeds the wall temperature by 5 to 15 K.

The $T(r)$ profile in this case coincides with the profile of heat release and defines the radial thermophoretic force acting on a microparticle (Fig. 5). In treating thermophoresis, we did not include the heat release associated with charged particle fluxes attracted by the dust cloud, because, in our case, these fluxes are much lower than both the radial ambipolar flux to the wall and the conduction current. One can see in Fig. 5 that the forces F_E and F_T are comparable in magnitude, act in different directions, and exhibit different dependences on the radius.

3.3. Ionic Drift and Its Effect on Particles

An ambipolar diffusion of electrons and ions occurs in a diffusion-controlled glow discharge. In so doing, the fluxes of ions and electrons and, consequently, their drift velocities coincide in the radial direction. Ion and electron fluxes affect a dust particle. The main contribution to this force is made by the ion momentum. The momentum transfer from ions to dust particles is defined by the velocity of ions in the vicinity of the surface of a charged particle, and the difference between ion fluxes affecting the particle is defined by the drift velocity of ions in a radial ambipolar field; therefore, the force with which a particle affected is

$$F_i = n_i m_i v_r v_i \sigma, \tag{8}$$

where n_i is the concentration of ions in a stratum, m_i is the mass of ions, v_r is the radial drift velocity, v_i is the ion velocity in the vicinity of a particle, and σ is the effective cross section of capture of ions by a charged microparticle [7].

The force of ionic entrainment (8) is an order of magnitude less than the force of radial electric field (4) and the thermophoretic force (5). Because the thermophoretic and ionic entrainment forces exhibit the same dependences on the particle size and on the discharge tube radius, one can assume that a dust structure is largely affected by the temperature gradient force. The thermophoretic force attempts to pull the particles held at the stratum center to the tube walls; however, it is inhibited by the force of radial ambipolar electric field.

3.4. Particle Energy

The motion of a dust particle in the radial direction is defined by its potential energy. Because the radial forces are proportional to the respective gradients, one can introduce the concept of potential energy for each one of these forces,

$$F(r) = -\frac{dU}{dr}. \tag{9}$$

Then, the total potential energy is

$$\begin{aligned} U(r) &= -\int_0^r [F_T(r) + F_E(r)]dr \\ &= -4PLa^2\ln\frac{T(0)}{T(r)} - qe\varphi(r). \end{aligned} \tag{10}$$

In our experiments, Joule loss is small, and the temperature on the tube axis $T(0)$ only slightly exceeds the wall temperature, $T(0) - T_w \ll T_w$; therefore,

$$U(r) = -\alpha(1 - J_0(r/\Lambda)) - \beta\ln(J_0(r/\Lambda)), \tag{11}$$

where

$$\alpha = \frac{1}{2}\frac{PLa^2 iE}{\lambda T_W}, \quad \beta = qkT_e. \tag{12}$$

In the vicinity of the axis, where $r/\Lambda \ll 1$,

$$U(r) \sim (\beta - \alpha) r^2/\Lambda^2.$$

If $\alpha/\beta \geq 1$, the root of $J_0(r_m/\Lambda) = \beta/\alpha$ defines the minimum of the function $U(r)$ (11),

$$U(r_m) = \beta(1 - \ln(\beta/\alpha)) - \alpha. \tag{13}$$

At the minimum, the value of $U(r_m)$ is always negative, $U(r_m) < 0$. In the vicinity of the axis, the potential energy is $U(r) < 0$, and then passes through the minimum. The value of potential energy (11) in the vicinity of the wall in view of φ_W is

$$U_W = -\alpha + \beta \ln(R/\lambda_e). \tag{14}$$

In the case when the electric forces prevail over the thermophoretic forces, $\beta > \alpha$, the potential energy $U(r) > 0$ has no local minimum, and the particles assemble in the vicinity of the axis. The behavior of the function $U(r)$ depending on r defines the region of finite motion of dust particles or departure to the wall. There are two characteristic regions, namely, $U(r) > 0$ and $U(r) < 0$. The particles go to the wall if their total energy is

$$\varepsilon = \sum_i \frac{m_i v_i^2}{2} + U(r) > U_W. \tag{15}$$

The summation is made over all particles in the layer in the entire region of motion. For finite motion, the total energy (15) is negative, and $U(r_m) < \varepsilon < U_W$. The particles "slide down" to the potential minimum.

Hence, it follows that two options are possible of radial structures of dust clouds in a glow discharge, namely,

(1) the particles take up either the central region or the entire cross section of the discharge tube, except for a small region of $\sim\lambda_e$ in the vicinity of the tube walls, and

(2) the particles are located at the potential minimum and form a space structure in the form of a ring in the neighborhood of the wall.

4. ANALYSIS AND COMPARISON WITH EXPERIMENT

We will determine the dependence of the forces of radial electric field and temperature gradient on the conditions of discharge, gas pressure, tube radius, current, and particle size. The force of radial electric field depends on the particle charge (proportionally), electron temperature, and radial electric field which increases with radius. The temperature gradient force is proportional to the cross-sectional area of particles and

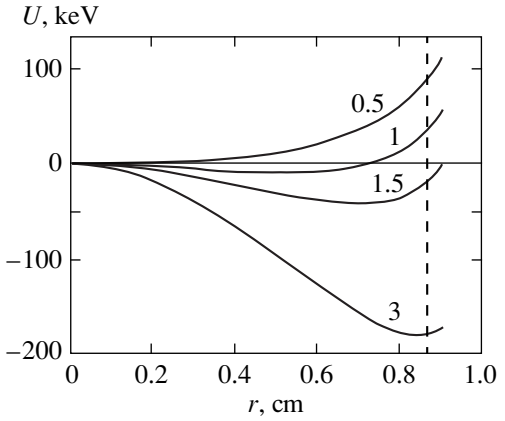

Fig. 6. The potential energy of charged dust particles 10 μm in diameter as a function of the distance from the tube center for different values of discharge current (0.5, 1, 1.5, and 3 mA).

increases with current and pressure, because the longitudinal electric field and heat release increase with pressure and decrease as the tube radius increases.

We will treat an ordered dust structure under conditions of variation of the discharge conditions. Figure 6 shows the variation of the potential energy of particles 10 μm in diameter as a function of discharge current. In the case of currents of up to 1 mA, it is advantageous for the particles to be in the central region of the tube, because a potential well is present at the tube center, as is observed in the experiment: the particles take up mainly the central part of the tube and form well-ordered structures. When the current increases (1–2 mA), the minimum of potential well shifts toward the walls; however, the potential energy of particles in the vicinity of the tube walls is higher than that at the center. In so doing, the particles may both form ring structures and take up almost the entire cross section of the tube except for the region in the vicinity of the tube walls. In this case, the size of the region taken up by the particles depends on the number of horizontal layers of dust particles that may be contained by the longitudinal electric field in a stratum. In order to calculate the space dimension of a dust structure, one must take into account the number of layers that may be contained in a stratum and, accordingly, solve a two-dimensional problem. The characteristic depth of potential well in the axial direction is defined by the potential drop on the stratum and, in our case, amounts to 20–40 V, and the radial potential drop $e\varphi_W$ does not exceed several values of T_e. Therefore, the depth of potential well in the longitudinal direction is much greater than that in the radial direction. In the case of weak fields in a stratum, the number of layers is small, and particles are absent from the central region; i.e., a wide ring structure is formed with the particles absent from a very narrow central part of the tube. In the case of fairly strong longitudinal fields in a stratum, when the number of layers is large, the particles take up almost the entire cross section of

the tube. In so doing, excess particles from the structure spill down. As the current continues to rise, the thermophoretic forces start playing the decisive part, the potential energy of particles at the tube center starts exceeding the energy in the vicinity of the tube walls, and the particles form only ring structures. The departure of excess particles also depends on the longitudinal electric field, and the particles either spill down or depart to the walls. The higher the current, the narrower the ring formed and the greater its inside diameter. At some value of the current, the potential barrier in the vicinity of the tube wall disappears and all particles in the stratum depart to the wall. The disintegration of a structure in the vicinity of the wall may occur because of a different reason. In the vicinity of the tube wall, the concentration of ions exceeds considerably that of electrons; this leads to a reduction of the particle charge, and the particles are not contained by the longitudinal field.

When the particle size changes, the values of critical current at which the structure disintegrates vary approximately inversely proportionally to the particle size. One can say that the less the particle size, the less the effect of the thermophoretic forces on the particles in the structure. The structure becomes more stable, the number of particles in the structure increases, and the range of currents at which the structure exists becomes wider. The effect of the gas pressure on the structure is defined by the variation of longitudinal field in the stratum and by heat release. The higher the pressure, the stronger the longitudinal electric field and, accordingly, the narrower the range of current values at which the structure may exist. At values of the gas pressure that are too low, the structure may not exist either, because a decrease in pressure leads to a reduction of the longitudinal field necessary to contain particles and to a reduction of the electron concentration in the discharge. Therefore, an optimal range of pressure exists for a "good" ordered structure, which depends on the tube radius and on the particle size.

The suggested model describes qualitatively all of the experimentally observed transitions between different forms of plasma-dust formations for dust particles of different sizes. A numerical comparison of predicted and experimentally obtained values of current, at which the observed transitions occur, has revealed a good agreement at a pressure of 0.8 to 1 torr. At a pressure from 0.3 to 0.5 torr, the qualitative pattern is maintained; however, a numerical discrepancy is observed. The numerical difference between experiment and theory is associated with other effects observed in the strata of a glow discharge, such as an inhomogeneity of electric and thermal fields along the tube axis, end effects, and the effect made on the plasma by the structure of dust particles proper. We are inclined to attribute the discrepancy between the prediction and experimental data to the effect of the cuplike shape of the stratum and, accordingly, to the presence of stronger radial fields than those in our calculations. In this case, the problem is two-dimensional. This assumption is favored by the fact that a transition to a ring structure was observed for particles 5 μm in size with a close-to-predicted value of current only at a pressure on the order of 1 torr when the shape of stratum became flatter.

5. CONCLUSION

In a bounded plasma with current, the thermophoretic forces associated with heat release may be of the same order of magnitude as the electric forces and, together with the electric forces, play the main part in the construction, stability, and disintegration of plasma-dust structures; in numerous cases, they define the structure and shape of dust formations and the conditions of their existence. The effect of thermophoresis may be used for removal and deposition (for example, onto a substrate) of charged particles and ordered structures, for separation of particles, in microelectronics, and for other applications. The forces of temperature gradient may be used in developing traps for containment of charged microparticles, for example, under conditions of microgravitation.

REFERENCES

1. A. P. Nefedov, O. F. Petrov, and V. E. Fortov, Usp. Fiz. Nauk **167**, 1215 (1997) [Phys. Usp. **40**, 1163 (1997)].
2. V. I. Molotkov, A. P. Nefedov, V. M. Torchinskiĭ, *et al.*, Zh. Éksp. Teor. Fiz. **115**, 837 (1999) [JETP **88**, 460 (1999)].
3. V. E. Fortov, A. P. Nefedov, V. I. Vladimirov, *et al.*, Phys. Lett. A **258**, 305 (1999).
4. V. E. Fortov, V. I. Molotkov, A. P. Nefedov, and O. F. Petrov, Phys. Plasmas **6**, 1759 (1999).
5. N. Sato, G. Uchida, R. Ozaki, *et al.*, in *Frontiers in Dusty Plasmas*, Ed. by Y. Nakamura, T. Yokota, and P. K. Shukla (Elsevier, Amsterdam, 2000).
6. A. P. Nefedov, L. M. Vasilyak, S. P. Vetchinin, and D. N. Polyakov, in *Proceedings of the Second International Conference on the Physics of Dusty Plasma*, Hakone, Japan, 1999, p. 104.
7. V. N. Tsytovich, Usp. Fiz. Nauk **167**, 57 (1997) [Phys. Usp. **40**, 53 (1997)].
8. A. M. Ignatov, Fiz. Plazmy **24**, 731 (1998) [Plasma Phys. Rep. **24**, 677 (1998)].
9. *Tables of Physical Quantities. Handbook*, Ed. by I. K. Kikoin (Atomizdat, Moscow, 1976).
10. Yu. P. Raĭzer, *The Physics of Gas Discharge* (Nauka, Moscow, 1992).
11. A. V. Nedospasov, Usp. Fiz. Nauk **94**, 439 (1968) [Sov. Phys. Usp. **11**, 174 (1968)].
12. B. M. Smirnov, *Physics of Weakly Ionized Gases* (Nauka, Moscow, 1972; Mir, Moscow, 1981).
13. H. L. Green and W. R. Lane, *Particulate Clouds: Dusts, Smokes and Mists* (Spon, London, 1964; Mir, Moscow, 1969).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 92, No. 1, 2001, pp. 93–99.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 1, 2001, pp. 107–114.

SOLIDS
Structure

Correlation Effects in the Heat Capacity of Separating ^{3}He–^{4}He Solid Solutions¶

T. N. Antsygina*, K. A. Chishko, and V. A. Slusarev

Institute for Low Temperature Physics and Engineering National Academy of Sciences of Ukraine, Kharkov, 310164 Ukraine
**e-mail: antsygina@ilt.kharkov.ua*

Received April 25, 2000

Abstract—We propose to describe the impurity heat capacity of solid ^{3}He–^{4}He mixtures both below and above the phase separation temperature T_s by an extension of the Bethe–Guggenheim approximation for the lattice gas model. It is shown that at $T > T_s$, the temperature behavior of the heat capacity is completely defined by correlation effects in the impurity subsystem. The developed theory enables us to explain from the common standpoint the experimental data by Edwards, McWilliams, and Daunt for all concentrations of ^{4}He and make some conclusions about the structure of second phase nuclei. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Solid mixtures of helium isotopes are known [1–3] to decompose into two (^{3}He-rich and ^{4}He-rich) phases on cooling below the phase separation temperature T_s, which depends on both the pressure and the mixture composition. The separation at $T = T_s$ is a first-order transition occurring with mass transport. It is clear that structure transformations of this type significantly affect thermodynamic properties of solid mixtures. In particular, a major contribution to the total heat capacity at $T < T_s$ is due to additional degrees of freedom of a heterophase system.

The heat capacity of separating solid ^{3}He–^{4}He mixtures has been experimentally investigated in close detail by several groups [1, 2, 4–7]. The most complete experimental results for a wide range of temperatures and concentrations n_0 of ^{4}He were reported by Edwards, McWilliams, and Daunt (EMD) [1, 2, 4]. Other concentrations were measured in [5–7].

In all the experiments, temperature dependences of the total heat capacity $C(T)$ are similar in character. At $T < T_s$, the experimental data for all impurity concentrations fall on a common curve. In this temperature region, $C(T)$ increases with the temperature. The total heat capacity drops sharply just above T_s and is completely determined by the lattice term $C_{\text{lat}} \propto T^3$ at $T > 0.3$ K. For $T_s < T < 0.3$ K, a deviation from the Debye T^3 law has been observed.

It is known that at $T < T_s$ (in the two-phase region), the thermodynamics of solid helium mixtures can be described very successfully within the regular solution model (the mean field approximation, MFA) [2, 8], but the behavior of $C(T)$ at $T > T_s$ cannot be explained by this model. In [9], an effort was made to take fluctuations of the local concentration n into account and an expression for the heat capacity of ^{3}He–^{4}He mixtures above T_s was obtained. The theory in [9] is in agreement with some experimental data in [1, 2], but the approach used in this work is not quite consistent. It has been critically discussed in great detail by Edwards and Pettersen (see [4]). Recently, the contribution of fluctuations to the heat capacity was calculated [10] in the framework of the random phase approximation (RPA) [11]. This allowed an adequate interpretation of the behavior of the concentrated mixtures in the entire temperature range. However, some additional considerations have been used to treat the limit of dilute solutions in a proper way [10]. Thus, it is clear that further improvements in the theoretical description of the thermodynamics of quantum solid ^{3}He–^{4}He mixtures remain urgent.

The purpose of the present work is to develop a rigorous theory that describes the temperature dependences of the heat capacity of separating helium solid solutions with arbitrary concentrations both below and above T_s. The proposed method is based on the quasichemical approximation [12, 13]. It enables us to properly take the contribution of fluctuations into account and make some conclusions about the structure of the second-phase nuclei. As shown in Section 4, our theory is in good agreement with the experimental data [1, 2]. It is significant that this agreement can be obtained without invoking any additional concepts about extraneous centers of a new phase nucleation such as dislocations, grain boundaries, and other lattice defects. We note that suggestions of this type were used in [14] to explain experimental results in [7].

2. GENERAL FORMALISM

The basis of the present theory is the lattice gas model [12]. Some simplifications can be made in applying this

¶This article was submitted by the authors in English.

model to quantum solid mixtures of helium isotopes (which is in contrast to classical systems). First, we are interested in effects that occur at temperatures considerably below the Debye temperature Θ_D. This allows us to neglect thermal vibrations of helium atoms and consider parameters of the interaction between impurities as temperature independent. Second, despite the low temperatures, the system approaches the thermodynamic equilibrium rather rapidly because of an intense quantum motion of helium atoms.

A. Hamiltonian

For definiteness, we hereafter treat the ^{4}He component as an impurity subsystem in the mixture. The Hamiltonian of the system has the form

$$H = \varepsilon_0 \sum_{\mathbf{f}} n_{\mathbf{f}} - \frac{1}{2} \sum_{\mathbf{ff'}} V_{\mathbf{ff'}} n_{\mathbf{f}} n_{\mathbf{f'}}, \quad (1)$$

where $n_{\mathbf{f}}$ is the occupation number of ^{4}He impurity atoms at a lattice site $\mathbf{f}$, ε_0 is the energy of a ^{4}He atom in the ^{3}He matrix, and $V_{\mathbf{ff'}} = V(|\mathbf{f} - \mathbf{f'}|)$ is the interaction energy of impurities placed at the sites $\mathbf{f}$ and $\mathbf{f'}$. Hamiltonian (1) can be represented in a somewhat different form for the following reason. In the n_0 versus T coordinates, where n_0 is the average ^{4}He concentration, the phase separation curve of solid ^{3}He–^{4}He mixtures is practically symmetrical about $n_0 = 0.5$ (see [4]) and has the maximum T_c at this point. The critical temperature T_c is the highest temperature at which phase separation occurs [3]. This means that the Hamiltonian must be explicitly invariant under the replacement $n_{\mathbf{f}} \longrightarrow 1 - n_{\mathbf{f}}$, and we can therefore rewrite Eq. (1) as

$$H = \frac{1}{2} \sum_{\mathbf{ff'}} V_{\mathbf{ff'}} n_{\mathbf{f}} (1 - n_{\mathbf{f'}}), \quad \varepsilon_0 = \frac{1}{2} \sum_{\mathbf{f}} V_{\mathbf{ff'}}. \quad (2)$$

As can be seen from (2), it is irrelevant whether ^{3}He or ^{4}He is chosen as the impurity subsystem. We also note that Hamiltonian (2) can be reduced to the Ising model with the spin 1/2 [12].

B. Interaction between Impurities

Because of a difference in the atomic volumes of helium isotopes, a ^{4}He impurity is the dilatation center in the ^{3}He matrix [4]. It is known [15] that the interaction between two dilatation centers is long-range; at a large separation r, it decreases as $1/r^3$ for cubic and hexagonal lattices. Moreover, it is anisotropic with a complicated angular dependence [16–18]. If many impurities exist in the matrix, their elastic fields overlap and the effective interaction becomes essentially isotropic but remains long-range. To describe the interaction $V_{\mathbf{ff'}}$ in the mixtures with arbitrary concentrations of ^{4}He, we use a rather simple model well known in the literature [11]. Namely, the interaction of an impurity atom placed at a lattice site $\mathbf{f}$ with another impurity atom is assumed to be constant (equal to V_0) if the distance between them is less than R_0 and equal to zero otherwise. Thus,

$$V_{\mathbf{ff'}} = \begin{cases} V_0 & \text{if} \quad |\mathbf{f} - \mathbf{f'}| \le R_0 \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

We introduce the effective coordination number z that can be considered as the number of impurities interacting with the given impurity atom. The effective radius R_0 of the impurity–impurity interaction is related to z as, $R_0 = (3\omega_0 z/4\pi)^{1/3}$ (where ω_0 is the volume per atom). In the present theory, z is considered as a fitting parameter.

C. Quasi-Chemical Approximation

Although the exact solution for the lattice gas model is not available in the three-dimensional case, there are several rather efficient approximate methods for its analysis. To analyze the model, we use the Bethe–Guggenheim method [12, 13], also known as the quasi-chemical approximation (QCA). It provides a sufficiently accurate description for the system under study. In this approximation, the lattice symmetry is of significance and the final results only depend on the effective coordination number z. This method is precise in the limit of dilute solutions. It also gives an accurate result for artificial models of the Bethe lattice type.

In the QCA framework, a two-component solid mixture involving A-type and B-type atoms can be represented as a set of independent atomic pairs of the AA, BB, and AB types. The partition function has the form [12]

$$Z_0 = \sum_N y^N \sum_{N_{AB}} g(N, N_{AB}, \mathcal{B}) \exp(-E(N_{AB})/T), \quad (4)$$

where $\mathcal{B}$ is the total number of lattice sites, N is the number of sites occupied by A-type atoms (^{4}He impurities), $y = \exp(\mu/T)$, and μ is the chemical potential of the impurity subsystem. We choose the units such that the Boltzmann constant $k_B = 1$. The function $g(N, N_{AB}, \mathcal{B})$ is the number of configurations containing N impurities and N_{AB} pairs. Each of these configurations has the energy

$$E(N_{AB}) = \frac{V_0}{2} N_{AB}.$$

Within the QCA, the function g is assumed to be

$$g = G \frac{(z\mathcal{B}/2)!}{N_{AA}! N_{BB}! [(N_{AB}/2)!]^2}, \quad (5)$$

where

$$N_{AA} = \frac{1}{2}(zN - N_{AB}), \quad N_{BB} = \frac{1}{2}[z(\mathcal{B} - N) - N_{AB}]$$

are the respective total numbers of the AA-type and BB-type pairs. The constant G is determined by the obvious relation

$$\sum_{N_{AB}} g(N, N_{AB}, \mathcal{B}) = \frac{\mathcal{B}!}{N!(\mathcal{B}-N)!}.$$

Setting the number of unlike pairs N_{AB} entering Eqs. (4) and (5) equal to its average value in the mixture, we obtain the well-known mean-field approximation [12]. In the limit of $\mathcal{B}, N, N_{AB} \longrightarrow \infty$ with $N/\mathcal{B}$ and $N_{AB}/\mathcal{B}$ constant, the sum in Eq. (4) can be replaced by its maximum term. We let

$$\frac{N}{\mathcal{B}} \longrightarrow n, \quad \frac{2N_{AB}}{z\mathcal{B}} \longrightarrow a.$$

The quantities n and a are chosen such that they vary from zero to unity. We now introduce the long-range order parameter σ related to the local impurity concentration n by $\sigma = 1 - 2n$. We note that the quantity a is related to the short-order parameter x introduced in [19] by $a = (1 + x)/2$.

Using the Stirling formula, we obtain the thermodynamic potential Ω_0 (per site) as

$$\Omega_0 = -\frac{T}{\mathcal{B}}\ln Z_0 = E - TS - \frac{\mu}{2}(1-\sigma), \tag{6}$$

where E is the internal energy per site and S is the entropy per site. In accordance with the above-mentioned approximations, we find

$$E = \frac{V_0}{4}za, \tag{7}$$

$$\begin{aligned} S = \frac{1}{\mathcal{B}}\ln g = \left(1-\frac{z}{2}\right)\ln 2 + \frac{z-1}{2} \\ \times[(1-\sigma)\ln(1-\sigma) + (1+\sigma)\ln(1+\sigma)] \\ -\frac{z}{4}[(1-\sigma-a)\ln(1-\sigma-a) \\ + (1+\sigma-a)\ln(1+\sigma-a) + 2a\ln a]. \end{aligned} \tag{8}$$

The parameters σ and a satisfy the set of equations

$$(z-1)\ln\frac{1-\sigma}{1+\sigma} + \frac{z}{2}\ln\frac{1+\sigma-a}{1-\sigma-a} + \frac{\mu}{T} = 0, \tag{9}$$

$$\ln[(1-a)^2 - \sigma^2] - 2\ln a - \frac{V_0}{T} = 0. \tag{10}$$

If we put $\mu = 0$, the system of Eqs. (9) and (10) is invariant under the inversion $\sigma \longrightarrow -\sigma$, which means that the two-phase mixture occurs. One phase (^{4}He-weak) corresponds to $\sigma > 0$ ($0 < n < 1/2$) and the other (^{4}He-rich), to $\sigma < 0$ ($1/2 < n < 1$). It follows from Eq. (10) that

$$\begin{aligned} a = \frac{1-\sigma^2}{\gamma+1}, \\ \gamma = \left\{1 + (1-\sigma^2)\left[\exp\left(\frac{V_0}{T}\right) - 1\right]\right\}^{1/2}. \end{aligned} \tag{11}$$

Inserting (11) in (9), we obtain

$$y \equiv \exp(\mu/T) = \frac{1-\sigma}{1+\sigma}\left[\frac{(1+\sigma)(\gamma-\sigma)}{(1-\sigma)(\gamma+\sigma)}\right]^{z/2}, \tag{12}$$

where y is equal to unity for all temperatures below T_s.

As the temperature is increased, the long-range order parameter decreases and becomes equal to $\sigma_0 = 1 - 2n_0$ at $T = T_s$, where n_0 is the equilibrium concentration given by the coexistence curve. It follows from (12) that the phase separation temperature T_s is given by

$$T_s^{-1} = \frac{2}{V_0}\ln\left[\frac{1-u^z}{u(1-u^{z-2})}\right], \quad u = \left(\frac{1-\sigma_0}{1+\sigma_0}\right)^{1/z}. \tag{13}$$

For $z \longrightarrow \infty$, $V_0 \longrightarrow 0$, and $V_0 z \longrightarrow$ const, Eq. (13) reduces to a well-known expression for T_s resulting from the regular solution model [11]. At $T > T_s$, the mixture becomes homogeneous with the constant long-range order parameter σ_0. Equations (9) and (10) then determine the temperature dependences of the chemical potential μ and the parameter a.

3. FLUCTUATION EFFECTS

To make the results of the quasi-chemical approximation more precise, spatially inhomogeneous fluctuations in parameters σ and a must be taken into account. Assuming that the spatial scale of fluctuations is much larger than the lattice parameter, we consider these fluctuations in the continuum approximation. We let

$$\sigma \longrightarrow \sigma + \eta(\mathbf{r}), \quad a \longrightarrow a + \alpha(\mathbf{r}), \tag{14}$$

where the variations $\eta(\mathbf{r})$ and $\alpha(\mathbf{r})$ are functions of the coordinate $\mathbf{r}$ satisfying the conditions

$$\int d\mathbf{r}\,\eta(\mathbf{r}) = 0, \quad \int d\mathbf{r}\,\alpha(\mathbf{r}) = 0. \tag{15}$$

Assuming that the variations of the long-range order parameter and the parameter a from their equilibrium values are sufficiently small, we expand (6) in $\eta(\mathbf{r})$ and $\alpha(\mathbf{r})$. The partition function of the system can then be written as

$$Z = Z_0 Z_{fl},$$

where Z_0 is the partition function in the QCA given by Eq. (4). The contribution from fluctuations Z_{fl} is given by the functional integral

$$Z_{fl} = A\int\int \mathcal{D}[\eta(\mathbf{r})]\mathcal{D}[\alpha(\mathbf{r})]\exp\left(-\frac{\delta\Omega}{T}\right), \tag{16}$$

where the normalization factor A is determined below. The quantity $\delta\Omega$ consists of two terms,

$$\delta\Omega = \delta\Omega_1 + \delta\Omega_2. \tag{17}$$

The first one is the series expansion of (6) to the second order in $\eta(\mathbf{r})$ and $\alpha(\mathbf{r})$. It is given by

$$\delta\Omega_1 = \frac{Tz}{4\omega_0 ab}\int d\mathbf{r}\{ f_{11}(\sigma, a)\eta^2(\mathbf{r}) + 2 f_{12}(\sigma, a)\eta(\mathbf{r})\alpha(\mathbf{r}) + f_{22}(\sigma, a)\alpha^2(\mathbf{r})\}, \tag{18}$$

where

$$f_{11} = \frac{ab}{1-\sigma^2}(2 - z + 2a^2 f_{22}), \qquad f_{22} = 1 - a - \sigma^2, \quad f_{12} = a\sigma, \tag{19}$$

and $b = (1-a)^2 - \sigma^2$. The second term in (17), $\delta\Omega_2$, was introduced for the following reason. Because of spatial inhomogeneity, an expansion of the thermodynamic potential must contain not only powers of $\eta(\mathbf{r})$ and $\alpha(\mathbf{r})$ but also their spatial derivatives. These latter are introduced by $\delta\Omega_2$ to suppress short-wave spatial fluctuations that have no physical meaning. To understand the structure of $\delta\Omega_2$, we confine ourselves to the simplest case. Namely, assuming that the form of $\delta\Omega_2$ is identical to the MFA one, we obtain

$$\delta\Omega_2 = -\frac{1}{8\omega_0^2}\int d\mathbf{r}\int d\mathbf{r}' V(\mathbf{r}-\mathbf{r}')\eta(\mathbf{r})\eta(\mathbf{r}'). \tag{20}$$

However, nonlocal term (20) written in this form gives a nonzero contribution to $\delta\Omega$ in the spatially homogeneous case, i.e., for $\eta(\mathbf{r}) = \text{const}$. Thus, expression (16) includes fluctuations of the homogeneous system that have already been taken into account in deriving formulas of Section 2. In Eq. (20), we must therefore separate the part that vanishes at $\eta(\mathbf{r}) = \text{const}$ and, consequently, contains the contribution from spatially inhomogeneous fluctuations only. Thus, $\delta\Omega_2$ can finally be written as

$$\delta\Omega_2 = \frac{1}{16\omega_0^2}\int d\mathbf{r}\int d\mathbf{r}' V(\mathbf{r}-\mathbf{r}')[\eta(\mathbf{r}) - \eta(\mathbf{r}')]^2. \tag{21}$$

We note that fluctuations of the long-order and short-order parameters are related to each other. Therefore, although Eq. (21) only describes nonlocality in the long-range order parameter, the spatial fluctuations of the short-order parameter are also smoothed out. Because of conditions (15), $\delta\Omega$ does not contain linear terms in $\eta(\mathbf{r})$ and $\alpha(\mathbf{r})$.

In view of (18) and (21), the right-hand side of Eq. (16) represents a Gaussian functional integral, which can be easily evaluated [20]. The integration yields

$$Z_{fl} = A\left(\frac{8bQ}{\gamma z}\right)^{\mathcal{B}z/2}\prod_{\mathbf{q}}\left[1 - Q\frac{v(\mathbf{q})}{T}\right]^{-1/2}, \tag{22}$$

where $v(\mathbf{q})$ is the Fourier transform of $V(\mathbf{r})$ and we have introduced the notation

$$Q = \frac{T}{V_0 z\lambda}, \qquad \lambda = 1 + \frac{2T}{V_0\gamma(1-\sigma^2)}\left(1 - \gamma\frac{z-2}{z}\right). \tag{23}$$

For $z \gg 1$, Eq. (22) goes into the well-known expression for Z_{fl} corresponding to the random-phase approximation [11]. This allows us to determine the factor A in (22) as

$$A = \left(\frac{\gamma z}{8bQ}\right)^{\mathcal{B}z/2}.$$

We thus obtain the contribution to the thermodynamic potential $\Delta\Omega$ related to fluctuations,

$$\Delta\Omega = -\frac{T}{\mathcal{B}}\ln Z_{fl} = \frac{T}{2\mathcal{B}}\sum_{\mathbf{q}}\ln\left[1 - Q\frac{v(\mathbf{q})}{T}\right].$$

The contribution of fluctuations to the internal energy is given by

$$\Delta E = -T^2\frac{\partial}{\partial T}\left(\frac{\Delta\Omega}{T}\right) = -\frac{1}{2T\mathcal{B}}\sum_{\mathbf{q}}\frac{Q^2 v^2(\mathbf{q})}{1 - Q v(\mathbf{q})/T}. \tag{24}$$

In writing Eq. (24), we used the condition $\sum_{\mathbf{q}} v(\mathbf{q}) = 0$ to eliminate the terms that are responsible for the self-action of impurities.

In our model of the impurity–impurity interaction, the Fourier transform $v(\mathbf{q})$ is given by

$$v(\mathbf{q}) = V_0 z f(qR_0), \quad f(x) = \frac{3}{x^3}(\sin x - x\cos x).$$

On the interval from 0 to the fist zero of $f(x)$, this function can be approximated to a sufficient accuracy as

$$f(x) \approx 1 - \left(\frac{x}{4}\right)^2,$$

and $f(x) = 0$ otherwise. This gives

$$v(\mathbf{q}) = \begin{cases} V_0 z[1 - (q/q_0)^2], & q \le q_0, \\ 0, & q > q_0. \end{cases} \tag{25}$$

The quantity q_0 is related to the effective coordination number z by

$$q_0 = \frac{4}{R_0} = 4\left(\frac{4\pi}{3z\omega_0}\right)^{1/3}. \tag{26}$$

Replacing summation by integration in Eq. (24) and using (25) and (26), we obtain

$$\Delta E = \frac{64T}{9\pi z} \times\left\{1-3\lambda\left(1-\sqrt{\lambda-1}\arctan\frac{1}{\sqrt{\lambda-1}}\right)+\frac{2}{5\lambda}\right\}, \quad (27)$$

where λ is defined by Eq. (23).

4. RESULTS AND DISCUSSION

The total heat capacity of the mixture is equal to

$$C = C_{\rm imp} + C_{\rm lat} + C_{\rm ex}, \quad C_{\rm imp} = C_0 + \Delta C,$$

where

$$C_0 = \frac{V_0 z}{4T\gamma}\left(\frac{V_0 b}{2T} - T\sigma\frac{d\sigma}{dT}\right) \quad (28)$$

is the heat capacity in the quasi-chemical approximation. The contribution of fluctuations ΔC is given by

$$\Delta C = \frac{\Delta E}{T} + \left[\left(\frac{3}{2}-\frac{1}{\lambda}\right)\frac{\Delta E}{T} + \frac{256}{45\pi z\lambda^2}\right] \times\left\{1+\frac{1}{\gamma(\lambda-1)(1-\sigma^2)}\left[1-\frac{\sigma^2}{\gamma^2}\right.\right. \quad (29)$$

$$\left.\left.+\frac{4\sigma T^2}{V_0(1-\sigma^2)}\left(\frac{3}{2}-\gamma\frac{z-2}{z}-\frac{1}{2\gamma^2}\right)\frac{d\sigma}{dT}\right]\right\}.$$

Using (12), we find

$$T\frac{d\sigma}{dT} = \begin{cases}-\sigma/[\gamma(\lambda-1)], & T<T_s,\\ 0, & T>T_s.\end{cases}$$

To obtain $C(T)$, we have added two terms to the impurity contribution $C_{\rm imp}$. The first term is the lattice heat capacity $C_{\rm lat} = (12\pi^4/5)(T/\Theta_D)^3$, and the second term is associated with the exchange interaction between ^{3}He atoms. It can be represented as [21]

$$C_{\rm ex} = \frac{3}{4}(1+\sigma)^2\left(\frac{I}{T}\right)^2\left[1-\frac{I}{T}+\frac{7}{4}\left(\frac{I}{T}\right)^2\right], \quad (30)$$

where I is the exchange integral. Term (30) is only essential at very low temperatures. In comparing our results with the experimental data, we always take $C_{\rm lat}$ and $C_{\rm ex}$ into account. We note that because term (30) is small, we neglect the pressure dependence of I and set I equal to its maximum value ≈1 mK at the molar volume 24.2 cm^3/mol (see [21]). In addition, we neglect the temperature dependence of Θ_D because the corresponding correction is inessential for the temperature range where the relevant effects take place. We put $\Theta_D = 17.3$ K in accordance with [21]. This value of Θ_D provides the best agreement with the experimental data [1, 2] above $T \approx 0.3$ K, where the contribution $C_{\rm lat}$ prevails.

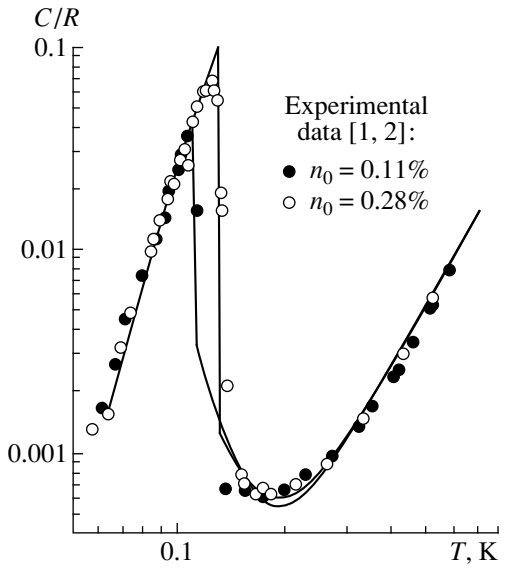

Fig. 1. Heat capacity of dilute ^{3}He–^{4}He mixtures as a function of temperature. The solid curves correspond to the present theory.

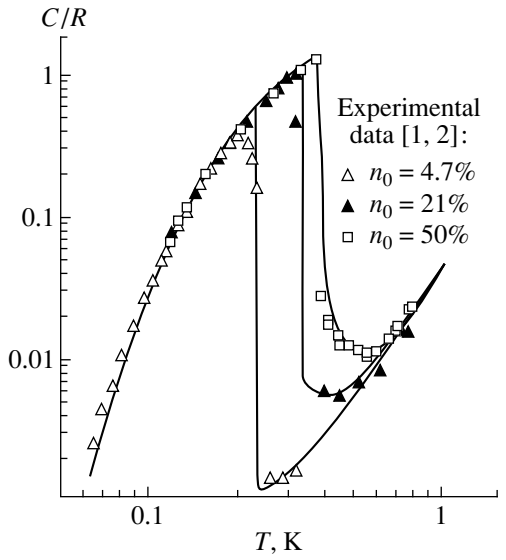

Fig. 2. Heat capacity of concentrated ^{3}He–^{4}He solid mixtures as a function of temperature. The solid curves correspond to the present theory.

The present calculations are compared with the EMD experimental data in Figs. 1 and 2. As can be seen from the figures, the agreement between the theory and experiment is very good. To make the theory fit the data in [1, 2], the parameters of potential (3) must be specified. The potential is determined by the intensity V_0 and the effective interaction radius R_0 (or by z directly

related to R_0). Both quantities must be treated as adjustable parameters. We note, however, that it is appropriate to choose the first fitting parameter as the product $V_0 z = v(0)$ rather than V_0. This choice is dictated by the following reasons. At low temperatures, the impurity heat capacity can be obtained as a low-temperature series expansion [12, 22] without using any theoretical approximations. The leading term of this expansion is given by

$$C_{\text{imp}} \approx \left(\frac{V_0 z}{2T}\right)^2 \exp\left(-\frac{V_0 z}{2T}\right). \quad (31)$$

Expression (31) is valid for all the mixtures irrespective of their concentration. It follows from Eq. (31) that the quantity C_{imp} depends only on the product $V_0 z$. Furthermore, as shown in [1, 2], the thermodynamics of the system at $T < T_s$ can be satisfactorily described within the MFA, which corresponds to the limit as $z \longrightarrow \infty$ and $V_0 \longrightarrow 0$ with $V_0 z \longrightarrow$ const (see, e.g., [11, 12]). In this approximation, the quantity $V_0 z$ is also the only parameter determining the behavior of the system at $T < T_s$. Thus, it is evident that the product $V_0 z$ plays the role of a universal parameter of the theory.[1] For all the mixtures, the value $V_0 z$ is assumed to be equal to 1.51 K, which is determined by the condition that the impurity heat capacity (31) fit the experimental data. The magnitude of $V_0 z$ is in complete agreement with the Edwards–Balibar constant $A = V_0 z/2$ that was found empirically to be 0.76 K [2–4]. The theoretical value of A and its pressure dependence were calculated by Mullin [8].

As the second independent adjustable parameter, we choose the value z. Precisely this parameter characterizes the spatial correlation scale between impurities. Numerical analysis shows that the heat capacity $C_{\text{imp}}(T)$ is highly sensitive to the magnitude of z in the fluctuation region (at $T > T_s$). This makes it possible to unambiguously determine the value of the effective coordination number for both dilute and concentrated mixtures.

Our analysis shows that the entire concentration interval ($n_0 \leq 50\%$) considered in [1, 2] can be conveniently divided into two parts: large concentrations $n_0 \geq 4.7\%$ and small concentrations $n_0 = 0.11\%$, 0.28%. The fitting parameter z is equal to 250 for all mixtures of the first group. This implies that the nearest neighbor approximation is inadequate for interpreting the thermodynamics of mixtures with large concentrations. In this case, the long-range interaction is therefore very significant. The value $z = 250$ implies that the radius of interaction between impurities spreads over four or five coordination spheres. Within the QCA, the critical temperature T_c is equal to [12]

$$T_c = -\frac{V_0}{2\ln(1-2/z)} \approx \frac{V_0 z}{4} \approx 0.38 \text{ K},$$

which is in excellent agreement with the experimental value.

For mixtures with small concentrations, the agreement between the theory and the experimental data can be reached for a unique choice of z, namely $z = 3$ for $n = 0.11\%$ and $z = 5$ for $n = 0.28\%$. We note that these values of z do not correspond to any regular three-dimensional crystal structure. This suggests that solid ^{3}He–^{4}He mixtures with small n_0 separate to yield the second (^{4}He-rich) phase in the form of a fractal (dendritic) structure.

There are several arguments supporting this assumption. The analysis shows that for low-concentration mixtures, the contribution ΔC of fluctuations is negligible. Thus, the heat capacity $C_0(T)$ obtained in the QCA can be considered as a rather accurate result for these mixtures. On the other hand, QCA is a precision method for structures of the Bethe lattice (or Cayley tree) type [11, 12]. It is therefore reasonable to assume that at low concentrations, the separation of the mixture occurs with formation of a new phase in the form of a fractal structure. For concentrated mixtures, the second phase nuclei are formed as three-dimensional precipitates.

In our opinion, this is not surprising because the difference between the effective coordination numbers z corresponding to the mixtures with small and large concentrations is related to the specific character of the impurity–impurity interaction. As mentioned above, the potential $V(\mathbf{r})$ decreases as $1/r^3$ at large distances r and has a complicated angular dependence [16–18]. In particular, it is attractive along some directions and repulsive along others. For small n_0, with the average distance between impurities by far exceeding the lattice parameter, the new phase can therefore grow as a fractal structure of the Cayley tree type such that the order of a node is determined by both the concentration and the number of crystallographic directions along which attraction occurs. In the opposite case of concentrated mixtures, elastic fields of impurities overlap and the effective interaction becomes essentially isotropic. As the result, the nuclei of the new phase grow in a compact form.

5. CONCLUSIONS

The results of the present work clearly demonstrate that thermodynamic properties of quantum solid ^{3}He–^{4}He mixtures can be successfully interpreted within the quasi-chemical approximation. The QCA has an essential advantage over the MFA because it accounts for the details of short-range correlations along with the long-range order. The QCA is therefore of primary importance for the description of the temperature region

[1] It is remarkable that in considering various approximations for the theory of solid mixtures, Fowler and Guggenheim introduce the parameter $T = V_0 z/2$ (*called The energy of unmixing*) from the outset. They denote the pair interaction energy by $2w/z$ (see [13, p. 570]).

above T_s, where the main contribution to the heat capacity is due to fluctuations in the impurity subsystem. Another important point of the theory is that it takes the long-range nature of the impurity–impurity interaction into account, which is crucial for understanding the behavior of the system. It is significant, however, that we must not know this interaction in every detail. Rather simple model (3) is quite sufficient to properly describe the thermodynamics of solid ^{3}He–^{4}He mixtures.

The theory is in good agreement with the experimental data in [1, 2] for both dilute and concentrated solutions. Unfortunately, we do not know works where the heat capacity of mixtures with intermediate concentrations ($0.3\% < n_0 < 5\%$) has been measured. The availability of these results would provide more distinctive conclusions about the change of nucleation mechanisms as a function of the solid mixture composition.

We would like to express our thanks to Prof. D.O. Edwards for discussions that led to the appearance of this work. We are indebted to Prof. G. Frossati, Prof. A.F. Andreev, Prof. V.N. Grigor'ev, and Prof. E.Ya. Rudavskiĭ for helpful conversations.

REFERENCES

1. D. O. Edwards, A. S. McWilliams, and J. G. Daunt, Phys. Lett. **1**, 218 (1961).
2. D. O. Edwards, A. S. McWilliams, and J. G. Daunt, Phys. Rev. Lett. **9**, 195 (1962).
3. D. O. Edwards and S. Balibar, Phys. Rev. B **39**, 4083 (1989).
4. D. O. Edwards and M. S. Pettersen, J. Low Temp. Phys. **87**, 473 (1992).
5. P. M. Tedrow and D. M. Lee, Phys. Rev. **181**, 399 (1969).
6. G. O. Zimmermann, in *Proceedings of LT-9 Conference, Columbus, 1964*, Vol. 1, p. 244.
7. R. Schrenk, O. Friz, Y. Fujii, E. Syskakis, and F. Pobell, J. Low Temp. Phys. **84**, 155 (1991).
8. W. J. Mullin, Phys. Rev. Lett. **20**, 254 (1968).
9. P. Kumar and M. Bernier, J. Low Temp. Phys. **79**, 1 (1990).
10. T. N. Antsygina, V. A. Slusarev, and K. A. Chishko, Fiz. Tverd. Tela (Leningrad) **40**, 355 (1998) [Phys. Solid State **40**, 325 (1998)].
11. R. Brout, *Phase Transitions* (Benjamin, New York, 1965; Mir, Moscow, 1967).
12. T. L. Hill, *Statistical Mechanics* (McGraw-Hill, New York, 1956; Inostrannaya Literatura, Moscow, 1960).
13. R. H. Fowler and E. A. Guggenheim, *Statistical Thermodynamics* (Cambridge Univ. Press, Cambridge, 1939; Inostrannaya Literatura, Moscow, 1949).
14. T. N. Antsygina, V. A. Slusarev, and K. A. Chishko, Fiz. Nizk. Temp. **21**, 583 (1995) [Low Temp. Phys. **21**, 453 (1995)].
15. C. Teodosiu, *Elastic Models of Crystal Defects* (Springer-Verlag, Berlin, 1982).
16. J. R. Hardy and R. Bullough, Philos. Mag. **15**, 237 (1967).
17. R. Bullough and J. R. Hardy, Philos. Mag. **17**, 833 (1968).
18. V. A. Slusarev, M. A. Strzhemechny, and I. A. Burakhovich, Fiz. Nizk. Temp. **3**, 1229 (1977) [Sov. J. Low Temp. Phys. **3**, 591 (1977)].
19. I. M. Lifshits, Zh. Éksp. Teor. Fiz. **42**, 1354 (1962) [Sov. Phys. JETP **15**, 939 (1962)].
20. R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965; Mir, Moscow, 1968).
21. S. Trickey, W. Kirk, and E. Adams, Rev. Mod. Phys. **44**, 668 (1972).
22. R. Kubo, *Statistical Mechanics* (North-Holland, Amsterdam, 1965; Mir, Moscow, 1967).